Springer Handbook of Nanotechnology

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

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

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Springer Handbook of Nanotechnology

Bharat Bhushan (Ed.)

With 972 Figures and 71 Tables

Springer

Professor Bharat Bhushan
Nanotribology Laboratory
for Information Storage and MEMS/NEMS
The Ohio State University
206 W. 18th Avenue
Columbus, Ohio 43210-1107
USA

Library of Congress Cataloging-in-Publication Data
Springer handbook of nanotechnology / Bharat Bhushan (ed.)
p. cm.
Includes bibliographical references and index
ISBN 3-540-01218-4 (alk. paper)
1. Nanotechnology--Handbooks, manuals, etc. I. Bhushan, Bharat; 1949-

T174.7S67 2003
620′.5--dc22 2003064953

ISBN 3-540-01218-4
Spinger-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September, 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

©Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

The use of designations, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Product liability: The publisher cannot guarantee the accuracy of any information about dosage and application contained in this book. In every individual case the user must check such information by consulting the relevant literature.

Production and typesetting: LE-TeX GbR, Leipzig
Handbook coordinator: Dr. W. Skolaut, Heidelberg
Typography, layout and illustrations: schreiberVIS, Seeheim
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: *design&production* GmbH, Heidelberg
Printing and binding: Stürtz AG, Würzburg

Printed on acid-free paper
SPIN 10890790 62/3141/YL 5 4 3 2 1 0

Foreword by Neal Lane

In a January 2000 speech at the California Institute of Technology, former President W. J. Clinton talked about the exciting promise of "nanotechnology" and the importance of expanding research in nanoscale science and engineering and in the physical sciences, more broadly. Later that month, he announced in his State of the Union Address an ambitious $ 497 million federal, multi-agency national nanotechnology initiative (NNI) in the fiscal year 2001 budget; and he made the NNI a top science and technology priority within a budget that emphasized increased investment in U.S. scientific research. With strong bipartisan support in Congress, most of this request was appropriated, and the NNI was born.

Nanotechnology is the ability to manipulate individual atoms and molecules to produce nanostructured materials and sub-micron objects that have applications in the real world. Nanotechnology involves the production and application of physical, chemical and biological systems at scales ranging from individual atoms or molecules to about 100 nanometers, as well as the integration of the resulting nanostructures into larger systems. Nanotechnology is likely to have a profound impact on our economy and society in the early 21st century, perhaps comparable to that of information technology or advances in cellular and molecular biology. Science and engineering research in nanotechnology promises breakthroughs in areas such as materials and manufacturing, electronics, medicine and healthcare, energy and the environment, biotechnology, information technology and national security. It is widely felt that nanotechnology will lead to the next industrial revolution.

Nanometer-scale features are built up from their elemental constituents. Micro- and nanosystems components are fabricated using batch-processing techniques that are compatible with integrated circuits and range in size from micro- to nanometers. Micro- and nanosystems include Micro/NanoElectroMechanical Systems (MEMS/NEMS), micromechatronics, optoelectronics, microfluidics and systems integration. These systems can sense, control, and activate on the micro/nanoscale and can function individually or in arrays to generate effects on the macroscale. Due to the enabling nature of these systems and the significant impact they can have on both the commercial and defense applications, industry as well as the federal government have taken special interest in seeing growth nurtured in this field. Micro- and nanosystems are the next logical step in the "silicon revolution".

The discovery of novel materials, processes, and phenomena at the nanoscale and the development of new experimental and theoretical techniques for research provide fresh opportunities for the development of innovative nanosystems and nanostructured materials. There is an increasing need for a multidisciplinary, systems-oriented approach to manufacturing micro/nanodevices which function reliably. This can only be achieved through the cross-fertilization of ideas from different disciplines and the systematic flow of information and people among research groups.

Prof. Neal Lane

University Professor
Department of Physics and Astronomy and James A. Baker III Institute for Public Policy Rice University Houston, Texas USA

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

Nanotechnology is a broad, highly interdisciplinary, and still evolving field. Covering even the most important aspects of nanotechnology in a single book that reaches readers ranging from students to active researchers in academia and industry is an enormous challenge. To prepare such a wide-ranging book on nanotechnology, Professor Bhushan has harnessed his own knowledge and experience, gained in several industries and universities, and has assembled about 90 internationally recognized authors from three continents to write 38 chapters. The authors come from both academia and industry.

Professor Bharat Bhushan's comprehensive book is intended to serve both as a textbook for university courses as well as a reference for researchers. It is a timely addition to the literature on nanotechnology, which I anticipate will stimulate further interest in this important new field and serve as an invaluable resource to members of the international scientific and industrial community.

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

Foreword by James R. Heath

Nanotechnology has become an increasingly popular buzzword over the past five years or so, a trend that has been fueled by a global set of publicly funded nanotechnology initiatives. Even as researchers have been struggling to demonstrate some of the most fundamental and simple aspects of this field, the term nanotechnology has entered into the public consciousness through articles in the popular press and popular fiction. As a consequence, the expectations of the public are high for nanotechnology, even while the actual public definition of nanotechnology remains a bit fuzzy.

Why shouldn't those expectations be high? The late 1990's witnessed a major information technology (IT) revolution and a minor biotechnology revolution. The IT revolution impacted virtually every aspect of life in the western world. I am sitting on an airplane at 30,000 feet at the moment, working on my laptop, as are about half of the other passengers on this plane. The plane itself is riddled with computational and communications equipment. As soon as we land, many of us will pull out cell phones, others will check email via wireless modem, some will do both. This picture would be the same if I was landing in Los Angeles, Beijing, or Capetown. I will probably never actually print this text, but will instead submit it electronically. All of this was unthinkable a dozen years ago. It is therefore no wonder that the public expects marvelous things to happen quickly. However, the science that laid the groundwork for the IT revolution dates back 60 years or more, with its origins in the fundamental solid state physics.

By contrast, the biotech revolution was relatively minor and, at least to date, not particularly effective. The major diseases that plagued mankind a quarter century ago are still here. In some third world countries, the average lifespan of individuals has actually decreased from where it was a full century ago. While the costs of electronics technologies have plummeted, health care costs have continued to rise. The biotech revolution may have a profound impact, but the task at hand is substantially more difficult to what was required for the IT revolution. In effect, the IT revolution was based on the advanced engineering of two-dimensional digital circuits constructed from relatively simple components – extended solids. The biotech revolution is really dependent upon the ability to reverse engineer three-dimensional analog systems constructed from quite complex components – proteins. Given that the basic science behind biotech is substantially younger than the science that has supported IT, it is perhaps not surprising that the biotech revolution has not really been a proper revolution yet, and it likely needs at least another decade or so to come to fruition.

Prof. James R. Heath

Department of Chemistry
Mail Code: 127-72
California Institute of Technology
Pasadena, CA 91125, USA

Worked in the group of Nobel Laureate Richard E. Smalley at Rice University (1984–88) and co-invented Fullerene molecules which led to a revolution in Chemistry including the realization of nanotubes. The work on Fullerene molecules was cited for the 1996 Nobel Prize in Chemistry. Later he joined the University of California at Los Angeles (1994–2002), and co-founded and served as a Scientific Director of The California Nanosystems Institute.

Where does nanotechnology fit into this picture? In many ways, nanotechnology depends upon the ability to engineer two- and three-dimensional systems constructed from complex components such as macromolecules, biomolecules, nanostructured solids, etc. Furthermore, in terms of patents, publications, and other metrics that can be used to gauge the birth and evolution of a field, nanotech lags some 15–20 years behind biotech. Thus, now is the time that the fundamental science behind nanotechnology is being explored and developed. Nevertheless, progress with that science is moving forward at a dramatic pace. If the scientific community can keep up this pace and if the public sector will continue to support this science, then it is possible, and perhaps even likely, that in 20 years from now we may be speaking of the nanotech revolution.

The Nanotechnology Handbook is timely in assembling chapters in the broad field of nanotechnology with an emphasis on reliability. The handbook should be a valuable reference for experienced researchers as well as for a novice in the field.

Preface

On December 29, 1959 at the California Institute of Technology, Nobel Laureate Richard P. Feynman gave a talk at the Annual meeting of the American Physical Society that has become one classic science lecture of the 20th century, titled "There's Plenty of Room at the Bottom." He presented a technological vision of extreme miniaturization in 1959, several years before the word "chip" became part of the lexicon. He talked about the problem of manipulating and controlling things on a small scale. Extrapolating from known physical laws, Feynman envisioned a technology using the ultimate toolbox of nature, building nanoobjects atom by atom or molecule by molecule. Since the 1980s, many inventions and discoveries in fabrication of nanoobjects have been a testament to his vision. In recognition of this reality, in a January 2000 speech at the same institute, former President W. J. Clinton talked about the exciting promise of "nanotechnology" and the importance of expanding research in nanoscale science and engineering. Later that month, he announced in his State of the Union Address an ambitious $497 million federal, multi-agency national nanotechnology initiative (NNI) in the fiscal year 2001 budget, and made the NNI a top science and technology priority. Nanotechnology literally means any technology done on a nanoscale that has applications in the real world. Nanotechnology encompasses production and application of physical, chemical and biological systems at size scales, ranging from individual atoms or molecules to submicron dimensions as well as the integration of the resulting nanostructures into larger systems. Nanofabrication methods include the manipulation or self-assembly of individual atoms, molecules, or molecular structures to produce nanostructured materials and sub-micron devices. Micro- and nanosystems components are fabricated using top-down lithographic and nonlithographic fabrication techniques. Nanotechnology will have a profound impact on our economy and society in the early 21st century, comparable to that of semiconductor technology, information technology, or advances in cellular and molecular biology. The research and development in nanotechnology will lead to potential breakthroughs in areas such as materials and manufacturing, nanoelectronics, medicine and healthcare, energy, biotechnology, information technology and national security. It is widely felt that nanotechnology will lead to the next industrial revolution.

Reliability is a critical technology for many micro- and nanosystems and nanostructured materials. No book exists on this emerging field. A broad based handbook is needed. The purpose of this handbook is to present an overview of nanomaterial synthesis, micro/nanofabrication, micro- and nanocomponents and systems, reliability issues (including nanotribology and nanomechanics) for nanotechnology, and industrial applications. The chapters have been written by internationally recognized experts in the field, from academia, national research labs and industry from all over the world.

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

We embarked on this project in February 2002, and we worked very hard to get all the chapters to the publisher in a record time of about 1 year. I wish to sincerely thank the authors for offering to write comprehensive chapters on a tight schedule. This is generally an added responsibility in the hectic work schedules of researchers today. I depended on a large number of reviewers who provided critical reviews. I would like to thank Dr. Phillip J. Bond, Chief of Staff and Under Secretary for Technology, US Department of Commerce, Washington, D.C. for suggestions for chapters as well as authors in the handbook. I would also like to thank my colleague, Dr. Huiwen Liu, whose efforts during the preparation of this handbook were very useful.

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

September 2003

Bharat Bhushan
Editor

Editors Vita

Dr. Bharat Bhushan received an M.S. in mechanical engineering from the Massachusetts Institute of Technology in 1971, an M.S. in mechanics and a Ph.D. in mechanical engineering from the University of Colorado at Boulder in 1973 and 1976, respectively, an MBA from Rensselaer Polytechnic Institute at Troy, NY in 1980, Doctor Technicae from the University of Trondheim at Trondheim, Norway in 1990, a Doctor of Technical Sciences from the Warsaw University of Technology at Warsaw, Poland in 1996, and Doctor Honouris Causa from the Metal-Polymer Research Institute of National Academy of Sciences at Gomel, Belarus in 2000. He is a registered professional engineer (mechanical). He is presently an Ohio Eminent Scholar and The Howard D. Winbigler Professor in the Department of Mechanical Engineering, Graduate Research Faculty Advisor in the Department of Materials Science and Engineering, and the Director of the Nanotribology Laboratory for Information Storage & MEMS/NEMS (NLIM) at the Ohio State University, Columbus, Ohio. He is an internationally recognized expert of tribology on the macro- to nanoscales, and is one of the most prolific authors in the field. He is considered by some a pioneer of the tribology and mechanics of magnetic storage devices and a leading researcher in the fields of nanotribology and nanomechanics using scanning probe microscopy and applications to micro/nanotechnology. He has authored 5 technical books, 45 handbook chapters, more than 450 technical papers in referred journals, and more than 60 technical reports, edited more than 25 books, and holds 14 U.S. patents. He is founding editor-in-chief of World Scientific Advances in Information Storage Systems Series, CRC Press Mechanics and Materials Science Series, and Microsystem Technologies – Micro- & Nanosystems and Information Storage & Processing Systems (formerly called Journal of Information Storage and Processing Systems). He has given more than 250 invited presentations on five continents and more than 60 keynote/plenary addresses at major international conferences.

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

Dr. Bhushan has previously worked for the R & D Division of Mechanical Technology Inc., Latham, NY; the Technology Services Division of SKF Industries Inc., King of Prussia, PA; the General Products Division Laboratory of IBM Corporation, Tucson, AZ; and the Almaden Research Center of IBM Corporation, San Jose, CA.

List of Authors

Chong H. Ahn
University of Cincinnati
Department of Electrical and Computer
Engineering and Computer Science
814 Rhodes Hall
Cincinnati, OH 45221-0030, USA
e-mail: *chong.ahn@uc.edu*

Boris Anczykowski
nanoAnalytics GmbH
Gievenbecker Weg 11
48149 Münster, Germany
e-mail: *anczykowski@nanoanalytics.com*

Massood Z. Atashbar
Western Michigan University
Department of Electrical
and Computer Engineering
200 Union Street SE
Kalamazoo, MI 49008-5329, USA
e-mail: *massood.atashbar@wmich.edu*

Wolfgang Bacsa
Université Paul Sabatier
Laboratoire de Physique des Solides (LPST)
118 Route de Narbonne
31062 Toulouse Cedex 4, France
e-mail: *bacsa@lpst.ups-tlse.fr*

William Sims Bainbridge
National Science Foundation
Division of Information and Intelligent Systems
4201 Wilson Boulevard
Arlington, VA 22230, USA
e-mail: *wbainbri@nsf.gov*

Antonio Baldi
Institut de Microelectronica de Barcelona (IMB)
Centro National Microelectrónica (CNM-CSIC)
Campus Universitat Autonoma de Barcelona
08193 Barcelona, Spain
e-mail: *Antoni.baldi@cnm.es*

Philip D. Barnes
Ohio State University
Biomedical Engineering Center
1080 Carmack Road
Columbus, OH 43210, USA
e-mail: *d_skill@yahoo.com*

James D. Batteas
National Institute of Standards and Technology
Surface and Microanalysis Science Division
100 Bureau Drive Mailstop 8372
Gaithersburg, MD 20899-8372, USA
e-mail: *james.batteas@nist.gov*

Roland Bennewitz
McGill University
Physics Department
3600 rue University
Montreal, QC H3A 2T8, Canada
e-mail: *roland@physics.mcgill.ca*

Alan D. Berman
Monitor Venture Enterprises
241 S. Figueroa St. Suite 300
Los Angeles, CA 90012, USA
e-mail: *alan.berman.2001@anderson.ucla.edu*

Bharat Bhushan
The Ohio State University
Nanotribology Laboratory for Information Storage
and MEMS/NEMS
206 W. 18th Avenue
Columbus, OH 43210-1107, USA
e-mail: *bhushan.2@osu.edu*

Gerd K. Binnig
IBM Zurich Research Laboratory
Micro-/Nanomechanics
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *gbi@zurich.ibm.com*

Marcie R. Black
Massachusetts Institute of Technology
Department of Electrical Engineering
and Computer Science
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: *marcie@alum.mit.edu*

Jean-Marc Broto
University Toulouse III
Laboratoire National
des Champs Magnétiques Pulsés (LNCMP)
143 Avenue de Rangueil
31432 Toulouse Cedex 4, France
e-mail: *broto@insa-tlse.fr*

Robert W. Carpick
University of Wisconsin-Madison
Department of Engineering Physics
1500 Engineering Drive
Madison, WI 53706-1687, USA
e-mail: *carpick@engr.wisc.edu*

Tsung-Lin Chen
National Chiao Tung University
Department of Mechanical Engineering
30050 Shin Chu, Taiwan
e-mail: *tsunglin@mail.nctu.edu.tw*

Yu-Ting Cheng
National Chiao Tung University
Department of Electronics Engineering
& Institute of Electronics
1001, Ta-Hsueh Road
300 HsinChu, Taiwan
e-mail: *ytcheng@faculty.nctu.edu.tw*

Giovanni Cherubini
IBM Zurich Research Laboratory
Storage Technologies
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *cbi@zurich.ibm.com*

Jin-Woo Choi
Louisiana State University
Department of Electrical
and Computer Engineering
102 South Campus Drive
Baton Rouge, LA 70803-5901, USA
e-mail: *choi@ece.lsu.edu*

Shawn J. Cunningham
WiSpry, Inc.
Colorado Springs Design Center
7150 Campus Drive, Suite 255
Colorado Springs, CO 80920, USA
e-mail: *shawn.cunningham@wispry.com*

Michel Despont
IBM Zurich Research Laboratory
Micro-/Nanomechanics
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *dpt@zurich.ibm.com*

Gene Dresselhaus
Massachusetts Institute of Technology
Francis Bitter Magnet Laboratory
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: *gene@mgm.mit.edu*

Mildred S. Dresselhaus
Massachusetts Institute of Technology
Department of Electrical Engineering
and Computer Science and Department of Physics
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: *millie@mgm.mit.edu*

Martin L. Dunn
University of Colorado at Boulder
Department of Mechanical Engineering
Campus Box 427
Boulder, CO 80309, USA
e-mail: *martin.dunn@colorado.edu*

Urs T. Dürig
IBM Zurich Research Laboratory
Micro-/Nanomechanics
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *drg@zurich.ibm.com*

Evangelos Eleftheriou
IBM Zurich Research Laboratory
Storage Technologies
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *ele@zurich.ibm.com*

Mauro Ferrari
Ohio State University
Biomedical Engineering Center
1080 Carmack Road
Columbus, OH 43210-1002, USA
e-mail: *Ferrari.5@osu.edu*

Emmanuel Flahaut
Université Paul Sabatier
CIRIMAT (Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux)
118 Route de Narbonne
31062 Toulouse Cedex 04, France
e-mail: *flahaut@chimie.ups-tlse.fr*

Lásló Forró
Swiss Federal Institute of Technology (EPFL)
Institute of Physics of Complex Matter
Ecublens
1015 Lausanne, Switzerland
e-mail: *laszlo.forro@epfl.ch*

Jane Frommer
IBM Almaden Research Center
Department of Science and Technology
650 Harry Road
San Jose, CA 95120, USA
e-mail: *frommer@Almaden.ibm.com*

Harald Fuchs
Universität Münster
Physikalisches Institut
Wilhelm-Klemm-Straße 10
48149 Münster, Germany
e-mail: *fuchsh@uni-muenster.de*

Franz J. Giessibl
Universität Augsburg
Lehrstuhl für Experimentalphysik VI
Universitätsstraße 1
86135 Augsburg, Germany
e-mail: *franz.giessibl@physik.uni-augsburg.de*

Enrico Gnecco
University of Basel
Department of Physics
Klingelbergstraße 82
4056 Basel, Switzerland
e-mail: *Enrico.Gnecco@unibas.ch*

Gérard Gremaud
Swiss Federal Institute of Technology (EPFL)
Institute of Physics of Complex Matter
Ecublens
1015 Lausanne, Switzerland
e-mail: *gremaud@epfl.ch*

Jason H. Hafner
Rice University
Department of Physics & Astronomy
PO BOX 1892
Houston, TX 77251-1892, USA
e-mail: *hafner@rice.edu*

Stefan Hengsberger
University of Applied Science of Fribourg
Bd de Pérolles
1705 Fribourg, Switzerland
e-mail: *stefan.hengsberger@eif.ch*

Peter Hinterdorfer
Johannes Kepler University of Linz
Institute for Biophysics
Altenbergerstraße 69
4040 Linz, Austria
e-mail: *peter.hinterdorfer@jku.at*

Roberto Horowitz
University of California at Berkeley
Department of Mechanical Engineering
5121 Etcheverry Hall
Berkeley, CA 94720-1742, USA
e-mail: *horowitz@me.berkeley.edu*

Hirotaka Hosoi
Japan Science and Technology Corporation
Innovation Plaza, Hokkaido
060-0819 Sapporo, Japan
e-mail: *hosoi@sapporo.jst-plaza.jp*

Jacob N. Israelachvili
University of California
Department of Chemical Engineering
and Materials Department
Santa Barbara, CA 93106, USA
e-mail: *Jacob@engineering.ucsb.edu*

Ghassan E. Jabbour
University of Arizona
Optical Sciences Center
1630 East University Boulevard
Tucson, AZ 85721, USA
e-mail: *gej@optics.arizona.edu*

Harold Kahn
Case Western Reserve University
Department of Materials Science and Engineering
10900 Euclid Avenue
Cleveland, OH 44106-7204, USA
e-mail: *kahn@cwru.edu*

András Kis
Swiss Federal Institute of Technology (EPFL)
Institute of Physics of Complex Matter
Ecublens
1015 Lausanne, Switzerland
e-mail: *andras@igahpse.epfl.ch*

Jané Kondev
Brandeis University
Physics Department
Waltham, MA 02454, USA
e-mail: *kondev@brandeis.edu*

Andrzej J. Kulik
Swiss Federal Institute of Technology (EPFL)
Institute of Physics of Complex Matter
1015 Lausanne, Switzerland
e-mail: *andrzej.kulik@epfl.ch*

Christophe Laurent
Université Paul Sabatier
CIRIMAT (Centre Interuniversitaire de Recherche
et d'Ingénierie des Matériaux)
118 Route de Narbonne
31062 Toulouse Cedex 04, France
e-mail: *laurent@chimie.ups-tlse.fr*

Stephen C. Lee
Ohio State University
Biomedical Engineering Center
1080 Carmack Road
Columbus, OH 43210-1002, USA
e-mail: *Lee@bme.ohio-state.edu*

Yunfeng Li
University of California at Berkeley
Department of Mechanical Engineering
5121 Etcheverry Hall
Berkeley, CA 94720-1740, USA
e-mail: *yunfeng@me.Berkeley.edu*

Liwei Lin
University of California at Berkeley
Mechanical Engineering Department
5126 Etcheverry
Berkeley, CA 94720-1740, USA
e-mail: *lwlin@me.berkeley.edu*

Yu-Ming Lin
Massachusetts Institute of Technology
Department of Electrical Engineering
and Computer Science
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: *yming@mgm.mit.edu*

Huiwen Liu
Ohio State University
Nanotribology Laboratory for Information Storage and MEMS/NEMS
3070 St John Ct 7
Columbus, OH 43210-1107, USA
e-mail: *liu.403@osu.edu*

Adrian B. Mann
Rutgers University
Department of Ceramics and Materials Engineering
607 Taylor Road
Piscataway, NJ 08854, USA
e-mail: *abmann@rci.rutgers.edu*

Othmar Marti
University of Ulm
Department of Experimental Physics
Albert-Einstein-Allee 11
89069 Ulm, Germany
e-mail: *Othmar.Marti@physik.uni-ulm.de*

Jack Martin
Analog Devices, Inc.
Micromachined Products Division
21 Osborn Street
Cambridge, MA 02139, USA
e-mail: *jack.martin@analog.com*

Brendan McCarthy
University of Arizona
Optical Sciences Center
1630 East University Boulevard
Tucson, AZ 85721, USA
e-mail: *bmccarthy@optics.arizona.edu*

Mehran Mehregany
Case Western Reserve University
Department of Electrical Engineering and Computer Science
188 Bingham Building
Cleveland, OH 44106, USA
e-mail: *mxm31@cwru.edu*

Ernst Meyer
University of Basel
Institute of Physics
Klingelbergstraße 82
4056 Basel, Switzerland
e-mail: *Ernst.Meyer@unibas.ch*

Marc Monthioux
UPR A-8011 CNRS
Centre d'Elaboration des Matériaux et d'Etudes Structurales (CEMES)
29 Rue Jeanne Marvig
31055 Toulouse Cedex 4, France
e-mail: *monthiou@cemes.fr*

Markus Morgenstern
University of Hamburg
Institute of Applied Physics
Jungiusstraße 11
20355 Hamburg, Germany
e-mail: *mmorgens@physnet.uni-hamburg.de*

Seizo Morita
Osaka University
Department of Electronic Engineering
Yamada-Oka 2-1
565-0871 Suita-Citiy, Osaka, Japan
e-mail: *smorita@ele.eng.osaka-u.ac.jp*

Koichi Mukasa
Hokkaido University
Nanoelectronics Laboratory
Nishi-8, Kita-13, Kita-ku
060-8628 Sapporo, Japan
e-mail: *mukasa@nano.eng.hokudai.ac.jp*

Martin H. Müser
University of Western Ontario
Department of Applied Mathematics
WSC 139, Faculty of Science
London, Ontario N6A 5B7, Canada
e-mail: *mmuser@uwo.ca*

Kenn Oldham
University of California at Berkeley
Department of Mechanical Engineering
5121 Etcheverry Hall
Berkeley, CA 94720-1740, USA
e-mail: *oldham@newton.berkeley.edu*

Hiroshi Onishi
Kanagawa Academy of Science and Technology
Surface Chemistry Laboratory
KSP East 404, 3-2-1 Sakado, Takatsu-ku,
Kawasaki-shi
213-0012 Kanagawa, Japan
e-mail: *oni@net.ksp.or.jp*

René M. Overney
University of Washington
Department of Chemical Engineering
Seattle, WA 98195-1750, USA
e-mail: *roverney@u.Washington.edu*

Alain Peigney
Université Paul Sabatier
CIRIMAT (Centre Inter-universitaire de Recherches
et d'Ingénierie des Matériaux) – UMR CNRS 5085
118 Route de Narbonne
31062 Toulouse Cedex 4, France
e-mail: *peigney@chimie.ups-tlse.fr*

Oliver Pfeiffer
University of Basel
Institute of Physics
Klingelbergstraß 82
4056 Basel, Switzerland
e-mail: *Oliver.Pfeiffer@stud.unibas.ch*

Rob Phillips
California Institute of Technology
Mechanical Engineering and Applied Physics
1200 California Boulevard
Pasadena, CA 91125, USA
e-mail: *phillips@aero.caltech.edu*

Haralampos Pozidis
IBM Zurich Research Laboratory
Storage Technologies
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *hap@zurich.ibm.com*

Prashant K. Purohit
California Institute of Technology
Mechanical Engineering
1200 California Boulevard
Pasadena, CA 91125, USA
e-mail: *prashant@caltech.edu*

Oded Rabin
Massachusetts Institute of Technology
Department of Chemistry
77 Massachusetts Avenue
Cambridge, MA 02139, USA
e-mail: *oded@mgm.mit.edu*

Françisco M. Raymo
University of Miami
Department of Chemistry
1301 Memorial Drive
Coral Gables, FL 33146-0431, USA
e-mail: *fraymo@miami.edu*

Manitra Razafinimanana
Université Paul Sabatier
Centre de Physique des Plasmas
et leurs Applications (CPPAT)
118 Route de Narbonne
31062 Toulouse Cedex, France
e-mail: *razafinimanana@cpat.ups-tlse.fr*

Mark O. Robbins
Johns Hopkins University
Department of Physics and Astronomy
3400 North Charles Street
Baltimore, MD 21218, USA
e-mail: *mr@jhu.edu*

John A. Rogers
University of Illinois
Department of Materials Science and Engineering
1304 W. Green Street
Urbana, IL 61801, USA
e-mail: *jrogers@uiuc.edu*

Mark Ruegsegger
Ohio State University
Biomedical Engineering Center
1080 Carmack Road
Columbus, OH 43210, USA
e-mail: *mark@bme.ohio-state.edu*

Marina Ruths
Åbo Akademi University
Department of Physical Chemistry
Porthansgatan 3–5
20500 Åbo, Finland
e-mail: *mruths@abo.fi*

Dror Sarid
University of Arizona
Optical Sciences Center
1630 East University Boulevard
Tucson, AZ 85721, USA
e-mail: *sarid@optics.arizona.edu*

Akira Sasahara
Kanagawa Academy of Science and Technology
Surface Chemistry Laboratory
KSP East 404, 3-2-1 Sakado, Takatsu-ku,
Kawasaki-shi
213-0012 Kanagawa, Japan
e-mail: *ryo@net.ksp.or.jp*

André Schirmeisen
University of Münster
Institute of Physics
Whilhem-Klemm-Straße 10
48149 Münster, Germany
e-mail: *schira@uni-muenster.de*

Alexander Schwarz
University of Hamburg
Institute of Applied Physics
Jungiusstraße 11
20355 Hamburg, Germany
e-mail: *aschwarz@physnet.uni-hamburg.de*

Udo D. Schwarz
Yale University
Department of Mechanical Engineering
15 Prospect Street
New Haven, CT 06510, USA
e-mail: *udo.schwarz@yale.edu*

Philippe Serp
Ecole Nationale Supèrieure d'Ingénieurs
en Arts Chimiques et Technologiques
Laboratoire de Catalyse, Chimie Fine et Polymères
118 Route de Narbonne
31077 Toulouse, France
e-mail: *Philippe.Serp@ensiacet.fr*

Bryan R. Smith
Ohio State University
Biomedical Engineering Center
1080 Carmack Road
Columbus, OH 43210, USA
e-mail: *bryan@bme.ohio-state.edu*

Anisoara Socoliuc
University of Basel
Institute of Physics
Klingelbergstraße 82
4056 Basel, Switzerland
e-mail: *A.Socoliuc@unibas.ch*

Yasuhiro Sugawara
Osaka University
Department of Applied Physics
Yamada-Oka 2-1
565-0871 Suita, Japan
e-mail: *sugawara@ap.eng.osaka-u.ac.jp*

George W. Tyndall
IBM Almaden Research Center
Science and Technology
6369 Didion Ct.
San Jose, CA 95123, USA
e-mail: *tyndallgw@netscape.net*

Peter Vettiger
IBM Zurich Research Laboratory
Manager Micro-/Nanomechanics
Säumerstraße 4
8803 Rüschlikon, Switzerland
e-mail: *pv@zurich.ibm.com*

Darrin J. Young
Case Western Reserve University
Electrical Engineering and Computer Science
10900 Euclid Avenue
Cleveland, OH 44106, USA
e-mail: *djy@po.cwru.edu*

Babak Ziaie
University of Minnesota
Department of Electrical
and Computer Engineering
200 Union Street SE
Minneapolis, MN 55455, USA
e-mail: *ziaie@ece.umn.edu*

Christian A. Zorman
Case Western Reserve University
Department of Electrical Engineering
and Computer Science
719 Glennan Building
Cleveland, OH 44106, USA
e-mail: *caz@po.cwru.edu*

Philippe K. Zysset
Technische Universität Wien
Institut für Leichtbau und Flugzeugbau (ILFB)
Gußhausstraße 27–29
1040 Wien, Austria
e-mail: *philippe.zysset@epfl.ch*

Contents

List of Tables XXIX
List of Abbreviations XXXIII

1 **Introduction to Nanotechnology** 1
1.1 Background and Definition of Nanotechnology 1
1.2 Why Nano? 2
1.3 Lessons from Nature 2
1.4 Applications in Different Fields 3
1.5 Reliability Issues of MEMS/NEMS 4
1.6 Organization of the Handbook 5
References 5

Part A Nanostructures, Micro/Nanofabrication, and Micro/Nanodevices

2 **Nanomaterials Synthesis and Applications: Molecule-Based Devices** 9
2.1 Chemical Approaches to Nanostructured Materials 10
2.2 Molecular Switches and Logic Gates 14
2.3 Solid State Devices 22
2.4 Conclusions and Outlook 35
References 36

3 **Introduction to Carbon Nanotubes** 39
3.1 Structure of Carbon Nanotubes 40
3.2 Synthesis of Carbon Nanotubes 45
3.3 Growth Mechanisms of Carbon Nanotubes 59
3.4 Properties of Carbon Nanotubes 63
3.5 Carbon Nanotube-Based Nano-Objects 68
3.6 Applications of Carbon Nanotubes 73
References 86

4 **Nanowires** 99
4.1 Synthesis 100
4.2 Characterization and Physical Properties of Nanowires 110
4.3 Applications 131
4.4 Concluding Remarks 138
References 138

5 **Introduction to Micro/Nanofabrication** 147
5.1 Basic Microfabrication Techniques 148
5.2 MEMS Fabrication Techniques 159

5.3 Nanofabrication Techniques 170
References 180

6 **Stamping Techniques for Micro and Nanofabrication: Methods and Applications** 185
6.1 High Resolution Stamps 186
6.2 Microcontact Printing 187
6.3 Nanotransfer Printing 190
6.4 Applications 193
6.5 Conclusions 200
References 200

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems** 203
7.1 Silicon 203
7.2 Germanium-Based Materials 210
7.3 Metals 211
7.4 Harsh Environment Semiconductors 212
7.5 GaAs, InP, and Related III-V Materials 217
7.6 Ferroelectric Materials 218
7.7 Polymer Materials 219
7.8 Future Trends 220
References 221

8 **MEMS/NEMS Devices and Applications** 225
8.1 MEMS Devices and Applications 227
8.2 NEMS Devices and Applications 246
8.3 Current Challenges and Future Trends 249
References 250

9 **Microfluidics and Their Applications to Lab-on-a-Chip** 253
9.1 Materials for Microfluidic Devices and Micro/Nano Fabrication Techniques 254
9.2 Active Microfluidic Devices 257
9.3 Smart Passive Microfluidic Devices 262
9.4 Lab-on-a-Chip for Biochemical Analysis 270
References 276

10 **Therapeutic Nanodevices** 279
10.1 Definitions and Scope of Discussion 280
10.2 Synthetic Approaches: "top-down" versus "bottom-up" Approaches for Nanotherapeutic Device Components 285
10.3 Technological and Biological Opportunities 288
10.4 Applications for Nanotherapeutic Devices 307
10.5 Concluding Remarks: Barriers to Practice and Prospects 315
References 317

Part B Scanning Probe Microscopy

11 **Scanning Probe Microscopy – Principle of Operation, Instrumentation, and Probes** 325
11.1 Scanning Tunneling Microscope 327
11.2 Atomic Force Microscope 331
11.3 AFM Instrumentation and Analyses 347
References 364

12 **Probes in Scanning Microscopies** 371
12.1 Atomic Force Microscopy 372
12.2 Scanning Tunneling Microscopy 382
References 383

13 **Noncontact Atomic Force Microscopy and Its Related Topics** 385
13.1 Principles of Noncontact Atomic Force Microscope (NC-AFM) 386
13.2 Applications to Semiconductors 391
13.3 Applications to Insulators 397
13.4 Applications to Molecules 404
References 407

14 **Low Temperature Scanning Probe Microscopy** 413
14.1 Microscope Operation at Low Temperatures 414
14.2 Instrumentation 415
14.3 Scanning Tunneling Microscopy and Spectroscopy 419
14.4 Scanning Force Microscopy and Spectroscopy 433
References 442

15 **Dynamic Force Microscopy** 449
15.1 Motivation: Measurement of a Single Atomic Bond 450
15.2 Harmonic Oscillator: A Model System for Dynamic AFM 454
15.3 Dynamic AFM Operational Modes 455
15.4 *Q*-Control 464
15.5 Dissipation Processes Measured with Dynamic AFM 468
15.6 Conclusion 471
References 471

16 **Molecular Recognition Force Microscopy** 475
16.1 Ligand Tip Chemistry 476
16.2 Fixation of Receptors to Probe Surfaces 478
16.3 Single-Molecule Recognition Force Detection 479
16.4 Principles of Molecular Recognition Force Spectroscopy 482
16.5 Recognition Force Spectroscopy: From Isolated Molecules to Biological Membranes 484
16.6 Recognition Imaging 489
16.7 Concluding Remarks 491
References 492

Part C Nanotribology and Nanomechanics

17 **Micro/Nanotribology and Materials Characterization Studies Using Scanning Probe Microscopy** 497
17.1 Description of AFM/FFM and Various Measurement Techniques 499
17.2 Friction and Adhesion 507
17.3 Scratching, Wear, Local Deformation, and Fabrication/Machining 518
17.4 Indentation 526
17.5 Boundary Lubrication 530
17.6 Closure 538
References 539

18 **Surface Forces and Nanorheology of Molecularly Thin Films** 543
18.1 Introduction: Types of Surface Forces 544
18.2 Methods Used to Study Surface Forces 546
18.3 Normal Forces Between Dry (Unlubricated) Surfaces 550
18.4 Normal Forces Between Surfaces in Liquids 554
18.5 Adhesion and Capillary Forces 564
18.6 Introduction: Different Modes of Friction and the Limits of Continuum Models 569
18.7 Relationship Between Adhesion and Friction Between Dry (Unlubricated and Solid Boundary Lubricated) Surfaces 571
18.8 Liquid Lubricated Surfaces 580
18.9 Role of Molecular Shape and Surface Structure in Friction 591
References 594

19 **Scanning Probe Studies of Nanoscale Adhesion Between Solids in the Presence of Liquids and Monolayer Films** 605
19.1 The Importance of Adhesion at the Nanoscale 605
19.2 Techniques for Measuring Adhesion 606
19.3 Calibration of Forces, Displacements, and Tips 610
19.4 The Effect of Liquid Capillaries on Adhesion 612
19.5 Self-Assembled Monolayers 618
19.6 Concluding Remarks 624
References 624

20 **Friction and Wear on the Atomic Scale** 631
20.1 Friction Force Microscopy in Ultra-High Vacuum 632
20.2 The Tomlinson Model 636
20.3 Friction Experiments on Atomic Scale 638
20.4 Thermal Effects on Atomic Friction 642
20.5 Geometry Effects in Nanocontacts 646
20.6 Wear on the Atomic Scale 649
20.7 Molecular Dynamics Simulations of Atomic Friction and Wear 651
20.8 Energy Dissipation in Noncontact Atomic Force Microscopy 654
20.9 Conclusion 656
References 657

21 **Nanoscale Mechanical Properties – Measuring Techniques and Applications** 661
21.1 Local Mechanical Spectroscopy by Contact AFM 662
21.2 Static Methods – Mesoscopic Samples 667
21.3 Scanning Nanoindentation: An Application to Bone Tissue 674
21.4 Conclusions and Perspectives 682
References 682

22 **Nanomechanical Properties of Solid Surfaces and Thin Films** 687
22.1 Instrumentation 688
22.2 Data Analysis 694
22.3 Modes of Deformation 702
22.4 Thin Films and Multilayers 707
22.5 Developing Areas 711
References 712

23 **Atomistic Computer Simulations of Nanotribology** 717
23.1 Molecular Dynamics 718
23.2 Friction Mechanisms at the Atomic Scale 723
23.3 Stick–Slip Dynamics 732
23.4 Conclusions 734
References 735

24 **Mechanics of Biological Nanotechnology** 739
24.1 Science at the Biology–Nanotechnology Interface 740
24.2 Scales at the Bio–Nano Interface 746
24.3 Modeling at the Nano–Bio Interface 752
24.4 Nature's Nanotechnology Revealed: Viruses as a Case Study 755
24.5 Concluding Remarks 760
References 761

25 **Mechanical Properties of Nanostructures** 763
25.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures 765
25.2 Experimental Results and Discussion 770
25.3 Finite Element Analysis of Nanostructures with Roughness and Scratches 778
25.4 Closure 785
References 786

Part D Molecularly Thick Films for Lubrication

26 **Nanotribology of Ultrathin and Hard Amorphous Carbon Films** 791
26.1 Description of Commonly Used Deposition Techniques 795
26.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties 800

26.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques 805
References 827

27 **Self-Assembled Monolayers for Controlling Adhesion, Friction and Wear** 831
27.1 A Primer to Organic Chemistry 834
27.2 Self-Assembled Monolayers: Substrates, Head Groups, Spacer Chains, and End Groups 839
27.3 Tribological Properties of SAMs 841
27.4 Closure 856
References 857

28 **Nanoscale Boundary Lubrication Studies** 861
28.1 Lubricants Details 862
28.2 Nanodeformation, Molecular Conformation, and Lubricant Spreading 864
28.3 Boundary Lubrication Studies 866
28.4 Closure 880
References 881

29 **Kinetics and Energetics in Nanolubrication** 883
29.1 Background: From Bulk to Molecular Lubrication 885
29.2 Thermal Activation Model of Lubricated Friction 887
29.3 Functional Behavior of Lubricated Friction 888
29.4 Thermodynamical Models Based on Small and Nonconforming Contacts 890
29.5 Limitation of the Gaussian Statistics – The Fractal Space 891
29.6 Fractal Mobility in Reactive Lubrication 892
29.7 Metastable Lubricant Systems in Large Conforming Contacts 894
29.8 Conclusion 895
References 895

Part E Industrial Applications and Microdevice Reliability

30 **Nanotechnology for Data Storage Applications** 899
30.1 Current Status of Commercial Data Storage Devices 901
30.2 Opportunities Offered by Nanotechnology for Data Storage 907
30.3 Conclusion 918
References 919

31 **The "Millipede" – A Nanotechnology-Based AFM Data-Storage System** 921
31.1 The Millipede Concept 923
31.2 Thermomechanical AFM Data Storage 924
31.3 Array Design, Technology, and Fabrication 926

31.4 Array Characterization 927
31.5 *x*/*y*/*z* Medium Microscanner 929
31.6 First Write/Read Results with the 32×32 Array Chip 931
31.7 Polymer Medium 932
31.8 Read Channel Model 939
31.9 System Aspects 943
31.10 Conclusions 948
References 948

32 **Microactuators for Dual-Stage Servo Systems in Magnetic Disk Files** 951
32.1 Design of the Electrostatic Microactuator 952
32.2 Fabrication 962
32.3 Servo Control Design of MEMS Microactuator Dual-Stage Servo Systems 968
32.4 Conclusions and Outlook 978
References 979

33 **Micro/Nanotribology of MEMS/NEMS Materials and Devices** 983
33.1 Introduction to MEMS 985
33.2 Introduction to NEMS 988
33.3 Tribological Issues in MEMS/NEMS 989
33.4 Tribological Studies of Silicon and Related Materials 995
33.5 Lubrication Studies for MEMS/NEMS 1003
33.6 Component-Level Studies 1009
References 1017

34 **Mechanical Properties of Micromachined Structures** 1023
34.1 Measuring Mechanical Properties of Films on Substrates 1023
34.2 Micromachined Structures for Measuring Mechanical Properties 1024
34.3 Measurements of Mechanical Properties 1034
References 1037

35 **Thermo- and Electromechanics of Thin-Film Microstructures** 1039
35.1 Thermomechanics of Multilayer Thin-Film Microstructures 1041
35.2 Electromechanics of Thin-Film Microstructures 1061
35.3 Summary and Mention of Topics not Covered 1078
References 1078

36 **High Volume Manufacturing and Field Stability of MEMS Products** 1083
36.1 Manufacturing Strategy 1086
36.2 Robust Manufacturing 1087
36.3 Stable Field Performance 1102
References 1106

37 **MEMS Packaging and Thermal Issues in Reliability** 1111
37.1 MEMS Packaging 1111
37.2 Hermetic and Vacuum Packaging and Applications 1116
37.3 Thermal Issues and Packaging Reliability 1122
37.4 Future Trends and Summary 1128
References 1129

Part F Social and Ethical Implication

38 **Social and Ethical Implications of Nanotechnology** 1135
38.1 Applications and Societal Impacts 1136
38.2 Technological Convergence 1139
38.3 Major Socio-technical Trends 1141
38.4 Sources of Ethical Behavior 1143
38.5 Public Opinion 1145
38.6 A Research Agenda 1148
References 1149

Acknowledgements 1153
About the Authors 1155
Detailed Contents 1171
Subject Index 1189

List of Tables

Part A Nanostructures, Micro/Nanofabrication, and Micro/Nanodevices

3 **Introduction to Carbon Nanotubes**
Table 3.1 Different carbon morphologies 50
Table 3.2 Carbon nanofilament morphologies and some basic synthesis conditions 62
Table 3.3 Adsorption properties and sites of SWNTs and MWNTs 65
Table 3.4 Preparation and catalytic performances of some nanotube-supported catalysts..... 77
Table 3.5 Applications for nanotube-based multifunctional materials ... 84

4 **Nanowires**
Table 4.1 Selected syntheses of nanowires by material 101

5 **Introduction to Micro/Nanofabrication**
Table 5.1 Typical dry etch chemistries..... 157
Table 5.2 Surface micromachined sacrificial/structural combinations 165

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems**
Table 7.1 Selected materials for MEMS and NEMS..... 204

9 **Microfluidics and Their Applications to Lab-on-a-Chip**
Table 9.1 Overview of the different polymer micro/nano fabrication techniques 257

10 **Therapeutic Nanodevices**
Table 10.1 Ideal characteristics of nanodevices..... 282
Table 10.2 Orthogonal conjugation chemistries 295

Part B Scanning Probe Microscopy

11 **Scanning Probe Microscopy – Principle of Operation, Instrumentation, and Probes**
Table 11.1 Comparison of Various Conventional Microscopes with SPMs ... 326
Table 11.2 Relevant properties of materials used for cantilevers..... 339
Table 11.3 Measured vertical spring constants and natural frequencies of triangular (V-shaped) cantilevers made of PECVD Si_3N_4 341
Table 11.4 Vertical and torsional spring constants of rectangular cantilevers made of Si and PECVD..... 341
Table 11.5 Noise in Interferometers 353

13 **Noncontact Atomic Force Microscopy and Its Related Topics**
Table 13.1 Comparison between the adatom heights observed in an AFM image and the variety of properties for inequivalent adatoms 393

Part C Nanotribology and Nanomechanics

17 **Micro/Nanotribology and Materials Characterization Studies Using Scanning Probe Microscopy**
Table 17.1 Comparison of typical operating parameters in SFA, STM, and AFM/FFM used for micro/nanotribological studies..... 498
Table 17.2 Surface roughness and coefficients of friction of various samples in air..... 518

18 **Surface Forces and Nanorheology of Molecularly Thin Films**
Table 18.1 Types of surface forces in vacuum vs. in liquid (colloidal forces) 545
Table 18.2 Van der Waals interaction energy and force between macroscopic bodies of different geometries 551
Table 18.3 Electrical "double-layer" interaction energy and force between macroscopic bodies..... 556
Table 18.4 The three main tribological regimes characterizing the changing properties of liquids..... 570
Table 18.5 Effect of molecular shape and short-range forces on tribological properties..... 592

19 **Scanning Probe Studies of Nanoscale Adhesion Between Solids in the Presence of Liquids and Monolayer Films**
Table 19.1 Interactions evaluated by atomic force microscopy adhesion measurements..... 620

21 **Nanoscale Mechanical Properties – Measuring Techniques and Applications**
Table 21.1 Summary of the mechanical properties of carbon nanotubes.. 672
Table 21.2 Absolute and relative changes of the indentation modulus with respect to the initial values under wet conditions 680

22 **Nanomechanical Properties of Solid Surfaces and Thin Films**
Table 22.1 Results for some experimental studies of multilayer hardness 710

24 **Mechanics of Biological Nanotechnology**
Table 24.1 Sizes of entities in comparison to the size of *E. coli* 748
Table 24.2 Time scales of various events in biological nanotechnology 750

25 **Mechanical Properties of Nanostructures**
Table 25.1 Hardness, elastic modulus, fracture toughness, and critical load results of the bulk single-crystal Si(100) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au 770

Table 25.2 Summary of measured parameters from quasi-static bending tests 775
Table 25.3 Stresses and displacements for materials that are elastic, elastic-plastic, or elastic-perfectly plastic 784

Part D Molecularly Thick Films for Lubrication

26 **Nanotribology of Ultrathin and Hard Amorphous Carbon Films**
Table 26.1 Summary of the most commonly used deposition techniques . 796
Table 26.2 Experimental results from EELS and Raman spectroscopy 801
Table 26.3 Experimental results of FRS analysis 802
Table 26.4 Experimental results of physical properties 802
Table 26.5 Hardness, elastic modulus, fracture toughness, fatigue life, critical load during scratch, coefficient of friction of various DLC coatings 809

27 **Self-Assembled Monolayers for Controlling Adhesion, Friction and Wear**
Table 27.1 Relative electronegativity of selected elements 835
Table 27.2 Names and formulas of selected hydrocarbons 836
Table 27.3 Names and formulas of selected alcohols, ethers, phenols, and thiols 836
Table 27.4 Names and formulas of selected aldehydes and ketones 837
Table 27.5 Names and formulas of selected carboxylic acids and esters ... 838
Table 27.6 Names and formulas of selected organic nitrogen compounds 838
Table 27.7 Some examples of polar and nonpolar groups 839
Table 27.8 Organic groups listed in the increasing order of polarity 839
Table 27.9 Selected substrates and precursors used for formation of SAMs 840
Table 27.10 The roughness, thickness, tilt angles, and spacer chain lengths of SAMs 846
Table 27.11 Melting point of typical organic compounds similar to HDT and BPT SAMs 851
Table 27.12 Calculated and measured relative heights of HDT and MHA self-assembled monolayers 855
Table 27.13 Calculated and measured residual film thickness for SAMs under critical load 855
Table 27.14 Bond strengths of the chemical bonds in SAMs 856

28 **Nanoscale Boundary Lubrication Studies**
Table 28.1 Typical properties of Z-15 and Z-DOL 863

Part E Industrial Applications and Microdevice Reliability

30 **Nanotechnology for Data Storage Applications**
Table 30.1 A comparison of the parameters of HDD and ODD 903

Table 30.2 Perceived technical limits for hard disk drive technology..... 904
Table 30.3 A table of probe storage operating parameters 909
Table 30.4 Probe storage projected performance 909

31 **The "Millipede" – A Nanotechnology-Based AFM Data-Storage System**
Table 31.1 Areal density and storage capacity 947

32 **Microactuators for Dual-Stage Servo Systems in Magnetic Disk Files**
Table 32.1 Parameters of the electrostatic microactuator 960

33 **Micro/Nanotribology of MEMS/NEMS Materials and Devices**
Table 33.1 Dimensions and masses in perspective 985
Table 33.2 Selected bulk properties of SiC and Si(100)..... 996
Table 33.3 Surface roughness and micro- and macroscale coefficients of friction of selected samples..... 997
Table 33.4 RMS, microfriction, microscratching/microwear, and nanoindentation hardness data for various virgin, coated, and treated silicon samples 999
Table 33.5 Summary of micro/nanotribological properties of the sample materials 1001
Table 33.6 Surface roughness parameters and microscale coefficient of friction 1009
Table 33.7 Coefficient of static friction measurements of MEMS devices and structures 1015

35 **Thermo- and Electromechanics of Thin-Film Microstructures**
Table 35.1 Stresses, curvature, and strains in a film/substrate system..... 1045
Table 35.2 Predicted and measured curvature per unit negative temperature change for gold/polysilicon microstructures as a function of the polysilicon thickness. 1053

37 **MEMS Packaging and Thermal Issues in Reliability**
Table 37.1 Summary of bonding mechanisms..... 1116
Table 37.2 The Maximum Likelihood Estimation for Mean Time To Failure (MTTF) 1127

List of Abbreviations

μCP	microcontact printing
μTAS	micro-total analysis systems
2-DEG	two-dimensional electron gas

A

A	adenine
ABS	air-bearing surface
ADEPTS	antibody directed enzyme-prodrug therapy
AFAM	atomic force acoustic microscopy
AFM	atomic force microscope/microscopy
AIDCN	2-amino-4,5-imidazoledicarbonitrile
AM	amplitude modulation
APCVD	atmospheric pressure chemical vapor deposition
ASA	anti-stiction agent
ATP	adenosine triphosphat

B

BE	boundary element
BioMEMS	biological or biomedical microelectromechanical systems
BP	bit pitch
BPI	bits per inch
bpsi	bits per square inch
BSA	bovine serum albumin

C

C	cytosine
CA	constant amplitude
CBA	cantilever beam array
CCVD	catalytic chemical vapor deposition
CDS	correlated double sampling
CDW	charge density wave
CE	capillary electrophoresis
CE	constant excitation mode
CFM	chemical force microscopy
CG	controlled geometry
CNT	carbon nanotube
COC	cyclic olefin copolymers
COF	chip-on-flex
COGs	cost of goods
CSM	continuous stiffness measurement
CTE	coefficient of thermal expansion
CTLs	cytoxic T-lymphocytes
CVD	chemical vapor deposition

D

DAS	dimer adatom stacking
DBR	distributed Bragg reflector
DCs	dendritic cells
DFB	distributed feedback
DFM	dynamic force microscopy
DFT	density functional theory
DLC	diamond-like carbon
DLP	digital light processing
DLVO	Derjaguin–Landau–Verwey–Overbeek
DMD	digital micromirror device
DMT	Derjaguin–Muller–Toporov
DOS	density of states
DPN	dip-pen nanolithography
DRIE	deep reactive ion etching
DSC	differential scanning calorimetry
DSP	digital signal processor
DT	diphteria toxin
DWNTs	double-wall nanotubes

E

EAM	embedded atom method
EBD	electron beam deposition
ECR-CVD	electron cyclotron resonance chemical vapor deposition
EDC	1-ethyl-3-(3-diamethylaminoprophyl) carbodiimide
EDP	ethylene diamine pyrocatechol
EDS	energy dispersive X-ray spectrometer
EELS	electron energy loss spectrometer/spectroscopy
EFC	electrostatic force constant
EFM	electric field gradient microscopy
EHD	electrohydrodynamic
EL	electro-luminiscence
EO	electro-osmosis
EOF	electro-osmotic flow
EPR	enhanced permeability and retention effect
ESD	electrostatic discharge

F

FAD	flavin adenine dinucleotide
FC	flip chip technique
FCA	filtered cathodic arc
FCP	force calibration plot
FD	finite difference
FE	finite element
FEM	finite element method/modeling

FET	field-effect transistor
FFM	friction force microscope/microscopy
FIB	focused ion beam
FID	free induction decay
FIM	field-ion microscope/microscopy
FKT	Frenkel-Kontorova-Tomlinson
FM	frequency modulation
FM-AFM	frequency modulation AFM
FMEA	failure mode effect analysis
FMM	force modulation mode
FM-SFM	frequency-modulation SFM
FS	force spectroscopy

G

G	guanine
GIO	grazing impact oscillator
GMR	giant magnetoresistance
GOD	glucose oxidase
Gox	flavoenzyme glucose oxidase

H

HARMEMS	high-aspect-ratio MEMS
HDD	hard disk drive
HF	hydrofluoric acid
HOP	highly oriented pyrolytic
HOPG	highly oriented pyrolytic graphite
HPMA	hydroxyl polymethacrylamide
HtBDC	hexa-tert-butyl-decacyclene
HTCS	high temperature superconductivity

I

IBD	ion beam deposition
IC	integrated circuit
ICAM-1	intercellular adhesion molecule-1
IFM	interfacial force microscope
ISE	indentation size effect
ITO	indium tin oxide

J

JKR	Johnson–Kendall–Roberts

K

KPFM	Kelvin probe force microscopy

L

LB	Langmuir–Blodgett
LCC	leadless chip carrier
LDOS	local density of states
LEDs	light-emitting diodes
LFA-1	leukocyte function-associated antigen-1
LFM	lateral force microscope
LFT	linear fractional transformation
LN	liquid nitrogen
LPCVD	low pressure chemical vapor deposition
LQG	linear quadratic Gaussian
LTR	loop-transfer recovery
LTSPM	low-temperature SPM
LTSTS	low-temperature scanning tunneling spectroscopy
LVDT	linear variable differential transformers

M

MAP	manifold absolute pressure
MD	molecular dynamics
MDS	molecular dynamics simulation
ME	metal evaporated
MEMS	microelectromechanical systems
MFM	magnetic field microscope/microscopy
MHA	16-mercaptohexadecanoic acid thiol
MHC	major histocompatibility complex
MHD	magnetohydrodynamic
MIM	metal/insulator/metal
MLE	maximum likelihood estimator
MOCVD	metalorganic CVD
MP	metal particle
MPTMS	mercaptopropyltrimethoxysilane
MRAM	magnetoresistive RAM
MRFM	magnetic resonance force microscopy
MRFM	molecular recognition force microscopy
MRI	magnetic resonance imaging
MTTF	mean time to failure
MWCNT	multiwall carbon nanotube

N

NA	nucleic acid
NBMN	3-nitrobenzal malonitrile
NC-AFM	noncontact atomic force microscopy
NCS	neocarzinostatin
NEMS	nanoelectromechanical systems
NMP	no moving part
NNI	National Nanotechnology Initiative
NSOM	near-field scanning optical microscope/microscopy
NTA	nitrilotriacetate
nTP	nanotransfer printing
NVRAM	nonvolatile random access memories

O

ODD	optical disk drives
OMVPE	organometallic vapor phase epitaxy
OT	optical tweezers
OTE	octadecyltrimethoxysilane
OTS	octadecyltrichlorosilane

OUM	Ovonyx unified memory

P

P/W	power to weight
PA	plasminogen activator
PAMAM	poly(amido) amine
PAPP	p-aminophenyl phosphate
PC	ploycarbonate
PC-RAM	phase change RAM
pDA	1,4-phenylenediamine
PDMS	polydimethylsiloxane
PDP	2-pyridyldithiopropionyl
PE	polyethylene
PECVD	plasma enhanced CVD
PEG	poly(ethylene glycol)
PES	photoemission spectroscopy
PES	position error signal
PET	poly(ethylene terephthalate)
PFDA	perfluorodecanoic acid
PFPE	perfluoropolyether
PL	photoluminescence
PMMA	poly(methylmethacrylate)
PS	polystyrene
PSG	phosphorus-doped glass
PSGL-1	P-selection glycoprotein ligand-1
PTFE	polytetrafluoroethylene
PZT	lead zirconate titanate

Q

QCM	quartz-crystal microbalance

R

RES	reticuloendothelial system
RF	radiofrequency
RH	relative humidity
RICM	reflection interface contrast microscopy
RIE	reactive ion etching
RLS	recursive least square algorithm
RPES	relative position error signal
RTP	rapid thermal processing

S

SACA	static advancing contact angle
SAED	selected area electron diffraction
SAM	self-assembling monolayer
SAM	scanning acoustic microscopy
SCM	scanning capitance microscopy
SCPM	scanning chemical potential
SEcM	scanning electrochemical microscopy
SEFM	scanning electrostatic force microscopy
SEM	scanning electron microscope/microscopy
SFA	surface forces apparatus
SFAM	scanning force acoustic microscopy
SFD	shear flow detachment
SFM	scanning force microscopy
SFS	scanning force spectroscopy
SICM	scanning ion conductance microscopy
SIMO	single-input–multi-output
SISO	single-input–single-output
SKPM	scanning Kelvin probe microscopy
SLAM	scanning local-acceleration microscopy
SMA	shape memory alloy
SMANCS	S-Methacryl-neocarzinostatin
SMM	scanning magnetic microscopy
SN	scanning nanoindentation
SNOM	scanning near-field optical microscopy
SPM	scanning probe microscopy
sPROMS	structurally programmable microfluidic system
SPS	spark plasma sintering
SRAM	static random access memory
ssDNA	single stranded DNA
SSNA	single-stranded nucleic acid molecule
SThM	scanning thermal microscopy
STM	scanning tunneling microscope/microscopy
SWCNT	single-wall carbon nanotubes
SWNT	single-wall nanotubes

T

T	thymine
TAAs	tumor associated antigens
TEM	transmission electron microscopy
TESP	tapping-mode etched silicon probe
TGA	thermo-gravimetric analysis
TIRM	total internal reflection microscopy
TMAH	tetramethyl ammonium hydroxide
TMAH	tetramethyl-aluminium hydroxide
TP	track pitch
TPI	tracks per inch
TRM/TMR	track mis-registration
T-SLAM	variable-temperature SLAM
TTF	tetrathiofulvane

U

UHV	ultrahigh vacuum

V

VCM	voice coil motor
VCO	voltage-controlled oscillator
VLS	vapor-liquid-solid
VLSI	very large-scale integration

X

XRD	X-ray diffraction

1. Introduction to Nanotechnology

A biological system can be exceedingly small. Many of the cells are very tiny, but they are very active; they manufacture various substances; they walk around; they wiggle; and they do all kinds of marvelous things – all on a very small scale. Also, they store information. Consider the possibility that we too can make a thing very small which does what we want – that we can manufacture an object that maneuvers at that level.

(From the talk "There's Plenty of Room at the Bottom", delivered by Richard P. Feynman at the annual meeting of the American Physical Society at the California Institute of Technology, Pasadena, CA, on December 29, 1959.)

by B. Bhushan

1.1 Background and Definition of Nanotechnology 1
1.2 Why Nano? 2
1.3 Lessons from Nature 2
1.4 Applications in Different Fields 3
1.5 Reliability Issues of MEMS/NEMS 4
1.6 Organization of the Handbook 5
References 5

1.1 Background and Definition of Nanotechnology

On Dec. 29, 1959, at the California Institute of Technology, Nobel Laureate Richard P. Feynman gave a talk at the annual meeting of the American Physical Society that has become one of the twentieth century's classic science lectures, titled "There's Plenty of Room at the Bottom" [1.1]. He presented a technological vision of extreme miniaturization several years before the word "chip" became part of the lexicon. He talked about the problem of manipulating and controlling things on a small scale. Extrapolating from known physical laws, Feynman envisioned a technology using the ultimate toolbox of nature, building nanoobjects atom by atom or molecule by molecule. Since the 1980s, many inventions and discoveries in the fabrication of nanoobjects have become a testament to his vision. In recognition of this reality, the National Science and Technology Council (NSTC) of the White House created the Interagency Working Group on Nanoscience, Engineering and Technology (IWGN) in 1998. In a January 2000 speech at the same institute, former President William J. Clinton talked about the exciting promise of nanotechnology and, more generally, the importance of expanding research in nanoscale science and technology. Later that month, he announced in his State of the Union Address an ambitious $497 million federal, multi-agency National Nanotechnology Initiative (NNI) in the fiscal year 2001 budget, and made it a top science and technology priority [1.2, 3]. The objective of this initiative was to form a broad-based coalition in which academe, the private sector, and local, state, and federal governments would work together to push the envelope of nanoscience and nanoengineering to reap nanotechnology's potential social and economic benefits.

Nanotechnology literally means any technology performed on a nanoscale that has applications in the real world. Nanotechnology encompasses the production and application of physical, chemical, and biological systems at scales ranging from individual atoms or molecules to submicron dimensions, as well as the integration of the resulting nanostructures into larger systems. Nanotechnology is likely to have a profound impact on our economy and society in the early twenty-first century, comparable to that of semiconductor technology, information technology, or cellular and molecular biology. Science and technology research in nanotechnology promises breakthroughs in such areas as materials and manufacturing, nanoelectronics, medicine and healthcare, energy, biotechnology, information technology, and national security. It is widely felt that nanotechnology will be the next industrial revolution.

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

Nanometer-scale features are mainly built up from their elemental constituents. Chemical synthesis – the spontaneous self-assembly of molecular clusters (molecular self-assembly) from simple reagents in solution – or biological molecules (e.g., DNA) are used as building blocks for the production of three-dimensional nanostructures, including quantum dots (nanocrystals) of arbitrary diameter (about 10 to 10^5 atoms). A variety of vacuum deposition and nonequilibrium plasma chemistry techniques are used to produce layered nanocomposites and nanotubes. Atomically controlled structures are produced using molecular beam epitaxy and organo-metallic vapor phase epitaxy. Micro- and nanosystem components are fabricated using top-down lithographic and nonlithographic fabrication techniques and range in size from micro- to nanometers. Continued improvements in lithography for use in the production of nanocomponents have resulted in line widths as small as 10 nanometers in experimental prototypes. The nanotechnology field, in addition to the fabrication of nanosystems, provides the impetus to development of experimental and computational tools.

The micro- and nanosystems include micro/nanoelectromechanical systems (MEMS/NEMS) (e.g., sensors, actuators, and miniaturized systems comprising sensing, processing, and/or actuating functions), micromechatronics, optoelectronics, microfluidics, and systems integration. These systems can sense, control, and activate on the micro/nanoscale and function individually or in arrays to generate effects on the macroscale. The microsystems market in 2000 was about $ 15 billion, and, with a projected 10–20 % annual growth rate, it is expected to increase to more than $ 100 billion by the end of this decade. The nanosystems market in 2001 was about $ 100 million and the integrated nanosystems market is expected to be more than $ 25 billion by the end of this decade. Due to the enabling nature of these systems, and because of the significant impact they can have on the commercial and defense applications, venture capitalists, industry, as well as the federal government have taken a special interest in nurturing growth in this field. Micro- and nanosystems are likely to be the next logical step in the "silicon revolution."

1.2 Why Nano?

The discovery of novel materials, processes, and phenomena at the nanoscale, as well as the development of new experimental and theoretical techniques for research provide fresh opportunities for the development of innovative nanosystems and nanostructured materials. Nanosystems are expected to find various unique applications. Nanostructured materials can be made with unique nanostructures and properties. This field is expected to open new venues in science and technology.

1.3 Lessons from Nature

Nanotechnology is a new word, but it is not an entirely new field. Nature has many objects and processes that function on a micro- to nanoscale [1.2, 4]. The understanding of these functions can guide us in imitating and producing nanodevices and nanomaterials.

Billions of years ago, molecules began organizing themselves into the complex structures that could support life. Photosynthesis harnesses solar energy to support plant life. Molecular ensembles are present in plants, which include light harvesting molecules, such as chlorophyll, arranged within the cells on the nanometer to micrometer scales. These structures capture light energy, and convert it into the chemical energy that drives the biochemical machinery of plant cells. Live organs use chemical energy in the body. The flagella, a type of bacteria, rotates at over 10,000 RPM [1.5]. This is an example of a biological molecular machine. The flagella motor is driven by the proton flow caused by the electrochemical potential differences across the membrane. The diameter of the bearing is about 20–30 nm, with an estimated clearance of about 1 nm.

In the context of tribology, some biological systems have anti-adhesion surfaces. First, many plant leaves (such as lotus leaf) are covered by a hydrophobic cuticle, which is composed of a mixture of large hydrocarbon molecules that have a strong hydrophobia. Second, the surface is made of a unique roughness distribution [1.6, 7]. It has been reported that for some leaf surfaces, the roughness of the hydrophobic leaf surface decreases wetness, which is reflected in a greater contact angle of water droplets on such surfaces.

1.4 Applications in Different Fields

Science and technology continue to move forward in making the fabrication of micro/nanodevices and systems possible for a variety of industrial, consumer, and biomedical applications. A range of MEMS devices have been produced, some of which are commercially used [1.4, 8–12]. A variety of sensors are used in industrial, consumer, and biomedical applications. Various microstructures or microcomponents are used in micro-instruments and other industrial applications, such as micromirror arrays. Two of the largest "killer" industrial applications are accelerometers (about 85 million units in 2002) and digital micromirror devices (about $400 million in sales in 2001). Integrated capacitive-type, silicon accelerometers have been used in airbag deployment in automobiles since 1991 [1.13, 14]. Accelerometer technology was about a billion-dollar-a-year industry in 2001, dominated by Analog Devices followed by Motorola and Bosch. Commercial digital light processing (DLP) equipment using digital micromirror devices (DMD) were launched in 1996 by Texas Instruments for digital projection displays in portable and home theater projectors, as well as table-top and projection TVs [1.15, 16]. More than 1.5 million projectors were sold before 2002. Other major industrial applications include pressure sensors, inkjet printer heads, and optical switches. Silicon-based piezoresistive pressure sensors for manifold absolute pressure sensing for engines were launched in 1991 by NovaSensor, and their annual sales were about 25 million units in 2002. Annual sales of inkjet printer heads with microscale functional components were about 400 million units in 2002. Capacitive pressure sensors for tire pressure measurements were launched by Motorola. Other applications of MEMS devices include chemical sensors; gas sensors; infrared detectors and focal plane arrays for earth observations; space science and missile defense applications; pico-satellites for space applications; and many hydraulic, pneumatic, and other consumer products. MEMS devices are also being pursued in magnetic storage systems [1.17], where they are being developed for super-compact and ultra-high recording-density magnetic disk drives. Several integrated head/suspension microdevices have been fabricated for contact recording applications [1.18, 19]. High-bandwidth, servo-controlled microactuators have been fabricated for ultrahigh track-density applications, which serve as the fine-position control element of a two-stage, coarse/fine servo system, coupled with a conventional actuator [1.20–23]. Millimeter-sized wobble motors and actuators for tip-based recording schemes have also been fabricated [1.24].

BIOMEMS are increasingly used in commercial and defense applications (e.g., [1.4,25–28]). Applications of BIOMEMS include biofluidic chips (otherwise known as microfluidic chips, bioflips, or simply biochips) for chemical and biochemical analyses (biosensors) in medical diagnostics (e.g., DNA, RNA, proteins, cells, blood pressure and assays, and toxin identification) and implantable pharmaceutical drug delivery. The biosensors, also referred to as lab-on-a-chip, integrate sample handling, separation, detection, and data analysis onto one platform. Biosensors are designed to either detect a single or class of (bio)chemicals or system-level analytical capabilities for a broad range of (bio)chemical species known as micro total analysis systems (μTAS). The chips rely on microfluidics and involve the manipulation of tiny amounts of fluids in microchannels using microvalves for various analyses. The test fluid is pumped into the chip generally using an external pump for analyses. Some chips have been designed with an integrated electrostatically actuated diaphragm-type micropump. Silicon-based, disposable blood-pressure sensor chips were introduced in the early 1990s by NovaSensor for blood pressure monitoring (about 20 million units in 2002). A variety of biosensors are manufactured by various companies, including ACLARA, Agilent Technologies, Calipertech, and I-STAT.

After the tragedy of Sept. 11, 2001, concern over biological and chemical warfare has led to the development of handheld units with bio- and chemical sensors for the detection of biological germs, chemical or nerve agents, mustard agents, and chemical precursors to protect subways, airports, the water supply, and the population [1.29].

Other BIOMEMS applications include minimal invasive surgery, such as endoscopic surgery, laser angioplasty, and microscopic surgery. Implantable artificial organs can also be produced.

Micro-instruments and micro-manipulators are used to move, position, probe, pattern, and characterize nanoscale objects and nanoscale features. Miniaturized analytical equipment includes gas chromatography and mass spectrometry. Other instruments include micro-STM, where STM stands for scanning tunneling microscope.

Examples of NEMS include nanocomponents, nanodevices, nanosystems, and nanomaterials, such as

microcantilever with integrated sharp nanotips for STM and atomic force microscopy (AFM), AFM array (millipede) for data storage, AFM tips for nanolithography, dip-pen nanolithography for printing molecules, biological (DNA) motors, molecular gears, molecularly thick films (e.g., in giant magneto-resistive or GMR heads and magnetic media), nanoparticles, (e.g., nanomagnetic particles in magnetic media), nanowires, carbon nanotubes, quantum wires (QWRs), quantum boxes (QBs), and quantum transistors [1.30–34]. BIONEMS include nanobiosensors – a microarray of silicon nanowires, roughly a few nm in size, to selectively bind and detect even a single biological molecule, such as DNA or protein, by using nanoelectronics to detect the slight electrical charge caused by such binding, or a microarray of carbon nanotubes to electrically detect glucose, implantable drug-delivery devices – e.g., micro/nanoparticles with drug molecules encapsulated in functionized shells for a site-specific targeting application, and a silicon capsule with a nanoporous membrane filled with drugs for long term delivery, nanodevices for sequencing single molecules of DNA in the Human Genome Project, cellular growth using carbon nanotubes for spinal cord repair, nanotubes for nanostructured materials for various applications, such as spinal fusion devices, organ growth, and growth of artificial tissues using nanofibers.

Nanoelectronics can be used to build computer memory, using individual molecules or nanotubes to store bits of information, as well as molecular switches, molecular or nanotube transistors, nanotube flat-panel displays, nanotube integrated circuits, fast logic gates, switches, nanoscopic lasers, and nanotubes as electrodes in fuel cells.

1.5 Reliability Issues of MEMS/NEMS

There is an increasing need for a multidisciplinary, system-oriented approach to manufacturing micro/nanodevices that function reliably. This can only be achieved through the cross-fertilization of ideas from different disciplines and the systematic flow of information and people among research groups. Common potential failure mechanisms for MEMS/NEMS that need to be addressed in order to increase reliability are: adhesion, friction, wear, fracture, fatigue, and contamination. Surface micro/nanomachined structures often include smooth and chemically active surfaces. Due to the large surface area to volume ratio in MEMS/NEMS, they are particularly prone to stiction (high static friction) as part of normal operation. Fracture occurs when the load on a microdevice is greater than the strength of the material. Fracture is a serious reliability concern, particularly for the brittle materials used in the construction of these components, since it can immediately, or eventually, lead to catastrophic failures. Additionally, debris can be formed from the fracturing of microstructures, leading to other failure processes. For less brittle materials, repeated loading over a long period causes fatigue that would also lead to the breaking and fracturing of the device. In principle, this failure mode is relatively easy to observe and simple to predict. However, the materials properties of thin films are often not known, making fatigue predictions prone to error.

Many MEMS/NEMS devices operate near their thermal dissipation limit. They may encounter hot spots that can cause failures, particularly in weak structures such as diaphragms or cantilevers. Thermal stressing and relaxation caused by thermal variations can create material delamination and fatigue in cantilevers. In large temperature changes, as experienced in the space environment, bimetallic beams will also experience warping due to mismatched coefficients of thermal expansion. Packaging has been a big problem. The contamination, which probably happens in packaging and during storage, can also strongly influence the reliability of MEMS/NEMS. For example, a particulate dust landed on one of the electrodes of a comb drive can cause catastrophic failure. There are no MEMS/NEMS fabrication standards, which makes it difficult to transfer fabrication steps in MEMS/NEMS between foundaries.

Obviously, studies of determination and suppression of active failure mechanisms affecting this new and promising technology are critical to the high reliability of MEMS/NEMS and are determining factors in successful practical application.

Mechanical properties are known to exhibit a dependence on specimen size. Mechanical property evaluation of nanometer-scaled structures is carried out to help design reliable systems, since good mechanical properties are of critical importance in such applications. Some of the properties of interest are: Young's mod-

ulus of elasticity, hardness, bending strength, fracture toughness, and fatigue life. Finite element modeling is carried out to study the effects of surface roughness and scratches on stresses in nanostructures. When nanostructures are smaller than a fundamental physical length scale, conventional theory may no longer apply, and new phenomena may emerge. Molecular mechanics is used to simulate the behavior of a nano-object.

1.6 Organization of the Handbook

The handbook integrates knowledge from the fabrication, mechanics, materials science, and reliability points of view. Organization of the book is straightforward. The handbook is divided into six parts. This first part introduces the nanotechnology field, including an introduction to nanostructures, micro/nanofabrication and, micro/nanodevices. The second part introduces scanning probe microscopy. The third part provides an overview of nanotribology and nanomechanics, which will prepare the reader to understand the tribology and mechanics of industrial applications. The fourth part provides an overview of molecularly thick films for lubrication. The fifth part focuses on industrial applications and microdevice reliability. And the last part focuses on the social and ethical implications of nanotechnology.

References

1.1 R. P. Feynmann: There's plenty of room at the bottom, Eng. Sci. **23** (1960) 22–36, and www.zyvex.com/nanotech/feynman.html (1959)

1.2 I. Amato: Nanotechnology, www.ostp.gov/nstc/html/iwgn/iwgn.public.brochure/welcome.htm or www.nsf.gov/home/crssprgm/nano/nsfnnireports.htm (2000)

1.3 Anonymous: National nanotechnology initiative, www.ostp.gov/nstc/html/iwgn.fy01budsuppl/nni.pdf or www.nsf.gov/home/crssprgm/nano/nsfnnireports.htm (2000)

1.4 I. Fujimasa: *Micromachines: A New Era in Mechanical Engineering* (Oxford Univ. Press, Oxford 1996)

1.5 C. J. Jones, S. Aizawa: The bacterial flagellum and flagellar motor: Structure, assembly, and functions, Adv. Microb. Physiol. **32** (1991) 109–172

1.6 V. Bergeron, D. Quere: Water droplets make an impact, Phys. World **14** (May 2001) 27–31

1.7 M. Scherge, S. Gorb: *Biological Micro- and Nanotribology* (Springer, Berlin, Heidelberg 2000)

1.8 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

1.9 G. T. A. Kovacs: *Micromachined Transducers Sourcebook* (WCB McGraw-Hill, Boston 1998)

1.10 S. D. Senturia: *Microsystem Design* (Kluwer, Boston 2001)

1.11 T. R. Hsu: *MEMS and Microsystems* (McGraw-Hill, Boston 2002)

1.12 M. Madou: *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd edn. (CRC, Boca Raton 2002)

1.13 T. A. Core, W. K. Tsang, S. J. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. **36** (October 1993) 39–47

1.14 J. Bryzek, K. Peterson, W. McCulley: Micromachines on the march, IEEE Spectrum (May 1994) 20–31

1.15 L. J. Hornbeck, W. E. Nelson: Bistable deformable mirror device, OSA Technical Digest **8** (1988) 107–110

1.16 L. J. Hornbeck: A digital light processing(tm) update – Status and future applications (invited), Proc. Soc. Photo-Opt. Eng. **3634** (1999) 158–170

1.17 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)

1.18 H. Hamilton: Contact recording on perpendicular rigid media, J. Mag. Soc. Jpn. **15** (Suppl. S2) (1991) 481–483

1.19 T. Ohwe, Y. Mizoshita, S. Yonoeka: Development of integrated suspension system for a nanoslider with an MR head transducer, IEEE Trans. Magn. **29** (1993) 3924–3926

1.20 D. K. Miu, Y. C. Tai: Silicon micromachined scaled technology, IEEE Trans. Industr. Electron. **42** (1995) 234–239

1.21 L. S. Fan, H. H. Ottesen, T. C. Reiley, R. W. Wood: Magnetic recording head positioning at very high track densities using a microactuator-based, two-stage servo system, IEEE Trans. Ind. Electron. **42** (1995) 222–233

1.22 D. A. Horsley, M. B. Cohn, A. Singh, R. Horowitz, A. P. Pisano: Design and fabrication of an angular

microactuator for magnetic disk drives, J. Microelectromech. Syst. **7** (1998) 141–148

1.23 T. Hirano, L. S. Fan, D. Kercher, S. Pattanaik, T. S. Pan: HDD tracking microactuator and its integration issues, Proc. ASME Int. Mech. Eng. Congress, MEMS, New York 2000, ed. by A. P. Lee, J. Simon, F. K. Foster, R. S. Keynton (ASME, New York 2000) 449–452

1.24 L. S. Fan, S. Woodman: Batch fabrication of mechanical platforms for high-density data storage, 8th Int. Conf. Solid State Sensors and Actuators (Transducers '95)/Eurosensors IX, Stockholm (June, 1995) 434–437

1.25 P. Gravesen, J. Branebjerg, O. S. Jensen: Microfluidics – A review, J. Micromech. Microeng. **3** (1993) 168–182

1.26 C. Lai Poh San, E. P. H. Yap (Eds.): *Frontiers in Human Genetics* (World Scientific, Singapore 2001)

1.27 C. H. Mastrangelo, H. Becker (Eds.): *Microfluidics and BioMEMS*, Proc. SPIE **4560** (SPIE, Bellingham 2001)

1.28 H. Becker, L. E. Locascio: Polymer microfluidic devices, Talanta **56** (2002) 267–287

1.29 M. Scott: MEMS and MOEMS for national security applications, , Reliability, Testing, and Characterization of MEMS/MOEMS II, Proc. SPIE **4980** (SPIE, Bellingham 2003)

1.30 K. E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)

1.31 G. Timp (Ed.): *Nanotechnology* (Springer, Berlin, Heidelberg 1999)

1.32 E. A. Rietman: *Molecular Engineering of Nanosystems* (Springer, Berlin, Heidelberg 2001)

1.33 H. S. Nalwa (Ed.): *Nanostructured Materials and Nanotechnology* (Academic, San Diego 2002)

1.34 W. A. Goddard, D. W. Brenner, S. E. Lyshevski, G. J. Iafrate: *Handbook of Nanoscience, Engineering, and Technology* (CRC, Boca Raton 2003)

Part A Nanostruc

Part A Nanostructures, Micro/Nanofabrication, and Micro/Nanodevices

2 Nanomaterials Synthesis and Applications: Molecule-Based Devices
Françisco M. Raymo, Coral Gables, USA

3 Introduction to Carbon Nanotubes
Marc Monthioux, Toulouse, France
Philippe Serp, Toulouse, France
Emmanuel Flahaut, Toulouse, France
Manitra Razafinimanana, Toulouse, France
Christophe Laurent, Toulouse, France
Alain Peigney, Toulouse, France
Wolfgang Bacsa, Toulouse, France
Jean-Marc Broto, Toulouse, France

4 Nanowires
Mildred S. Dresselhaus, Cambridge, USA
Yu-Ming Lin, Cambridge, USA
Oded Rabin, Cambridge, USA
Marcie R. Black, Cambridge, USA
Gene Dresselhaus, Cambridge, USA

5 Introduction to Micro/Nanofabrication
Babak Ziaie, Minneapolis, USA
Antonio Baldi, Barcelona, Spain
Massood Z. Atashbar, Kalamazoo, USA

6 Stamping Techniques for Micro and Nanofabrication: Methods and Applications
John A. Rogers, Urbana, USA

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

8 MEMS/NEMS Devices and Applications
Darrin J. Young, Cleveland, USA
Christian A. Zorman, Cleveland, USA
Mehran Mehregany, Cleveland, USA

9 Microfluidics and Their Applications to Lab-on-a-Chip
Chong H. Ahn, Cincinnati, USA
Jin-Woo Choi, Baton Rouge, USA

10 Therapeutic Nanodevices
Stephen C. Lee, Columbus, USA
Mark Ruegsegger, Columbus, USA
Philip D. Barnes, Columbus, USA
Bryan R. Smith, Columbus, USA
Mauro Ferrari, Columbus, USA

2. Nanomaterials Synthesis and Applications: Molecule-Based Devices

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

2.1 **Chemical Approaches to Nanostructured Materials** 10
2.1.1 From Molecular Building Blocks to Nanostructures 10
2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins 10
2.1.3 Chemical Synthesis of Artificial Nanostructures 12
2.1.4 From Structural Control to Designed Properties and Functions 12

2.2 **Molecular Switches and Logic Gates** 14
2.2.1 From Macroscopic to Molecular Switches 15
2.2.2 Digital Processing and Molecular Logic Gates 15
2.2.3 Molecular AND, NOT, and OR Gates 16
2.2.4 Combinational Logic at the Molecular Level 17
2.2.5 Intermolecular Communication 18

2.3 **Solid State Devices** 22
2.3.1 From Functional Solutions to Electroactive and Photoactive Solids 22
2.3.2 Langmuir–Blodgett Films 23
2.3.3 Self-Assembled Monolayers 27
2.3.4 Nanogaps and Nanowires 31

2.4 **Conclusions and Outlook** 35

References 36

2.1 Chemical Approaches to Nanostructured Materials

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

2.1.1 From Molecular Building Blocks to Nanostructures

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

2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins

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

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

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

Nature's operating principles to fabricate nanostructures are not limited to nucleic acids. Proteins are also built joining simple molecular building blocks, the amino acids, by strong covalent bonds [2.1]. More precisely, nature relies on 20 amino acids differing in their side chains to assemble linear polymers, called polypeptides, incorporating an extended backbone of robust [C−N] and [C−C] bonds (Fig. 2.2a). For a single polymer strand of 100 repeating amino acid units, a total of 20^{100} unique combinations of polypeptide sequences are possible. Considering that proteins can incorporate more than one polypetide chain with over 4,000 amino acid residues each, it is obvious that nature can assemble an enormous number of different biomolecules relying on the same fabrication strategy and a relatively small pool of building blocks.

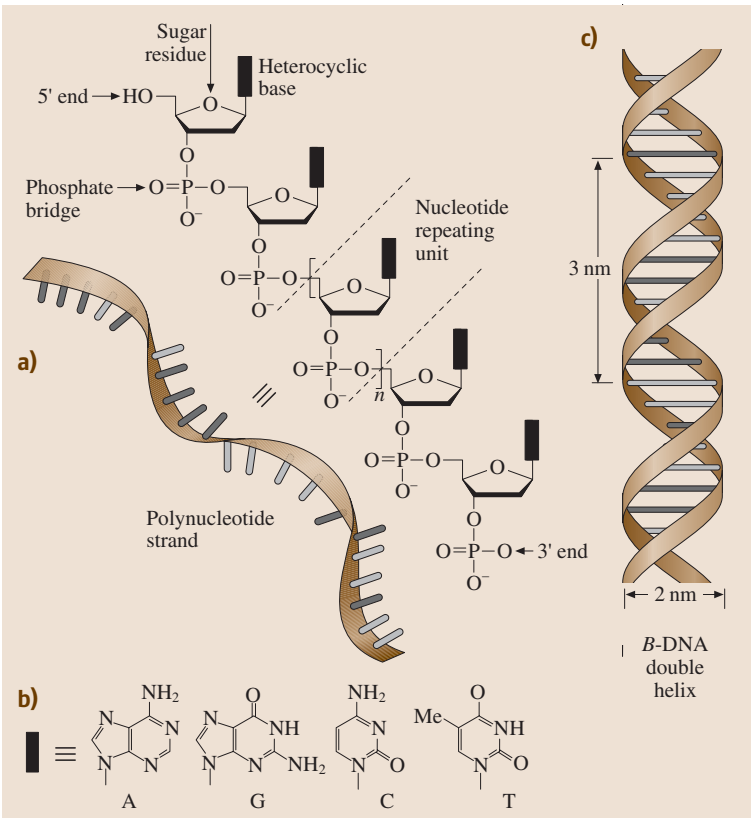


Fig. 2.1a–c A polynucleotide strand **(a)** incorporates alternating phosphate and sugar residues joined by covalent bonds. Each sugar carries one of four heterocyclic bases **(b)**. Noncovalent interactions between complementary bases in two independent polynucleotide strands encourage the formation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

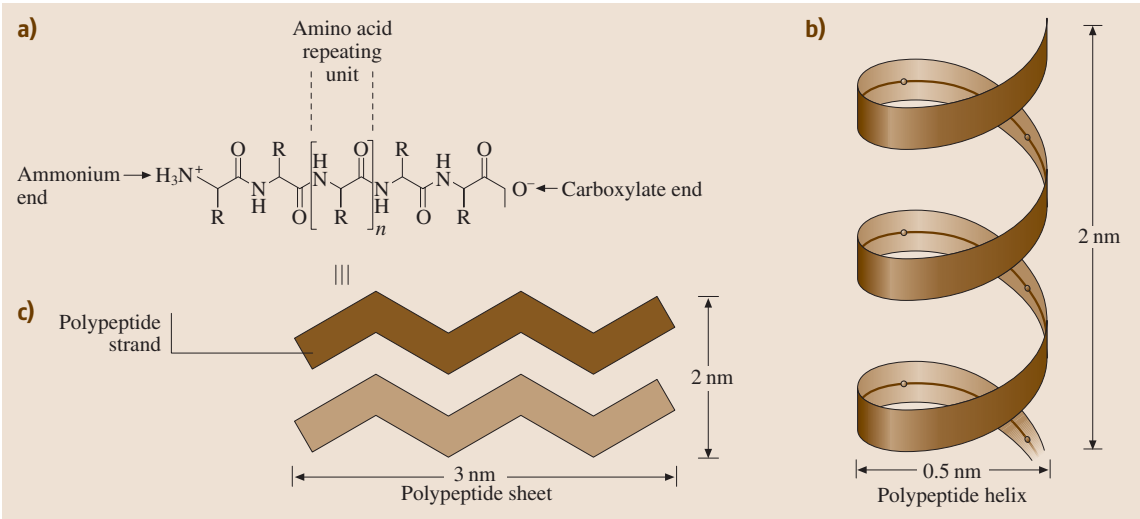


Fig. 2.2a–c A polypeptide strand **(a)** incorporates amino acid residues differing in their side chains and joined by covalent bonds. Hydrogen bonding interactions curl a single polypeptide strand into a helical arrangement **(b)** or lock pairs of strands into nanoscaled sheets **(c)**

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

2.1.3 Chemical Synthesis of Artificial Nanostructures

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

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

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

The chemical synthesis of nanostructures can borrow nature's design criteria as well as its molecular building blocks. Amino acids, the basic components of proteins, can be assembled into artificial macrocycles. In the example of Fig. 2.4, eight amino acid residues are joined through the formation of [C−N] bonds in multiple synthetic steps [2.7]. The resulting covalent backbone defines a circular cavity with a diameter of ca. 0.8 nm [2.8]. In analogy to the polypeptide chains of proteins, the amino acid residues of this artificial oligopeptide can sustain hydrogen bonding interactions. It follows that multiple macrocycles can pile on top of each other to form tubular nanostructures. The walls of the resulting nanotubes are maintained in position by the cooperative action of at least eight primary hydrogen bonding contacts per macrocycle. These noncovalent interactions maintain the mean planes of independent macrocycles in an approximately parallel arrangement with a plane-to-plane separation of ca. 0.5 nm.

2.1.4 From Structural Control to Designed Properties and Functions

The examples in Figs. 2.3 and 2.4 demonstrate that modular building blocks can be assembled into target compounds with precise structural control at the picometer level through programmed sequences of synthetic steps. Indeed, modern chemical synthesis offers access to complex molecules with nanoscaled dimensions and, thus, provides cost-effective strategies for the production

and characterization of billions of engineered nanostructures in parallel. Furthermore, the high degree of structural control is accompanied by the possibility of designing specific properties into the target nanostructures. Electroactive and photoactive components can be integrated chemically into functional molecular machines [2.9]. Extensive electrochemical investigations have demonstrated that inorganic and organic compounds can exchange electrons with macroscopic electrodes [2.10]. These studies have unraveled the processes responsible for the oxidation and reduction of numerous functional groups and indicated viable design criteria to adjust the ability of molecules to accept or donate electrons [2.11]. Similarly, detailed photochemical and photophysical investigations have elucidated the mechanisms responsible for the absorption and emission of photons at the molecular level [2.12]. The vast knowledge established on the interactions between light and molecules offers the opportunity to engineer chromophoric and fluorophoric functional groups with defined absorption and emission properties [2.11, 13].

The power of chemical synthesis to deliver functional molecules is, perhaps, better illustrated by the molecular motor shown in Fig. 2.5. The preparation of this [2]rotaxane requires 12 synthetic steps starting from known precursors [2.14]. This complex molecule incorporates a Ru(II)-trisbipyridine stopper bridged to a linear tetracationic fragment by a rigid triaryl spacer. The other end of the tetracationic portion is terminated by a bulky tetraarylmethane stopper. The bipyridinium unit of this dumbbell-shaped compound is encircled by a macrocyclic polyether. No covalent bonds join the macrocyclic and linear components.

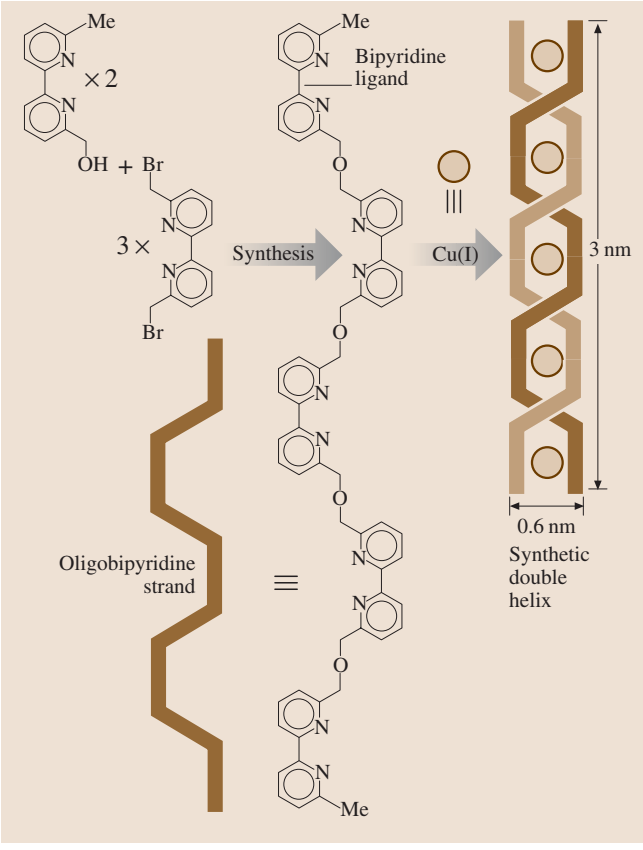


Fig. 2.3 An oligobipyridine strand can be synthesized joining five bipyridine subunits by covalent bonds. The tetrahedral coordination of pairs of bipyridine ligands by Cu(I) ions encourages the assembly two oligobipyridine strands into a double helical arrangement

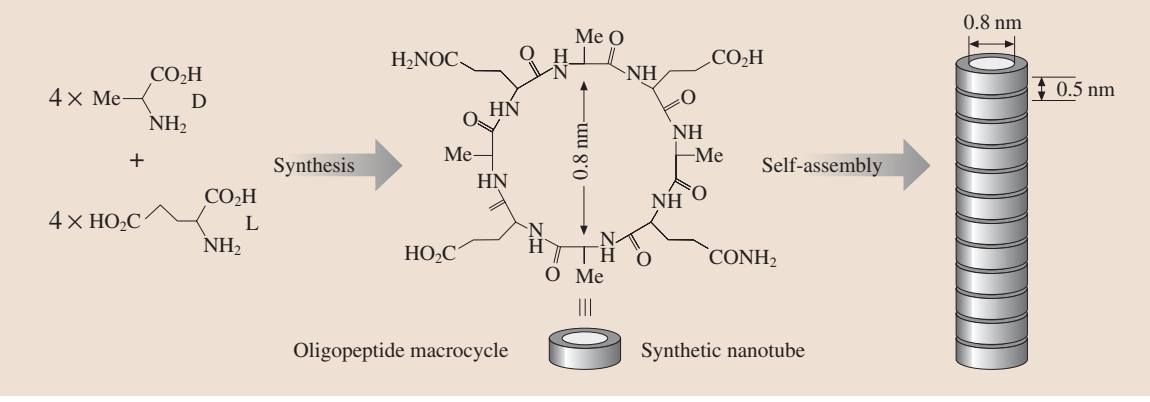


Fig. 2.4 Cyclic oligopeptides can be synthesized joining eight amino acid residues by covalent bonds. The resulting macrocycles self-assemble into nanoscaled tube-like arrays

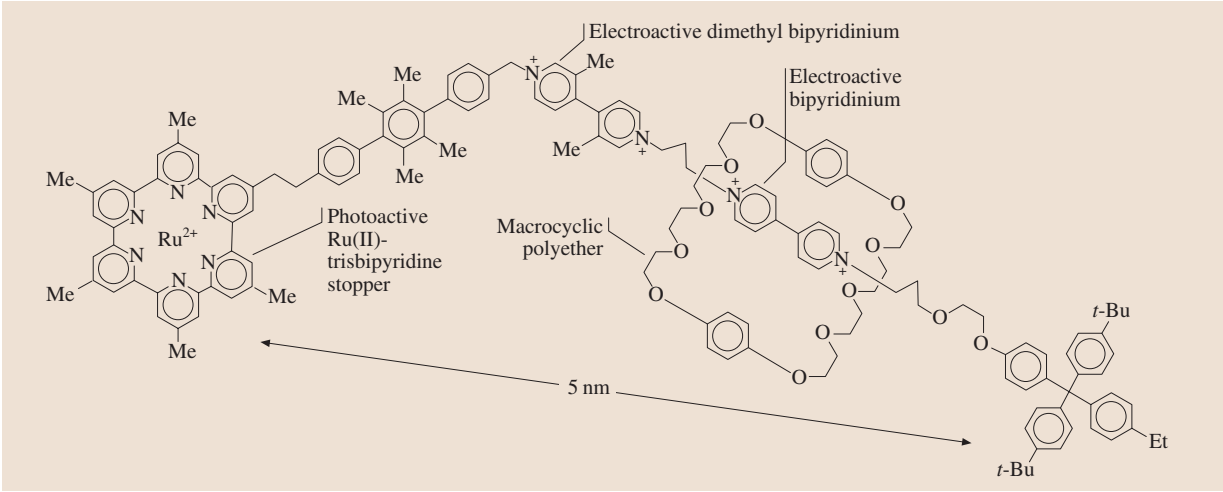


Fig. 2.5 This nanoscaled [2]rotaxane incorporates a photoactive Ru(II)-trisbipyridine stopper and two electroactive bipyridinium units. Photoinduced electron transfer from the photoactive stopper to the encircled electroactive unit forces the macrocyclic polyether to shuttle to the adjacent bipyridinium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

Rather, hydrogen bonding and $[\pi \cdots \pi]$ stacking interactions maintain the macrocyclic polyether around the bipyridinium unit. In addition, mechanical constrains associated with the bulk of the two terminal stoppers prevent the macrocycle to slip off the thread. The approximate end-to-end distance for this [2]rotaxane is ca. 5 nm.

The bipyridinium and the 3,3′-dimethyl bipyridinium units within the dumbbell-shaped component undergo two consecutive and reversible monoelectronic reductions [2.14]. The two methyl substituents on the 3,3′-dimethyl bipyridinium dication make this electroactive unit more difficult to reduce. In acetonitrile, its redox potential is ca. 0.29 V more negative than that of the unsubstituted bipyridinium dication. Under irradiation at 436 nm in degassed acetonitrile, the excitation of the Ru(II)-trisbipyridine stopper is followed by electron transfer to the unsubstituted bipyridinium unit. In the presence of a sacrificial electron donor (triethanolamine) in solution, the photogenerated hole in the photoactive stopper is filled, and undesired back electron transfer is suppressed. The permanent and light-induced reduction of the dicationic bipyridinium unit to a radical cation depresses significantly the magnitude of the noncovalent interactions holding the macrocyclic polyether in position. As a result, the macrocycle shuttles from the reduced unit to the adjacent dicationic 3,3′-dimethyl bipyridinum. After the diffusion of molecular oxygen into the acetonitrile solution, oxidation occurs restoring the dicationic form of the bipyridinium unit and its ability to sustain strong noncovalent bonds. As a result, the macrocyclic polyether shuttles back to its original position. This amazing example of a molecular shuttle reveals that dynamic processes can be controlled reversibly at the molecular level relying on the clever integration of electroactive and photoactive fragments into functional and nanoscaled molecules.

2.2 Molecular Switches and Logic Gates

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

Overall, the switch transduces a mechanical input into an electrical output.

2.2.1 From Macroscopic to Molecular Switches

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

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

Certain organic molecules adjust their structural and electronic properties when stimulated with chemical, electrical, or optical inputs. Generally, the change is accompanied by an electrochemical or spectroscopic response. Overall, these nanostructures transduce input stimulations into detectable outputs and, appropriately, are called molecular switches [2.15, 16]. The chemical transformations associated with these switching processes are often reversible. The chemical system returns to the original state when the input signal is turned off. The interconverting states of a molecular switch can be isomers, an acid and its conjugated base, the oxidized and reduced forms of a redox active molecule, or even the complexed and uncomplexed forms of a receptor [2.9, 15, 16]. The output of a molecular switch can be a chemical, electrical, and/or optical signal that varies in intensity with the interconversion process. For example, changes in absorbance, fluorescence, pH, or redox potential can accompany the reversible transformation of a molecular switch.

2.2.2 Digital Processing and Molecular Logic Gates

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

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

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

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

to a variety of input stimulations producing specific outputs and can, therefore, be exploited to implement logic functions [2.20, 21].

2.2.3 Molecular AND, NOT, and OR Gates

More than a decade ago, researchers proposed a potential strategy to execute logic operations at the molecular level [2.22]. Later, the analogy between molecular switches and logic gates was recognized in a seminal article [2.23], in which it was demonstrated that AND, NOT, and OR operations can be reproduced with fluorescent molecules. The pyrazole derivative **1** (Fig. 2.6) is a molecular NOT gate. It imposes an inverse relation between a chemical input (concentration of H^+) and an optical output (emission intensity). In a mixture of methanol and water, the fluorescence quantum yield of **1** is 0.13 in the presence of only 0.1 equivalents of H^+ [2.23]. The quantum yield drops to 0.003 when the equivalents of H^+ are 1,000. Photoinduced electron transfer from the central pyrazoline unit to the pendant benzoic acid quenches the fluorescence of the protonated form. Thus, a change in H^+ concentration (I) from a low to a high value switches the emission intensity (O) from a high to a low value. The inverse relationship between the chemical input I and the optical output O translates into the truth table of a NOT operation if a positive logic convention (low = 0, high = 1) is applied to both signals. The emission intensity is high (O = 1) when the concentration of H^+ is low (I = 0). The emission intensity is low (O = 0) when the concentration of H^+ is high (I = 1).

The anthracene derivative **2** (Fig. 2.6) is a molecular OR gate. It transduces two chemical inputs

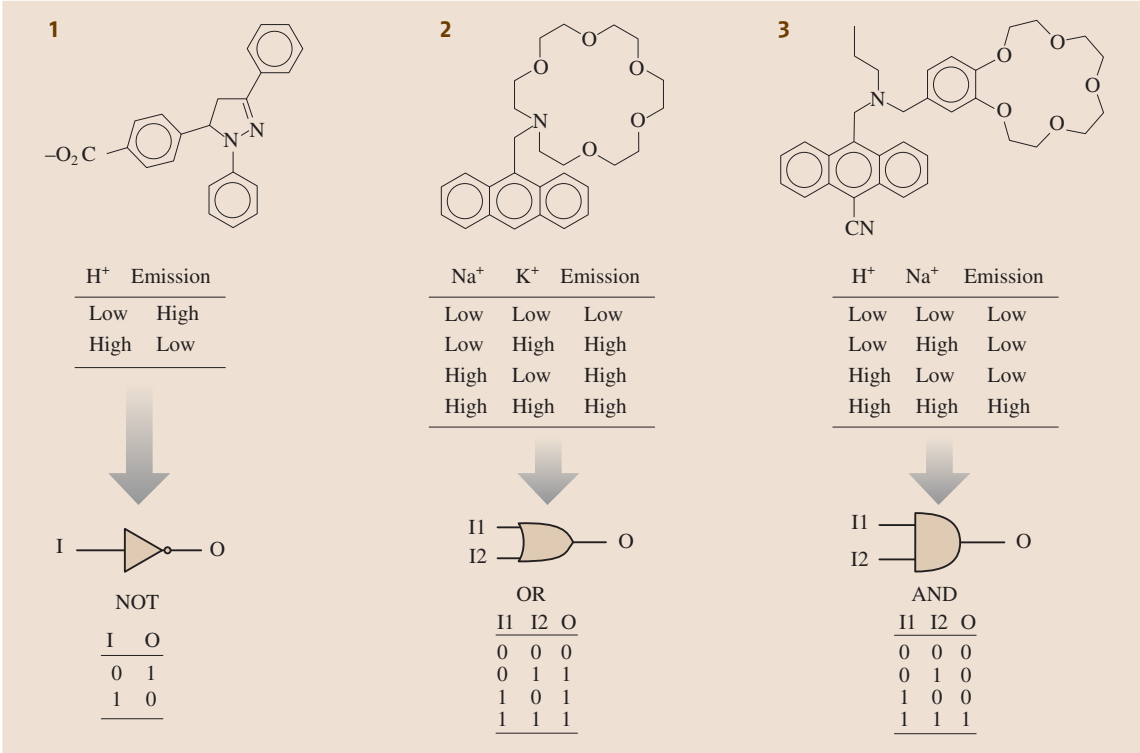


Fig. 2.6 The fluorescence intensity of the pyrazoline derivative **1** is high when the concentration of H^+ is low, and vice versa. The fluorescence intensity of the anthracene derivative **2** is high when the concentration of Na^+ and/or K^+ is high. The emission is low when both concentrations are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

(concentrations of Na^+ and K^+) into an optical output (emission intensity). In methanol, the fluorescence quantum yield is only 0.003 in the absence of metal cations [2.23]. Photoinduced electron transfer from the nitrogen atom of the azacrown fragment to the anthracene fluorophore quenches the emission. After the addition of 1,000 equivalents of either Na^+ or K^+, the quantum yield raises to 0.053 and 0.14, respectively. Similarly, the quantum yield is 0.14 when both metal cations are present in solution. The complexation of one of the two metal cations inside the azacrown receptor depresses the efficiency of the photoinduced electron transfer enhancing the fluorescence. Thus, changes in the concentrations of Na^+ (I1) and/or K^+ (I2) from low to high values switch the emission intensity (O) from a low to a high value. The relationship between the chemical inputs I1 and I2 and the optical output O translates into the truth table of an OR operation if a positive logic convention (low = 0, high = 1) is applied to all signals. The emission intensity is low (O = 0) only when the concentration of Na^+ and K^+ are low (I1 = 0, I2 = 0). The emission intensity is high (O = 1) for the other three input combinations.

The anthracene derivative **3** (Fig. 2.6) is a molecular AND gate. It transduces two chemical inputs (concentrations of H^+ and Na^+) into an optical output (emission intensity). In a mixture of methanol and *iso*-propanol, the fluorescence quantum yield is only 0.011 in the absence of H^+ or Na^+ [2.23]. Photoinduced electron transfer from either the tertiary amino group or the catechol fragment to the anthracene fluorophore quenches the emission. After the addition of either 100 equivalents of H^+ or 1,000 equivalents of Na^+, a modest change of the quantum yield to 0.020 and 0.011, respectively, is observed. Instead, the quantum yield increases to 0.068 when both species are present in solution. The protonation of the amino group and the insertion of the metal cation in the benzocrown ether receptor depress the efficiency of the photoinduced electron transfer processes enhancing the fluorescence. Thus, changes in the concentrations of H^+ (I1) and Na^+ (I2) from low to high values switch the emission intensity (O) from a low to a high value. The relationship between the chemical inputs I1 and I2 and the optical output O translates into the truth table of an AND operation if a positive logic convention (low = 0, high = 1) is applied to all signals. The emission intensity is high (O = 1) only when the concentration of H^+ and Na^+ are high (I1 = 1, I2 = 1). The emission intensity is low (O = 0) for the other three input combinations.

2.2.4 Combinational Logic at the Molecular Level

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

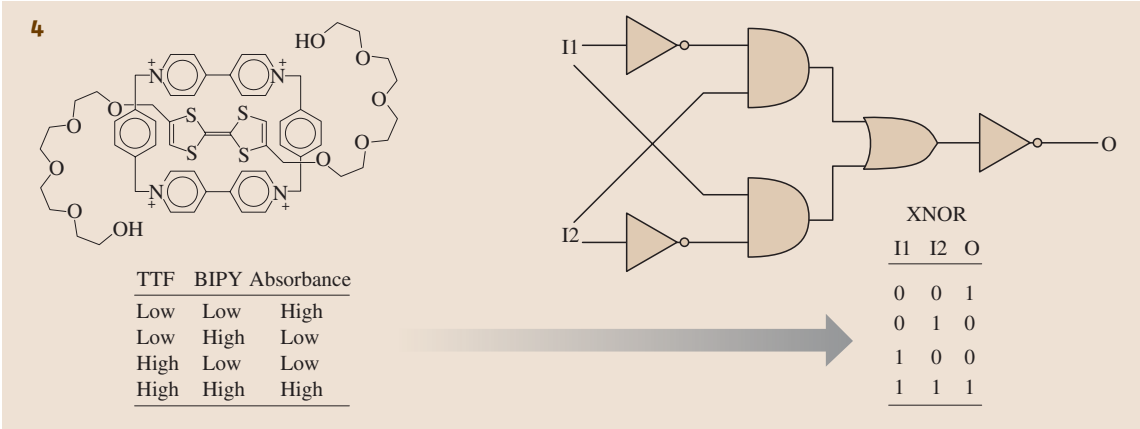


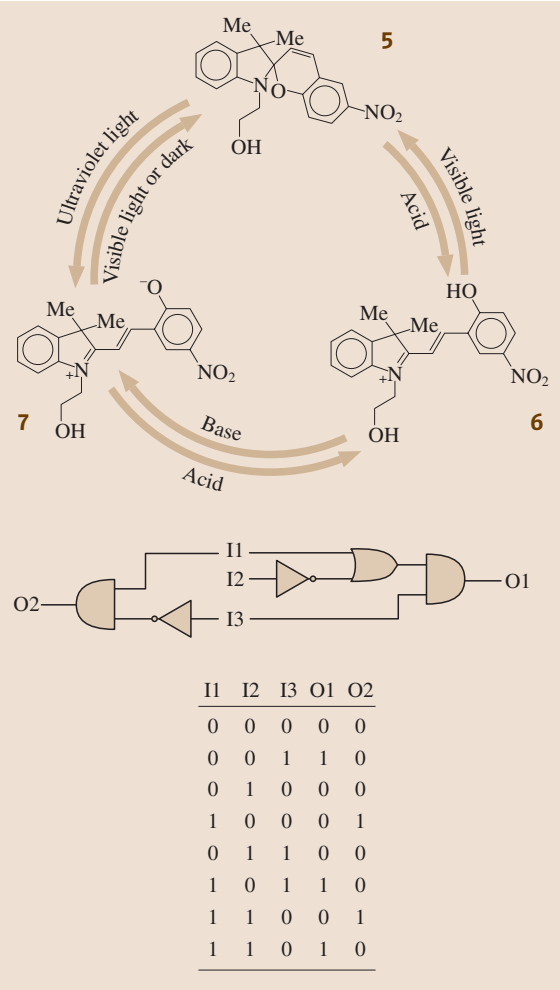
Fig. 2.7 The charge-transfer absorbance of the complex **4** is high when the voltage input addressing the tetrathiafulvalene (TTF) unit is low and that stimulating the bipyridinium (BIPY) units is high and vice versa. If a positive logic convention is applied to the TTF input and to the absorbance output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

is high (I1 = 0, I2 = 0) or vice versa (I1 = 1, I2 = 1). It is important to note that the input string with both I1 and I2 equal to 1 implies that input potentials of +0.5 and −0.3 V are applied simultaneously to a solution containing the supramolecular assembly **4** and not to an individual complex. Of course, the concomitant oxidation of the TTF guest and reduction of the BIPY units in the very same complex would be unrealistic. In bulk solution, instead, some complexes are oxidized while others are reduced, leaving the average solution composition unaffected. Thus, the XNOR operation executed by this supramolecular system is a consequence of bulk properties and not a result of unimolecular signal transduction.

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

2.2.5 Intermolecular Communication

The combinational logic circuits in Figs. 2.7 and 2.8 are arrays of interconnected AND, NOT, and OR operators. The digital communication between these basic logic elements ensures the execution of a sequence of simple logic operations that results in the complex logic function processed by the entire circuit. It follows that



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

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

the logic function of a given circuit can be adjusted altering the number and type of basic gates and their interconnection protocol [2.17]. This modular approach to combinational logic circuits is extremely powerful. Any logic function can be implemented connecting the appropriate combination of simple AND, NOT, and OR gates.

The strategies followed so far to implement complex logic functions with molecular switches are based on the careful design of the chemical system and on the judicious choice of the inputs and outputs [2.20, 21]. A specific sequence of AND, NOT, and OR operations is programmed in a single molecular switch. No digital communication between distinct gates is needed since they are built in the same molecular entity. Though extremely elegant, this strategy does not have the same versatility of a modular approach. A different molecule has to be designed, synthesized, and analyzed every single time a different logic function has to be realized. In addition, the degree of complexity that can be achieved with only one molecular switch is fairly limited. The connection of the input and output terminals of independent molecular AND, NOT, and OR operators, instead, would offer the possibility of assembling any combinational logic circuit from three basic building blocks.

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

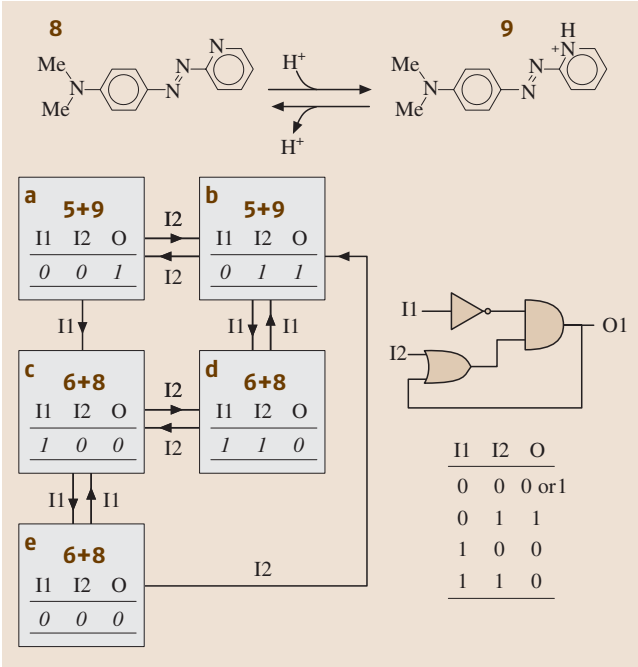
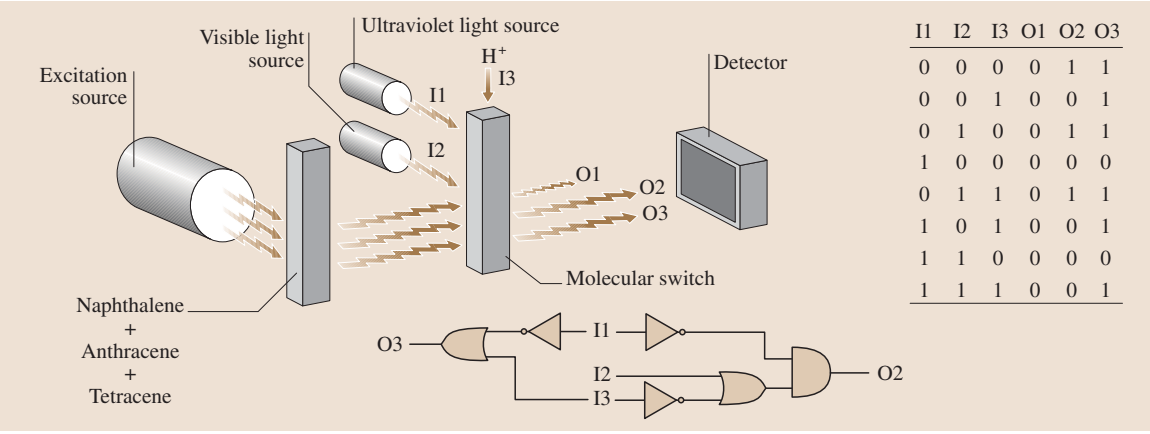


Fig. 2.9 The concentration of H^+ controls the reversible interconversion between the two states **8** and **9**. In response to ultraviolet (I1) and visible (I2) inputs, the three-state molecular switch in Fig. 2.7 modulates the ratio between these two forms and the absorbance (O) of **9** through photoinduced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

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

The logic behavior of the two communicating molecular switches is significantly different from those of the chemical systems illustrated in Figs. 2.6, 2.7, and 2.8 [2.26]. The truth table in Fig. 2.9 lists the four possible combinations of two-digit input strings and the corresponding one-digit output. The output digit O for the input strings 01, 10, and 11 can take only one value. In fact, the input string 01 is transduced into a 1, and the input strings 10 and 11 are converted into 0. Instead, the output digit O for the input string 00 can be either 0 or 1. The sequence of events leading to the input string 00 determines the value of the output. The boxes ***a***–***e*** in Fig. 2.9 illustrates this effect. They correspond to the five three-digit input/output strings. The transformation of one box into any of the other four is achieved in one or two steps by changing the values of I1 and/or I2. In two instances (***a*** and ***b***), the two-state molecular switch is in state **9**, and the output signal is high (O = 1). In the other three cases (***c***, ***d***, and ***e***), the two-state molecular switch is in state **8**, and the output signal is low (O = 0). The strings 000 (***e***) and 001 (***a***) correspond to the first entry of the truth table. They share the same input digits but differ in the output value. The string 000 (***e***) can be obtained only from the string 100 (***c***) varying the value of I1. Similarly, the string 001 (***a***) can be accessed only from the string 011 (***b***) varying the value of I2. In both transformations, the output digit remains unchanged. Thus, the value of O1 in the parent string is memorized and maintained in the daughter string when both inputs become 0. This memory effect is the fundamental operating principle of sequential logic circuits [2.17], which are used extensively to assemble the memory elements of modern microprocessors. The sequential logic circuit equivalent to the truth table of the two communicating molecular switches is also shown in Fig. 2.9. In this circuit, the input data I1 and I2 are combined through NOT, OR, and AND operators. The output of the AND gate O is also an input of the OR gate and controls, together with I1 and I2, the signal transduction behavior.

The other strategy for digital transmission between molecules is based on the communication of optical signals between the three-state molecular switch (Fig. 2.8) and fluorescent compounds [2.27,28]. In the optical network of Fig. 2.10, three optical signals travel from an excitation source to a detector after passing through two quartz cells. The first cell contains an equimolar acetonitrile solution of naphthalene, anthracene, and tetracene. The second cell contains an acetonitrile solu-

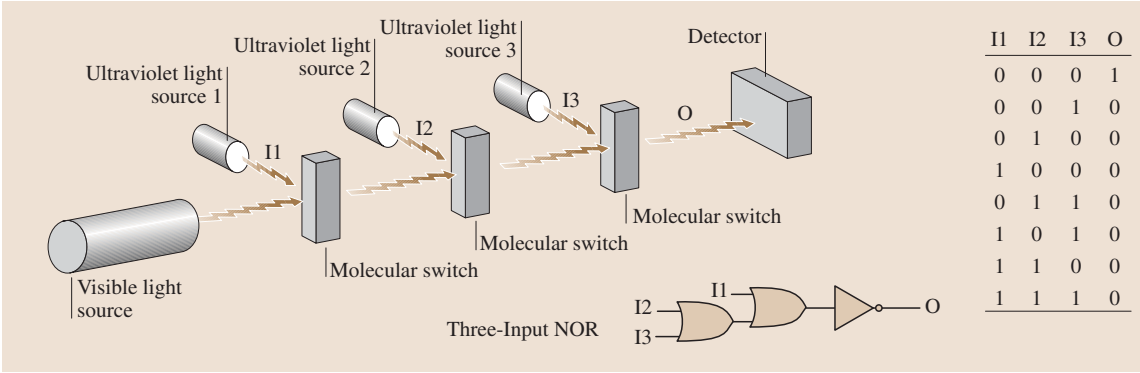


I1	I2	I3	O1	O2	O3
0	0	0	0	1	1
0	0	1	0	0	1
0	1	0	0	1	1
1	0	0	0	0	0
0	1	1	0	1	1
1	0	1	0	0	1
1	1	0	0	0	0
1	1	1	0	0	1

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

tion of the three-state molecular switch. The excitation source sends three consecutive monochromatic light beams to the first cell stimulating the emission of the three fluorophores. The light emitted in the direction perpendicular to the exciting beam reaches the second cell. When the molecular switch is in state **5**, the naphthalene emission at 335 nm is absorbed and a low intensity output (O1) reaches the detector. Instead, the anthracene and tetracene emissions at 401 and 544 nm, respectively, pass unaffected and high intensity outputs (O2 and O3) reach the detector. When the molecular switch is in state **6**, the naphthalene and anthracene emissions are absorbed and only the tetracene emission reaches the detector (O1 = 0, O2 = 0, O3 = 1). When the molecular switch is state **7**, the emission of all three fluorophores is absorbed (O1 = 0, O2 = 0, O3 = 0). The interconversion of the molecular switch between the three states is induced addressing the second cell with ultraviolet (I1), visible (I2) and H^+ (I3) inputs. Thus, three independent optical outputs (O1, O2 and O3) can be modulated stimulating the molecular switch with two optical and one chemical input. The truth table in Fig. 2.10 illustrates the relation between the three inputs (I1, I2 and I3) and the three outputs (O1, O2 and O3), when positive logic conventions are applied to all signals. The equivalent logic circuit shows that all three inputs control the anthracene channel O2, but only I1 and I3 influence the tetracene channel O3. Instead, the intensity of the naphthalene channel O1 is always low, and it is not affected by the three inputs.

The operating principles of the optical network in Fig. 2.10 can be simplified to implement all-optical logic gates. The chemical input inducing the formation of the protonated form **6** of the molecular switch can be eliminated. The interconversion between the remaining two states **5** and **7** can be controlled relying exclusively on ultraviolet inputs. Indeed, ultraviolet irradiation induces the isomerization of the colorless form **5** to the colored species **7**, which reisomerizes to the original state in the dark. Thus, a single ultraviolet source is sufficient to control the switching from **5** to **7** and vice versa. On the basis of these considerations, all-optical NAND, NOR, and NOT gates can be implemented operating sequentially or in parallel from one to three independent switching elements [2.28]. For example, the all-optical network illustrated in Fig. 2.11 is a three-input NOR gate. A monochromatic optical signal travels from a visible source to a detector. Three switching elements are aligned along the path of the traveling light. They are quartz cells containing an acetonitrile solution of the mo-



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

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

lecular switch shown in Fig. 2.8. The interconversion of the colorless form **5** into the purple isomer **7** is induced stimulating the cell with an ultraviolet input. The reisomerization from **7** to **5** occurs spontaneously, as the ultraviolet sources is turned off. Using three distinct ultraviolet sources, the three switching elements can be controlled independently.

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

2.3 Solid State Devices

The fascinating chemical systems illustrated in Figs. 2.6–2.11 demonstrate that logic functions can be implemented relying on the interplay between designed molecules and chemical, electrical and/or optical signals [2.20, 21].

2.3.1 From Functional Solutions to Electroactive and Photoactive Solids

These molecular switches, however, are operated exclusively in solution and remain far from potential applications in information technology at this stage. The integration of liquid components and volatile organic solvents in practical digital devices is hard to envisage. Furthermore, the logic operations executed by these chemical systems rely on bulk addressing. Although the individual molecular components have nanoscaled dimensions, macroscopic collections of them are employed for digital processing. In some instances, the operating principles cannot even be scaled down to the unimolecular level. Often bulk properties are responsible for signal transduction. For example, a single fluorescent compound **2** cannot execute an OR operation. Its azacrown appendage can accommodate only one metal cation. As a result, an individual molecu-

lar switch can respond to only one of the two chemical inputs. It is a collection of numerous molecular switches dissolved in an organic solvent that responds to both inputs enabling an OR operation.

The development of miniaturized molecule-based devices requires the identification of methods to transfer the switching mechanisms developed in solution to the solid state [2.29]. Borrowing designs and fabrication strategies from conventional electronics, researchers are starting to explore the integration of molecular components into functional circuits and devices [2.30–33]. Generally, these strategies combine lithography and surface chemistry to assemble nanometer-thick organic films on the surfaces of microscaled or nanoscaled electrodes. Two main approaches for the deposition of organized molecular arrays on inorganic supports have emerged so far. In one instance, amphiphilic molecular building blocks are compressed into organized monolayers at air/water interfaces. The resulting films can be transferred on supporting solids employing the Langmuir–Blodgett technique [2.34]. Alternatively, certain molecules can be designed to adsorb spontaneously on the surfaces of compatible solids from liquid or vapor phases. The result is the self-assembly of organic layers on inorganic supports [2.35].

2.3.2 Langmuir–Blodgett Films

Films of amphiphilic molecules can be deposited on a variety of solid supports employing the Langmuir–Blodgett technique [2.34]. This method can be extended to electroactive compounds incorporating hydrophilic and hydrophobic groups. For example, the amphiphile **10** (Fig. 2.12) has a hydrophobic hexadecyl tail attached to a hydrophilic bipyridinium dication [2.36, 37]. This compound dissolves in mixtures of chloroform and methanol, but it is not soluble in moderately concentrated aqueous solutions of sodium perchlorate.

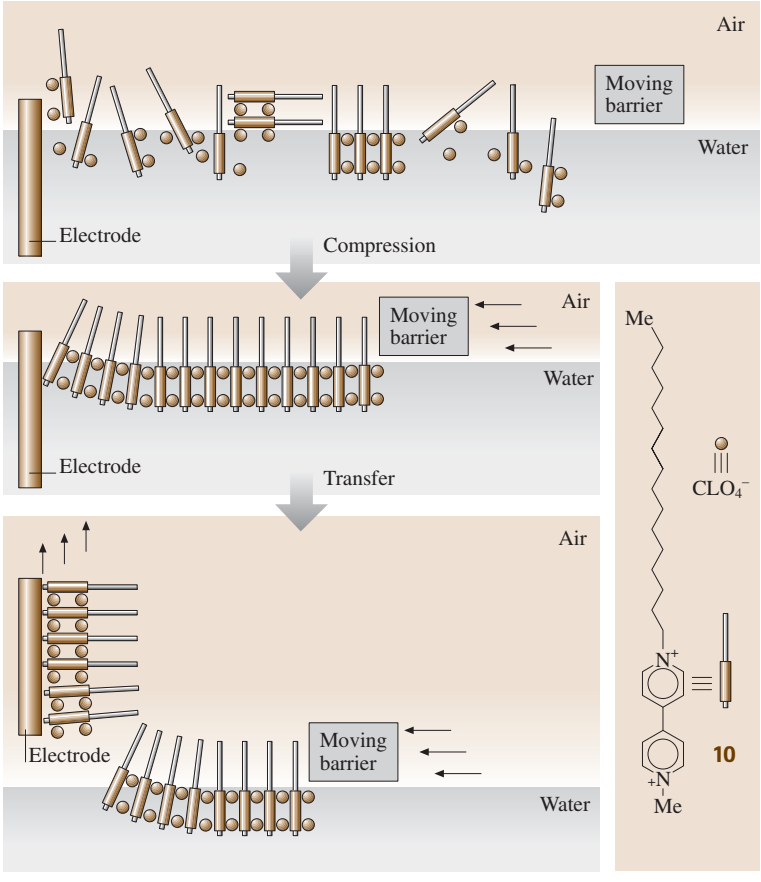


Fig. 2.12 The compression of the amphiphilic dication **10** with a moving barrier results in the formation of a packed monolayer at the air/water interface. The lifting of an electrode pre-immersed in the aqueous subphase encourages the transfer of part of the monolayer on the solid support

Thus the spreading of an organic solution of **10** on an aqueous sodium perchlorate subphase affords a collection of disorganized amphiphiles floating on the water surface (Fig. 2.12), after the organic solvent has evaporated. The molecular building blocks can be compressed into a monolayer with the aid of a moving barrier. The hydrophobic tails align away from the aqueous phase. The hydrophilic dicationic heads and the accompanying perchlorate counterions pack to form an organized monolayer at the air/water interface. The compression process can be monitored recording the surface pressure (π)-area per molecule (A) isotherm, which indicates a limiting molecular area of ca. 50 Å^2. This value is larger than the projected area of an oligomethylene chain. It correlates reasonably, however, with the overall area of a bipyridinium dication plus two perchlorate anions.

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

These seminal experiments demonstrate that electroactive amphiphiles can be organized into uniform monolayers at the air/water interface and then transferred efficiently on the surface of appropriate substrates to produce electrode/monolayer junctions. The resulting electroactive materials can become the functional components of molecule-based devices. For example, bipyridinium-based photodiodes can be fabricated following this approach [2.38, 39]. Their operating principles rely on photoinduced electron transfer from chromophoric units to bipyridinium acceptors. The electroactive and photoactive amphiphile **11** (Fig. 2.13) incorporates hydrophobic ferrocene and pyrene tails and a hydrophilic bipyridinium head. Chloroform solutions of **11** containing ten equivalents of arachidic acid can be spread on an aqueous calcium chloride subphase in a Langmuir trough. The amphiphiles can be compressed into a mixed monolayer, after the evaporation of the organic solvent. Pronounced steps in the corresponding π–A isotherm suggest that the bulky ferrocene and pyrene groups are squeezed away from the water surface. In the final arrangement, both photoactive groups align above the hydrophobic dication.

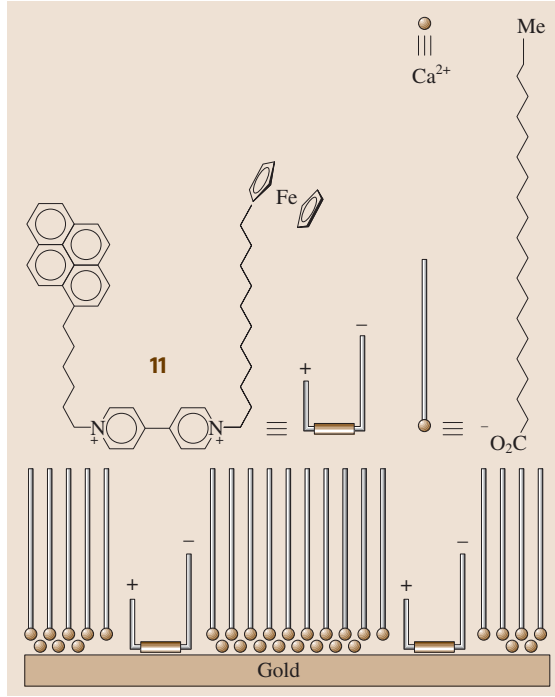


Fig. 2.13 Mixed monolayers of the amphiphile **11** and arachidic acid can be transferred from the air/water interface to the surface of an electrode to generate a molecule-based photodiode

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

donor to the oxidized pyrene fills its photogenerated hole. Overall, a unidirectional flow of electrons across the monolayer/electrode junction is established under the influence of light.

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

The [2]rotaxane **12** (Fig. 2.14) incorporates a macrocyclic polyether threaded onto a bipyridinium-based backbone [2.40, 41]. The two bipyridinium dications are bridged by a m-phenylene spacer and terminated by tetraarylmethane appendages. These two bulky groups trap mechanically the macrocycle preventing its dis-

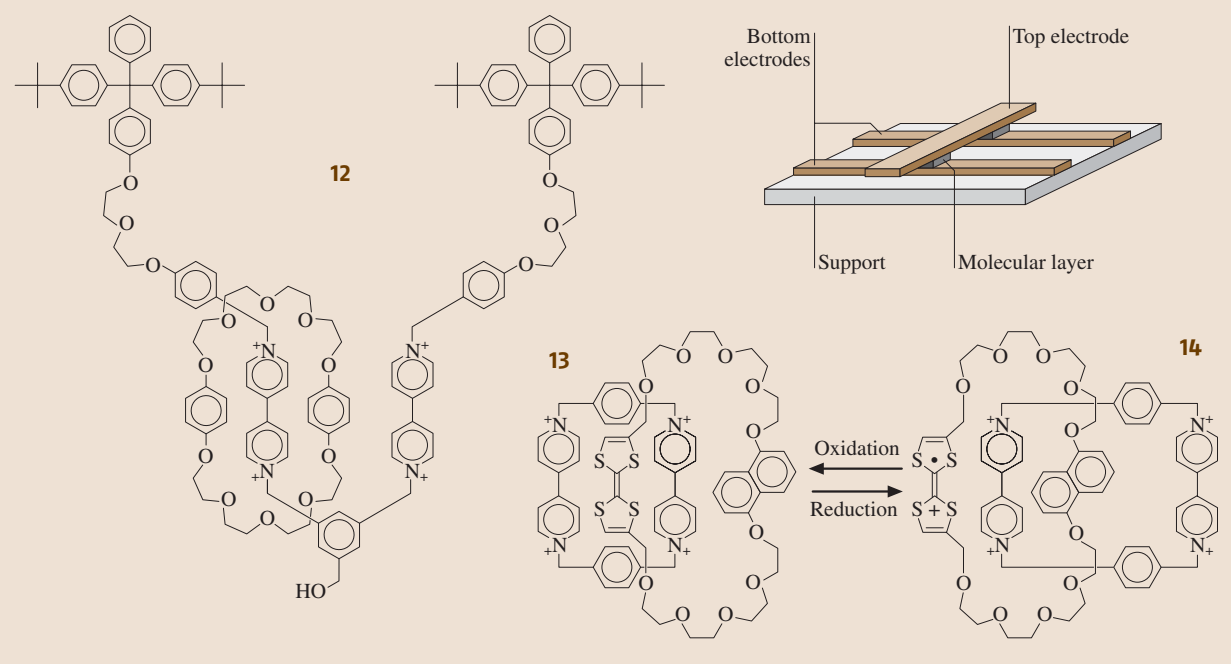


Fig. 2.14 The [2]rotaxane **12** and the [2]catenane **13** can be compressed into organized monolayers at air/water interfaces. The resulting monolayers can be transferred on the bottom electrodes of a patterned silicon support. After the deposition of a top electrode, electrode/monolayer/electrode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

sociation from the tetracationic backbone. In addition, their hydrophobicity complements the hydrophilicity of the two bipyridinium dications imposing amphiphilic character on the overall molecular assembly. This compound does not dissolve in aqueous solutions and can be compressed into organized monolayers at air/water interfaces. The corresponding π–A isotherm reveals a limiting molecular area of ca. 130 Å^2. This large value is a consequence of the bulk associated with the hydrophobic tetraarylmethane tails and the macrocycle encircling the tetracationic backbone.

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

The device incorporating the [2]rotaxane **13** can be exploited to implement simple logic operations [2.40]. The two bottom electrodes can be stimulated with voltage inputs (I1 and I2) while measuring a current output (O) at the common top electrode. When at least one of the two inputs is high (0 V), the output is low (<0.7 nA). When both inputs are low (-2 V), the output is high (ca. 4 nA). If a negative logic convention is applied to the voltage inputs (low = 1, high = 0) and a positive logic convention is applied to the current output (low = 0, high = 1), the signal transduction behavior translates into the truth table of an AND gate. The output O is 1 only when both inputs are 1. Instead, an OR operation can be executed if the logarithm of the current is considered as the output. The logarithm of the current is -12 when both voltage inputs are 0 V. It raises to ca. -9 when one or both voltage inputs are lowered to -2 V. This signal transduction behavior translates into the truth table of an OR gate if a negative logic convention is applied to the voltage inputs (low = 1, high = 0) and a positive logic convention is applied to the current output (low = 0, high = 1). The output O is 1 when at least one of the two inputs is 1.

The [2]catenane **13** (Fig. 2.14) incorporates a macrocyclic polyether interlocked with a tetracationic cyclophane [2.42, 43]. Organic solutions of the hexafluorophosphate salt of this [2]catenane and six equivalents of the sodium salt of dimyristoylphosphatidic acid can be co-spread on the water surface of a Langmuir trough [2.44]. The sodium hexafluorophosphate formed dissolves in the supporting aqueous phase, while the hydrophilic bipyridinium cations and the amphiphilic anions remain at the interface. Upon compression, the anions align their hydrophobic tails away from the water surface forming a compact monolayer above the cationic bipyridinium derivatives. The corresponding π–A isotherm indicates limiting molecular areas of ca. 125 Å^2. This large value is a consequence of the bulk associated with the two interlocking macrocycles.

Monolayers of the [2]catenane **13** can be transferred from the air/water interface to the surfaces of the bottom n-doped silicon/silicon dioxide electrodes of a patterned silicon chip with the hydrophobic tails of the amphiphilic anions pointing away from the supporting substrate [2.45, 46]. The subsequent assembly of a top titanium/aluminum electrode affords electrode/monolayer/electrode arrays. Their junction resistance can be probed grounding the top electrode and maintaining the potential of the bottom electrode at $+0.1$ V. If a voltage pulse of $+2$ V is applied to the bottom electrode before the measurement, the junction resistance probed is ca. 0.7 GΩ. After a pulse of -2 V applied to the bottom electrode, the junction resistance probed at $+0.1$ V drops ca. 0.3 GΩ. Thus, alternat-

ing positive and negative voltage pulses can switch reversibly the junction resistance between high and low values. This intriguing behavior is a result of the redox and dynamic properties of the [2]catenane **13**.

Extensive spectroscopic and crystallographic studies [2.42, 43] demonstrated that the tetrathiafulvalene unit resides preferentially inside the cavity of the tetracationic cyclophane of the [2]catenane **13** (Fig. 2.14). Attractive $[\pi \cdots \pi]$ stacking interactions between the neutral tetrathiafulvalene and the bipyridinium dications are responsible for this co-conformation. Oxidation of the tetrathiafulvalene generates a cationic form that is expelled from the cavity of the tetracationic cyclophane. After the circumrotation of the macrocyclic polyether, the oxidized tetrathiafulvalene is exchanged with the neutral 1,5-dioxynaphthalene producing the [2]catenane **14** (Fig. 2.14). The reduction of the tetrathiafulvalene back to its neutral state is followed by the circumrotation of the macrocyclic polyether, which restores the original state **14**. The voltage pulses applied to the bottom electrode of the electrode/monolayer/junction oxidize and reduce the tetrathiafulvalene unit inducing the interconversion between the forms **13** and **14**. The difference in the stereoelectronic properties of these two states translates into distinct current/voltage signatures. Indeed, their ability to mediate the tunneling of electrons across the junction differs significantly. As a result, the junction resistance probed at a low voltage after an oxidizing pulse is significantly different from that determined under the same conditions after a reducing pulse.

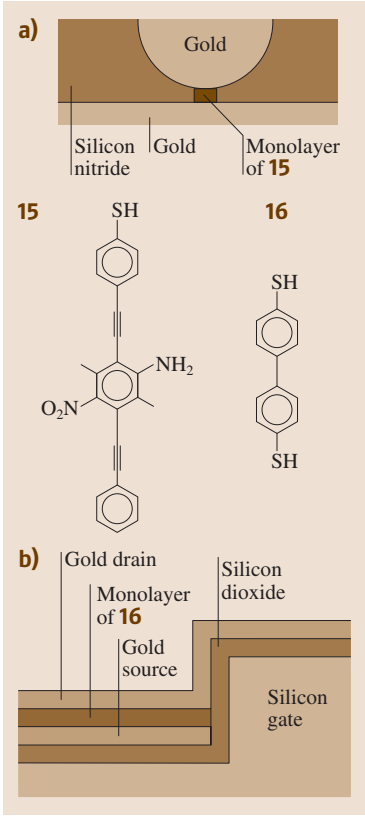


Fig. 2.15
(a) A monolayer of the thiol **15** is embedded between two gold electrodes maintained in position by a silicon nitride support. **(b)** A monolayer of the bisthiol **16** is located between two gold electrodes and stimulated by a silicon gate

2.3.3 Self-Assembled Monolayers

In the examples illustrated in Figs. 2.12–2.15, monolayers of amphiphilic and electroactive derivatives are assembled at air/water interfaces and then transferred on the surfaces of appropriate substrates. An alternative strategy to coat electrodes with molecular layers relies on the ability of certain compounds to adsorb spontaneously on solid supports from liquid or vapor phases [2.35]. In particular, the affinity of certain sulfurated functional groups for gold can be exploited to encourage the self-assembly of organic molecules on microscaled and nanoscaled electrodes.

The electrode/monolayer/electrode junction in Fig. 2.15a incorporates a molecular layer between two gold electrodes mounted on a silicon nitride support. This device can be fabricated combining chemical vapor deposition, lithography, anisotropic etching, and self-assembly [2.47]. Initially, a silicon wafer is coated with a 50 nm thick layer of silicon nitride by low pressure chemical vapor deposition. Then, a square of $400 \times 400\,\mu m$ is patterned on one side of the coated wafer by optical lithography and reactive ion etching. Anisotropic etching of the exposed silicon up to the other side of the wafer leaves a suspended silicon nitride membrane of $40 \times 40\,\mu m$. Electron beam lithography and reactive ion etching can be used to carve a bowl-shaped hole (diameter = 30–50 nm) in the membrane. Evaporation of gold on the membrane fills the pore producing a bowl-shaped electrode. Immersion of the substrate in a solution of the thiol **15** results in the self-assembly of a molecular layer on the narrow part of the bowl-shaped electrode. The subsequent evaporation of a gold film on the organic monolayer produces an electrode/monolayer/electrode junction (Fig. 2.15a) with a contact area of less than $2{,}000\,nm^2$ and approximately 1,000 molecules.

Under the influence of voltage pulses applied to one of the two gold electrodes in Fig. 2.15a, the conductivity of the sandwiched monolayer switches reversibly be-

tween low and high values [2.48]. In the initial state, the monolayer is in a low conducting mode. A current output of only 30 pA is detected, when a probing voltage of +0.25 V is applied to the bowl-shaped electrode. If the same electrode is stimulated with a short voltage pulse of +5 V, the monolayer switches to a high conducting mode. Now a current output of 150 pA is measured at the same probing voltage of +0.25 V. Repeated probing of the current output at various intervals of time indicates that the high conducting state is memorized by the molecule-based device, and it is retained for more than 15 min. The low conducting mode is restored after either a relatively long period of time or the stimulation of the bowl-shaped electrode with a reverse voltage pulse of −5 V. Thus the current output switches from a low to a high value, if a high voltage input is applied. It switches from a high to a low value, under the influence of a low voltage pulse. This behavior offers the opportunity to store and erase binary data in analogy to a conventional random access memory [2.17]. Binary digits can be encoded on the current output of the molecule-based device applying a positive logic convention (low = 0, high = 1). It follows that a binary 1 can be stored in the molecule-based device applying a high voltage input, and it can be erased applying a low voltage input [2.48].

The electrode/monolayer/electrode junctions in Fig. 2.14 and in Fig. 2.15a are two-terminal devices. A voltage input is applied to one terminal, and a current output is measured at the other. The implementation of molecule-based transistors, however, requires three terminals. A third electrode is needed to gate the current flowing between the source and drain terminals. These types of devices are much more versatile than their two-terminal counterparts for the implementation of logic functions. But the incorporation of molecular components in a three-terminal configuration is considerably more challenging. A remarkable example is the composite assembly in Fig. 2.15b [2.49]. This field-effect transistor integrates a monolayer of the bisthiol **16** and three electrodes. Conventional lithography and anisotropic etching are employed to define a vertical step in a doped silicon substrate. The resulting structure is covered with an insulating layer of silicon dioxide with a thickness of ca. 30 nm. Then gold is evaporated on the bottom part of the step to allow the self-assembly the bisthiol **16**. Under the influence of the gate voltage, the molecular components embedded in this device modulate the current flowing between the source and drain electrodes. The intensity of the drain current increases by five orders of magnitude if the gate voltage is lowered below −0.2 V. It is important to note, however, that only the molecules within 5 nm from the monolayer/insulator interface are responsible for conduction. Thus the current/voltage behavior of this three-terminal device is determined only by few thousand molecules.

Conventional transistors are the basic building blocks of digital circuits [2.19]. They are routinely interconnected with specific configurations to execute designed logic functions. The molecule-based transistor in Fig. 2.15b is no exception. It can be used to assemble real electronic circuits incorporating molecular components. The connection of the drain terminals of two molecule-based transistors affords an inverter [2.49]. In this basic circuit, an output voltage switches from a high (0 V) to a low value (−2 V) when an input voltage varies from a low (−2 V) to a high value (0 V) and vice versa. This signal transduction behavior translates into the truth table of a NOT gate, if a positive logic convention (low = 0, high = 1) is applied to the input and output voltages.

The ability of thiols to self-assemble on the surface of gold can be exploited to fabricate nanocomposite materials integrating organic and inorganic components. For example, the bisthiol **17** forms monolayers (Fig. 2.16a) on gold electrodes with surface coverages of ca. 4.1×10^{10} mol cm^2 [2.50, 51]. The formation of a thiolate–gold bond at one of the two thiol ends of **17** is responsible for adsorption. The remaining thiol group points away from the supporting surface and can be exploited for further functionalization. Gold nanoparticles adsorb on the molecular layer (Fig. 2.16b), once again, as a result of thiolate–gold bond formation. The immersion of the resulting material in a methanol solution of **17** encourages the adsorption of an additional organic layer (Fig. 2.16c) on the composite material. Following these procedures, up to ten alternating organic and inorganic layers can be deposited on the electrode surface. The resulting assembly can mediate the unidirectional electron transfer from the supporting electrode to redox active species in solution. For example, the cyclic voltammogram of the $[Ru(NH_3)_6]^{3+/2+}$ couple recorded with a bare gold electrode reveals a reversible reduction process. In the presence of ten alternating molecular and nanoparticle layers on the electrode surface, the reduction potential shifts by ca. −0.2 V and the back oxidation wave disappears. The pronounced potential shift indicates that $[Ru(NH_3)_6]^{3+}$ accepts electrons only after the surface-confined bipyridinium dications have been reduced. The lack of reversibility indicates that the back oxidation to the bipyridinium dications inhibits the transfer of elec-

trons from the $[Ru(NH_3)_6]^{2+}$ to the electrode. Thus the electroactive multilayer allows the flow of electrons in one direction only in analogy to conventional diodes.

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

Similar nanostructured materials, combining molecular and nanoparticles layers, can be prepared on layers on indium-tin oxide electrodes following multistep procedures [2.53]. The hydroxylated surfaces of indium-tin oxide supports can be functionalized with 3-ammoniumpropylysilyl groups and then exposed to gold nanoparticles having a diameter of ca. 13 nm [2.54, 55]. Electrostatic interactions promote the adsorption of the nanoparticles on the organic layer (Fig. 2.17a). The treatment of the composite film with the bipyridinium cyclophane **18** produces an organic layer on the gold nanoparticles (Fig. 2.18b). Following this approach, alternating layers of inorganic nanoparticles and organic building blocks can be assembled on the indium-tin oxide support. Cyclic voltammograms of the resulting materials show the oxidation of the gold nanoparticles and the reduction of the bipyridinium units. The peak current for both processes increases with the number of alternating layers. Comparison of these values indicates that the ratio between the number of tetracationic cyclophanes and that of the nanoparticles is ca. 100 : 1.

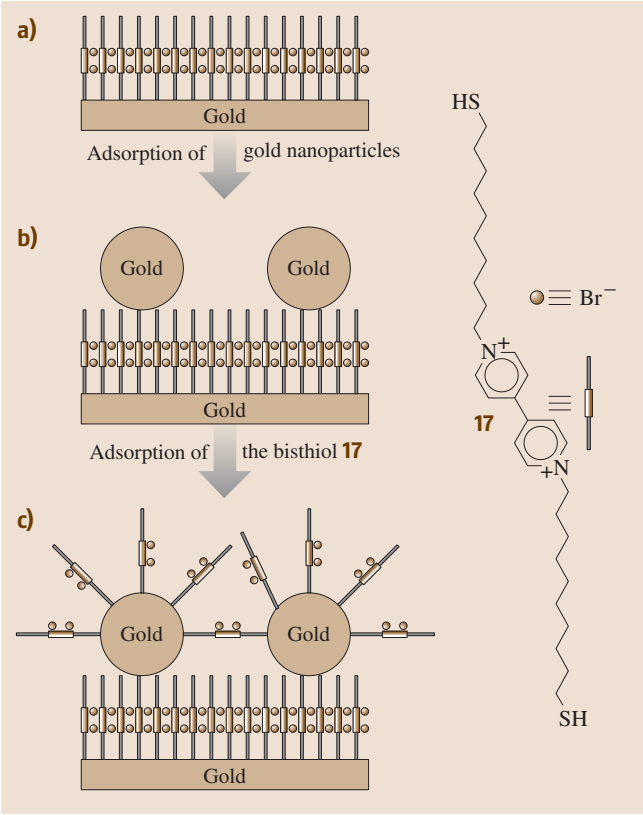


Fig. 2.16 **(a)** The bisthiol **18** self-assembles on gold electrodes as a result of thiolate–gold bond formation. **(b)** Gold nanoparticles adsorb spontaneously on the molecular layer. **(c)** Exposure of the composite assembly to a solution of **18** results in the formation of an additional molecular layer on the surface of the gold nanoparticles

The tetracationic cyclophane **18** binds dioxyarenes in solution [2.56, 57]. Attractive supramolecular forces between the electron deficient bipyridinium units and the electron rich guests are responsible for complexation. This recognition motif can be exploited to probe the ability of the composite films in Fig. 2.17b,c to sense electron rich analytes. In particular, hydroquinone is expected to enter the electron deficient cavities of the surface-confined cyclophanes. Cyclic voltammograms consistently reveal the redox waves associated with the reversible oxidation of hydroquinone even when very small amounts of the guest (ca. 1×10^5 M) are added to the electrolyte solution [2.54, 55]. No redox response can be detected with a bare indium–tin oxide electrode under otherwise identical conditions. The supramolecular association of the guest and the surface confined

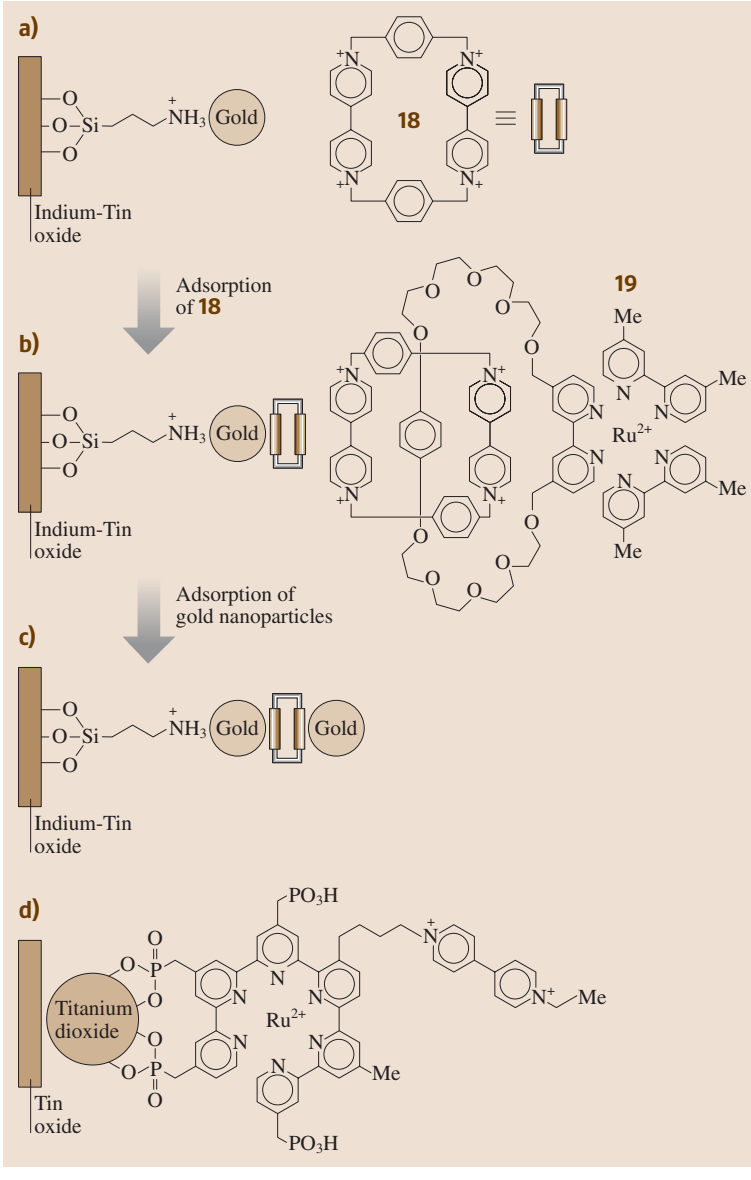


Fig. 2.17 **(a)** Gold nanoparticles assemble spontaneously on pre-functionalized indium-tin oxide electrodes. **(b)** Electrostatic interactions encourage the adsorption of the tetracationic cyclophane **18** on the surface-confined nanoparticles. **(c)** An additional layer of nanoparticles assembles on the cationic organic coating. Similar composite films can be prepared using the tetracationic [2]catenane **19** instead of the cyclophane **18**. **(d)** Phosphonate groups can be used to anchor molecular building blocks to titanium dioxide nanoparticles

cyclophanes increases the local concentration of hydroquinone at the electrode/solution interface enabling its electrochemical detection.

Following a related strategy, the [2]catenane **19** (Fig. 2.17) can be incorporated into similar composite arrays [2.58,59]. This interlocked molecule incorporates a Ru(II)/trisbipyridine sensitizer and two bipyridinium acceptors. Upon irradiation of the composite material at 440 nm, photoinduced electron transfer from the sensitizer to the appended acceptors occurs. The photogenerated hole in the sensitizer is filled after the transfer of an electron from a sacrificial electron donor present in the electrolyte solution. Under a positive voltage bias applied to the supporting electrode, an electron flow from the bipyridinium acceptors to the indium-tin oxide support is established. The resulting current switches between high and low values as the light source is turned on and off.

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

2.3.4 Nanogaps and Nanowires

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

A promising approach to unimolecular devices relies on the fabrication of nanometer-sized gaps in metallic features followed by the insertion of individual molecules between the terminals of the gap. This strategy permits the assembly of nanoscaled three-terminal devices equivalent to conventional transistors [2.62–64]. A remarkable example is illustrated in Fig. 2.18a [2.62]. It incorporates a single molecule in the nanogap generated between two gold electrodes. Initially electron beam lithography is used to pattern a gold wire on a doped silicon wafer covered by an insulating silicon dioxide layer. Then the gold feature is broken by electromigration to generate the nanogap. The lateral size of the separated electrodes is ca. 100 nm and their thickness is ca. 15 nm. Scanning electron microcopy indicates that the facing surfaces of the separated electrodes are not uniform and that tiny gaps between their protrusions are formed. Current/voltage measurements suggest that the size of the smallest nanogap is ca. 1 nm. When the

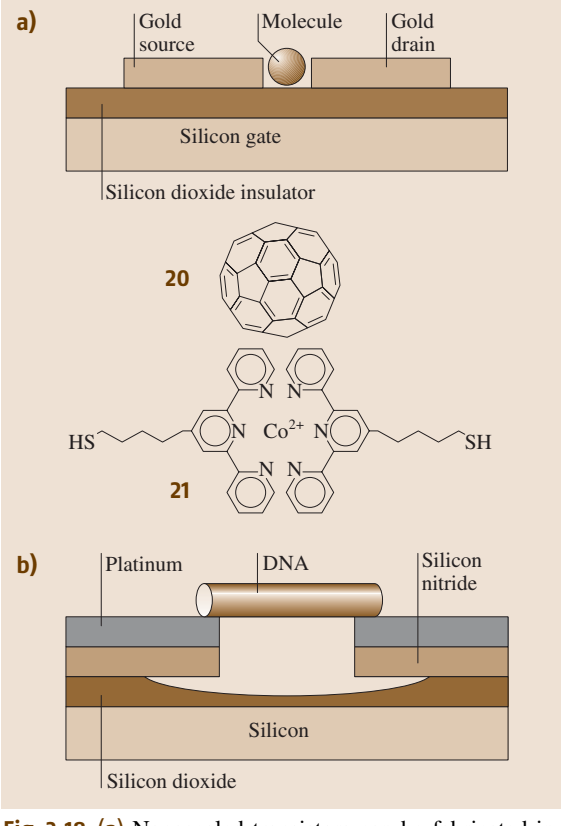


Fig. 2.18 **(a)** Nanoscaled transistors can be fabricated inserting a single molecule (**20** or **21**) between source and drain electrodes mounted on a silicon/silicon dioxide support. **(b)** A DNA nanowire can bridge nanoelectrodes suspended above a silicon dioxide support

breakage of the gold feature is preceded by the deposition of a dilute toluene solution of C_{60} (**20**), junctions with enhanced conduction are obtained. This particular molecule has a diameter of ca. 0.7 nm and can insert in the nanogap facilitating the flow of electrons across the junction.

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

A similar strategy can be employed to fabricate a nanoscaled transistor incorporating the Co(II) complex **21** shown in Fig. 2.18 [2.64]. In this instance, a silicon dioxide layer with a thickness of ca. 30 nm is grown thermally on a doped silicon substrate. Then a gold wire with a width of ca. 200 nm and a thicknesses of ca. 10–15 nm is patterned on the silicon dioxide overlayer by electron beam lithography. After extensive washing of the substrate with acetone and methylene chloride and cleaning with oxygen plasma, the gold wire is exposed to a solution of the bisthiol **21**. The formation of thiolate–gold bonds promotes the self-assembly of the molecular building block on the gold surface. At this point, electromigration-induced breakage produces a gap of 1–2 nm in the gold wire. The surface-confined bisthiol **21** is only 0.24 nm long and, therefore, it can insert in the nanogap producing an electrode/molecule/electrode junction.

The cobalt center in **21** can be oxidized/reduced reversibly between Co(II) and Co(III) [2.64]. When this electroactive molecule is inserted in a nanogap (Fig. 2.18a), its ability to accept and donate electrodes dictates the current/voltage profile of the resulting electrode/molecule/electrode junction. More precisely, no current flows across the junction below a certain voltage threshold. As the source voltage is raised above this particular value, the drain current increases in steps. The threshold associated with the source voltage varies in magnitude with the gate voltage. This intriguing behavior is a consequence of the finite energy necessary to oxidize/reduce the cobalt center and of a change in the relative stabilities of the oxidized and reduced forms Co(II) and Co(III) with the gate voltage. In summary, the conduction of the electrode/molecule/electrode junction can be tuned adjusting the voltage of the silicon support. The behavior of this molecule-based nanoelectronic device is equivalent to that of a conventional transistor [2.19]. In both instances, the gate voltage regulates the current flowing from the source to the drain.

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

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

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

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

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

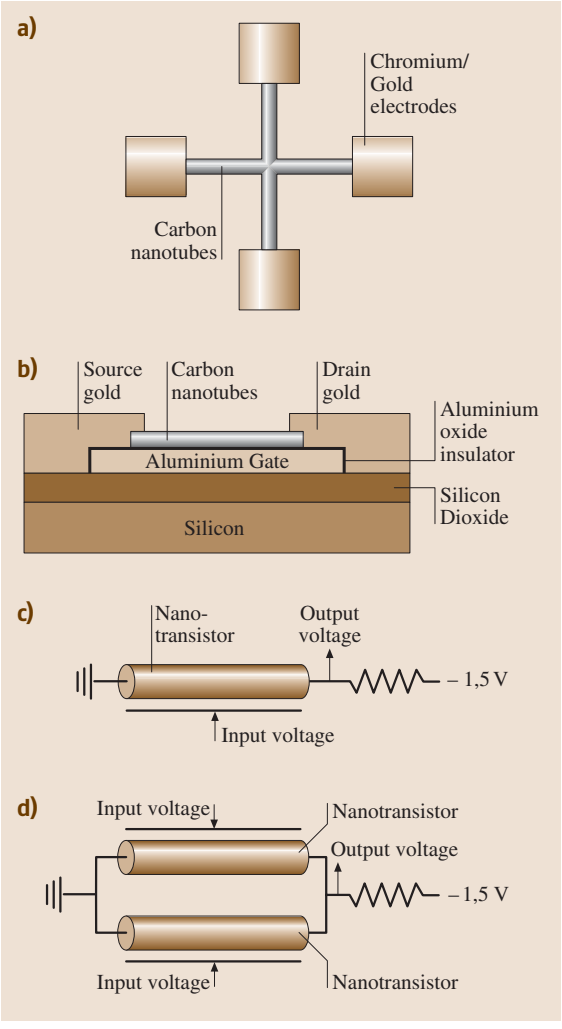


Fig. 2.19 **(a)** Nanoscaled junctions can be assembled on silicon/silicon dioxide supports crossing pairs of orthogonally arranged single-wall carbon nanotubes with chromium/gold electrical contacts at their ends. **(b)** Nanotransistors can be fabricated contacting the two ends of a single-wall carbon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

The fabrication of this device involves three main steps. First, alignment marks for the electrodes are patterned on a silicon/silicon dioxide support by electron beam lithography. Then the substrate is exposed to a dichloromethane suspension of single-wall SWNT carbon nanotubes. After washing with isopropanol, crosses of carbon nanotubes in an appropriate alignment relative to the electrode marks are identified by tapping mode atomic force microscopy. Finally chromium/gold electrodes are fabricated on top of the nanotube ends, again, by electron beam lithography. The conductance of individual nanotubes can be probed by exploiting the two electric contacts at their ends. These two-terminal measurements reveal that certain nanotubes have metallic behavior, while others are semiconducting. It follows that three distinct types of cross junctions differing in the nature of their constituent nanotubes can be identified on the silicon/silicon dioxide support. Four terminal current/voltage measurements indicate that junctions formed by two metallic nanotubes have high conductance and ohmic behavior. Similarly, high junction conductance and ohmic behavior is observed when two semiconducting nanotubes cross. The current/voltage signature of junctions formed when a metallic nanotube crosses a semiconducting one are, instead, completely different. The metallic nanotube depletes the semiconducting one at the junction region producing a nanoscaled Schottky barrier with a pronounced rectifying behavior.

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

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

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

2.4 Conclusions and Outlook

Nature builds nanostructured biomolecules relying on a highly modular approach [2.1]. Small building blocks are connected by robust chemical bonds to generate long strands of repeating units. The synergism of a multitude of attractive supramolecular forces determines the three-dimensional arrangement of the resulting polymeric chains and controls the association of independent strands into single and well-defined entities. Nucleic acids and proteins are two representative classes of biomolecules assembled with subnanometer precision through the subtle interplay of covalent and noncovalent bonds starting from a relatively small pool of nucleotide and amino acid building blocks.

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

The chemical construction of nanoscaled molecules from modular building blocks also offers the opportunity for engineering specific properties in the resulting assemblies. In particular, electroactive and photoactive fragments can be integrated into single molecules. The ability of these functional subunits to accept/donate electrons and photons can be exploited to design nanoscaled electronic and photonic devices. Indeed, molecules that respond to electrical and optical stimulations producing detectable outputs have been designed already [2.16]. These chemical systems can be employed to control the interplay of input and output signals at the molecular level. Their conceptual analogy with the signal transduction operated by conventional logic gates in digital circuits is evident. In fact, electroactive and photoactive molecules able to reproduce AND, NOT, and OR operations as well as simple combinational of these basic logic functions are already a reality [2.20, 21].

Most of the molecular switches for digital processing developed so far rely on bulk addressing. In general, relatively large collections of functional molecules are addressed simultaneously in solution. The realization of molecule-based devices with reduced dimensions as well as practical limitations associated with liquid phases in potential applications are encouraging a transition from the solution to the solid state. The general strategy followed so far relies on the deposition of functional molecules on the surfaces of appropriate electrodes following either the Langmuir–Blodgett methodology [2.34] or self-assembly processes [2.35]. The combination of these techniques with the nanofabrication of insulating, metallic, and semiconducting features on appropriate supports has already allowed the realization of fascinating molecule-based devices [2.30–33, 53]. The resulting assemblies integrate inorganic and organic components and, in some instances, even biomolecules to execute specific functions. They can convert optical stimulations into electrical signals. They can execute irreversible and reversible switching operations. They can sense qualitatively and quantitatively specific analytes. They can reproduce the functions of conventional rectifiers and transistors. They can be integrated within functioning nanoelectronic devices capable of simple logic operations.

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

References

2.1 D. Voet, J. G. Voet: *Biochemistry* (Wiley, New York 1995)
2.2 K. C. Nicolau, E. C. Sorensen: *Classics in Total Synthesis* (VCH, Weinheim 1996)
2.3 J.-M. Lehn: *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim 1995)
2.4 M. M. Harding, U. Koert, J.-M. Lehn, A. Marquis-Rigault, C. Piguet, J. Siegel: Synthesis of unsubstituted and 4,4′-substituted oligobipyridines as ligand strands for helicate self-assembly, Helv. Chim. Acta **74** (1991) 594–610
2.5 J.-M. Lehn, A. Rigault, J. Siegel, B. Harrowfield, B. Chevrier, D. Moras: Spontaneous assembly of double-stranded helicates from oligobipyridine ligands and copper(I) cations: Structure of an inorganic double helix, Proc. Natl. Acad. Sci. USA **84** (1987) 2565–2569
2.6 J.-M. Lehn, A. Rigault: Helicates: Tetra- and pentanuclear double helix complexes of Cu(I) and poly(bipyridine) strands, Angew. Chem. Int. Ed. Engl. **27** (1988) 1095–1097
2.7 J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. R. Ghadiri: Self-assembling peptide nanotubes, J. Am. Chem. Soc. **118** (1996) 43–50
2.8 M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee, N. Khazanovich: Self-assembling organic nanotubes based on a cyclic peptide architecture, Nature **366** (1993) 324–327
2.9 V. Balzani, A. Credi, F. M. Raymo, J. F. Stoddart: Artificial molecular machines, Angew. Chem. Int. Ed. **39** (2000) 3348–3391
2.10 A. J. Bard, L. R. Faulkner: *Electrochemical Methods: Fundamentals and Applications* (Wiley, New York 2000)
2.11 V. Balzani (Ed.): *Electron Transfer in Chemistry* (Wiley-VCH, Weinheim 2001)
2.12 J. D. Coyle: *Principles and Applications of Photochemistry* (Wiley, New York 1988)
2.13 J. F. Stoddart (Ed.): Molecular machines, Acc. Chem. Res. **34** (2001) 409–522
2.14 P. R. Ashton, R. Ballardini, V. Balzani, A. Credi, K. R. Dress, E. Ishow, C. J. Kleverlaan, O. Kocian, J. A. Preece, N. Spencer, J. F. Stoddart, M. Venturi, S. Wenger: A photochemically driven molecular-level abacus, Chem. Eur. J. **6** (2000) 3558–3574
2.15 B. L Feringa (Ed.): Photochromism: memories and switches, Chem. Rev. **100** (2000) 1683–1890
2.16 B. L. Feringa (Ed.): *Molecular Switches* (Wiley-VCH, Weinheim 2001)
2.17 R. J. Mitchell: *Microprocessor Systems: An Introduction* (Macmillan, London 1995)
2.18 D. R. Smith: *Digital Transmission Systems* (Van Nostrand Reinhold, New York 1993)
2.19 S. Madhu: *Electronics: Circuits and Systems* (SAMS, Indianapolis 1985)
2.20 A. P. de Silva, N. D. McClenaghan, C. P. McCoy: Logic Gates. In: *Electron Transfer in Chemistry*, ed. by V. Balzani (Wiley-VCH, Weinheim 2001) pp. 156–185
2.21 F. M. Raymo: Digital processing and communication with molecular switches, Adv. Mater. **14** (2002) 401–414
2.22 A. Aviram: Molecules for memory, logic and amplification, J. Am. Chem. Soc. **110** (1988) 5687–5692
2.23 A. P. de Silva, H. Q. N. Gunaratne, C. P. McCoy: A molecular photoionic AND gate based on fluorescent signaling, Nature **364** (1993) 42–44
2.24 M. Asakawa, P. R. Ashton, V. Balzani, A. Credi, G. Mattersteig, O. A. Matthews, M. Montalti, N. Spencer, J. F. Stoddart, M. Venturi: Electrochemically induced molecular motions in pseudorotaxanes: A case of dual-mode (oxidative and reductive) dethreading, Chem. Eur. J. **3** (1997) 1992–1996
2.25 F. M. Raymo, S. Giordani: Signal processing at the molecular level, J. Am. Chem. Soc. **123** (2001) 4651–4652
2.26 F. M. Raymo, S. Giordani: Signal communication between molecular switches, Org. Lett. **3** (2001) 3475–3478
2.27 F. M. Raymo, S. Giordani: Digital communication through intermolecular fluorescence modulation, Org. Lett. **3** (2001) 1833–1836
2.28 F. M. Raymo, S. Giordani: All-optical processing with molecular switches, Proc. Natl. Acad. Sci. USA **99** (2002) 4941–4944
2.29 A. J. Bard: *Integrated Chemical Systems: A Chemical Approach to Nanotechnology* (Wiley, New York 1994)
2.30 R. M. Metzger: Electrical rectification by a molecule: The advent of unimolecular electronic devices, Acc. Chem. Res. **32** (1999) 950–957
2.31 C. Joachim, J. K. Gimzewski, A. Aviram: Electronics using hybrid-molecular and mono-molecular devices, Nature **408** (2000) 541–548
2.32 J. M. Tour: Molecular electronics. Synthesis and testing of components, Acc. Chem. Res. **33** (2000) 791–804
2.33 A. R. Pease, J. O. Jeppesen, J. F. Stoddart, Y. Luo, C. P. Collier, J. R. Heath: Switching devices based on interlocked molecules, Acc. Chem. Res. **34** (2001) 433–444
2.34 M. C. Petty: *Langmuir–Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge 1996)
2.35 A. Ulman: *An Introduction to Ultrathin Organic Films* (Academic, Boston 1991)
2.36 C. Lee, A. J. Bard: Comparative electrochemical studies of *N*-Methyl-*N*′-hexadecyl viologen monomolecular films formed by irreversible adsorption and the Langmuir–Blodgett method, J. Electroanal. Chem. **239** (1988) 441–446
2.37 C. Lee, A. J. Bard: Cyclic voltammetry and Langmuir film isotherms of mixed monolayers of *N*-docosoyl-

N′-methyl viologen with arachidic acid, Chem. Phys. Lett. **170** (1990) 57–60
2.38 M. Fujihira, K. Nishiyama, H. Yamada: Photoelectrochemical responses of optically transparent electrodes modified with Langmuir–Blodgett films consisting of surfactant derivatives of electron donor, acceptor and sensitizer molecules, Thin Solid Films **132** (1985) 77–82
2.39 M. Fujihira: Photoelectric conversion with Langmuir–Blodgett films. In: *Nanostructures Based on Molecular Materials*, ed. by W. Göpel, C. Ziegler (VCH, Weinheim 1992) pp. 27–46
2.40 C. P. Collier, E. W. Wong, M. Belohradsky, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, J. R. Heath: Electronically configurable molecular-based logic gates, Science **285** (1999) 391–394
2.41 E. W. Wong, C. P. Collier, M. Belohradsky, F. M. Raymo, J. F. Stoddart, J. R. Heath: Fabrication and transport properties of single-molecule-thick electrochemical junctions, J. Am. Chem. Soc. **122** (2000) 5831–5840
2.42 M. Asakawa, P. R. Ashton, V. Balzani, A. Credi, C. Hamers, G. Mattersteig, M. Montalti, A. N. Shipway, N. Spencer, J. F. Stoddart, M. S. Tolley, M. Venturi, A. J. P. White, D. J. Williams: A chemically and electrochemically switchable [2]catenane incorporating a tetrathiafulvalene unit, Angew. Chem. Int. Ed. **37** (1998) 333–337
2.43 V. Balzani, A. Credi, G. Mattersteig, O. A. Matthews, F. M. Raymo, J. F. Stoddart, M. Venturi, A. J. P. White, D. J. Williams: Switching of pseudorotaxanes and catenanes incorporating a tetrathiafulvalene unit by redox and chemical inputs, J. Org. Chem. **65** (2000) 1924–1936
2.44 M. Asakawa, M. Higuchi, G. Mattersteig, T. Nakamura, A. R. Pease, F. M. Raymo, T. Shimizu, J. F. Stoddart: Current/Voltage characteristics of monolayers of redox-switchable [2]catenanes on gold, Adv. Mater. **12** (2000) 1099–1102
2.45 C. P. Collier, G. Mattersteig, E. W. Wong, Y. Luo, K. Beverly, J. Sampaio, F. M. Raymo, J. F. Stoddart, J. R. Heath: A [2]catenane based solid-state electronically reconfigurable switch, Science **289** (2000) 1172–1175
2.46 C. P. Collier, J. O. Jeppesen, Y. Luo, J. Perkins, E. W. Wong, J. R. Heath, J. F. Stoddart: Molecular-based electronically switchable tunnel junction devices, J. Am. Chem. Soc. **123** (2001) 12632–12641
2.47 J. Chen, M. A. Reed, A. M. Rawlett, J. M. Tour: Large on-off ratios and negative differential resistance in a molecular electronic device, Science **286** (1999) 1550–1552
2.48 M. A. Reed, J. Chen, A. M. Rawlett, D. W. Price, J. M. Tour: Molecular random access memory cell, Appl. Phys. Lett. **78** (2001) 3735–3737
2.49 J. H. Schön, H. Meng, Z. Bao: Self-assembled monolayer organic field-effect transistors, Nature **413** (2001) 713–716
2.50 D. I. Gittins, D. Bethell, R. J. Nichols, D. J. Schiffrin: Redox-controlled multilayers of discrete gold particles: A novel electroactive nanomaterial, Adv. Mater. **9** (1999) 737–740
2.51 D. I. Gittins, D. Bethell, R. J. Nichols, D. J. Schiffrin: Diode-like electron transfer across nanostructured films containing a redox ligand, J. Mater. Chem. **10** (2000) 79–83
2.52 D. I. Gittins, D. Bethell, D. J. Schiffrin, R. J. Nichols: A nanometer-scale electronic switch consisting of a metal cluster and redox-addressable groups, Nature **408** (2000) 67–69
2.53 A. N. Shipway, M. Lahav, I. Willner: Nanostructured gold colloid electrodes, Adv. Mater. **12** (2000) 993–998
2.54 A. N. Shipway, M. Lahav, R. Blonder, I. Willner: Bis-bipyridinium cyclophane receptor–Au nanoparticle superstructure for electrochemical sensing applications, Chem. Mater. **11** (1999) 13–15
2.55 M. Lahav, A. N. Shipway, I. Willner, M. B. Nielsen, J. F. Stoddart: An enlarged bis-bipyridinum cyclophane–Au nanoparticle superstructure for selective electrochemical sensing applications, J. Electroanal. Chem. **482** (2000) 217–221
2.56 R. E. Gillard, F. M. Raymo, J. F. Stoddart: Controlling self-assembly, Chem. Eur. J. **3** (1997) 1933–1940
2.57 F. M. Raymo, J. F. Stoddart: From supramolecular complexes to interlocked molecular compounds, Chemtracts – Organic Chemistry **11** (1998) 491–511
2.58 M. Lahav, T. Gabriel, A. N. Shipway, I. Willner: Assembly of a Zn(II)-porphyrin-bipyridinium dyad and Au-nanoparticle superstructures on conductive surfaces, J. Am. Chem. Soc. **121** (1999) 258–259
2.59 M. Lahav, V. Heleg-Shabtai, J. Wasserman, E. Katz, I. Willner, H. Durr, Y. Hu, S. H. Bossmann: Photoelectrochemistry with integrated photosensitizer-electron acceptor Au-nanoparticle arrays, J. Am. Chem. Soc. **122** (2000) 11480–11487
2.60 G. Will, S. N. Rao, D. Fitzmaurice: Heterosupramolecular optical write-read-erase device, J. Mater. Chem. **9** (1999) 2297–2299
2.61 A. Merrins, C. Kleverlann, G. Will, S. N. Rao, F. Scandola, D. Fitzmaurice: Time-resolved optical spectroscopy of heterosupramolecular assemblies based on nanostructured TiO_2 films modified by chemisorption of covalently linked ruthenium and viologen complex components, J. Phys. Chem. B **105** (2001) 2998–3004
2.62 H. Park, J. Park, A. K. L. Lim, E. H. Anderson, A. P. Alivisatos, P. L. McEuen: Nanomechanical oscillations in a single C_{60} transistor, Nature **407** (2000) 57–60
2.63 W. Liang, M. P. Shores, M. Bockrath, J. R. Long, H. Park: Kondo resonance in a single-molecule transistor, Nature **417** (2002) 725–729
2.64 J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna,

H. D. Abruna, P. L. McEuen, D. C. Ralph: Coulomb blockade and the Kondo effect in single-atom transistors, Nature **417** (2002) 722–725

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

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

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

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

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

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

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

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

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

2.74 M. S. Fuhrer, J. Nygård, L. Shih, M. Forero, Y.-G. Yoon, M. S. C. Mazzoni, H. J. Choi, J. Ihm, S. G. Louie, A. Zettl, P. L. McEuen: Crossed nanotube junctions, Science **288** (2000) 494–497

2.75 T. Rueckes, K. Kim, E. Joselevich, G. Y. Tseng, C.-L. Cheung, C. M. Lieber: Carbon nanotube-based nonvolatile random access memory for molecular computing, Science **289** (2000) 94–97

2.76 A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker: Logic circuits with carbon nanotube transistors, Science **294** (2001) 1317–1320

3. Introduction to Carbon Nanotubes

Introduction

Carbon nanotubes are among the amazing objects that science sometimes creates by accident, without meaning to, but that will likely revolutionize the technological landscape of the century ahead. Our society stands to be significantly influenced by carbon nanotubes, shaped by nanotube applications in every aspect, just as silicon-based technology still shapes society today. The world already dreams of space-elevators tethered by the strongest of cables, hydrogen-powered vehicles, artificial muscles, and so on – feasts that would be made possible by the emerging carbon nanotube science.

Of course, nothing is set in stone yet. We are still at the stage of possibilities and potential. The recent example of fullerenes – molecules closely related to nanotubes, whose importance was so anticipated that their discovery in 1985 brought a Nobel Prize to their finders in 1996 although few related applications have actually yet reached the market – teaches us to play the game of enthusiastic predictions with some caution. But in the case of carbon nanotubes, expectations are high. Taking again the example of electronics, the miniaturization of chips is about to reach its lowest limits. Are we going to accept that our video camera, computers, and cellular phones no longer decrease in size and increase in memory every six months? Surely not. Always going deeper, farther, smaller, higher is a characteristic unique to humankind, and which helps to explain its domination of the living world on earth. Carbon nanotubes can help us fulfill our expectation of constant technological progress as a source of better living.

In this chapter, after the structure, synthesis methods, growth mechanisms, and properties of carbon nanotubes will be described, an entire section will be devoted to nanotube-related nano-objects. Indeed, should pristine nanotubes reach any limitation in some area, their ready and close association to foreign atoms, molecules, and compounds offers the prospect of an even magnified set of properties. Finally, we will describe carbon nanotube applications supporting the idea that the future for the science and technology of carbon nanotubes looks very promising.

3.1 **Structure of Carbon Nanotubes** 40
- 3.1.1 Single-Wall Nanotubes 40
- 3.1.2 Multiwall Nanotubes 43

3.2 **Synthesis of Carbon Nanotubes** 45
- 3.2.1 Solid Carbon Source-Based Production Techniques for Carbon Nanotubes 45
- 3.2.2 Gaseous Carbon Source-Based Production Techniques for Carbon Nanotubes 52
- 3.2.3 Miscellaneous Techniques 57
- 3.2.4 Synthesis of Aligned Carbon Nanotubes 58

3.3 **Growth Mechanisms of Carbon Nanotubes** 59
- 3.3.1 Catalyst-Free Growth 59
- 3.3.2 Catalytically Activated Growth 60

3.4 **Properties of Carbon Nanotubes** 63
- 3.4.1 Variability of Carbon Nanotube Properties 63
- 3.4.2 General Properties 63
- 3.4.3 SWNT Adsorption Properties 63
- 3.4.4 Transport Properties 65
- 3.4.5 Mechanical Properties 67
- 3.4.6 Reactivity 67

3.5 **Carbon Nanotube-Based Nano-Objects** ... 68
- 3.5.1 Hetero-Nanotubes 68
- 3.5.2 Hybrid Carbon Nanotubes 68
- 3.5.3 Functionalized Nanotubes 71

3.6 **Applications of Carbon Nanotubes** 73
- 3.6.1 Current Applications 73
- 3.6.2 Expected Applications Related to Adsorption 76

References 86

Carbon nanotubes have been synthesized for a long time as products from the action of a catalyst over the gaseous species originating from the thermal decomposition of hydrocarbons (see Sect. 3.2). Some of the first evidence that the nanofilaments thus produced were actually nanotubes – i. e., exhibiting an inner cavity – can be found in the transmission electron microscope micrographs published by *Hillert* et al. in 1958 [3.1]. This was of course related to and made possible by the progress in transmission electron microscopy. It is then likely that the carbon filaments prepared by *Hughes* and *Chambers* in 1889 [3.2], reported by *Maruyama* et al. [3.3] as probably the first patent ever deposited in the field, and whose preparation method was also based on the catalytically enhanced thermal cracking of hydrocarbons, were already carbon nanotube-related morphologies. The preparation of vapor-grown carbon fibers had actually been reported as early as more than one century ago [3.4, 5]. Since then, the interest in carbon nanofilaments/nanotubes has been recurrent, though within a scientific area almost limited to the carbon material scientist community. The reader is invited to consult the review published by *Baker* et al. [3.6] regarding the early works. The worldwide enthusiasm came unexpectedly in 1991, after the catalyst-free formation of nearly perfect concentric multiwall carbon nanotubes (c-MWNTs, see Sect. 3.1) was reported [3.7] as by-products of the formation of fullerenes by the electric-arc technique. But the real breakthrough occurred two years later, when attempts in situ to fill the nanotubes with various metals (see Sect. 3.5) led to the discovery – again unexpected – of single-wall carbon nanotubes (SWNTs) simultaneously by *Iijima* et al. [3.8] and *Bethune* et al. [3.9]. Single-wall carbon nanotubes were really new nano-objects many of whose properties and behaviors are quite specific (see Sect. 3.4). They are also beautiful objects for fundamental physics as well as unique molecules for experimental chemistry, still keeping some mystery since their formation mechanisms are a subject of controversy and are still debated (see Sect. 3.3). Potential applications seem countless, although few have reached a marketable status so far (see Sect. 3.6). Consequently, about five papers a day with carbon nanotubes as the main topic are currently published by research teams from around the world, an illustration of how extraordinarily active – and highly competitive – is this field of research. It is a quite unusual situation, similar to that of fullerenes, which, by the way, are again carbon nano-objects structurally closely related to nanotubes.

This is not, however, only about scientific exaltation. Economical aspects are leading the game to a greater and greater extent. According to experts, the world market is predicted to be more than 430 M $ in 2004 and estimated to grow to several b $ before 2009. That is serious business, and it will be closely related to how scientists and engineers will be able to deal with the many challenges found on the path from the beautiful, ideal molecule to the reliable – and it is hoped, cheap – manufactured product.

3.1 Structure of Carbon Nanotubes

It is simple to imagine a single-wall carbon nanotube (SWNT). Ideally, it is enough to consider a perfect graphene sheet (a graphene being the same polyaromatic mono-atomic layer made of an hexagonal display of sp^2 hybridized carbon atoms that genuine graphite is built up with), to roll it into a cylinder (Fig. 3.1) paying attention that the hexagonal rings put in contact join coherently, then to close the tips by two caps, each cap being a hemi-fullerene with the appropriate diameter (Fig. 3.2a–c).

3.1.1 Single-Wall Nanotubes

Geometrically, there is no restriction regarding the tube diameter. But calculations have shown that collapsing the single-wall tube into a flattened two-layer ribbon is energetically more favorable than maintaining the tubular morphology beyond a diameter value of ~ 2.5 nm [3.10]. On the other hand, it intuitively comes to mind that the shorter the radius of curvature, the higher the stress and the energetic cost, although SWNTs with diameters as low as 0.4 nm have been successfully synthesized [3.11]. A suitable energetic compromise is thus reached for ~ 1.4 nm, the most frequent diameter encountered regardless of the synthesis techniques (at least those based on solid carbon source) when conditions for high SWNT yields are used. There is no such restriction for nanotube length, which only depends on limitations brought by the preparation method and the specific conditions used for the synthesis (thermal gradients, res-

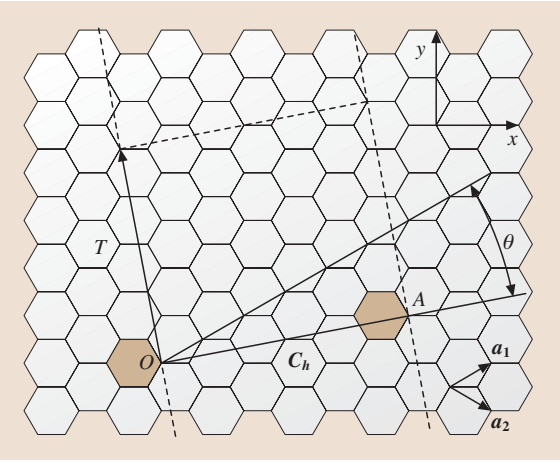


Fig. 3.1 Sketch of the way to make a single-wall carbon nanotube, starting from a graphene sheet (adapted from [3.12])

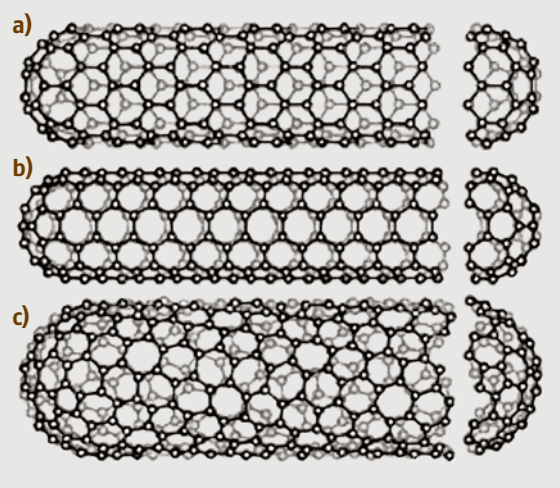


Fig. 3.2a–c Sketch of three different SWNT structures as examples for **(a)** a zig-zag-type nanotube, **(b)** an armchair-type nanotube, **(c)** a helical nanotube (adapted from [3.13])

idence time, etc.). Experimental data are consistent with these statements, since SWNTs wider than 2.5 nm are scarcely reported in literature, whatever the preparation method, while SWNT length can be in the micrometer or the millimeter range. These features make single-wall carbon nanotubes a unique example of single molecules with huge aspect ratios.

Two important consequences derive from the SWNT structure as described above:

1. All carbon atoms are involved in hexagonal aromatic rings only and are therefore in equivalent position, except at the nanotube tips where $6 \times 5 = 30$ atoms at each tip are involved in pentagonal rings (considering that adjacent pentagons are unlikely) – though not more, not less, as a consequence of the Euler's rule that also governs the fullerene structure. In case SWNTs are ideally perfect, their chemical reactivity will therefore be highly favored at the tube tips, at the very location of the pentagonal rings.
2. Though carbon atoms are involved in aromatic rings, the C=C bond angles are no longer planar as they should ideally be. This means that the hybridization of carbon atoms are no longer pure sp^2 but get some percentage of the sp^3 character, in a proportion that increases as the tube radius of curvature decreases. For example, the effect is the same as for the C_{60} fullerene molecules, whose radius of curvature is 0.35 nm, and the subsequent sp^3 character proportion about 30%. On the one hand, this is supposed to make the SWNT surface (though consisting of aromatic ring faces) a bit more reactive than regular, planar graphene, relatively speaking. On the other hand, this somehow induces a variable overlapping of the bands of density of states, thereby inducing a unique versatile electronic behavior (see Sect. 3.4).

As illustrated by Fig. 3.2, there are many ways to roll a graphene into a single-wall nanotube, some of the resulting nanotubes enabling symmetry mirrors both parallel and perpendicular to the nanotube axis (such as the SWNTs from Fig. 3.2a and 3.2b), some others not (such as the SWNT from Fig. 3.2c). By correspondence with the terms used for molecules, the latter are commonly called "chiral" nanotubes. "Helical" should be preferred, however, in order to respect the definition of chirality, which makes all chiral molecules unable to be superimposed on their own image in a mirror. The various ways to roll graphene into tubes are therefore mathematically defined by the vector of helicity $\boldsymbol{C}_\mathrm{h}$, and the angle of helicity θ, as follows (referring to Fig. 3.1):

$$\boldsymbol{OA} = \boldsymbol{C}_\mathrm{h} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$$

with

$$\boldsymbol{a}_1 = \frac{a\sqrt{3}}{2}\boldsymbol{x} + \frac{a}{2}\boldsymbol{y} \quad \text{and} \quad \boldsymbol{a}_2 = \frac{a\sqrt{3}}{2}\boldsymbol{x} - \frac{a}{2}\boldsymbol{y}$$

$$\text{where} \quad a = 2.46\,\text{Å}$$

and

$$\cos\theta = \frac{2n+m}{2\sqrt{n^2+m^2+nm}}$$

where n and m are the integers of the vector $\boldsymbol{OA}$ considering the unit vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.

The vector of helicity $\boldsymbol{C}_\mathrm{h}(=\boldsymbol{OA})$ is perpendicular to the tube axis, while the angle of helicity θ is taken with respect to the so-called zig-zag axis, i. e., the vector of helicity that makes nanotubes of the "zig-zag" type (see below). The diameter D of the corresponding nanotube is related to $\boldsymbol{C}_\mathrm{h}$ by the relation:

$$D = \frac{|\boldsymbol{C}_\mathrm{h}|}{\pi} = \frac{a_\mathrm{CC}\sqrt{3(n^2+m^2+nm)}}{\pi} ,$$

where

$$\underset{\text{(graphite)}}{1.41\,\text{Å}} \le a_{\mathrm{C{=}C}} \le \underset{(\mathrm{C}_{60})}{1.44\,\text{Å}} .$$

The C−C bond length is actually elongated by the curvature imposed, taking the average value for the C_{60} fullerene molecule as a reasonable upper limit, and the value for flat graphenes in genuine graphite as the lower limit (corresponding to an infinite radius of curvature). Since $\boldsymbol{C}_\mathrm{h}$, θ, and D are all expressed as a function of the integers n and m, they are sufficient to define any SWNT specifically, by noting them (n, m). Obtaining the values of n and m for a given SWNT is simple by counting the number of hexagons that separate the extremities of the $\boldsymbol{C}_\mathrm{h}$ vector following the unit vector $\boldsymbol{a}_1$ first, then $\boldsymbol{a}_2$ [3.12]. In the example of Fig. 3.1, the SWNT that will be obtained by rolling the graphene so that the two shaded aromatic cycles superimpose exactly is a (4,2) chiral nanotube. Similarly, SWNTs from Fig. 3.2a to 3.2c are (9,0), (5,5), and (10,5) nanotubes respectively, thereby providing examples of zig-zag–type SWNT (with an angle of helicity $= 0°$), armchair-type SWNT (with an angle of helicity of 30°) and a chiral SWNT, respectively. This also illustrates why the term "chiral" is sometimes inappropriate and should preferably be replaced by "helical". Armchair (n, n) nanotubes, though definitely achiral from a standpoint of symmetry, exhibit a "chiral angle" different from 0. "Zig-zag" and "armchair" qualifications for achiral nanotubes refer to the way carbon atoms are displayed at the edge of the nanotube cross section (Fig. 3.2a and 3.2b). Generally speaking, it is clear from Figs. 3.1 and 3.2a that having the vector of helicity perpendicular to any of the three overall C=C bond directions will provide zig-zag–type SWNTs, noted $(n,0)$, while having the vector of helicity parallel to one of the three C=C bond directions will provide armchair-type SWNTs, noted (n, n). On the other hand, because of the sixfold symmetry of the graphene sheet, the angle of helicity θ for the chiral (n, m) nanotubes is such as $0 < \theta < 30°$. Figure 3.3 provides two examples of what chiral SWNTs look like, as seen by means of atomic force microscopy.

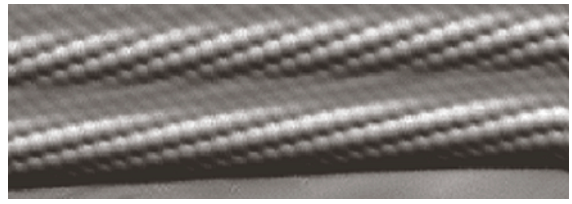


Fig. 3.3 Image of two neighboring chiral SWNTs within a SWNT bundle as seen by high resolution scanning tunneling microscopy (by courtesy of Prof. Yazdani, University of Illinois at Urbana, USA)

Planar graphenes in graphite have π electrons, which are satisfied by the stacking of graphenes that allows van der Waals forces to develop. Similar reasons make fullerenes gather and order into fullerite crystals and SWNTs into SWNT ropes (Fig. 3.4a). Spontaneously, SWNTs in ropes tend to arrange into an hexagonal array, which corresponds to the highest compactness achievable (Fig. 3.4b). This feature brings new periodicities

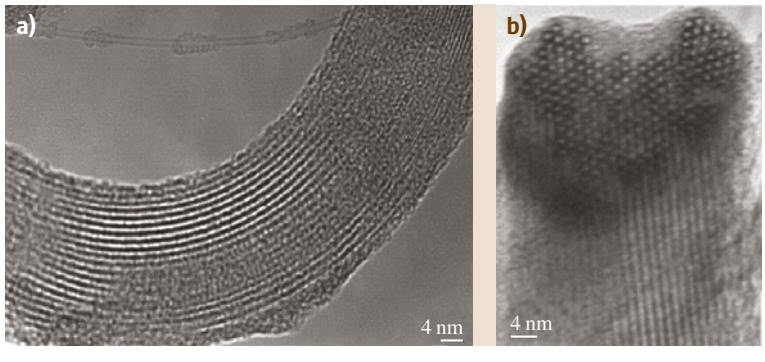


Fig. 3.4a,b High resolution transmission electron microscopy images of a SWNT rope **(a)** longitudinal view. At the top of the image an isolated single SWNT also appears. **(b)** cross section view (from [3.14])

with respect to graphite or turbostratic polyaromatic carbon crystals. Turbostatic structure corresponds to graphenes which are stacked with random rotations or translations instead of being piled up following sequential ABAB positions, as in graphite structure. This implies that no lattice atom plane exists anymore other than the graphene planes themselves (corresponding to the (001) atom plane family). These new periodicities make diffraction patterns specific and quite different from that of other sp^2-carbon-based crystals, although hk reflections, which account for the hexagonal symmetry of graphene plane, are still present. On the other hand, $00l$ reflections, which account for the stacking sequence of graphenes in regular, "multilayered" polyaromatic crystals that does not exist in SWNT ropes, are absent. Such a hexagonal packing of SWNTs within the ropes requires that SWNTs exhibit similar diameters, which is the common case for SWNTs prepared by the electric arc or the laser vaporization processes. SWNTs prepared these ways are actually about 1.35 nm wide (diameter value for a (10,10) tube, among others), for reasons still unclear but related to the growth mechanisms specific to the conditions provided by these techniques (see Sect. 3.3).

3.1.2 Multiwall Nanotubes

Building multiwall carbon nanotubes is a little bit more complex, since it has to deal with the various ways graphenes can be displayed and mutually arranged within a filament morphology. Considering the usual textural versatility of polyaromatic solids, a similar versatility can be expected. Likewise, diffraction patterns no longer differentiate from that of anisotropic polyaromatic solids. The easiest MWNT to imagine is the concentric type (c-MWNT), in which SWNTs with regularly increasing diameters are coaxially displayed according to a Russian-doll model into a multiwall nanotube (Fig. 3.5). Such nanotubes are generally formed either by the electric-arc technique (without need of any catalyst) or by catalyst-enhanced thermal cracking of gaseous hydrocarbons or CO disproportionation (see Sect. 3.2). The number of walls (or number of coaxial tubes) can be anything, starting from two, with no upper limit. The intertube distance is approximately that of the intergraphene distance in turbostratic, polyaromatic solids, i. e. 0.34 nm (as opposed to 0.335 nm in genuine graphite), since the increasing radius of curvature imposed on the concentric graphenes prevents the carbon atoms from being displayed as in graphite, i. e., each of the carbon atoms from a graphene facing alternatively either a ring center or a carbon atom from the neighboring graphene. But two cases allow such a nanotube to reach – totally or partially – the 3-D crystal periodicity of graphite. One is to consider a high number of concentric graphenes, i. e. concentric graphenes with long radius of curvature. In that case, the shift of the relative positions of carbon atoms from superimposed graphenes is so small with respect to that in graphite that some commensurability is possible. It may result in MWNTs with the association of both structures, i. e. turbostratic in the core and graphitic in the outer part [3.15]. The other case occurs for c-MWNTs exhibiting faceted morphologies, originating either from the synthesis process or more likely from

Fig. 3.5 High resolution transmission electron microscopy image (longitudinal view) of a concentric multiwall carbon nanotube (c-MWNT) prepared by electric arc. In *insert*, sketch of the Russian-doll-like display of graphenes

subsequent heat-treatments at high temperatures (e.g., 2,500 °C) in inert atmospheres. Obviously, facets allow graphenes to get back a flat arrangement of atoms – except at the junction between neighboring facets – in which the specific stacking sequence of graphite can develop.

Another frequent inner texture for multiwall carbon nanotubes is the so-called herringbone texture (h-MWNTs), in which graphenes make an angle with respect to the nanotube axis (Fig. 3.6). The angle value varies upon the processing conditions (e.g., the catalyst morphology or the atmosphere composition), from 0 (in which case the texture turns into that of a c-MWNT) to 90° (in that case, the filament is no longer a tube, see below), and the inner diameter varies so that the tubular feature can be lost [3.19], justifying that the latter can be called nanofibers rather than nanotubes. h-MWNTs are exclusively obtained by processes involving catalysts, generally catalyst-enhanced thermal cracking of hydrocarbons or CO disproportionation. One unresolved question is whether the herringbone texture, which actually describes the texture projection rather than the overall three-dimensional texture, originates from the scroll-like spiral arrangement of a single graphene ribbon or from the stacking of independent truncated-cone-like graphenes. Although the former is more likely for energetic reasons since providing a minimal and constant amount of transitorily unsatisfied bonds during the nanotube growth (similar to the well-known growth mechanism from a screw dislocation), the question is still debated.

Another common feature is the occurrence, at a variable frequency, of a limited amount of graphenes oriented perpendicular to the nanotube axis, thus forming the so-called "bamboo" texture. It cannot be a texture by its own but affects in a variable extent either c-MWNT (bc-MWNT) or h-MWNT (bh-MWNT) textures (Figs. 3.6 and 3.7). The question may be addressed whether such filaments, though hollowed, should still be called nanotubes, since the inner cavity is no longer opened all along the filament as it is for a genuine tube. They therefore are sometimes referred as "nanofibers" in literature.

Nanofilaments that definitely cannot be called nanotubes are those built from graphenes oriented perpendicular to the filament axis and stacked as piled-up plates. Although they actually correspond to h-MWNTs with the graphene/MWNT axis angle = 90°, the occurrence of inner cavity is no longer possible, and such filaments are therefore most often referred to as "platelet-nanofibers" in literature [3.19].

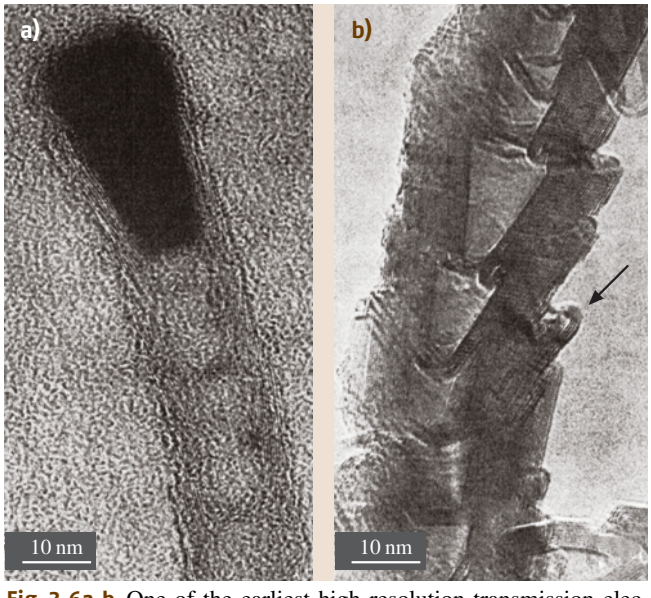


Fig. 3.6a,b One of the earliest high resolution transmission electron microscopy image of a herringbone (and bamboo) multi-wall nanotube (bh-MWNT, longitudinal view) prepared by CO disproportionation on Fe-Co catalyst. **(a)** as-grown. The nanotube surface is made of free graphene edges. **(b)** after 2,900 °C heat-treatment. Both the herringbone and the bamboo textures have become obvious. Graphene edges from the surface have buckled with neighbors (*arrow*), closing the access to the intergraphene spacing (adapted from [3.16])

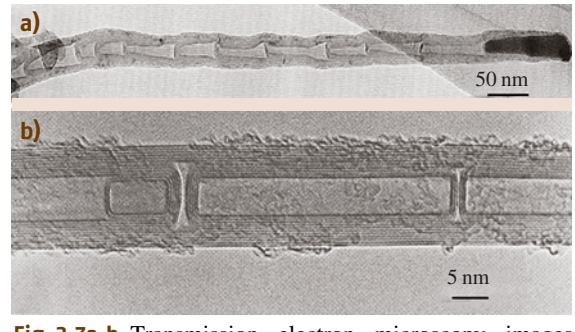


Fig. 3.7a,b Transmission electron microscopy images from bamboo-multi-wall nanotubes (longitudinal views). **(a)** low magnification of a bamboo-herringbone multi-wall nanotube (bh-MWNT) showing the nearly periodic feature of the texture, which is very frequent. (from [3.17]); **(b)** high resolution image of a bamboo-concentric multi-wall nanotube (bc-MWNT) (modified from [3.18])

As opposed to SWNTs, whose aspect ratio is so high that acceding the tube tips is almost impossible, aspect ratio for MWNTs (and carbon nanofibers) are generally lower and often allow one to image tube ends by transmission electron microscopy. Except for c-MWNTs from electric arc (see Fig. 3.5) that grow following a catalyst-free process, nanotube tips are frequently found associated with the catalyst crystal from which they have been formed.

MWNT properties (see Sect. 3.4) will obviously depend on the perfection and orientation of graphenes in the tube more than any other feature (e.g., the respective spiral angle of the constituting nanotubes for c-MWNTs has little importance). Graphene orientation is a matter of texture, as described above. Graphene perfection is a matter of nanotexture, which is commonly used to describe other polyaromatic carbon materials, and which is quantified by several parameters preferably obtained from high resolution transmission electron microscopy (Fig. 3.8). Both texture and nanotexture depend on the processing conditions. While the texture type is a permanent, intrinsic feature only able to go toward complete alteration upon severe degradation treatments (e.g., oxidation), however, nanotexture can be improved by subsequent thermal treatments at high temperatures (e.g., > 2,000 °C) or, reversibly, possibly degraded by chemical treatments (e.g., slight oxidation conditions).

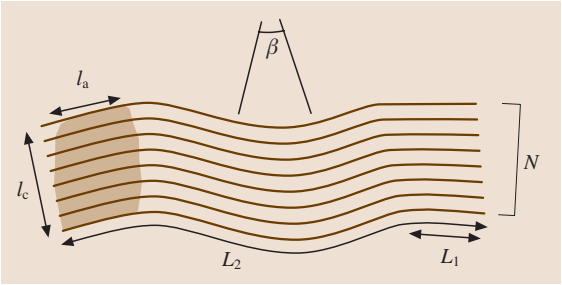


Fig. 3.8 Sketch explaining the various parameters obtained from high resolution (lattice fringe mode) transmission electron microscopy for quantifying nanotexture: L_1 is the average length of perfect (distortion-free) graphenes of coherent areas; N is the number of piled-up graphenes in the coherent (distortion-free) areas; L_2 is the average length of continuous though distorted graphenes within graphene stacks; β is the average distortion angle. L_1 and N are more or less related to L_a and L_c obtained from X-ray diffraction

3.2 Synthesis of Carbon Nanotubes

Producing carbon nanotubes so that the currently planned applications become marketable requires solving some problems that are more or less restrictive depending on the cases. One example is to specifically control the configuration (chirality), the purity, or the structural quality of SWNTs, with the production capacity adapted to the application. One condition would be to understand perfectly the mechanism of nanotube nucleation and growth, which remains the object of a lot of controversy in spite of an intense, worldwide experimental effort. This problem is partly explained by the lack of knowledge regarding several parameters controlling the synthesis conditions. For instance, neither the temperatures of nanotube condensation and formation nor the manner in which they are influenced by the synthesis parameters are known. Equally often unknown is the exact and accurate role of the catalysts in nanotube growth. Given the large number of experimental parameters and considering the large range of conditions that the several synthesis techniques correspond to, it is quite legitimate to think of more than one mechanism intervening in the nanotube formation.

3.2.1 Solid Carbon Source-Based Production Techniques for Carbon Nanotubes

Among the different SWNT production techniques, the three processes (laser ablation, solar energy, and electric arc) presented in this section have at least two common points: a high temperature ($1{,}000\,\mathrm{K} < T < 6{,}000\,\mathrm{K}$) medium and the fact that the carbon source originates from the erosion of solid graphite. In spite of these common points, both the morphologies of the carbon nanostructures and the SWNT yields can differ notably with respect to the experimental conditions.

Before being utilized for carbon nanotube synthesis, these techniques permitted the production of fullerenes. Laser vaporization of graphite was actually the very first method to demonstrate the existence of fullerenes, including the most popular one (because the most stable and therefore the most abundant), C_{60} [3.20]. On the other hand, the electric arc technique was (and still is) the first method to produce fullerenes in relatively large quantities [3.21]. As opposed to the fullerene formation, which requires the presence of carbon atoms

in high temperature media and the absence of oxygen, the utilization of these techniques for the synthesis of nanotubes requires an additional condition, i. e. the presence of catalysts in one of the electrode or the target.

In case of these synthesis techniques requiring relatively high temperatures to sublime graphite, the different mechanisms such as the carbon molecule dissociation and the atom recombination processes take place at different time scales, from nanosecond to microsecond and even millisecond. The formation of nanotubes and other graphene-based products appears afterward with a relatively long delay.

The methods of laser ablation, solar energy, and electric arc are based on an essential mechanism, i. e. the energy transfer resulting from the interaction between either the target material and an external radiation (laser beam or radiation emanating from solar energy) or the electrode and the plasma (in case of electric arc). This interaction is at the origin of the target or anode erosion leading to the formation of a plasma, i. e. an electrically neutral ionized gas, composed of neutral atoms, charged particles (molecules and ionized species), and electrons. The ionization degree of this plasma, defined by the ratio $(n_e/n_e + n_o)$, where n_e and n_o are the density of electrons and that of neutral atoms respectively, highlights the importance of energy transfer between the plasma and the material. The characteristics of this plasma and notably the fields of temperature and of concentration of the various species present in the plasma thereby depend not only on the nature and composition of the target or that of the electrode but also on the energy transferred.

One of the advantages of these different synthesis techniques is the possibility of varying a large number of parameters, which allow the composition of the high temperature medium to be modified and, consequently, the most relevant parameters to be determined, in order to define the optimal conditions for the control of carbon nanotube formation. But a major drawback of these techniques – as for any other technique for the production of SWNTs – is that SWNTs never come pure, i. e. they are associated with other carbon phases and catalyst remnants. Although purification processes are proposed in literature and by some commercial companies to rid these undesirable phases, they are all based on oxidation (e.g. acidic) processes that are likely to affect deeply the SWNT structure [3.14]. Subsequent thermal treatments at $\sim 1{,}200\,°C$ under inert atmosphere, however, succeed somewhat in recovering the former structure quality [3.24].

Laser Ablation

After the first laser was built in 1960, physicists immediately made use of it as a means of concentrating a large quantity of energy inside a very small volume within a relatively short time. The consequences of this energy input naturally depends on the characteristics of the device employed. During the interaction between the laser beam and the material, numerous phenomena superimpose and/or follow each other within the time range, each of these processes being sensitive to such different parameters as the laser beam characteristics, the incoming power density (also termed "fluence"), the nature of the target, and the environment surrounding it. For instance, the solid target can merely heat up, or melt, or vaporize depending on the power provided.

While this technique was used successfully to synthesize fullerene-related structures for the very first time [3.20], the synthesis of SWNTs by laser ablation came only ten years later [3.22].

Laser Ablation – Experimental Devices

Two types of laser devices are utilized nowadays for carbon nanotube production: lasers operating in pulsed mode on the one hand and lasers operating in continuous mode on the other hand, the latter generally providing a smaller influence.

An example of the layout indicating the principle of a laser ablation device is given in Fig. 3.9. A graphite pellet containing the catalyst is put in the middle of an inert gas-filled quartz tube placed in an oven maintained at a temperature of $1{,}200\,°C$ [3.22, 23]. The energy of the laser beam focused on the pellet permits it to vaporize and sublime the graphite by uniformly bombarding its surface. The carbon species swept by a flow of neutral gas are thereafter deposited as soot in different regions: on the conical water-cooled copper collector, on the quartz tube walls, and on the backside of the pellet.

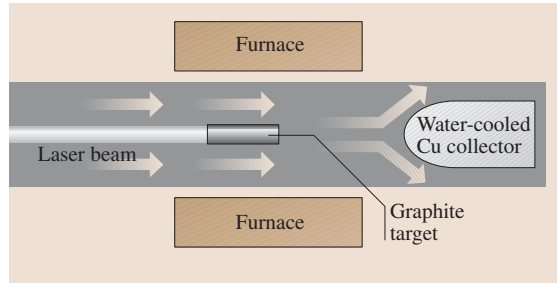


Fig. 3.9 Sketch of an early laser vaporization apparatus (adapted from [3.22, 23])

Various improvements have been made to this device in order to increase the production efficiency. For example, *Thess* et al. [3.25] employed a second pulsed laser that follows the initial impulsion but at a different frequency in order to ensure a more complete and efficient irradiation of the pellet. This second impulsion has the role of vaporizing the coarse aggregates issued from the first ablation and thereby making them participate in the active carbon feedstock involved in the nanotube growth. Other modifications were brought by *Rinzler* et al. [3.24], who inserted a second quartz tube of a smaller diameter coaxially disposed inside to the first one. This second tube has the role of reducing the vaporization zone and thereby permitting an increase in the quantity of sublimed carbon. They also arranged to place the graphite pellet on a revolving system so that the laser beam uniformly scans its whole surface.

Other groups have realized that, as far as the target contains both the catalyst and the graphite, the latter evaporates in priority and the pellet surface becomes more and more metal rich, resulting in a decrease of the efficiency in nanotube formation in the course of the process. To solve this problem, *Yudasaka* et al. [3.27] utilized two pellets facing each other, one entirely made from the graphite powder and the other from an alloy of transition metals (catalysts), and irradiated simultaneously.

A sketch of a synthesis reactor based on the vaporization of a target at a fixed temperature by a continuous CO_2 laser beam ($\lambda = 10.6\,\mu m$) is shown in Fig. 3.10 [3.26]. The power can be varied from 100 W to 1,600 W. The temperature of the target is measured with an optical pyrometer, and these measurements are used to regulate the laser power to maintain a constant vaporization temperature. The gas, heated by the contact with the target, acts as a local furnace and creates an extended hot zone, making an external furnace unnecessary. The gas is extracted through a silica pipe, and the solid products formed are carried away by the gas flow through the pipe and then collected on a filter. The synthesis yield is controlled by three parameters: the cooling rate of the medium where the active, secondary catalyst particles are formed, the residence time, and the temperature (in the 1,000–2,100 K range) at which SWNTs nucleate and grow [3.28].

But devices equipped with facilities to gather in situ data such as the target temperature are few and, generally speaking, among the numerous parameters of the laser ablation synthesis technique. The most studied are the nature of the target, the nature and concentration of the catalysts, the nature of the neutral gas flow, and the temperature of the outer oven (when any).

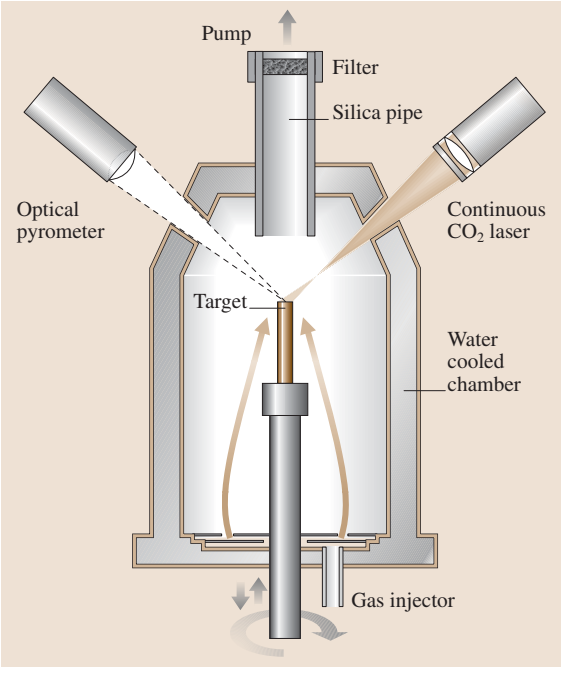


Fig. 3.10 Sketch of a synthesis reactor with a continuous CO_2 laser device (adapted from [3.26])

Laser Ablation – Results

In the absence of catalysts in the target, the soot collected mainly contains multiwall nanotubes (c-MWNTs). Their lengths can reach 300 nm. Their quantity and structure quality are dependent on the oven temperature. The best quality is obtained for an oven temperature set at 1,200 °C. At lower oven temperatures, the structure quality decreases, and the nanotubes start presenting many defects [3.23]. As soon as small quantities (few percents or less) of transition metals (Ni, Co) playing the role of catalysts are incorporated into the graphite pellet, products yielded undergo significant modifications, and SWNTs are formed instead of MWNTs. The yield of SWNTs strongly depends on the type of metal catalyst used and is seen to increase with furnace temperature, among other factors. The SWNTs have remarkably uniform diameter and they self-organize into rope-like crystallites 5–20 nm in diameter and tens to hundreds of micrometers in length (Fig. 3.11). The ends of all SWNTs appear to be perfectly closed with hemispherical end caps showing no evidence of any as-

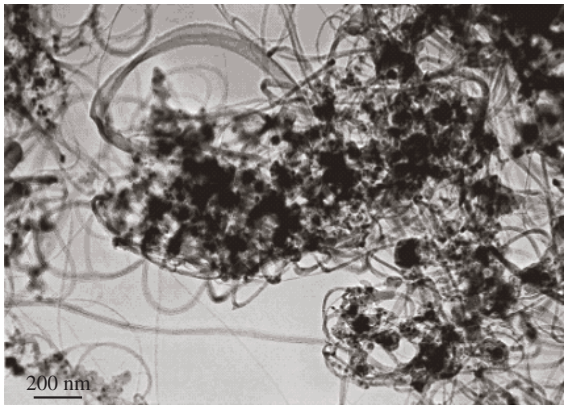


Fig. 3.11 Low magnification TEM images of a typical raw SWNT material obtained from the laser vaporization technique. Fibrous morphologies are SWNT bundles, and dark particles are catalyst remnants. Raw SWNT materials from electric arc exhibit a similar aspect (from [3.14])

sociated metal catalyst particle, although, as pointed out in Sect. 3.1, finding the two tips of a SWNT is rather challenging, considering the huge aspect ratio of the nanotube and their entanglement. Another feature of SWNTs produced with this technique is that they are supposedly "cleaner" than those produced employing other techniques, i. e., associated with a lower amount of an amorphous carbon phase, either coating the SWNTs or gathered into nano-particles. Such an advantage, however, stands only for synthesis conditions specifically set to ensure high quality SWNTs. It can no longer be true when conditions are such that high yields are preferred, and SWNTs from electric arc may appear cleaner than SWNTs from laser vaporization [3.14].

The laser vaporization technique is one of the three methods currently used to prepare SWNTs as commercial products. SWNTs prepared that way were first marketed by Carbon Nanotechnologies Inc. (Texas, USA), with prices as high as $ 1,000/g (raw materials) until December 2002. As a probable consequence of the impossibility to lower the amount of impurities in the raw materials, they have recently decided to focus on fabricating SWNTs by the HiPCo technique (see Sect. 3.2.2). Though laser-based method are generally considered not competitive in the long term for the low-cost production of SWNTs compared to CCVD-based methods (see Sect. 3.2.2), prediction of prices as low as $ 0.03 per gram of raw high concentration SWNT soot are expected in the near future (Acolt S.A., Switzerland).

Electric-Arc Method

Electric arcs between carbon electrodes have been studied as light sources and radiation standards for a very long time. Lately they have received renewed attention for their use in producing new fullerenes-related molecular carbon nanostructures such as genuine fullerenes or nanotubes. This technique was first brought to light by *Krätschmer* et al. [3.21] who utilized it to achieve the production of fullerenes in macroscopic quantities. In the course of investigating the other carbon nanostructures formed along with the fullerenes and more particularly the solid carbon deposit forming onto the cathode, Iijima [3.7] discovered the catalyst-free formation of perfect c-MWNT-type carbon nanotubes. Then, as mentioned in the introduction of this chapter, the catalyst-promoted formation of SWNTs was incidentally discovered after some amounts of transition metals were introduced into the anode in an attempt to fill the c-MWNTs with metals while they grow [3.8, 9]. Since then, a lot of work has been carried out by many groups using this technique to understand the mechanisms of nanotube growth as well as the role played by the catalysts (when any) for the synthesis of MWNTs and/or SWNTs [3.29–41].

Electric-Art Method – Experimental Devices

The principle of this technique is to vaporize carbon in the presence of catalysts (iron, nickel, cobalt, yttrium, boron, gadolinium, and so forth) under reduced atmosphere of inert gas (argon or helium). After the triggering of the arc between two electrodes, a plasma is formed consisting of the mixture of carbon vapor, the rare gas (helium or argon), and the vapors of catalysts. The vaporization is the consequence of the energy transfer from the arc to the anode made of graphite doped with catalysts. The anode erosion rate is more or less important depending on the power of the arc and also on the other experimental conditions. It is noteworthy that a high anode erosion does not necessarily lead to a high carbon nanotube production.

An example of a reactor layout is shown in Fig. 3.12. It consists of a cylinder of about 30 cm in diameter and about 1 m in height, equipped with diametrically opposed sapphire windows located so that they face the plasma zone in view of observing the arc. The reactor possesses two valves, one for carrying out the primary evacuation (0.1 Pa) of the chamber, the other permitting it to fill with a rare gas up to the desired working pressure.

Contrary to the solar energy technique, SWNTs are deposited (provided appropriate catalysts are used) in

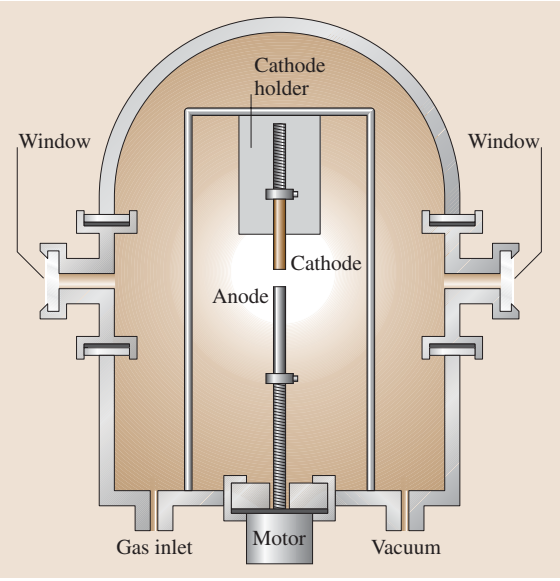


Fig. 3.12 Sketch of an electric arc reactor

different regions of the reactor: (1) the collaret, which forms around the cathode; (2) the web-like deposits found above the cathode; (3) the soot deposited all around the reactor walls and bottom. On the other hand, MWNTs are formed embedded in a hard deposit adherent to the cathode, whenever catalysts are used or not. The cathode deposits formed under the cathode. The formation of collaret and web is not systematic and depends on the experimental conditions as indicated in Table 3.1, as opposed to cathode deposit and soot that are systematically found.

Two graphite rods of few millimeters in diameter constitute the electrodes between which a potential difference is applied. The dimensions of these electrodes vary according to the authors. In certain cases, the cathode has a greater diameter than the anode in order to facilitate their alignment [3.32, 42]. Other authors utilize electrodes of the same diameter [3.41]. The whole device can be designed horizontally [3.33, 41] or vertically [3.34, 36–38]. The advantage of the latter is the symmetry brought by the verticality with respect to gravity, which facilitates computer modeling (regarding convection flows, for instance).

Two types of anodes can be utilized as soon as catalysts need to be introduced: (1) graphite anodes in which a coaxial hole is drilled several centimeters in length and in which catalyst and graphite powders are mixed; (2) graphite anodes in which the catalysts are homogeneously dispersed [3.43]. The former are by far the most popular, due to their ease of fabrication.

The optimization of the process regarding the nanotube yield and quality is attempted by studying the role of various parameters such as the type of doped anode (homogeneous or heterogeneous catalyst dispersion), the nature as well as the concentration of catalysts, the nature of the plasmagenic gas, the buffer gas pressure, the arc current intensity, and the distance between electrodes. Investigating the result of varying these parameters on the type and amount of carbon nanostructures formed is, of course, the preliminary work that has been done. Though electric arc reactors equipped with the related facilities are few (see Fig. 3.12), investigating the missing link, i. e. the effect of varying the parameters on the plasma characteristics (fields of concentration species and temperature), is likely to provide a more comprehensive understanding of the phenomena involved in nanotube formation. This has been recently developed by atomic and molecular optical emission spectroscopy [3.34, 36–39, 41].

Finally, mention has to be made of attempts to make the electric arc within liquid media such as liquid nitrogen [3.44] or water [3.45, 46], the goal being easier processing since they do not require pumping devices or a closed volume and therefore are likely to allow continuous synthesis. This adaptation has not, however, reached the state of mass production.

Electric-Arc Method – Results

In view of the numerous results obtained with this electric-arc technique, it appears clearly that both the nanotube morphology and the nanotube production efficiency strongly depend on the experimental conditions and, in particular, on the nature of the catalysts. It is worth noting that the products obtained do not consist solely in carbon nanotubes but also in nontubular forms of carbon such as nanoparticles, fullerene-like structures including C_{60}, poorly organized polyaromatic carbons, nearly amorphous nanofibers, multiwall shells, single-wall nanocapsules, and amorphous carbon as reported in Table 3.1 [3.35, 37, 38]. In addition, catalyst remnants are found all over the place, i. e. in the soot, collaret, web, and cathode deposit in various concentration. Generally, at pressure value of about 600 mbar of helium, for 80 A arc current and 1 mm electrode gap, the use of Ni/Y as coupled catalysts favors more particularly the synthesis of SWNTs [3.8, 33, 47]. For such conditions providing high SWNT yields, SWNT concentrations are higher in the collaret (in the range of 50–70 %), then in the web (∼ 50% or less), then in the soot. On the

Table 3.1 Different carbon morphologies obtained by changing the type of anode, the type of catalysts, and pressure value, in a series of arc-discharge experiments (electrode gap = 1 mm)

Catalyst (atom%) Arc conditions	0.6Ni + 0.6Co (homogeneous anode) $P \sim 60$ kPa $I \sim 80$ A	0.6Ni + 0.6Co (homogeneous anode) $P \sim 40$ kPa $I \sim 80$ A	0.5Ni + 0.5Co $P \sim 60$ kPa $I \sim 80$ A	4.2Ni + 1Y $P \sim 60$ kPa $I \sim 80$ A
Soot	• **MWNT** + **MWS** + **POPAC** or **Cn** ± catalysts $\phi \sim 3$–35 nm • NANF + catalysts • AC particles + catalysts • [*DWNT*], [*SWNT*], ropes or isolated, + POPAC	• **POPAC** and **AC** particles + **catalysts** $\phi \sim 2$–20 nm • NANF + catalysts $\phi \sim 5$–20 nm + MWS • [*SWNT*] $\phi \sim 1$–1.4 nm, distorted or damaged, isolated or ropes + Cn	• **AC** and POPAC particles + catalysts $\phi \sim 3$–35 nm • **NANF** + catalysts $\phi \sim 4$–15 nm • [*SWNT*] $\phi \sim 1.2$ nm, isolated or ropes	• **POPAC** and **AC** + particles + catalysts $\phi \leq 30$ nm • SWNT $\phi \sim 1.4$ nm, clean + Cn, short with tips, [*damaged*], isolated or ropes $\phi \leq 25$ nm • [*SWNC*] particles
Web	• [*MWNT*], **DWNT**, ϕ 2.7 – **4** – 5.7 nm SWNT ϕ 1.2–1.8 nm, isolated or **ropes** $\phi < 15$ nm, + POPAC ± **Cn** • AC particles + catalysts $\phi \sim 3$–40 nm + MWS • [*NANF*]	None	None	• **SWNT**, $\phi \sim 1.4$ nm, isolated or **ropes** $\phi \leq 20$ nm, + **AC** • POPAC and AC particles + catalysts $\phi \sim 3$ – **10** – 40 nm + MWS
Collaret	• **POPAC** and **SWNC** particles • Catalysts $\phi \sim 3$–250 nm, < **50** nm + MWS • SWNT ϕ 1–1.2 nm, [*opened*], **distorted**, isolated or **ropes** $\phi < 15$ nm, + **Cn** • [*AC*] particles	• **AC** and **POPAC** particles + catalysts $\phi \sim 3$–25 nm • SWNT $\phi \sim 1$–1.4 nm clean + Cn, [*isolated*] or ropes $\phi < 25$ nm • Catalysts $\phi \sim 5$–50 nm + MWS, • [SWNC]	• **Catalysts** $\phi \sim 3$–170 nm + MWS • AC or POPAC particles + **catalysts** $\phi \sim 3$–50 nm • SWNT $\phi \sim 1.4$ nm clean + Cn isolated or ropes $\phi < 20$ nm	• **SWNT** $\phi \sim 1.4$–2.5 nm, clean + Cn, [*damaged*], isolated or **ropes** $\phi < 30$ nm • POPAC or AC particles + catalysts $\phi \sim 3$–30 nm • [*MWS*] + catalysts or catalyst-free.
Cathode deposit	• **POPAC** and SWNC particles • **Catalysts** $\phi \sim$ 5–300 nm MWS • MWNT $\phi < 50$ nm • [*SWNT*] $\phi \sim 1.6$ nm clean + Cn, isolated or ropes	• **POPAC** and **SWNC** particles + Cn • Catalysts $\phi \sim 20$–100 nm + MWS	• **MWS**, catalyst-free • **MWNT** $\phi < 35$ nm • POPAC and PSWNC particles • [*SWNT*], isolated or ropes • [*Catalysts*] $\phi \sim 3$–30 nm	• **SWNT** $\phi \sim 1.4$–4.1 nm, clean + **Cn**, short with tips, isolated or ropes $\phi \leq 20$ nm. • **POPAC** or **AC** particles + catalysts $\phi \sim 3$–30 nm • MWS + catalysts $\phi < 40$ nm or catalyst-free • [*MWNT*]

Abundant – Present – [*Rare*]

Glossary: AC: amorphous carbon; POPAC: poorly organised polyaromatic carbon; Cn: fullerene-like structure, including C_{60}; NANF: nearly amorphous nanofiber; MWS: multiwall shell; SWNT: single-wall nanotube; DWNT: double-wall nanotube, MWNT: multiwall nanotube; SWNC: single-wall nanocapsule.

other hand, c-MWNTs are found in the cathode deposit. SWNT lengths are micrometric and, typically, outer diameters range around 1.4 nm. Taking the latter conditions as a reference experiment (Table 3.1, column 4), Table 3.1 illustrates the consequence of changing the parameters. For instance (Table 3.1, column 3), using Ni/Co instead of Ni/Y as catalysts prevents the formation of SWNTs. But when the Ni/Co catalysts are homogeneously dispersed in the anode (Table 3.1, column 1), the formation of nanotubes is promoted again, but MWNTs with two or three walls prevail over SWNTs, among which DWNTs (double-wall nanotubes) make a majority. But it is enough to decrease the ambient pressure from 60 to 40 kPa (Table 3.1, column 2) for getting back to conditions where the nanotube formation is impeached.

Recent works have attempted to replace graphite powder (sp^2 hybridized carbon) by diamond powder (sp^3 hybridized carbon) to mix with the catalyst powder and fill hollowed-type graphite anodes. The result was an unexpected but quite significant increase (up to +230%) in the SWNT yield [3.39,40]. Such experiments reveal, as for the comparison between the results from using homogeneous instead of heterogeneous anodes that the physical phenomena (charge and heat transfers) that occurred in the anode during the arc are of utmost importance, which was neglected until now.

It is clear that while the use of rare earth element (such as Y) as a single catalyst does not provide the conditions to grow SWNTs, associating it with a transition metal (such as Ni/Y) seems to correspond to the most appropriate combinations, leading to the highest SWNT yields [3.42]. On the other hand, using a single rare earth element may lead to unexpected results, such as the anticipated closure of graphene edges from a c-MWNT wall with the neighboring graphene edges from the same wall side, leading to the preferred formation of telescope-like and open c-MWNTs able to contain nested Gd crystals [3.36,38]. Such a need for bimetallic catalysts has just begun to be understood as a possible requirement to promote the transitory formation of nickel particles coated with yttrium carbide, whose lattice constants are somewhat commensurable with that of graphenes [3.48].

Figure 3.13 illustrates the kind of information provided by the analysis of the plasma by means of emission spectroscopy, i. e. radial temperature (Fig. 3.13a) or C_2 species concentration (Fig. 3.13b) profiles in the plasma. A common feature is that a huge vertical gradient ($\sim$ 500 K/mm) rapidly establishes (at $\sim$ 0.5 mm from the center in the radial direction) from the bottom to the top of the plasma, as a probable consequence of convection phenomena (Fig. 3.13a). The zone of actual SWNT formation is beyond the limit of the volume analyzable in the radial direction, corresponding to colder areas. The C_2 concentration increases dramatically from the anode to the cathode and decreases dramatically in the radial direction (Fig. 3.13b). This demonstrates that C_2 moieties are secondary products resulting from the recombination of primary species formed from the anode. It also suggests that C_2 moieties may be the building blocks for MWNTs (formed at the cathode) but not for SWNTs [3.38,40].

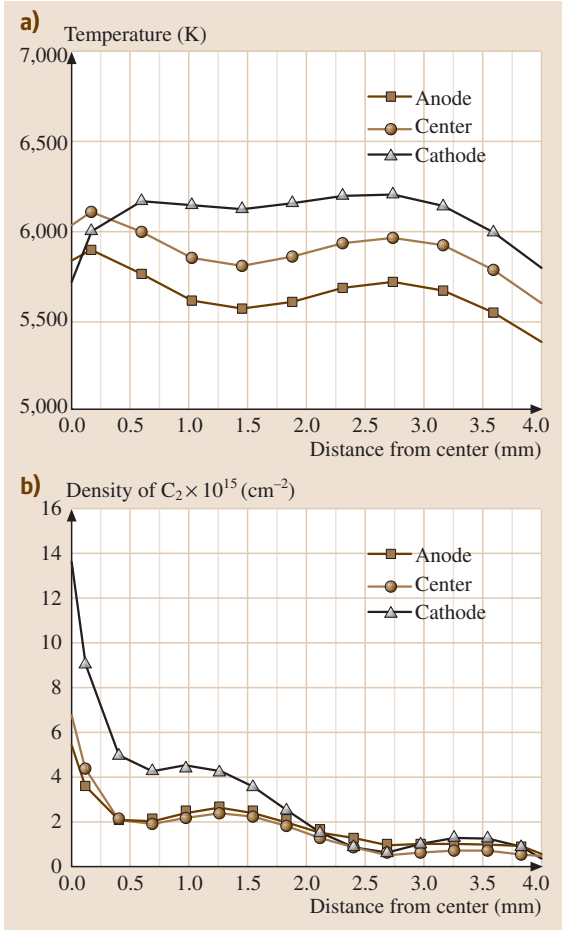


Fig. 3.13a,b Typical temperature **(a)** and C_2 concentration **(b)** profiles in the plasma at the anode surface (*square*), at the plasma center (*dot*), and at the cathode surface (*triangle*) for "standard" conditions (see text). Gradients are similar whatever the catalyst, although absolute value may vary

Although many aspects of it still need to be understood, the electric-arc method is one of the three methods currently used to produce SWNTs as commercial products. At the French company Nanoledge S.A. (Montpellier, France), for instance, current production (for 2003) reaches ∼ 20 kg/year (raw SWNTs, i. e. not purified), with a marketed price of ∼ 90 Euros/g, which is much cheaper than any other production method so far. A decrease to 2–5 Euros/g is expected for sometime before 2007, with a SWNT production capacity of 4–5 tons/year. Amazingly, raw SWNTs from electric arc proposed by Bucky USA (Texas, USA) are still proposed at a marketed price of $ 1,000/g.

Solar Furnace

Solar furnace devices were originally utilized by several groups to produce fullerenes [3.49–51]. *Heben* et al. [3.52] and *Laplaze* et al. [3.53] later modified their former devices in order to use them for carbon nanotube production. This modification consisted mainly in using more powerful ovens [3.54, 55].

Solar Furnace – Experimental Devices

The principle of this technique is again based on the sublimation in an inert gas of a mixture of graphite powder and catalysts formerly placed in a crucible. An example of such a device is shown in Fig. 3.14. The solar rays are collected by a plain mirror and reflected toward a parabolic mirror that focuses them directly onto a graphite pellet under controlled atmosphere. The high temperature of about 4,000 K permits both the carbon and the catalysts to vaporize. The vapors are then dragged by the neutral gas and condense onto the cold walls of the thermal screen. The reactor consists of a brass support cooled by water circulation, on which Pyrex® chambers with various shapes can be fixed (Fig. 3.14b). This support contains a watertight passage permitting the introduction of the neutral gas and a copper rod on which the target is mounted. The target is surrounded by a graphite tube that plays both the role of a thermal screen to reduce radiation losses (very important in the case of graphite) and the role of a duct to lead carbon vapors to a filter in order to avoid soot deposits on the Pyrex® chamber wall. A graphite crucible filled with powdered graphite (for fullerene synthesis) or a mixture of graphite and catalysts (for nanotube synthesis) is utilized in order to reduce the conduction losses.

These studies primarily investigated the target composition, the type and concentration of catalysts, the flow-rate, the composition and pressure of the plasmagenic gas inside the chamber, and the oven power. The objectives are similar to that of works carried out with the other solid-carbon-source-based processes. When possible, specific in situ diagnostics (pyrometry, optical emission spectroscopy, etc.) are also performed in order to understand better the role of the various parameters (temperature measurements at the crucible surface, along the graphite tube acting as thermal screen, C_2 radical concentration in the immediate vicinity of the crucible).

Solar Furnace – Results

Some results obtained by different groups concerning the influence of the catalysts can be summarized as follows. With Ni/Co and at low pressure, the sample collected contains mainly MWNTs with bamboo texture, carbon shells, and some bundles of SWNTs [3.54]. At higher pressures, only bundles of SWNTs are obtained with fewer carbon shells. With Ni/Y and at a high pressure, relatively long bundles of SWNTs are observed. With Co, SWNT bundles are obtained in the soot with SWNT diameters ranging from 1 to 2 nm. Laplaze et al. [3.54] observed very few nanotubes but a large quantity of carbon shells.

In order to proceed to a large-scale synthesis of single-wall carbon nanotubes, which is still a challenge for chemical engineers, *Flamant* et al. [3.56] recently demonstrated that solar energy–based synthesis is a versatile method to obtain SWNTs and can be scaled from 0.1–0.5 g/h to 10 g/h and then to 100 g/h productivity using existing solar furnaces. Experiments at medium scale (10-g/h, 50 kW solar power) have proven the feasibility of designing and building such a reactor and of the scaling-up method. Numerical simulation was meanwhile performed in order to improve the selectivity of the synthesis, in particular by controlling the carbon vapor cooling rate.

3.2.2 Gaseous Carbon Source-Based Production Techniques for Carbon Nanotubes

As mentioned in the introduction of this chapter, the catalysis-enhanced thermal cracking of gaseous carbon source (hydrocarbons, CO) – commonly referred to as catalytic chemical vapor deposition (CCVD) – has been known to produce carbon nanofilaments for a long time [3.4], so that reporting all the works published in the field since the beginning of the century is almost impossible. Until the 90s, however, carbon nanofilaments were mainly produced to act as a core substrate for the subsequent growth of larger (micro-

metric) carbon fibers – so-called vapor-grown carbon fibers – by means of thickening through mere catalyst-free CVD processes [3.57, 58]. We therefore are going to focus instead on the more recent attempts made to prepare genuine carbon nanotubes.

The synthesis of carbon nanotubes (either single- or multiwalled) by CCVD methods involves the catalytic decomposition of a carbon containing source on small metallic particles or clusters. This technique involves either an heterogeneous process if a solid substrate has a role or an homogeneous process if everything takes place in the gas phase. The metals generally used for these reactions are transition metals, such as Fe, Co, and Ni. It is a rather "low temperature" process compared to arc-discharge and laser-ablation methods, with the formation of carbon nanotubes typically occurring between 600 °C and 1,000 °C. Because of the low temperature, the selectivity of the CCVD method is generally better for the production of MWNTs with respect to graphitic particles and amorphous-like carbon, which remain an important part of the raw arc-discharge SWNT samples, for example. Both homogeneous and heterogeneous processes appear very sensitive to the nature and the structure of the catalyst used, as well as to the operating conditions. Carbon nanotubes prepared by CCVD methods are generally much longer (few tens to hundreds of micrometers) than those obtained by arc discharge (few micrometers). Depending on the experimental conditions, it is possible to grow dense arrays of nanotubes. It is a general statement that MWNTs from CCVD contain more structural defects (i. e., exhibit a lower nanotexture) than MWNTs from arc, due to the lower temperature reaction, which does not allow any structural rearrangements. These defects can be removed by subsequently applying heat treatments in vacuum or inert atmosphere to the products. Whether such a discrepancy is also true for SWNTs remains questionable. CCVD SWNTs are generally gathered into bundles that are generally of smaller diameter (few tens of nm) than their arc-discharge and laser-ablation counterparts (around 100 nm in diameter). CCVD provides reasonably good perspectives of large-scale and low-cost processes for the mass production of carbon nanotubes, a key point for their applications at the industrial scale.

A last word concerns the nomenclature. Because work in the field started more than one century ago, denomination of the carbon objects prepared by this method has changed with time with regard to the authors, research areas, and fashions. The same objects are found to be called vapor grown carbon fibers, nanofilaments, nanofibers, and from now on,

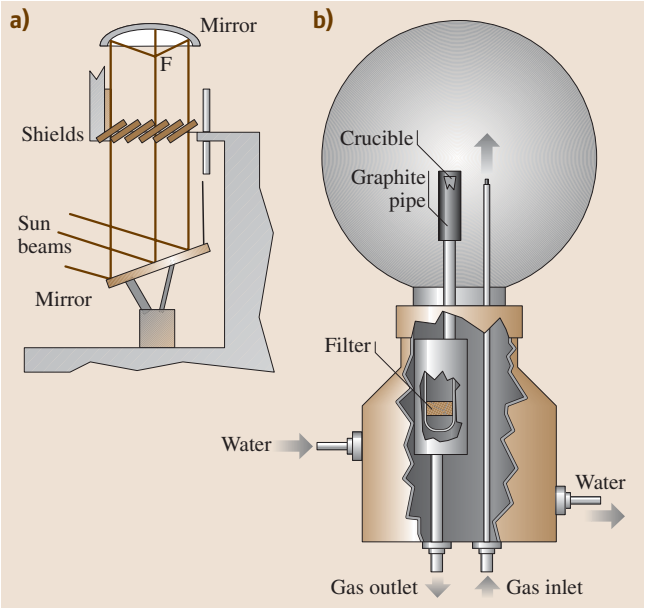


Fig. 3.14a,b Sketch of a solar energy reactor in use in Odeilho (France). **(a)** Gathering of sun rays, focused in F; **(b)** Example of Pyrex® chamber placed in **(a)** so that the graphite crucible is at the point F (adapted from [3.56])

nanotubes. Specifically regarding multilayered fibrous morphologies (since single-layered fibrous morphologies cannot be otherwise than SWNT anyway), the exact name should be vapor-grown carbon nanofilaments (VGCNF). Whether or not the filaments are then tubular is a matter of textural description, which should go with other textural features such as bamboo, herringbone, concentric, etc. (see Sect. 3.1.2). But the usage has decided otherwise. In the following, we will therefore use MWNTs for any hollowed nanofilament, whether they contain transversally oriented graphene walls or not. Any other nanofilament will be named "nanofiber."

Heterogeneous Processes

Heterogeneous CCVD processes are basically simple and consist, within a furnace heated to the desired temperature, in passing a gaseous flow containing a given proportion of a hydrocarbon (mainly CH_4, C_2H_2, C_2H_4, or C_6H_6, usually as a mixture with either H_2 or an inert gas such as Ar) over small transition metal particles (Fe, Co, Ni). The latter are previously deposited onto an inert substrate, for instance, by spraying a suspension of the metal particles on it, or any other method. The reac-

tion is chemically defined as catalysis-enhanced thermal cracking

$$C_xH_y \rightarrow x\,C + y/2\,H_2\,.$$

Catalysis-enhanced thermal cracking was used as early as the late 19th century (see beginning of chapter). Further extensive works before the 90s include those by Baker et al. [3.6, 59], or *Endo* et al. [3.60, 61]. Several review papers have been published since then, such as [3.62], in addition to many regular papers.

CO can be used instead of hydrocarbons, the reaction is then chemically defined as catalysis-enhanced disproportionation

$$2\,CO \rightarrow C + CO_2\,.$$

Heterogeneous Processes – Experimental Devices

The ability of catalysis-enhanced CO disproportionation in making carbon nanofilaments was reported by *Davis* et al. [3.63] as early as 1953, probably for the first time. Extensive following works were performed by *Boehm* [3.64], *Audier* et al. [3.16, 65–67], and *Gadelle* et al. [3.68–71].

Although formation mechanisms for SWNTs and MWNTs can be quite different (see Sect. 3.3, or refer to a review article such as [3.72]), many of the parameters of the catalytic processes have a similar and important role on the type of nanotubes formed: the temperature, the duration of the treatment, the gas composition and flow rate, and of course the catalyst nature and size. At a given temperature, depending mainly on the nature of both the catalyst and the carbon-containing gas, the catalytic decomposition will take place at the surface of the metal particles, followed by a mass-transport of the freshly produced carbon either by surface or volume diffusion until the carbon concentration reaches the solubility limit, and the precipitation starts.

It is now agreed that the formation of CCVD carbon nanotubes occurs on metal particles of a very small size, typically in the nanometer-size range [3.72]. These catalytic metal particles are prepared mainly by reduction of transition metals compounds (salts, oxides) by H_2, prior to the nanotube formation step (where the carbon containing gas is required). It is possible, however, to produce these catalytic metal particles in situ in presence of the carbon source, allowing for a one-step process [3.73]. Because the control of the metal particle size is the key point (they have to be kept at a nanometric size), their coalescence is generally avoided by supporting them on an inert support such as an oxide (Al_2O_3, SiO_2, zeolites, $MgAl_2O_4$, MgO, etc.), or more rarely on graphite. A low concentration of the catalytic metal precursor is required to limit the coalescence of the metal particles that may happen during the reduction step.

There are two main ways for the preparation of the catalyst: (a) the impregnation of a substrate with a solution of a salt of the desired transition metal catalyst, and (b) the preparation of a solid solution of an oxide of the chosen catalytic metal in a chemically inert and thermally stable host oxide. The catalyst is then reduced to form the metal particles on which the catalytic decomposition of the carbon source will lead to carbon nanotube growth. In most cases, the nanotubes can then be separated from the catalyst (Fig. 3.15).

Heterogeneous Processes – Results with CCVD Involving Impregnated Catalysts

A lot of work was already done in this area even before the discovery of fullerenes and carbon nanotubes, but although the formation of tubular carbon structures by catalytic processes involving small metal particles was clearly identified, the authors did not focus on the preparation of SWNTs or MWNTs with respect to the other carbon species. Some examples will be given here to illustrate the most striking improvements obtained.

With the impregnation method, the process generally involves four different and successive steps: (1) the impregnation of the support by a solution of a salt (nitrate, chloride) of the chosen metal catalyst; (2) the drying and calcination of the supported catalyst to get the oxide of the catalytic metal; (3) the reduction in a H_2-containing atmosphere to make the catalytic metal particles and at last (4) the decomposition of a carbon-containing gas over the freshly prepared metal particles that will lead to the nanotube growth. For example, *Ivanov* et al. [3.74] have prepared nanotubes by the decomposition of C_2H_2 (pure or in mixture with H_2) on well-dispersed transition metal particles (Fe, Co, Ni, Cu) supported on graphite or SiO_2. Co-SiO_2 was found to be the best catalyst/support combination for the preparation of MWNTs, but most of the other combinations led to carbon filaments, sometimes covered with amorphous-like carbon. The same authors have developed a precipitation-ion-exchange method that provides a better dispersion of metals on silica compared to the classical impregnation technique. The same group has then proposed the use of a zeolite-supported Co catalyst [3.75, 76], resulting in very finely dispersed metal particles (from 1 to 50 nm in diameter). Only on this catalyst could they observe MWNTs with a diameter around 4 nm and only two or three walls. *Dai* et al. [3.77] have prepared SWNTs by CO dispro-

portionation on nano-sized Mo particles. The diameters of the nanotubes obtained are closely related to that of the original particles and range from 1 to 5 nm. The nanotubes obtained by this method are free of amorphous carbon coating. It is also found that a synergetic effect occurs in the case of an alloy instead of the components alone, and one of the most striking examples is the addition of Mo to Fe [3.78] or Co [3.79].

Heterogeneous Processes – Results with CCVD Involving Solid-Solutions-Based Catalysts

A solid solution of two metal oxides is formed when some ions of one of the metals are found in substitution of the other metal ions. For example, Fe_2O_3 can be prepared in solid solution into Al_2O_3 to give a $Al_{2-2x}Fe_{2x}O_3$ solid solution. The use of a solid solution allows a perfect homogeneity of the dispersion of each oxide one in the other. These solid solutions can be prepared by different ways but the co-precipitation of mixed-oxalates and the combustion synthesis have been used mainly for the preparation of nanotubes. The synthesis of nanotubes by catalytic decomposition of CH_4 over $Al_{2-2x}Fe_{2x}O_3$ solid solutions was originated by *Peigney* et al. [3.73] and then studied extensively by the same group using different oxides such as spinel-based solid solutions ($Mg_{1-x}M_xAl_2O_4$ with M = Fe, Co, Ni, or a binary alloy [3.80, 81] or magnesia-based solid solutions [3.80, 82] ($Mg_{1-x}M_xO$, with M = Fe, Co or Ni)). Because of the very homogeneous dispersion of the catalytic oxide, it is possible to produce very small catalytic metal particles, at the high temperature required for the decomposition of CH_4 (which was chosen for its greater thermal stability compared to other hydrocarbons). The method proposed by these authors involves the heating of the solid solution from room temperature to a temperature between 850 °C and 1,050 °C in a mixture of H_2 and CH_4, typically containing 18 mol.% of CH_4. The nanotubes obtained depend clearly on the nature of both the transition metal (or alloy) used and the inert oxide (matrix), the latter because the Lewis acidity seems to play an important role [3.83]. For example, in the case of solid solutions containing around 10 wt% of Fe, the amount of carbon nanotubes obtained is decreasing in the following order depending of the matrixoxide: $MgO > Al_2O_3 > MgAl_2O_4$ [3.80]. In the case of MgO-based solid solutions the nanotubes can be very easily separated from the catalyst by dissolving it, in diluted HCl, for example [3.82]. The nanotubes obtained are typically gathered into small diameter bundles (less than 15 nm) with lengths up to 100 µm. The nanotubes are mainly SWNTs and DWNTs, with diameters ranging between 1 and 3 nm.

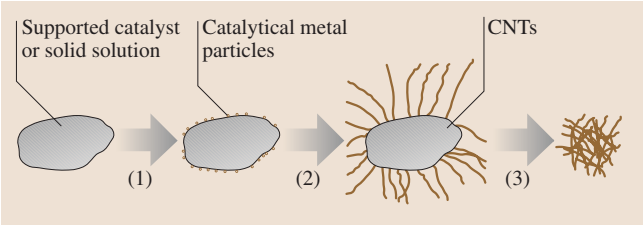


Fig. 3.15 Formation of nanotubes by the CCVD-based impregnation technique. (1) Formation of the catalytic metal particles by reduction of a precursor; (2) Catalytic decomposition of a carbon-containing gas, leading to the growth of carbon nanotubes (CNTs); (3) Removal of the catalyst to recover the CNTs (from [3.80])

Obtaining pure nanotubes by the CCVD method requires, as for all the other techniques, the removal of the catalyst. When a catalyst supported (impregnation) in a solid solution is used, the supporting – and catalytically inactive – oxide is the main impurity, both in weight and volume. When oxides such as Al_2O_3 or SiO_2 (or even combinations) are used, aggressive treatments involving hot caustic solutions (KOH, NaOH) for Al_2O_3 or the use of HF for SiO_2 are required. These treatments have no effect, however, on the other impurities such as other forms of carbon (amorphous-like carbon, graphitized carbon particles and shells, and so on). Oxidizing treatments (air oxidation, use of strong oxidants such as HNO_3, $KMnO_4$, H_2O_2) are thus required and allow for the removal of most of the unwanted forms of carbon but are resulting in a low yield of remaining carbon nanotubes, which are often quite damaged. *Flahaut* et al. [3.82] were the first to use a MgCoO solid solution to prepare SWNTs and DWNTs that could be easily separated without any damage by a fast and safe washing with an aqueous HCl solution.

In most cases, only very small quantities of catalyst (typically less than 500 mg) have been used, and most claims for "high yield" productions of nanotubes are based on laboratory experimental data, without taking into account all the technical problems related to the scaling up of a laboratory-scale CCVD reactor. At the present time, although the production of MWNTs is possible at an industrial scale, the production of SWNTs at an affordable cost is still a challenge.

Homogeneous Processes

The homogenous route, also called "floating catalyst method," differs from the other CCVD-based meth-

ods because it uses only gaseous species and does not require the presence of any solid phase in the reactor. The basic principle of this technique, similar to the other CCVD processes, is to decompose a carbon source (ethylene, xylene, benzene, carbon monoxide, etc.) on nanometric transition metallic (generally Fe, Co, or Ni) particles in order to obtain carbon nanotubes. The catalytic particles are formed directly in the reactor, however, and are not introduced before the reaction, as it happens in the supported CCVD, for instance.

Homogeneous Processes – Experimental Devices

The typical reactor used in this technique is a quartz tube placed in an oven to which the gaseous feedstock, containing the metal precursor, the carbon source, some hydrogen, and a vector gas (N_2, Ar, or He), is sent. The first zone of the reactor is kept at a lower temperature, and the second zone, where the formation of tubes occurs, is heated to 700–1,200 °C. The metal precursor is generally a metal-organic compound, such as a zero-valent carbonyl compound like $Fe(CO)_5$ [3.84], or a metallocene [3.85–87], for instance ferrocene, nickelocene or cobaltocene. It may be advantageous to make the reactor vertical, in order for the effect of gravity to be symmetrically dispatched on the gaseous volume inside the furnace and to help in maintaining for a while the solid products in fluidized bed.

Homogeneous Processes – Results

The metal-organic compound decomposes in the first zone of the reactor to generate the nanometric metallic particles that can catalyze the nanotubes formation. In the second part of the reactor, the carbon source is decomposed to atomic carbon which then is responsible for the formation of nanotubes.

This technique is quite flexible and both single-walled nanotubes [3.88] and multiwalled nanotubes [3.89] have been obtained depending on the carbon feedstock gas; it has also been exploited for some years in the production of vapor grown carbon nanofibers [3.90].

The main drawback of this type of process is again, as for heterogeneous processes, the difficulty to control the size of the metal nanoparticles, and thus the nanotube formation is often accompanied by the production of undesired carbon forms (amorphous carbon or polyaromatic carbon phases found as various phases or as coatings). In particular, encapsulated forms have been often found, as the result of the creation of metallic particles that are too big to be active for growing nanotubes (but are still effective for catalytically decomposing the carbon source) and be totally recovered by graphene layers.

The same kind of parameters as for heterogeneous processes have to be controlled in order to finely tune this process and obtain selectively the desired morphology and structure of the nanotubes formed, such as: the choice of the carbon source; the reaction temperature; the residence time; the composition of the incoming gaseous feedstock with a particular attention paid to the role played by hydrogen proportion, which can control the orientation of graphenes with respect to the nanotube axis thus switching from c-MWNT to h-MWNT [3.69]; and the ratio of the metallorganic precursor to the carbon source (for lower values, SWNTs are obtained [3.85]). As recently demonstrated, the overall process can be improved by adding other compounds such as ammonia or sulfur-containing species to the reactive gas phase. The former allows aligned nanotubes and mixed C-N nanotubes [3.91] to be obtained, the latter results in a significant increase in productivity [3.90].

It should be emphasized that only small productions have been achieved so far, and the scale up toward industrial exploitation seems quite difficult because of the large number of parameters that have to be considered. A critical one is to be able to increase the quantity of metallorganic compound that has to be sent in the reactor, as a requirement to increase the production, without obtaining too big particles. This problem has not yet been solved. An additional problem inherent in the process is the possibility of clogging the reactor due to the deposition of metallic nanoparticles on the reactor walls followed by carbon deposition.

A significant breakthrough concerning this technique could be the process developed at Rice University, the so-called HiPCo™ process to produce SWNTs of very high purity [3.92]. This gas phase catalytic reaction uses carbon monoxide to produce, from [$Fe(CO)_5$], a SWNT material claimed to be relatively free of by-products. The temperature and pressure conditions required are applicable to industrial plants. The company Carbon Nanotechnologies Inc. (Houston, TX, USA) actually sells raw SWNT materials prepared that way, at a marketed price of \$375/g, doubled if purified (2003 data). Other companies are more specialized in MWNTs such as Applied Sciences Inc. (Cedarville, Ohio, USA), which currently has a production facility of ~ 40 tons/year of ~ 100 nm large MWNTs (Pyrograf-III), or Hyperion Catalysis (Cambridge, MA, USA), which makes MWNT-based materials.

Templating

Another technique interesting to describe briefly, though definitely not suitable for mass production, is the templating technique. It is the second method only (the first being the electric-arc technique, when considering the formation of MWNTs on the cathode) able to synthesize carbon nanotubes without any catalyst. Any other work reporting the catalyst-free formation of nanotubes is likely to have been fooled by metallic impurities present in the reactor or by some other factors having brought a chemical gradient to the system. Another original aspect is that it allows aligned nanotubes to be obtained naturally, without the help of any subsequent alignment procedure. But recovering the nanotubes only requires the template to be removed (dissolved), which means the former alignment of the nanotubes is lost.

Templating – Experimental Devices

The principle of the technique is to deposit a solid carbon coating obtained from CVD method onto the walls of an appropriate porous substrate whose pores are displayed as parallel channels. The feedstock is again a hydrocarbon, as a common carbon source. The substrate can be alumina or zeolite for instance, however, which present natural channel pores, while the whole is heated to a temperature able to crack the hydrocarbon selected as carbon source (Fig. 3.16).

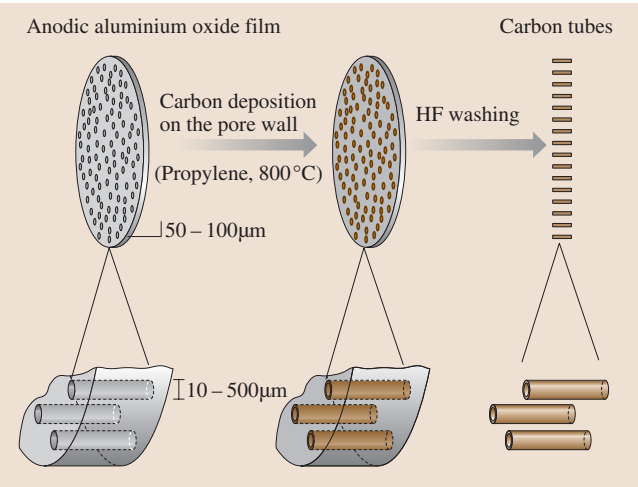


Fig. 3.16 Principle of the templating technique for the catalyst-free formation of single-walled or concentric-type multi-walled carbon nanotubes (from [3.93])

Templating – Results

Provided the chemical vapor deposition mechanism (which is actually better described as a chemical vapor infiltration mechanism) is well controlled, the synthesis results in the coating of the channel pore walls by a variable number of graphenes. Both MWNTs (exclusively concentric type) or SWNTs can be obtained. The smallest SWNTs (diameters ∼ 0.4 nm) ever obtained mentioned in Sect. 3.1 have actually been synthesized by this technique [3.11]. Lengths are directly determined by channel lengths, i. e. by the thickness of the substrate plate. A main advantage is the purity of the tubes (no catalyst remnants, and little other carbon phases). On the other hand, the nanotube structure is not closed at both ends, which can be an advantage or a drawback, depending on the application. For instance, recovering the tubes requires the porous matrix to dissolve using one of the chemical treatment previously cited. The fact that tubes are open makes them even more sensitive to the acid attack.

3.2.3 Miscellaneous Techniques

In addition to the major techniques described in Sects. 3.2.1 and 3.2.2, many attempts can be found in literature to produce nanotubes by various ways, with a generally specific goal, such as looking for a low-cost or a catalyst-free production process. None has been sufficiently convincing so far to be presented as a serious alternative to the major processes described previously. Some examples are provided in the following.

Hsu et al. [3.94] have succeeded in preparing MWNTs (including coiled MWNTs, a peculiar morphology resembling a spring) by a catalyst-free (although Li was present) electrolytic method, by running a 3–5 A current between two graphite electrodes (a graphite crucible as the anode and a graphite rod as the cathode). The graphite crucible was filled with lithium chloride, while the whole was heated in air or argon at ∼ 600 °C. As with many other techniques, by-products such as encapsulated metal particles, carbon shells, amorphous carbon, and so on, are formed.

Cho et al. [3.95] have proposed a pure chemistry route, by the polyesterification of citric acid onto ethylen glycol at 50 °C, followed by a polymerization at 135 °C, then carbonized at 300 °C under argon, then oxidized at 400 °C in air. Despite the latter oxidation step, the solid product surprisingly contains short MWNTs, although obviously with a poor nanotexture. By-products such as carbon shells and amorphous carbon are also formed.

Table 11.3 Measured vertical spring constants and natural frequencies of triangular (V-shaped) cantilevers made of PECVD Si_3N_4 (Data provided by Digital Instruments, Inc.)

Cantilever dimension	Spring constant (k_z) (N/m)	Natural frequency (ω_0) (kHz)
115-μm long, narrow leg	0.38	40
115-μm long, wide leg	0.58	40
193-μm long, narrow leg	0.06	13–22
193-μm long, wide leg	0.12	13–22

Table 11.4 Vertical (k_z), lateral (k_y), and torsional (k_{yT}) spring constants of rectangular cantilevers made of Si (IBM) and PECVD Si_3N_4 (Veeco Instruments, Inc.)

Dimensions/stiffness	Si cantilever	Si_3N_4 cantilever
Length (L) (μm)	100	100
Width (b) (μm)	10	20
Thickness (h) (μm)	1	0.6
Tip length (ℓ) (μm)	5	3
k_z (N/m)	0.4	0.15
k_y (N/m)	40	175
k_{yT} (N/m)	120	116
ω_0 (kHz)	∼ 90	∼ 65

Note: $k_z = Ebh^3/4L^3$, $k_y = Eb^3h/4\ell^3$, $k_{yT} = Gbh^3/3L\ell^2$, and $\omega_0 = [k_z/(m_c + 0.24bhL\rho)]^{1/2}$, where E is Young's modulus, G is the modulus of rigidity [$= E/2(1+\nu)$, ν is Poisson's ratio], ρ is the mass density of the cantilever, and m_c is the concentrated mass of the tip (∼ 4 ng) [11.94]. For Si, $E = 130$ GPa, $\rho g = 2{,}300\,\mathrm{kg/m^3}$, and $\nu = 0.3$. For Si_3N_4, $E = 150$ GPa, $\rho g = 3{,}100\,\mathrm{kg/m^3}$, and $\nu = 0.3$

tip side walls have a slope of 35 deg and the length of the edges of the tip at the cantilever base is about 4 μm.

An alternative to silicon nitride cantilevers with integrated tips are microfabricated single-crystal silicon cantilevers with integrated tips. Si tips are sharper than Si_3N_4 tips because they are directly formed by the anisotropic etch in single-crystal Si rather than using an etch pit as a mask for deposited materials [11.139]. Etched single-crystal n-type silicon rectangular cantilevers with square pyramidal tips with a lower radius of less than 10 nm for contact and tapping mode (tapping-mode etched silicon probe or TESP) AFMs are commercially available from Digital Instruments and Nanosensors GmbH, Aidlingen, Germany, Figs. 11.15b and 11.16. Spring constants and resonant frequencies are also presented in the Fig. 11.15b.

Commercial triangular Si_3N_4 cantilevers have a typical width-thickness ratio of 10 to 30 which results in 100 to 1000 times stiffer spring constants in the lateral direction compared to the normal direction. Therefore these cantilevers are not well suited for torsion. For friction measurements, the torsional spring constant should be minimized in order to be sensitive to the lateral forces. Rather long cantilevers with small thickness and large tip length are most suitable. Rectangular beams have lower torsional spring constants in comparison to the triangular (V-shaped) cantilevers. Table 11.4 lists the spring constants (with full length of the beam used) in three directions of the typical rectangular beams. We note that lateral and torsional spring constants are about two orders of magnitude larger than the normal spring constants. A cantilever beam required for the tapping mode is quite stiff and may not be sensitive enough for friction measurements. *Meyer* et al. [11.140] used a specially designed rectangular silicon cantilever with length = 200 μm, width = 21 μm, thickness = 0.4 μm, tip length = 12.5 μm, and shear modulus = 50 GPa, giving a normal spring constant of 0.007 N/m and torsional spring constant of 0.72 N/m which gives a lateral force sensitivity of 10 pN and an angle of resolution of 10^{-7} rad. With this particular geometry, sensitivity to lateral forces could be improved by about a factor of 100 compared with commercial V-shaped Si_3N_4 or rectangular Si or Si_3N_4 cantilevers used by *Meyer* and *Amer* [11.8] with torsional spring constant of ∼ 100 N/m. *Ruan* and *Bhushan* [11.136] and *Bhushan* and *Ruan* [11.9] used 115-μm long, wide-legged V-shaped cantilevers made of Si_3N_4 for friction measurements.

For scratching, wear, and indentation studies, single-crystal natural diamond tips ground to the shape of a three-sided pyramid with an apex angle of either 60° or 80° whose point is sharpened to a radius of about 100 nm are commonly used [11.4, 10] (Figs. 11.15c and 11.16). The tips are bonded with conductive epoxy to a gold-plated 304 stainless steel spring sheet (length = 20 mm, width = 0.2 mm, thickness = 20 to 60 μm) which acts as a cantilever. The free length of the spring is varied to change the beam stiffness. The normal spring con-

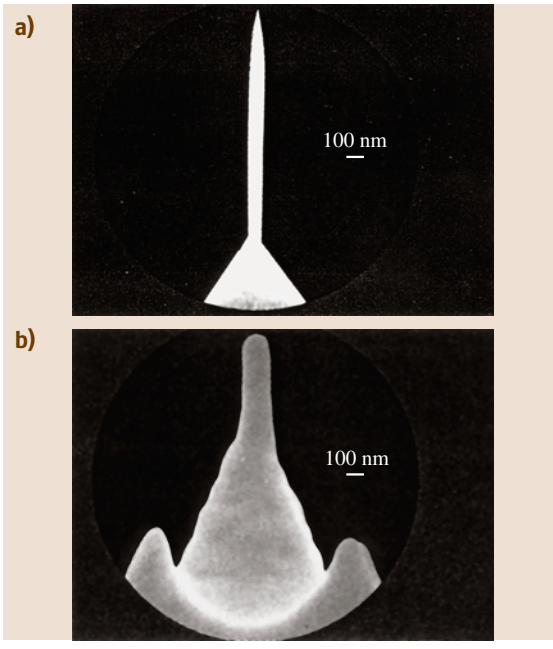


Fig. 11.17a,b Schematics of **(a)** HART Si_3N_4 probe, and **(b)** FIB milled Si_3N_4 probe

stant of the beam ranges from about 5 to 600 N/m for a 20 μm thick beam. The tips are produced by R-DEC Co., Tsukuba, Japan.

For imaging within trenches by an AFM, high aspect ratio tips are used. Examples of the two probes are shown in Fig. 11.17. The high-aspect ratio tip (Hart) probes are produced by starting with a conventional Si_3N_4 pyramidal probe. Through a combination of focused ion beam (FIB) and high resolution scanning electron microscopy (SEM) techniques, a thin filament is grown at the apex of the pyramid. The probe filament is approximately 1 μm long and 0.1 μm in diameter. It tapers to an extremely sharp point (radius better than the resolution of most SEMs). The long thin shape and sharp radius make it ideal for imaging within "vias" of microstructures and trenches (> 0.25 μm). Because of flexing of the probe, it is unsuitable for imaging structures at the atomic level since the flexing of the probe can create image artifacts. For atomic-scale imaging, a FIB-milled probe is used which is relatively stiff yet allows for closely spaced topography. These probes start out as conventional Si_3N_4 pyramidal probes but the pyramid is FIB milled until a small cone shape is formed which has a high aspect ratio with 0.2–0.3 μm in length. The milled probes allow nanostructure resolution without sacrificing rigidity. These types of probes are manufactured by various manufacturers including Materials Analytical Services, Raleigh, North Carolina.

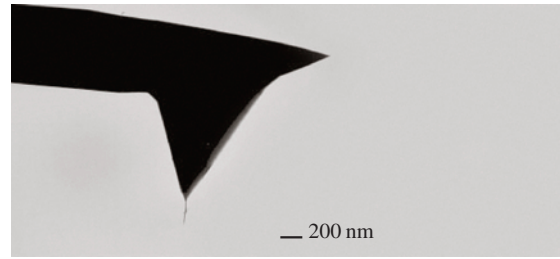


Fig. 11.18 SEM micrograph of a multi-walled carbon nanotube (MWNT) tip physically attached on the single-crystal silicon, square-pyramidal tip (Courtesy Piezomax Technologies, Inc.)

Carbon nanotube tips having small diameter and high aspect ratio are used for high resolution imaging of surfaces and of deep trenches, in the tapping mode or noncontact mode. Single-walled carbon nanotubes (SWNT) are microscopic graphitic cylinders that are 0.7 to 3 nm in diameter and up to many microns in length. Larger structures called multi-walled carbon nanotubes (MWNT) consist of nested, concentrically arranged SWNT and have diameters ranging from 3 to 50 nm. MWNT carbon nanotube AFM tips are produced by manual assembly [11.141], chemical vapor deposition (CVD) synthesis, and hybrid fabrication process [11.142]. Figure 11.18 shows TEM micrograph of a carbon nanotube tip, ProbeMax™, commercially produced by mechanical assembly by Piezomax Technologies, Inc., Middleton, Wisconsin. For production of these tips, MWNT nanotubes are produced by carbon arc. They are physically attached on the single-crystal silicon, square-pyramidal tips in the SEM using a manipulator and the SEM stage to control the nanotubes and the tip independently. Once the nanotube is attached to the tip, it is usually too long to image with. It is shortened by using an AFM and applying voltage between the tip and the sample. Nanotube tips are also commercially produced by CVD synthesis by NanoDevices, Santa Barbara, California.

11.2.4 Friction Measurement Methods

Based on the work by *Ruan* and *Bhushan* [11.136], the two methods for friction measurements are now described in more detail. (Also see [11.8].) A scanning angle is defined as the angle relative to the y-axis in Fig. 11.19a. This is also the long axis of the cantilever.

A zero degree scanning angle corresponds to the sample scanning in the y direction, and a 90 degree scanning angle corresponds to the sample scanning perpendicular to this axis in the xy plane (along x axis). If the scanning direction is in both the y and $-y$ directions, we call this "parallel scan". Similarly, a "perpendicular scan" means the scanning direction is in the x and $-x$ directions. The sample traveling direction for each of these two methods is illustrated in Fig. 11.19b.

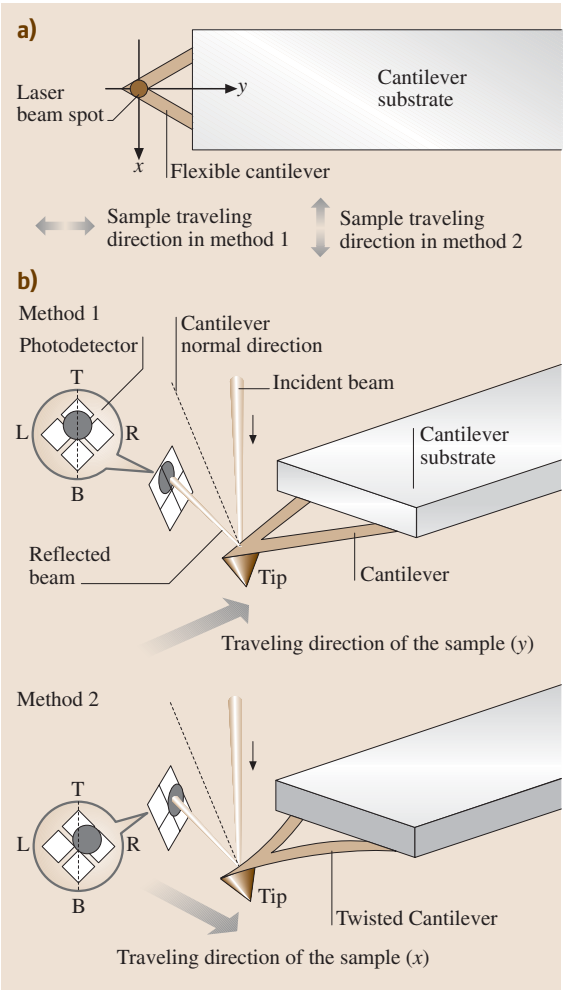


Fig. 11.19 **(a)** Schematic defining the x- and y-directions relative to the cantilever, and showing the sample traveling direction in two different measurement methods discussed in the text, **(b)** schematic of deformation of the tip and cantilever shown as a result of sliding in the x- and y-directions. A twist is introduced to the cantilever if the scanning is in the x-direction (**(b)**, *lower part*) [11.136]

In method 1 (using "height" mode with parallel scans) in addition to topographic imaging, it is also possible to measure friction force when the scanning direction of the sample is parallel to the y direction (parallel scan). If there were no friction force between the tip and the moving sample, the topographic feature would be the only factor which cause the cantilever to be deflected vertically. However, friction force does exist on all contact surfaces where one object is moving relative to another. The friction force between the sample and the tip will also cause a cantilever deflection. We assume that the normal force between the sample and the tip is W_0 when the sample is stationary (W_0 is typically in the range of 10 nN to 200 nN), and the friction force between the sample and the tip is W_f as the sample scans against the tip. The direction of friction force (W_f) is reversed as the scanning direction of the sample is reversed from positive (y) to negative ($-y$) directions ($\vec{W}_{f(y)} = -\vec{W}_{f(-y)}$).

When the vertical cantilever deflection is set at a constant level, it is the total force (normal force and friction force) applied to the cantilever that keeps the cantilever deflection at this level. Since the friction force is in opposite directions as the traveling direction of the sample is reversed, the normal force will have to be adjusted accordingly when the sample reverses its traveling direction, so that the total deflection of the cantilever will remain the same. We can calculate the difference of the normal force between the two traveling directions for a given friction force W_f. First, by means of a constant deflection, the total moment applied to the cantilever is constant. If we take the reference point to be the point where the cantilever joins the cantilever holder (substrate), point P in Fig. 11.20, we have the following relationship:

$$(W_0 - \Delta W_1)L + W_f \ell = (W_0 + \Delta W_2)L - W_f \ell \tag{11.1}$$

or

$$(\Delta W_1 + \Delta W_2)L = 2W_f \ell \,. \tag{11.2}$$

Thus

$$W_f = (\Delta W_1 + \Delta W_2)L/(2\ell) \,, \tag{11.3}$$

where ΔW_1 and ΔW_2 are the absolute value of the changes of normal force when the sample is traveling in $-y$ and y directions, respectively, as shown in Fig. 11.20; L is the length of the cantilever; ℓ is vertical distance between the end of the tip and point P. The coefficient of

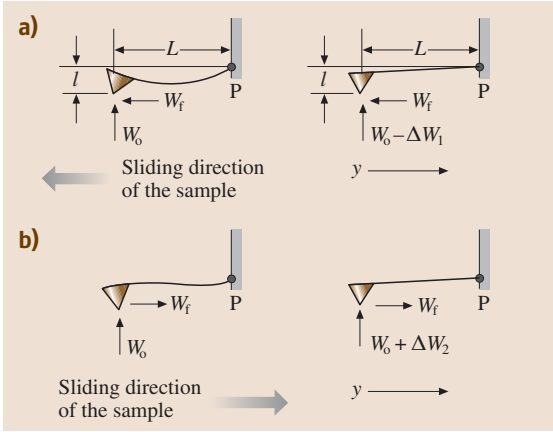


Fig. 11.20 **(a)** Schematic showing an additional bending of the cantilever – due to friction force when the sample is scanned in the y- or $-y$-direction (*left*). **(b)** This effect will be canceled by adjusting the piezo height by a feedback circuit (*right*) [11.136]

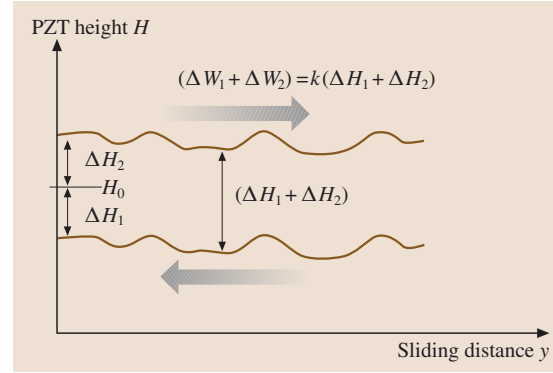


Fig. 11.21 Schematic illustration of the height difference of the piezoelectric tube scanner as the sample is scanned in y and $-y$ directions

friction (μ) between the tip and the sample is then given as

$$\mu = \frac{W_f}{W_0} = \left(\frac{(\Delta W_1 + \Delta W_2)}{W_0}\right)\left(\frac{L}{2\ell}\right) . \qquad (11.4)$$

In all circumstances, there are adhesive and interatomic attractive forces between the cantilever tip and the sample. The adhesive force can be due to water from the capillary condensation and other contaminants present at the surface which form meniscus bridges [11.4, 143, 144] and the interatomic attractive force includes van der Waals attraction [11.18]. If these forces (and indentation effect as well, which is usually small for rigid samples) can be neglected, the normal force W_0 is then equal to the initial cantilever deflection H_0 multiplied by the spring constant of the cantilever. $(\Delta W_1 + \Delta W_2)$ can be measured by multiplying the same spring constant by the height difference of the piezo tube between the two traveling directions (y and $-y$ directions) of the sample. This height difference is denoted as $(\Delta H_1 + \Delta H_2)$, shown schematically in Fig. 11.21. Thus, (11.4) can be rewritten as

$$\mu = \frac{W_f}{W_0} = \left(\frac{(\Delta H_1 + \Delta H_2)}{H_0}\right)\left(\frac{L}{2\ell}\right) . \qquad (11.5)$$

Since the piezo tube vertical position is affected by the surface topographic profile of the sample in addition to the friction force being applied at the tip, this difference has to be taken point by point at the same location on the sample surface as shown in Fig. 11.21. Subtraction of point by point measurements may introduce errors, particularly for rough samples. We will come back to this point later. In addition, precise measurement of L and ℓ (which should include the cantilever angle) are also required.

If the adhesive forces between the tip and the sample are large enough that it can not be neglected, one should include it in the calculation. However, there could be a large uncertainty in determining this force, thus an uncertainty in using (11.5). An alternative approach is to make the measurements at different normal loads and to use $\Delta(H_0)$ and $\Delta(\Delta H_1 + \Delta H_2)$ from the measurements in (11.5). Another comment on (11.5) is that, since only the ratio between $(\Delta H_1 + \Delta H_2)$ and H_0 comes into this equation, the piezo tube vertical position H_0 and its position difference $(\Delta H_1 + \Delta H_2)$ can be in the units of volts as long as the vertical traveling distance of the piezo tube and the voltage applied to it has a linear relationship. However, if there is a large nonlinearity between the piezo tube traveling distance and the applied voltage, this nonlinearity must be included in the calculation.

It should also be pointed out that (11.4) and (11.5) are derived under the assumption that the friction force W_f is the same for the two scanning directions of the sample. This is an approximation since the normal force is slightly different for the two scans and there may also be a directionality effect in friction. However, this difference is much smaller than W_0 itself. We can ignore the second order correction.

Method 2 ("aux" mode with perpendicular scan) to measure friction was suggested by *Meyer* and *Amer* [11.8]. The sample is scanned perpendicular to

the long axis of the cantilever beam (i. e., to scan along the x or $-x$ direction in Fig. 11.19a) and the output of the horizontal two quadrants of the photodiode-detector is measured. In this arrangement, as the sample moves under the tip, the friction force will cause the cantilever to twist. Therefore the light intensity between the left and right (L and R in Fig. 11.19b, right) detectors will be different. The differential signal between the left and right detectors is denoted as FFM signal $[(L-R)/(L+R)]$. This signal can be related to the degree of twisting, hence to the magnitude of friction force. Again, because of a possible error in determining normal force due to the presence of an adhesive force at the tip–sample interface, the slope of the friction data (FFM signal vs. normal load) needs to be taken for an accurate value of the coefficient of friction.

While friction force contributes to the FFM signal, friction force may not be the only contributing factor in commercial FFM instruments (for example, NanoScope IV). One can notice this fact by simply engaging the cantilever tip with the sample. Before engaging, the left and right detectors can be balanced by adjusting the position of the detectors so that the intensity difference between these two detectors is zero (FFM signal is zero). Once the tip is engaged with the sample, this signal is no longer zero even if the sample is not moving in the xy plane with no friction force applied. This would be a detrimental effect. It has to be understood and eliminated from the data acquisition before any quantitative measurement of friction force becomes possible.

One of the fundamental reasons for this observation is the following. The detectors may not have been properly aligned with respect to the laser beam. To be precise, the vertical axis of the detector assembly (the line joining T-B in Fig. 11.22) is not in the plane defined by the incident laser beam and the beam reflected from an untwisted cantilever (we call this plane the "beam plane"). When the cantilever vertical deflection changes due to a change of applied normal force (without having the sample scanned in the xy plane), the laser beam will be reflected up and down and form a projected trajectory on the detector. (Note that this trajectory is in the defined beam plane.) If this trajectory is not coincident with the vertical axis of the detector, the laser beam will not evenly bisect the left and right quadrants of the detectors, even under the condition of no torsional motion of the cantilever, see Fig. 11.22. Thus when the laser beam is reflected up and down due a change of the normal force, the intensity difference between the left and right detectors will also change. In other words, the FFM signal will change as the normal force applied to

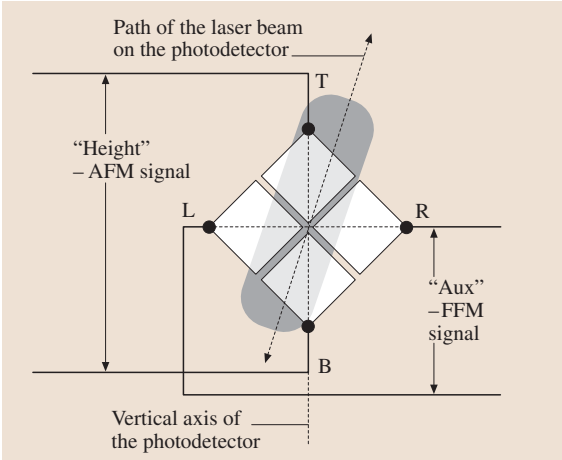


Fig. 11.22 The trajectory of the laser beam on the photodetectors in as the cantilever is vertically deflected (with no torsional motion) for a misaligned photodetector with respect to the laser beam. For a change of normal force (vertical deflection of the cantilever), the laser beam is projected at a different position on the detector. Due to a misalignment, the projected trajectory of the laser beam on the detector is not parallel with the detector vertical axis (the line joint T-B) [11.136]

the tip is changed, even if the tip is not experiencing any friction force. This (FFM) signal is unrelated to friction force or to the actual twisting of the cantilever. We will call this part of FFM signal "FFM_F", and the part which is truly related to friction force "FFM_T".

The FFM_F signal can be eliminated. One way of doing this is as follows. First the sample is scanned in both x and $-x$ directions and the FFM signal for scans in each direction is recorded. Since friction force reverses its directions when the scanning direction is reversed from x to $-x$ direction, the FFM_T signal will have opposite signs as the scanning direction of the sample is reversed ($FFM_T(x) = -FFM_T(-x)$). Hence the FFM_T signal will be canceled out if we take the sum of the FFM signals for the two scans. The average value of the two scans will be related to FFM_F due to the misalignment,

$$FFM(x) + FFM(-x) = 2FFM_F\,. \tag{11.6}$$

This value can therefore be subtracted from the original FFM signals of each of these two scans to obtain the true FFM signal (FFM_T). Or, alternately, by taking the difference of the two FFM signals, one directly gets the FFM_T value

$$\mathrm{FFM}(x) - \mathrm{FFM}(-x) = \mathrm{FFM_T}(x) - \mathrm{FFM_T}(-x) = 2\mathrm{FFM_T}(x)\,. \quad (11.7)$$

Ruan and *Bhushan* [11.136] have shown that the error signal ($\mathrm{FFM_F}$) can be very large compared to the friction signal $\mathrm{FFM_T}$, thus correction is required.

Now we compare the two methods. The method of using "height" mode and parallel scan (method 1) is very simple to use. Technically, this method can provide 3-D friction profiles and the corresponding topographic profiles. However, there are some problems with this method. Under most circumstances, the piezo scanner displays a hysteresis when the traveling direction of the sample is reversed. Therefore the measured surface topographic profiles will be shifted relative to each other along the y-axis for the two opposite (y and $-y$) scans. This would make it difficult to measure the local height difference of the piezo tube for the two scans. However, the average height difference between the two scans and hence the average friction can still be measured. The measurement of average friction can serve as an internal means of friction force calibration. Method 2 is a more desirable approach. The subtraction of $\mathrm{FFM_F}$ signal from FFM for the two scans does not introduce error to local friction force data. An ideal approach in using this method would be to add the average value of the two profiles in order to get the error component ($\mathrm{FFM_F}$) and then subtract this component from either profiles to get true friction profiles in either directions. By making measurements at various loads, we can get the average value of the coefficient of friction which then can be used to convert the friction profile to the coefficient of friction profile. Thus any directionality and local variations in friction can be easily measured. In this method, since topography data are not affected by friction, accurate topography data can be measured simultaneously with friction data and a better localized relationship between the two can be established.

11.2.5 Normal Force and Friction Force Calibrations of Cantilever Beams

Based on *Ruan* and *Bhushan* [11.136], we now discuss normal force and friction force calibrations. In order to calculate the absolute value of normal and friction forces in Newtons using the measured AFM and $\mathrm{FFM_T}$ voltage signals, it is necessary to first have an accurate value of the spring constant of the cantilever (k_c). The spring constant can be calculated using the geometry and the physical properties of the cantilever material [11.8,94,137]. However, the properties of the PECVD Si_3N_4 (used in fabricating cantilevers) could be different from those of the bulk material. For example, by using an ultrasonic measurement, we found the Young's modulus of the cantilever beam to be about 238 ± 18 GPa which is less than that of bulk Si_3N_4 (310 GPa). Furthermore the thickness of the beam is nonuniform and difficult to measure precisely. Since the stiffness of a beam goes as the cube of thickness, minor errors in precise measurements of thickness can introduce substantial stiffness errors. Thus one should experimentally measure the spring constant of the cantilever. *Cleveland* et al. [11.145] measured the normal spring constant by measuring resonant frequencies of the beams.

For normal spring constant measurement, *Ruan* and *Bhushan* [11.136] used a stainless steel spring sheet of known stiffness (width = 1.35 mm, thickness = 15 μm, free hanging length = 5.2 mm). One end of the spring was attached to the sample holder and the other end was made to contact with the cantilever tip during the measurement, see Fig. 11.23. They measured the piezo traveling distance for a given cantilever deflection. For a rigid sample (such as diamond), the piezo traveling distance Z_t (measured from the point where the tip touches the sample) should equal the cantilever deflection. To keep the cantilever deflection at the same level using a flexible spring sheet, the new piezo traveling distance $Z_{t'}$ would be different from Z_t. The difference between $Z_{t'}$ and Z_t corresponds to the deflection of the spring

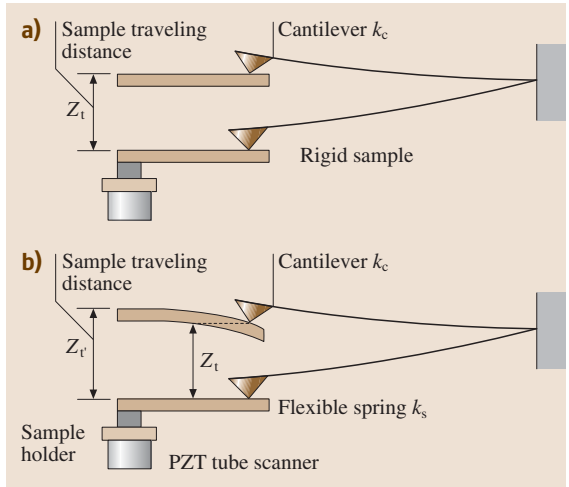


Fig. 11.23a,b Illustration showing the deflection of cantilever as it is pushed by **(a)** a rigid sample or by **(b)** a flexible spring sheet [11.136]

sheet. If the spring constant of the spring sheet is k_s, the spring constant of the cantilever k_c can be calculated by

$$(Z_{t'} - Z_t)k_s = Z_t k_c$$

or

$$k_c = k_s(Z_{t'} - Z_t)/Z_t\,. \tag{11.8}$$

The spring constant of the spring sheet (k_s) used in this study is calculated to be 1.54 N/m. For a wide-legged cantilever used in our study (length = 115 μm, base width = 122 μm, leg width = 21 μm, and thickness = 0.6 μm), k_c was measured to be 0.40 N/m instead of 0.58 N/m reported by its manufacturer – Digital Instruments Inc. To relate photodiode detector output to the cantilever deflection in nm, they used the same rigid sample to push against the AFM tip. Since for a rigid sample the cantilever vertical deflection equals the sample traveling distance measured from the point where the tip touches the sample, the photodiode output as the tip is pushed by the sample can be converted directly to cantilever deflection. For these measurements, they found the conversion factor to be 20 nm/V.

The normal force applied to the tip can be calculated by multiplying the cantilever vertical deflection by the cantilever spring constant for samples which have very small adhesive force with the tip. If the adhesive force between the sample and the tip is large, it should be included in the normal force calculation. This is particularly important in atomic-scale force measurement because in this region, the typical normal force that is measured is in the range of a few hundreds of nN to a few mN. The adhesive force could be comparable to the applied force.

The conversion of friction signal (from FFM_T to friction force) is not as straightforward. For example, one can calculate the degree of twisting for a given friction force using the geometry and the physical properties of the cantilever [11.53, 144]. One would need the information on the detectors such as the quantum efficiency of the detector, the laser power, the instrument's gain, etc. in order to be able convert the signal into the degree of twisting. Generally speaking, this procedure can not be accomplished without having some detailed information about the instrument. This information is not usually provided by the manufactures. Even if this information is readily available, error may still occur in using this approach because there will always be variations as a result of the instrumental set up. For example, it has been noticed that the measured FFM_T signal could be different for the same sample when different AFM microscopes of the same kind are used. The essence is that, one can not calibrate the instrument experimentally using this calculation. *O'Shea* et al. [11.144] did perform a calibration procedure in which the torsional signal was measured as the sample is displaced a known distance laterally while ensuring that the tip does not slide over the surface. However, it is difficult to verify that tip sliding does not occur.

Apparently, a new method of calibration is required. There is a more direct and simpler way of doing this. The first method described (method 1) to measure friction can directly provide an absolute value of the coefficient of friction. It can therefore be used just as an internal means of calibration for the data obtained using method 2. Or for a polished sample which introduces least error in friction measurement using method 1, method 1 can be used to obtain calibration for friction force for method 2. Then this calibration can be used for measurement on all samples using method 2. In method 1, the length of the cantilever required can be measured using an optical microscope; the length of the tip can be measured using a scanning electron microscope. The relative angle between the cantilever and the horizontal sample surface can be measured directly. Thus the coefficient of friction can be measured with few unknown parameters. The friction force can then be calculated by multiplying the coefficient of friction by the normal load. The FFM_T signal obtained using method 2 can then be converted into friction force. For their instrument, they found the conversion to be 8.6 nN/V.

11.3 AFM Instrumentation and Analyses

The performance of AFMs and the quality of AFM images greatly depend on the instrument available and the probes (cantilever and tips) in use. This section describes the mechanics of cantilevers, instrumentation and analysis of force detection systems for cantilever deflections, and scanning and control systems.

11.3.1 The Mechanics of Cantilevers

Stiffness and Resonances of Lumped Mass Systems

Any one of the building blocks of an AFM, be it the body of the microscope itself or the force measuring can-

tilevers, are mechanical resonators. These resonances can be excited either by the surroundings or by the rapid movement of the tip or the sample. To avoid problems due to building or air induced oscillations, it is of paramount importance to optimize the design of AFMs for high resonant frequencies. This usually means decreasing the size of the microscope [11.146]. By using cube-like or sphere-like structures for the microscope, one can considerably increase the lowest eigenfrequency. The fundamental natural frequency, ω_0, of any spring is given by

$$\omega_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m_{\text{eff}}}}\,, \tag{11.9}$$

where k is the spring constant (stiffness) in the normal direction and m_{eff} is the effective mass. The spring constant k of a cantilever beam with uniform cross section (Fig. 11.24) is given by [11.147]

$$k = \frac{3EI}{L^3}\,, \tag{11.10}$$

where E is the Young's modulus of the material, L is the length of the beam and I is the moment of inertia of the cross section. For a rectangular cross section with a width b (perpendicular to the deflection) and a height h one obtains an expression for I

$$I = \frac{bh^3}{12}\,. \tag{11.11}$$

Combining (11.9), (11.10) and (11.11) we get an expression for ω_0

$$\omega_0 = \sqrt{\frac{Ebh^3}{4L^3 m_{\text{eff}}}}\,. \tag{11.12}$$

The effective mass can be calculated using Raleigh's method. The general formula using Raleigh's method for the kinetic energy T of a bar is

$$T = \frac{1}{2}\int_0^L \frac{m}{L}\left(\frac{\partial z(x)}{\partial t}\right)^2 \mathrm{d}x\,. \tag{11.13}$$

For the case of a uniform beam with a constant cross section and length L one obtains for the deflection $z(x) = z_{\max}\left[1-(3x/2L)+(x^3/2L^3)\right]$. Inserting $z_{\max}$ into (11.13) and solving the integral gives

$$T = \frac{1}{2}\int_0^L \frac{m}{L}\left[\frac{\partial z_{\max}(x)}{\partial t}\left(1-\frac{3x}{2L}\right)+\left(\frac{x^3}{L^3}\right)\right]^2 \mathrm{d}x$$
$$= \frac{1}{2}m_{\text{eff}}(z_{\max}t)^2\,,$$

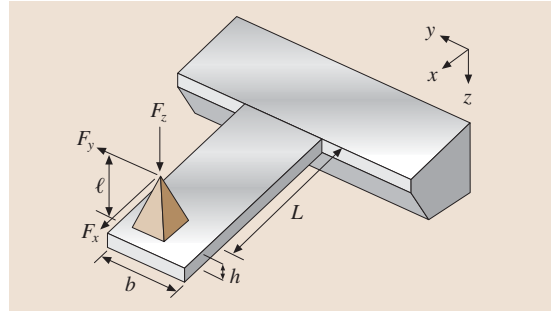


Fig. 11.24 A typical AFM cantilever with length L, width b, and height h. The height of the tip is ℓ. The material is characterized by Young's modulus E, the shear modulus G and a mass density ρ. Normal (F_z), axial (F_x), and lateral (F_y) forces exist at the end of the tip

which gives

$$m_{\text{eff}} = \frac{9}{20}m\,. \tag{11.14}$$

Substituting (11.14) into (11.12) and noting that $m = \rho Lbh$, where ρ is the mass density, one obtains the following expression

$$\omega_0 = \left(\frac{\sqrt{5}}{3}\sqrt{\frac{\mathrm{E}}{\rho}}\right)\frac{\mathrm{h}}{L^2}\,. \tag{11.15}$$

It is evident from (11.15), that one way to increase the natural frequency is to choose a material with a high ratio E/ρ; see Table 11.2 for typical values of $\sqrt{E/\rho}$ of various commonly used materials. Another way to increase the lowest eigenfrequency is also evident in (11.15). By optimizing the ratio h/L^2, one can increase the resonant frequency. However it does not help to make the length of the structure smaller than the width or height. Their roles will just be interchanged. Hence the optimum structure is a cube. This leads to the design rule, that long, thin structures like sheet metal should be avoided. For a given resonant frequency, the quality factor Q should be as low as possible. This means that an inelastic medium such as rubber should be in contact with the structure to convert kinetic energy into heat.

Stiffness and Resonances of Cantilevers

Cantilevers are mechanical devices specially shaped to measure tiny forces. The analysis given in the previous section is applicable. However, to understand better the intricacies of force detection systems we will discuss the

example of a cantilever beam with uniform cross section, Fig. 11.24. The bending of a beam due to a normal load on the beam is governed by the Euler equation [11.147]

$$M = EI\,(x)\,\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\,, \tag{11.16}$$

where M is the bending moment acting on the beam cross section. $I(x)$ the moment of inertia of the cross section with respect to the neutral axis defined by

$$I(x) = \int_z \int_y z^2\,\mathrm{d}y\,\mathrm{d}z\,. \tag{11.17}$$

For a normal force F_z acting at the tip,

$$M(x) = (L - x)\,F_z \tag{11.18}$$

since the moment must vanish at the endpoint of the cantilever. Integrating (11.16) for a normal force F_z acting at the tip and observing that EI is a constant for beams with a uniform cross section, one gets

$$z(x) = \frac{L^3}{6EI}\left(\frac{x}{L}\right)^2\left(3 - \frac{x}{L}\right)F_z\,. \tag{11.19}$$

The slope of the of the beam is

$$z'\,(x) = \frac{Lx}{2EI}\left(2 - \frac{x}{L}\right)F_z\,. \tag{11.20}$$

From (11.19) and (11.20), at the end of the cantilever, i. e. for $x = L$, for a rectangular beam, and by using an expression for I in (11.11), one gets,

$$z(L) = \frac{4}{Eb}\left(\frac{L}{h}\right)^3 F_z\,, \tag{11.21}$$

$$z'(L) = \frac{3}{2}\left(\frac{z}{L}\right)\,. \tag{11.22}$$

Now the stiffness in the normal (z) direction, k_z, is

$$k_z = \frac{F_z}{z(L)} = \frac{Eb}{4}\left(\frac{h}{L}\right)^3\,. \tag{11.23}$$

and a change in angular orientation of the end of cantilever beam is

$$\Delta\alpha = \frac{3}{2}\frac{z}{L} = \frac{6}{Ebh}\left(\frac{L}{h}\right)^2 F_z\,. \tag{11.24}$$

Now we ask what will, to first order, happen if we apply a lateral force F_y to the end of the tip (Fig. 11.24). The cantilever will bend sideways and it will twist. The stiffness in the lateral (y) direction, k_y, can be calculated with (11.23) by exchanging b and h

$$k_y = \frac{Eh}{4}\left(\frac{b}{L}\right)^3\,. \tag{11.25}$$

Therefore the bending stiffness in lateral direction is larger than the stiffness for bending in the normal direction by $(b/h)^2$. The twisting or torsion on the other hand is more complicated to handle. For a wide, thin cantilever ($b \gg h$) we obtain torsional stiffness along y axis, k_{yT}

$$k_{yT} = \frac{Gbh^3}{3L\ell^2}\,, \tag{11.26}$$

where G is the modulus of rigidity ($= E/2(1+\nu)$, where ν is the Poisson's ratio). The ratio of the torsional stiffness to the lateral bending stiffness is

$$\frac{k_{yT}}{k_y} = \frac{1}{2}\left(\frac{\ell b}{hL}\right)^2\,, \tag{11.27}$$

where we assume $\nu = 0.333$. We see that thin, wide cantilevers with long tips favor torsion while cantilevers with square cross sections and short tips favor bending. Finally we calculate the ratio between the torsional stiffness and the normal bending stiffness,

$$\frac{k_{yT}}{k_z} = 2\left(\frac{L}{\ell}\right)^2\,. \tag{11.28}$$

Equations (11.26) to (11.28) hold in the case where the cantilever tip is exactly in the middle axis of the cantilever. Triangular cantilevers and cantilevers with tips not on the middle axis can be dealt with by finite element methods.

The third possible deflection mode is the one from the force on the end of the tip along the cantilever axis, F_x (Fig. 11.24). The bending moment at the free end of the cantilever is equal to the $F_x\ell$. This leads to the following modification of (11.18) for the case of forces F_z and F_x

$$M\,(x) = (L - x)\,F_z + F_x\ell\,. \tag{11.29}$$

Integration of (11.16) now leads to

$$z(x) = \frac{1}{2EI}\left[Lx^2\left(1 - \frac{x}{3L}\right)F_z + \ell x^2 F_x\right] \tag{11.30}$$

and

$$z'\,(x) = \frac{1}{EI}\left[\frac{Lx}{2}\left(2 - \frac{x}{L}\right)F_z + \ell x F_x\right]\,. \tag{11.31}$$

Evaluating (11.30) and (11.31) at the end of the cantilever, we get the deflection and the tilt

$$z\,(L) = \frac{L^2}{EI}\left(\frac{L}{3}F_z - \frac{\ell}{2}F_x\right)\,,$$

$$z'\,(L) = \frac{L}{EI}\left(\frac{L}{2}F_z + \ell F_x\right)\,. \tag{11.32}$$

From these equations, one gets

$$F_z = \frac{12EI}{L^3}\left[z(L) - \frac{Lz'(L)}{2}\right],$$
$$F_x = \frac{2EI}{\ell L^2}\left[2Lz'(L) - 3z(L)\right]. \quad (11.33)$$

A second class of interesting properties of cantilevers is their resonance behavior. For cantilever beams one can calculate the resonant frequencies [11.147, 148]

$$\omega_n^{\text{free}} = \frac{\lambda_n^2}{2\sqrt{3}}\frac{h}{L^2}\sqrt{\frac{E}{\rho}} \quad (11.34)$$

with $\lambda_0 = (0.596864\ldots)\pi$, $\lambda_1 = (1.494175\ldots)\pi$, $\lambda_n \to (n+1/2)\pi$. The subscript n represents the order of the frequency, e.g., fundamental, second mode, and the nth mode.

A similar equation to (11.34) holds for cantilevers in rigid contact with the surface. Since there is an additional restriction on the movement of the cantilever, namely the location of its end point, the resonant frequency increases. Only the λ_n's terms change to [11.148]

$$\lambda_0' = (1.2498763\ldots)\pi, \quad \lambda_1' = (2.2499997\ldots)\pi,$$
$$\lambda_n' \to (n+1/4)\pi. \quad (11.35)$$

The ratio of the fundamental resonant frequency in contact to the fundamental resonant frequency not in contact is 4.3851.

For the torsional mode we can calculate the resonant frequencies as

$$\omega_0^{\text{tors}} = 2\pi\frac{h}{Lb}\sqrt{\frac{G}{\rho}}. \quad (11.36)$$

For cantilevers in rigid contact with the surface, we obtain the expression for the fundamental resonant frequency [11.148]

$$\omega_0^{\text{tors, contact}} = \frac{\omega_0^{\text{tors}}}{\sqrt{1+3(2L/b)^2}}. \quad (11.37)$$

The amplitude of the thermally induced vibration can be calculated from the resonant frequency using

$$\Delta z_{\text{therm}} = \sqrt{\frac{k_B T}{k}}, \quad (11.38)$$

where k_B is Boltzmann's constant and T is the absolute temperature. Since AFM cantilevers are resonant structures, sometimes with rather high Q, the thermal noise is not evenly distributed as (11.38) suggests. The spectral noise density below the peak of the response curve is [11.148]

$$z_0 = \sqrt{\frac{4k_B T}{k\omega_0 Q}} \quad (\text{in m}/\sqrt{\text{Hz}}), \quad (11.39)$$

where Q is the quality factor of the cantilever, described earlier.

11.3.2 Instrumentation and Analyses of Detection Systems for Cantilever Deflections

A summary of selected detection systems was provided in Fig. 11.8. Here we discuss in detail pros and cons of various systems.

Optical Interferometer Detection Systems

Soon after the first papers on the AFM [11.2] appeared, which used a tunneling sensor, an instrument based on an interferometer was published [11.149]. The sensitivity of the interferometer depends on the wavelength of the light employed in the apparatus. Figure 11.25 shows the principle of such an interferometeric design. The light incident from the left is focused by a lens on the cantilever. The reflected light is collimated by the same lens and interferes with the light reflected at the flat. To separate the reflected light from the incident light a $\lambda/4$ plate converts the linear polarized incident light to circular polarization. The reflected light is made linear polarized again by the $\lambda/4$-plate, but with a polarization orthogonal to that of the incident light. The polarizing

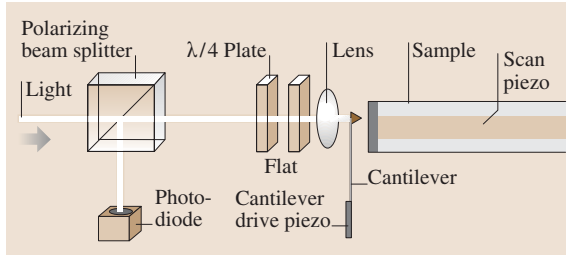


Fig. 11.25 Principle of an interferometric AFM. The light of the laser light source is polarized by the polarizing beam splitter and focused on the back of the cantilever. The light passes twice through a quarter wave plate and is hence orthogonally polarized to the incident light. The second arm of the interferometer is formed by the flat. The interference pattern is modulated by the oscillating cantilever

beam splitter then deflects the reflected light to the photo diode.

Homodyne Interferometer. To improve the signal to noise ratio of the interferometer the cantilever is driven by a piezo near its resonant frequency. The amplitude Δz of the cantilever as a function of driving frequency Ω is

$$\Delta z\,(\Omega) = \Delta z_0 \frac{\Omega_0^2}{\sqrt{\left(\Omega^2 - \Omega_0^2\right)^2 + \frac{\Omega^2\Omega_0^2}{Q^2}}}\,, \tag{11.40}$$

where Δz_0 is the constant drive amplitude and Ω_0 the resonant frequency of the cantilever. The resonant frequency of the cantilever is given by the effective potential

$$\Omega_0 = \sqrt{\left(k + \frac{\partial^2 U}{\partial z^2}\right)\frac{1}{m_{\text{eff}}}}\,, \tag{11.41}$$

where U is the interaction potential between the tip and the sample. Equation (11.41) shows that an attractive potential decreases Ω_0. The change in Ω_0 in turn results in a change of the Δz (see (11.40)). The movement of the cantilever changes the path difference in the interferometer. The light reflected from the cantilever with the amplitude $A_{\ell,0}$ and the reference light with the amplitude $A_{r,0}$ interfere on the detector. The detected intensity $I(t) = [A_\ell(t) + A_r(t)]^2$ consists of two constant terms and a fluctuating term

$$2A_\ell\,(t)\,A_r\,(t) = A_{\ell,0}A_{r,0}\sin\left[\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\right]\sin\,(\omega t)\ . \tag{11.42}$$

Here ω is the frequency of the light, λ is the wavelength of the light, δ is the path difference in the interferometer, and Δz is the instantaneous amplitude of the cantilever, given according to (11.40) and (11.41) as a function of Ω, k, and U. The time average of (11.42) then becomes

$$\begin{aligned}\langle 2A_\ell\,(t)\,A_r\,(t)\rangle_T &\propto \cos\left[\frac{4\pi\delta}{\lambda} + \frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\right]\\ &\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \sin\left[\frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\right]\\ &\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\ .\end{aligned} \tag{11.43}$$

Here all small quantities have been omitted and functions with small arguments have been linearized. The amplitude of Δz can be recovered with a lock-in technique. However, (11.43) shows that the measured amplitude is also a function of the path difference δ in the interferometer. Hence this path difference δ must be very stable. The best sensitivity is obtained when $\sin(4\delta/\lambda) \approx 0$.

Heterodyne Interferometer. This influence is not present in the heterodyne detection scheme shown in Fig. 11.26. Light incident from the left with a frequency ω is split in a reference path (upper path in Fig. 11.26) and a measurement path. Light in the measurement path is shifted in frequency to $\omega_1 = \omega + \Delta\omega$ and focused on the cantilever. The cantilever oscillates at the frequency Ω, as in the homodyne detection scheme. The reflected light $A_\ell(t)$ is collimated by the same lens and interferes on the photo diode with the reference light $A_r(t)$. The fluctuating term of the intensity is given by

$$2A_\ell\,(t)\,A_r\,(t) = A_{\ell,0}A_{r,0}\sin\left[(\omega + \Delta\omega)\,t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\right]\sin\,(\omega t)\ , \tag{11.44}$$

where the variables are defined as in (11.42). Setting the path difference $\sin(4\pi\delta/\lambda) \approx 0$ and taking the time average, omitting small quantities and linearizing functions with small arguments we get

$$\langle 2A_\ell\,(t)\,A_r\,(t)\rangle_{\text{T}} \propto \cos\left[\Delta\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\,\Delta z}{\lambda}\sin\,(\Omega t)\right]$$

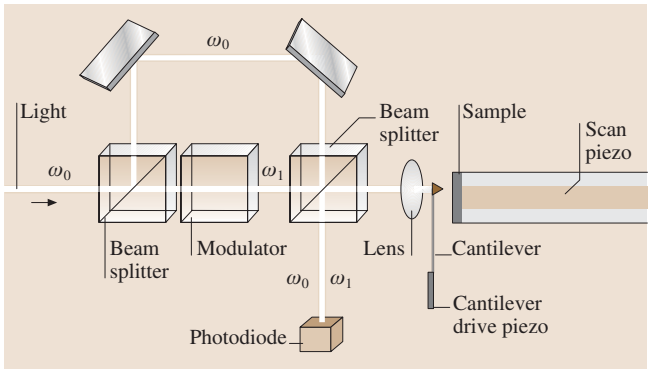


Fig. 11.26 Principle of a heterodyne interferometric AFM. Light with the frequency ω_0 is split into a reference path (upper path) and a measurement path. The light in the measurement path is frequency shifted to ω_1 by an acousto-optical modulator (or an electro-optical modulator) The light reflected from the oscillating cantilever interferes with the reference beam on the detector

$$
\begin{aligned}
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\cos\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\quad - \sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\approx \cos\left(\frac{4\pi\delta}{\lambda}\right) - \sin\left[\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)\right] \\
&\approx \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\left[1 - \frac{8\pi^2\Delta z^2}{\lambda^2}\sin(\Omega t)\right] \\
&\quad - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) - \frac{8\pi^2\Delta z^2}{\lambda^2}\cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) \\
&\quad \times \sin(\Omega t) - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) - \frac{4\pi^2\Delta z^2}{\lambda^2}\cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) \\
&\quad + \frac{4\pi^2\Delta z^2}{\lambda^2}\cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\cos(2\Omega t) \\
&\quad - \frac{4\pi\Delta z}{\lambda}\sin\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\sin(\Omega t) \\
&= \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right)\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \\
&\quad + \frac{2\pi^2\Delta z^2}{\lambda^2}\left\{\cos\left[(\Delta\omega + 2\Omega)t + \frac{4\pi\delta}{\lambda}\right]\right. \\
&\quad \left. + \cos\left[(\Delta\omega - 2\Omega)t + \frac{4\pi\delta}{\lambda}\right]\right\} \\
&\quad + \frac{2\pi\Delta z}{\lambda}\left\{\cos\left[(\Delta\omega + \Omega)t + \frac{4\pi\delta}{\lambda}\right]\right. \\
&\quad \left. + \cos\left[(\Delta\omega - \Omega)t + \frac{4\pi\delta}{\lambda}\right]\right\} . \qquad (11.45)
\end{aligned}
$$

Multiplying electronically the components oscillating at $\Delta\omega$ and $\Delta\omega + \Omega$ and rejecting any product except the one oscillating at Ω we obtain

$$
\begin{aligned}
A &= \frac{2\Delta z}{\lambda}\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right)\cos\left[(\Delta\omega + 2\Omega)t + \frac{4\pi\delta}{\lambda}\right] \\
&\quad \times \cos\left(\Delta\omega t + \frac{4\pi\delta}{\lambda}\right) \\
&= \frac{\Delta z}{\lambda}\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right)\left\{\cos\left[(2\Delta\omega + \Omega)t + \frac{8\pi\delta}{\lambda}\right]\right. \\
&\quad \left. + \cos(\Omega t)\right\} \\
&\approx \frac{\pi\Delta z}{\lambda}\cos(\Omega t) . \qquad (11.46)
\end{aligned}
$$

Unlike in the homodyne detection scheme the recovered signal is independent from the path difference δ of the interferometer. Furthermore a lock-in amplifier with the reference set $\sin(\Delta\omega t)$ can measure the path difference δ independent of the cantilever oscillation. If necessary, a feedback circuit can keep $\delta = 0$.

Fiber-optical Interferometer. The fiber-optical interferometer [11.129] is one of the simplest interferometers to build and use. Its principle is sketched in Fig. 11.27. The light of a laser is fed into an optical fiber. Laser diodes with integrated fiber pigtails are convenient light sources. The light is split in a fiber-optic beam splitter into two fibers. One fiber is terminated by index matching oil to avoid any reflections back into the fiber. The end of the other fiber is brought close to the cantilever in the AFM. The emerging light is partially reflected back into the fiber by the cantilever. Most of the light, however, is lost. This is not a big problem since only 4% of the light is reflected at the end of the fiber, at the glass-air interface. The two reflected light waves interfere with each other. The product is guided back into the fiber coupler and again split into two parts. One half is analyzed by the photodiode. The other half is fed back into the laser. Communications grade laser diodes are sufficiently resistant against feedback to be operated in this environment. They have, however, a bad coherence length, which in this case does not matter, since the optical path difference is in any case no larger than 5 μm. Again the end of the fiber has to be positioned on a piezo drive to set the distance between the fiber and the cantilever to $\lambda(n + 1/4)$.

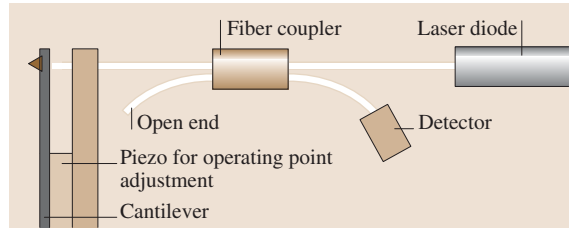


Fig. 11.27 A typical setup for a fiber optic interferometer readout

Nomarski-Interferometer. Another solution to minimize the optical path difference is to use the Nomarski interferometer [11.130]. Figure 11.28 shows a schematic of the microscope. The light of a laser is focused on the cantilever by lens. A birefringent crystal (for instance calcite) between the cantilever and the lens with its optical axis 45° off the polarization direction of the light

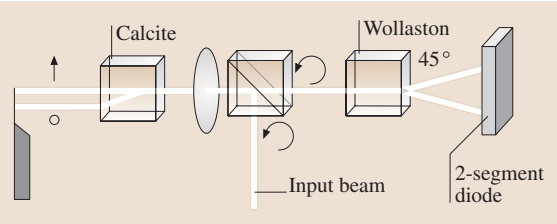


Fig. 11.28 Principle of Nomarski AFM. The circular polarized input beam is deflected to the left by a non-polarizing beam splitter. The light is focused onto a cantilever. The calcite crystal between the lens and the cantilever splits the circular polarized light into two spatially separated beams with orthogonal polarizations. The two light beams reflected from the lever are superimposed by the calcite crystal and collected by the lens. The resulting beam is again circular polarized. A Wollaston prism produces two interfering beams with $\pi/2$ phase shift between them. The minimal path difference accounts for the excellent stability of this microscope

splits the light beam into two paths, offset by a distance given by the length of the crystal. Birefringent crystals have varying indexes of refraction. In calcite, one crystal axis has a lower index than the other two. This means, that certain light rays will propagate at a different speed through the crystal than the others. By choosing a correct polarization, one can select the ordinary ray, the extraordinary ray or one can get any distribution of the intensity amongst those two rays. A detailed description of birefringence can be found in textbooks (e.g., [11.150]). A calcite crystal deflects the extraordinary ray at an angle of 6° within the crystal. By choosing a suitable length of the calcite crystal, any separation can be set.

The focus of one light ray is positioned near the free end of the cantilever while the other is placed close to the clamped end. Both arms of the interferometer pass through the same space, except for the distance between the calcite crystal and the lever. The closer the calcite crystal is placed to the lever, the less influence disturbances like air currents have.

Sarid [11.116] has given values for the sensitivity of the different interferometeric detection systems. Table 11.5 presents a summary of his results.

Optical Lever

The most common cantilever deflection detection system is the optical lever [11.53, 111]. This method, depicted in Fig. 11.29, employs the same technique as light beam deflection galvanometers. A fairly well collimated light beam is reflected off a mirror and projected to a receiving target. Any change in the angular position of the mirror will change the position, where the light ray hits the target. Galvanometers use optical path lengths of several meters and scales projected to the target wall as a read-out help.

For the AFM using the optical lever method a photodiode segmented into two (or four) closely spaced devices detects the orientation of the end of the cantilever. Initially, the light ray is set to hit the photodiodes in the middle of the two sub-diodes. Any deflection of

Table 11.5 Noise in Interferometers. F is the finesse of the cavity in the homodyne interferometer, P_i the incident power, P_d is the power on the detector, η is the sensitivity of the photodetector and RIN is the relative intensity noise of the laser. P_R and P_S are the power in the reference and sample beam in the heterodyne interferometer. P is the power in the Nomarski interferometer, $\delta\theta$ is the phase difference between the reference and the probe beam in the Nomarski interferometer. B is the bandwidth, e is the electron charge, λ is the wavelength of the laser, k the cantilever stiffness, ω_0 is the resonant frequency of the cantilever, Q is the quality factor of the cantilever, T is the temperature, and δi is the variation of current i

	Homodyne interferometer, fiber optic interferometer	Heterodyne interferometer	Nomarski interferometer
Laser noise $\langle \delta i^2 \rangle_L$	$\frac{1}{4}\eta^2 F^2 P_i^2$ RIN	$\eta^2 \left(P_R^2 + P_S^2\right)$ RIN	$\frac{1}{16}\eta^2 P^2 \delta\theta$
Thermal noise $\langle \delta i^2 \rangle_T$	$\frac{16\pi^2}{\lambda^2}\eta^2 F^2 P_i^2 \frac{4k_B TBQ}{\omega_0 k}$	$\frac{4\pi^2}{\lambda^2}\eta^2 P_d^2 \frac{4k_B TBQ}{\omega_0 k}$	$\frac{\pi^2}{\lambda^2}\eta^2 P^2 \frac{4k_B TBQ}{\omega_0 k}$
Shot Noise $\langle \delta i^2 \rangle_S$	$4e\eta P_d B$	$2e\eta\,(P_R + P_S)\,B$	$\frac{1}{2}e\eta PB$

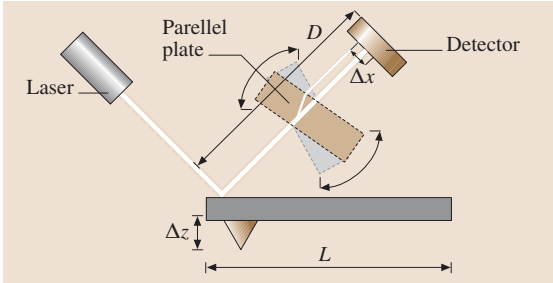


Fig. 11.29 The setup of optical lever detection microscope

the cantilever will cause an imbalance of the number of photons reaching the two halves. Hence the electrical currents in the photodiodes will be unbalanced too. The difference signal is further amplified and is the input signal to the feedback loop. Unlike the interferometeric AFMs, where often a modulation technique is necessary to get a sufficient signal to noise ratio, most AFMs employing the optical lever method are operated in a static mode. AFMs based on the optical lever method are universally used. It is the simplest method for constructing an optical readout and it can be confined in volumes smaller than 5 cm on the side.

The optical lever detection system is a simple yet elegant way to detect normal and lateral force signals simultaneously [11.7, 8, 53, 111]. It has the additional advantage that it is the fact that it is a remote detection system.

Implementations. Light from a laser diode or from a super luminescent diode is focused on the end of the cantilever. The reflected light is directed onto a quadrant diode that measures the direction of the light beam. A Gaussian light beam far from its waist is characterized by an opening angle β. The deflection of the light beam by the cantilever surface tilted by an angle α is 2α. The intensity on the detector then shifts to the side by the product of 2α and the separation between the detector and the cantilever. The readout electronics calculates the difference of the photocurrents. The photocurrents in turn are proportional to the intensity incident on the diode.

The output signal is hence proportional to the change in intensity on the segments

$$I_{\text{sig}} \propto 4\frac{\alpha}{\beta} I_{\text{tot}} \,. \tag{11.47}$$

For the sake of simplicity, we assume that the light beam is of uniform intensity with its cross section increasing proportional to the distance between the cantilever and the quadrant detector. The movement of the center of the light beam is then given by

$$\Delta x_{\text{Det}} = \Delta z \frac{D}{L} \,. \tag{11.48}$$

The photocurrent generated in a photodiode is proportional to the number of incoming photons hitting it. If the light beam contains a total number of N_0 photons then the change in difference current becomes

$$\Delta\left(I_{\text{R}} - I_{\text{L}}\right) = \Delta I = \text{const}\ \Delta z\ D\ N_0 \,. \tag{11.49}$$

Combining (11.48) and (11.49), one obtains that the difference current ΔI is independent of the separation of the quadrant detector and the cantilever. This relation is true, if the light spot is smaller than the quadrant detector. If it is greater, the difference current ΔI becomes smaller with increasing distance. In reality the light beam has a Gaussian intensity profile. For small movements Δx (compared to the diameter of the light spot at the quadrant detector), (11.49) still holds. Larger movements Δx, however, will introduce a nonlinear response. If the AFM is operated in a constant force mode, only small movements Δx of the light spot will occur. The feedback loop will cancel out all other movements.

The scanning of a sample with an AFM can twist the microfabricated cantilevers because of lateral forces [11.5, 7, 8] and affect the images [11.120]. When the tip is subjected to lateral forces, it will twist the cantilever and the light beam reflected from the end of the cantilever will be deflected perpendicular to the ordinary deflection direction. For many investigations this influence of lateral forces is unwanted. The design of the triangular cantilevers stems from the desire, to minimize the torsion effects. However, lateral forces open up a new dimension in force measurements. They allow, for instance, a distinction of two materials because of the different friction coefficient, or the determination of adhesion energies. To measure lateral forces the original optical lever AFM has to be modified. The only modification compared with Fig. 11.29 is the use of a quadrant detector photodiode instead of a two-segment photodiode and the necessary readout electronics, see Fig. 11.9a. The electronics calculates the following signals:

$$\begin{aligned} U_{\text{Normal Force}} &= \alpha\left[\left(I_{\text{Upper Left}} + I_{\text{Upper Right}}\right) \right. \\ &\quad \left. - \left(I_{\text{Lower Left}} + I_{\text{Lower Right}}\right)\right] \,, \\ U_{\text{Lateral Force}} &= \beta\left[\left(I_{\text{Upper Left}} + I_{\text{Lower Left}}\right) \right. \\ &\quad \left. - \left(I_{\text{Upper Right}} + I_{\text{Lower Right}}\right)\right] \,. \end{aligned} \tag{11.50}$$

The calculation of the lateral force as a function of the deflection angle does not have a simple solution

for cross-sections other than circles. An approximate formula for the angle of twist for rectangular beams is [11.151]

$$\theta = \frac{M_t L}{\beta G b^3 h}\,, \tag{11.51}$$

where $M_t = F_y \ell$ is the external twisting moment due to lateral force, F_y, and β a constant determined by the value of h/b. For the equation to hold, h has to be larger than b.

Inserting the values for a typical microfabricated cantilever with integrated tips

$$\begin{aligned} b &= 6\times 10^{-7}\,\mathrm{m}\,,\\ h &= 10^{-5}\,\mathrm{m}\,,\\ L &= 10^{-4}\,\mathrm{m}\,,\\ \ell &= 3.3\times 10^{-6}\,\mathrm{m}\,,\\ G &= 5\times 10^{10}\,\mathrm{Pa}\,,\\ \beta &= 0.333 \end{aligned} \tag{11.52}$$

into (11.51) we obtain the relation

$$F_y = 1.1\times 10^{-4}\,\mathrm{N}\times\theta\,. \tag{11.53}$$

Typical lateral forces are of order 10^{-10} N.

Sensitivity. The sensitivity of this setup has been calculated in various papers [11.116, 148, 152]. Assuming a Gaussian beam the resulting output signal as a function of the deflection angle is dispersion like. Equation (11.47) shows that the sensitivity can be increased by increasing the intensity of the light beam I_{tot} or by decreasing the divergence of the laser beam. The upper bound of the intensity of the light I_{tot} is given by saturation effects on the photodiode. If we decrease the divergence of a laser beam we automatically increase the beam waist. If the beam waist becomes larger than the width of the cantilever we start to get diffraction. Diffraction sets a lower bound on the divergence angle. Hence one can calculate the optimal beam waist w_{opt} and the optimal divergence angle β [11.148, 152]

$$\begin{aligned} w_{\mathrm{opt}} &\approx 0.36b\,,\\ \theta_{\mathrm{opt}} &\approx 0.89\frac{\lambda}{b}\,. \end{aligned} \tag{11.54}$$

The optimal sensitivity of the optical lever then becomes

$$\varepsilon\,[\mathrm{mW/rad}] = 1.8\frac{b}{\lambda} I_{\mathrm{tot}}\ [\mathrm{mW}]\,. \tag{11.55}$$

The angular sensitivity optical lever can be measured by introducing a parallel plate into the beam. A tilt of the parallel plate results in a displacement of the beam, mimicking an angular deflection.

Additional noise sources can be considered. Of little importance is the quantum mechanical uncertainty of the position [11.148, 152], which is, for typical cantilevers at room temperature

$$\Delta z = \sqrt{\frac{\hbar}{2m\omega_0}} = 0.05\,\mathrm{fm}\,, \tag{11.56}$$

where $\hbar$ is the Planck constant ($= 6.626\times 10^{-34}$ J s). At very low temperatures and for high frequency cantilevers this could become the dominant noise source. A second noise source is the shot noise of the light. The shot noise is related to the particle number. We can calculate the number of photons incident on the detector

$$n = \frac{I\tau}{\hbar\omega} = \frac{I\lambda}{2\pi B\hbar c} = 1.8\times 10^{9}\,\frac{I[\mathrm{W}]}{B[\mathrm{Hz}]}\,, \tag{11.57}$$

where I is the intensity of the light, τ the measurement time, $B = 1/\tau$ the bandwidth, and c the speed of light. The shot noise is proportional to the square root of the number of particles. Equating the shot noise signal with the signal resulting for the deflection of the cantilever one obtains

$$\Delta z_{\mathrm{shot}} = 68\frac{L}{w}\sqrt{\frac{B\,[\mathrm{kHz}]}{I\,[\mathrm{mW}]}}\,[\mathrm{fm}]\,, \tag{11.58}$$

where w is the diameter of the focal spot. Typical AFM setups have a shot noise of 2 pm. The thermal noise can be calculated from the equipartition principle. The amplitude at the resonant frequency is

$$\Delta z_{\mathrm{therm}} = 129\sqrt{\frac{B}{k\,[\mathrm{N/m}]\,\omega_0 Q}}\,[\mathrm{pm}]\,. \tag{11.59}$$

A typical value is 16 pm. Upon touching the surface, the cantilever increases its resonant frequency by a factor of 4.39. This results in a new thermal noise amplitude of 3.2 pm for the cantilever in contact with the sample.

Piezoresistive Detection

Implementation. An alternative detection system which is not as widely used as the optical detection schemes are piezoresistive cantilevers [11.125, 126, 132]. These cantilevers are based on the fact that the resistivity of certain materials, in particular of Si, changes with the applied stress. Figure 11.30 shows a typical implementation of a piezo-resistive cantilever. Four

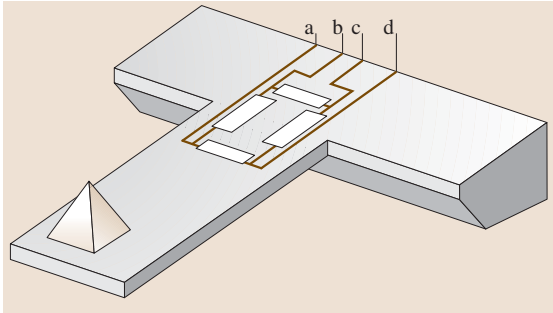


Fig. 11.30 A typical setup for a piezoresistive readout

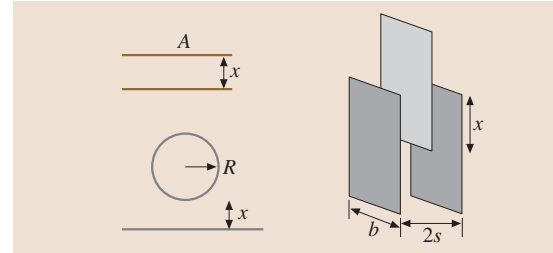


Fig. 11.31 Three possible arrangements of a capacitive readout. The *upper left* shows the cross section through a parallel plate capacitor. The *lower left* shows the geometry of sphere versus plane. The *right side* shows the more complicated, but linear capacitive readout

resistances are integrated on the chip, forming a Wheatstone bridge. Two of the resistors are in unstrained parts of the cantilever, the other two are measuring the bending at the point of the maximal deflection. For instance when an AC voltage is applied between terminals a and c one can measure the detuning of the bridge between terminals b and d. With such a connection the output signal varies only due to bending, but not due to changing of the ambient temperature and thus the coefficient of the piezoresistance.

Sensitivity. The resistance change is [11.126]

$$\frac{\Delta R}{R_0} = \Pi\delta\,, \tag{11.60}$$

where Π is the tensor element of the piezo-resistive coefficients, δ the mechanical stress tensor element and R_0 the equilibrium resistance. For a single resistor they separate the mechanical stress and the tensor element in longitudinal and transverse components

$$\frac{\Delta R}{R_0} = \Pi_\mathrm{t}\delta_\mathrm{t} + \Pi_\mathrm{l}\delta_\mathrm{l}\,. \tag{11.61}$$

The maximum value of the stress components are $\Pi_\mathrm{t} = -64.0\times10^{-11}\,\mathrm{m^2/N}$ and $\Pi_\mathrm{l} = -71.4\times10^{-11}\,\mathrm{m^2/N}$ for a resistor oriented along the (110) direction in silicon [11.126]. In the resistor arrangement of Fig. 11.30 two of the resistors are subject to the longitudinal piezo-resistive effect and two of them are subject to the transversal piezo-resistive effect. The sensitivity of that setup is about four times that of a single resistor, with the advantage that temperature effects cancel to first order. The resistance change is then calculated as

$$\frac{\Delta R}{R_0} = \Pi\frac{3Eh}{2L^2}\Delta z = \Pi\frac{6L}{bh^2}F_z\,, \tag{11.62}$$

where $\Pi = 67.7\times10^{-11}\,\mathrm{m^2/N}$ is the averaged piezo-resistive coefficient. Plugging in typical values for the dimensions (Fig. 11.24) ($L = 100\,\mu\mathrm{m}$, $b = 10\,\mu\mathrm{m}$, $h = 1\,\mu\mathrm{m}$) one obtains

$$\frac{\Delta R}{R_0} = \frac{4\times10^{-5}}{\mathrm{nN}}F_z\,. \tag{11.63}$$

The sensitivity can be tailored by optimizing the dimensions of the cantilever.

Capacitance Detection

The capacitance of an arrangement of conductors depends on the geometry. Generally speaking, the capacitance increases for decreasing separations. Two parallel plates form a simple capacitor (see Fig. 11.31, upper left), with the capacitance

$$C = \frac{\varepsilon\varepsilon_0 A}{x}\,, \tag{11.64}$$

where A is the area of the plates, assumed equal, and x is the separation. Alternatively one can consider a sphere versus an infinite plane (see Fig. 11.31, lower left). Here the capacitance is [11.116]

$$C = 4\pi\varepsilon_0 R\sum_{n=2}^{\infty}\frac{\sinh(\alpha)}{\sinh(n\alpha)} \tag{11.65}$$

where R is the radius of the sphere, and α is defined by

$$\alpha = \ln\left(1+\frac{z}{R}+\sqrt{\frac{z^2}{R^2}+2\frac{z}{R}}\right)\,. \tag{11.66}$$

One has to keep in mind that capacitance of a parallel plate capacitor is a nonlinear function of the separation. Using a voltage divider one can circumvent this problem.

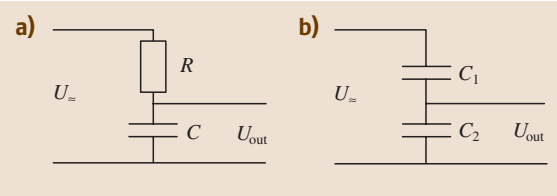


Fig. 11.32a,b Measuring the capacitance. The *left side*, **(a)**, shows a low pass filter, the *right side*, **(b)**, shows a capacitive divider. C (*left*) or C_2 (*right*) are the capacitances under test

Figure 11.32a shows a low pass filter. The output voltage is given by

$$U_{\text{out}} = U_{\approx}\frac{\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}} = U_{\approx}\frac{1}{j\omega CR+1} \cong \frac{U_{\approx}}{j\omega CR}\,. \tag{11.67}$$

Here C is given by (11.64), ω is the excitation frequency and j is the imaginary unit. The approximate relation in the end is true when $\omega CR \gg 1$. This is equivalent to the statement that C is fed by a current source, since R must be large in this setup. Plugging (11.64) into (11.67) and neglecting the phase information one obtains

$$U_{\text{out}} = \frac{U_{\approx}x}{\omega R\varepsilon\varepsilon_0 A}\,, \tag{11.68}$$

which is linear in the displacement x.

Figure 11.32b shows a capacitive divider. Again the output voltage U_{out} is given by

$$U_{\text{out}} = U_{\approx}\frac{C_1}{C_2+C_1} = U_{\approx}\frac{C_1}{\frac{\varepsilon\varepsilon_0 A}{x}+C_1}\,. \tag{11.69}$$

If there is a stray capacitance C_s then (11.69) is modified as

$$U_{\text{out}} = U_{\approx}\frac{C_1}{\frac{\varepsilon\varepsilon_0 A}{x}+C_s+C_1}\,. \tag{11.70}$$

Provided $C_s + C_1 \ll C_2$ one has a system which is linear in x. The driving voltage $U_{\approx}$ has to be large (more than 100 V) to have the output voltage in the range of 1 V. The linearity of the readout depends on the capacitance C_1 (Fig. 11.33).

Another idea is to keep the distance constant and to change the relative overlap of the plates (see Fig. 11.31, right side). The capacitance of the moving center plate versus the stationary outer plates becomes

$$C = C_s + 2\frac{\varepsilon\varepsilon_0 bx}{s}\,, \tag{11.71}$$

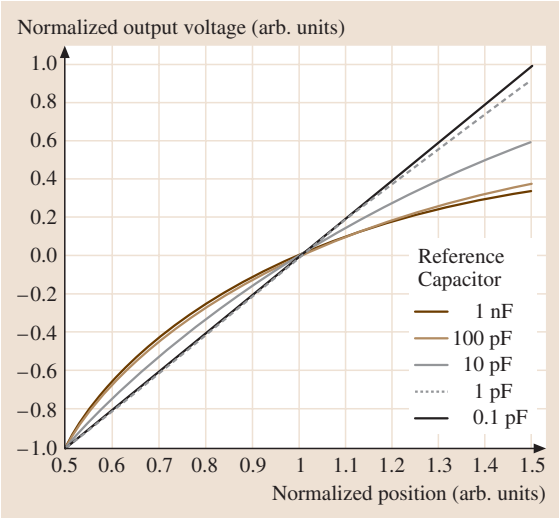


Fig. 11.33 Linearity of the capacitance readout as a function of the reference capacitor

where the variables are defined in Fig. 11.31. The stray capacitance comprises all effects, including the capacitance of the fringe fields. When the length x is comparable to the width b of the plates one can safely assume that the stray capacitance is constant and independent of x. The main disadvantage of this setup is that it is not as easily incorporated in a microfabricated device as the others.

Sensitivity. The capacitance itself is not a measure of the sensitivity, but its derivative is indicative of the signals one can expect. Using the situation described in Fig. 11.31, upper left, and in (11.64) one obtains for the parallel plate capacitor

$$\frac{\mathrm{d}C}{\mathrm{d}x} = -\frac{\varepsilon\varepsilon_0 A}{x^2}\,. \tag{11.72}$$

Assuming a plate area A of 20 µm by 40 µm and a separation of 1 µm one obtains a capacitance of 31 fF (neglecting stray capacitance and the capacitance of the connection leads) and a $\mathrm{d}C/\mathrm{d}x$ of 3.1×10^{-8} F/m = 31 fF/µm. Hence it is of paramount importance to maximize the area between the two contacts and to minimize the distance x. The latter however is far from being trivial. One has to go to the limits of microfabrication to achieve a decent sensitivity.

If the capacitance is measured by the circuit shown in Fig. 11.32 one obtains for the sensitivity

$$\frac{\mathrm{d}U_{\mathrm{out}}}{U_{\approx}} = \frac{\mathrm{d}x}{\omega R \varepsilon \varepsilon_0 A} . \tag{11.73}$$

Using the same value for A as above, setting the reference frequency to 100 kHz, and selecting $R = 1\,\mathrm{G\Omega}$, we get the relative change of the output voltage U_{out} to

$$\frac{\mathrm{d}U_{\mathrm{out}}}{U_{\approx}} = \frac{22.5 \times 10^{-6}}{\text{Å}} \times \mathrm{d}x . \tag{11.74}$$

A driving voltage of 45 V then translates to a sensitivity of 1 mV/Å. A problem in this setup is the stray capacitances. They are in parallel to the original capacitance and decrease the sensitivity considerably.

Alternatively one could build an oscillator with this capacitance and measure the frequency. RC-oscillators typically have an oscillation frequency of

$$f_{\mathrm{res}} \propto \frac{1}{RC} = \frac{x}{R \varepsilon \varepsilon_0 A} . \tag{11.75}$$

Again the resistance R must be of the order of $1\,\mathrm{G\Omega}$, when stray capacitances C_{s} are neglected. However C_{s} is of the order of 1 pF. Therefore one gets $R = 10\,\mathrm{M\Omega}$. Using these values the sensitivity becomes

$$\mathrm{d}f_{\mathrm{res}} = \frac{C\,\mathrm{d}x}{R\,(C + C_{\mathrm{s}})^2\,x} \approx \frac{0.1\,\mathrm{Hz}}{\text{Å}}\,\mathrm{d}x . \tag{11.76}$$

The bad thing is that the stray capacitances have made the signal nonlinear again. The linearized setup in Fig. 11.31 has a sensitivity of

$$\frac{\mathrm{d}C}{\mathrm{d}x} = 2\frac{\varepsilon \varepsilon_0 b}{s} . \tag{11.77}$$

Substituting typical values, $b = 10\,\mu\mathrm{m}$, $s = 1\,\mu\mathrm{m}$ one gets $\mathrm{d}C/\mathrm{d}x = 1.8 \times 10^{-10}\,\mathrm{F/m}$. It is noteworthy that the sensitivity remains constant for scaled devices.

Implementations. The readout of the capacitance can be done in different ways [11.123, 124]. All include an alternating current or voltage with frequencies in the 100 kHz to the 100 MHz range. One possibility is to build a tuned circuit with the capacitance of the cantilever determining the frequency. The resonance frequency of a high quality Q tuned circuit is

$$\omega_0 = (LC)^{-1/2} . \tag{11.78}$$

where L is the inductance of the circuit. The capacitance C includes not only the sensor capacitance but also the capacitance of the leads. The precision of a frequency measurement is mainly determined by the ratio of L and C

$$Q = \left(\frac{L}{C}\right)^{1/2} \frac{1}{R} . \tag{11.79}$$

Here R symbolizes the losses in the circuit. The higher the quality the more precise the frequency measurement. For instance a frequency of 100 MHz and a capacitance of 1 pF gives an inductance of 250 μH. The quality becomes then 2.5×10^8. This value is an upper limit, since losses are usually too high.

Using a value of $\mathrm{d}C/\mathrm{d}x = 31\,\mathrm{fF/\mu m}$ one gets $\Delta C/\text{Å} = 3.1\,\mathrm{aF/Å}$. With a capacitance of 1 pF one gets

$$\frac{\Delta\omega}{\omega} = \frac{1}{2}\frac{\Delta C}{C} ,$$

$$\Delta\omega = 100\,\mathrm{MHz} \times \frac{1}{2}\frac{3.1\mathrm{aF}}{1\,\mathrm{pF}} = 155\,\mathrm{Hz} . \tag{11.80}$$

This is the frequency shift for 1 Å deflection. The calculation shows, that this is a measurable quantity. The quality also indicates that there is no physical reason why this scheme should not work.

11.3.3 Combinations for 3-D-Force Measurements

Three dimensional force measurements are essential if one wants to know all the details of the interaction between the tip and the cantilever. The straightforward attempt to measure three forces is complicated, since force sensors such as interferometers or capacitive sensors need a minimal detection volume, which often is too large. The second problem is that the force-sensing tip has to be held by some means. This implies that one of the three Cartesian axes is stiffer than the others.

However by the combination of different sensors one can achieve this goal. Straight cantilevers are employed for these measurements, because they can be handled analytically. The key observation is, that the optical lever method does not determine the position of the end of the cantilever. It measures the orientation. In the previous sections, one has always made use of the fact, that for a force along one of the orthogonal symmetry directions at the end of the cantilever (normal force, lateral force, force along the cantilever beam axis) there is a one to one correspondence of the tilt angle and the deflection. The problem is, that the force along the cantilever beam axis and the normal force create a deflection in the same direction. Hence what is called the normal force component is actually a mixture of two forces. The deflection of the cantilever is the third quantity, which is not considered in most of the AFMs. A fiber optic interferometer in parallel to the optical lever measures the deflection. Three measured quantities then allow the separation of the three orthonormal force directions, as is evident from (11.27) and (11.33) [11.12–16].

Alternatively one can put the fast scanning direction along the axis of the cantilever. Forward and backward scans then exert opposite forces F_x. If the piezo movement is linearized, both force components in AFM based on the optical lever detection can be determined. In this case, the normal force is simply the average of the forces in the forward and backward direction. The force, F_x, is the difference of the forces measured in forward and backward direction.

11.3.4 Scanning and Control Systems

Almost all SPMs use piezo translators to scan the tip or the sample. Even the first STM [11.1, 103] and some of the predecessor instruments [11.153, 154] used them. Other materials or setups for nano-positioning have been proposed, but were not successful [11.155, 156].

Piezo Tubes

A popular solution is tube scanners (Fig. 11.34). They are now widely used in SPMs due to their simplicity and their small size [11.133, 157]. The outer electrode is segmented in four equal sectors of 90 degrees. Opposite sectors are driven by signals of the same magnitude, but opposite sign. This gives, through bending, a two-dimensional movement on, approximately, a sphere. The inner electrode is normally driven by the z signal. It is possible, however, to use only the outer electrodes for scanning and for the z-movement. The main drawback of applying the z-signal to the outer electrodes is, that the applied voltage is the sum of both the x- or y-movement and the z-movement. Hence a larger scan size effectively reduces the available range for the z-control.

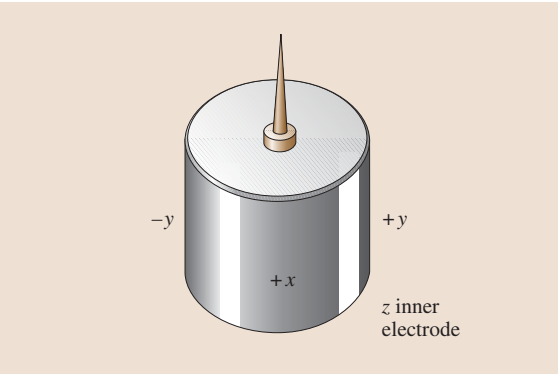


Fig. 11.34 Schematic drawing of a piezoelectric tube scanner. The piezo ceramic is molded into a tube form. The outer electrode is separated into four segments and connected to the scanning voltages. The z-voltage is applied to the inner electrode

Piezo Effect

An electric field applied across a piezoelectric material causes a change in the crystal structure, with expansion in some directions and contraction in others. Also, a net volume change occurs [11.132]. Many SPMs use the transverse piezo electric effect, where the applied electric field $\boldsymbol{E}$ is perpendicular to the expansion/contraction direction.

$$\Delta L = L\,(\boldsymbol{E}\cdot\boldsymbol{n})\,d_{31} = L\frac{V}{t}d_{31}\,, \tag{11.81}$$

where d_{31} is the transverse piezoelectric constant, V is the applied voltage, t is the thickness of the piezo slab or the distance between the electrodes where the voltage is applied, L is the free length of the piezo slab, and $\boldsymbol{n}$ is the direction of polarization. Piezo translators based on the transverse piezoelectric effect have a wide range of sensitivities, limited mainly by mechanical stability and breakdown voltage.

Scan Range

The calculation of the scanning range of a piezotube is difficult [11.157–159]. The bending of the tube depends on the electric fields and the nonuniform strain induced. A finite element calculation where the piezo tube was divided into 218 identical elements was used [11.158] to calculate the deflection. On each node the mechanical stress, stiffness, strain and piezoelectric stress was calculated when a voltage was applied on one electrode. The results were found to be linear on the first iteration and higher-order corrections were very small even for large electrode voltages. It was found that to first order the x- and z-movement of the tube could be reasonably well approximated by assuming that the piezo tube is a segment of a torus. Using this model one obtains

$$\mathrm{d}x = (V_+ - V_-)\,|d_{31}|\,\frac{L^2}{2t\,\mathrm{d}}\,, \tag{11.82}$$

$$\mathrm{d}z = (V_+ + V_- - 2V_z)\,|d_{31}|\,\frac{L}{2t}\,, \tag{11.83}$$

where $|d_{31}|$ is the coefficient of the transversal piezoelectric effect, L is tube's free length, t is tube's wall thickness, d is tube's diameter, V_+ is voltage on positive outer electrode while V_- is voltage of the opposite quadrant negative electrode, and V_z is voltage of the inner electrode.

The cantilever or sample mounted on the piezotube has an additional lateral movement because the point of

measurement is not in the end plane of the piezotube. The additional lateral displacement of the end of the tip is $\ell \sin\varphi \approx \ell\varphi$, where ℓ is the tip length and φ is the deflection angle of the end surface. Assuming that the sample or cantilever are always perpendicular to the end of the walls of the tube and calculating with the torus model one gets for the angle

$$\varphi = \frac{L}{R} = \frac{2\,\mathrm{d}x}{L}\,, \tag{11.84}$$

where R is the radius of curvature the piezo tube. Using the result of (11.84) one obtains for the additional x-movement

$$\begin{aligned}\mathrm{d}x_{\mathrm{add}} &= \ell\varphi = \frac{2\,\mathrm{d}x\ell}{L} \\ &= (V_+ - V_-)\,|d_{31}|\,\frac{\ell\,L}{td}\end{aligned} \tag{11.85}$$

and for the additional z-movement due to the x-movement

$$\begin{aligned}\mathrm{d}z_{\mathrm{add}} &= \ell - \ell\cos\varphi = \frac{\ell\varphi^2}{2} = \frac{2\ell\,(\mathrm{d}x)^2}{L^2} \\ &= (V_+ - V_-)^2\,|d_{31}|^2\,\frac{\ell L^2}{2t^2 d^2}\,.\end{aligned} \tag{11.86}$$

Carr [11.158] assumed for his finite element calculations that the top of the tube was completely free to move and, as a consequence, the top surface was distorted, leading to a deflection angle about half that of the geometrical model. Depending on the attachment of the sample or the cantilever this distortion may be smaller, leading to a deflection angle in-between that of the geometrical model and the one of the finite element calculation.

Nonlinearities and Creep

Piezo materials with a high conversion ratio, i. e. a large d_{31} or small electrode separations, with large scanning ranges are hampered by substantial hysteresis resulting in a deviation from linearity by more than 10%. The sensitivity of the piezo ceramic material (mechanical displacement divided by driving voltage) decreases with reduced scanning range, whereas the hysteresis is reduced. A careful selection of the material for the piezo scanners, the design of the scanners, and of the operating conditions is necessary to get optimum performance.

Passive Linearization: Calculation. The analysis of images affected by piezo nonlinearities [11.160–163] shows that the dominant term is

$$x = AV + BV^2\,, \tag{11.87}$$

where x is the excursion of the piezo, V the applied voltage and A and B two coefficients describing the sensitivity of the material. Equation (11.87) holds for scanning from $V = 0$ to large V. For the reverse direction the equation becomes

$$x = \tilde{A}V - \tilde{B}\,(V - V_{\max})^2\,, \tag{11.88}$$

where $\tilde{A}$ and $\tilde{B}$ are the coefficients for the back scan and $V_{\max}$ is the applied voltage at the turning point. Both equations demonstrate that the true x-travel is small at the beginning of the scan and becomes larger towards the end. Therefore images are stretched at the beginning and compressed at the end.

Similar equations hold for the slow scan direction. The coefficients, however, are different. The combined action causes a greatly distorted image. This distortion can be calculated. The data acquisition systems record the signal as a function of V. However the data is measured as a function of x. Therefore we have to distribute the x-values evenly across the image this can be done by inverting an approximation of (11.87). First we write

$$x = AV\left(1 - \frac{B}{A}V\right)\,. \tag{11.89}$$

For $B \ll A$ we can approximate

$$V = \frac{x}{A}\,. \tag{11.90}$$

We now substitute (11.90) into the nonlinear term of (11.89). This gives

$$\begin{aligned}x &= AV\left(1 + \frac{Bx}{A^2}\right)\,, \\ V &= \frac{x}{A}\,\frac{1}{(1 + Bx/A^2)} \approx \frac{x}{A}\left(1 - \frac{Bx}{A^2}\right)\,.\end{aligned} \tag{11.91}$$

Hence an equation of the type

$$\begin{aligned}&x_{\mathrm{true}} = x\,(\alpha - \beta x/x_{\max}) \\ &\text{with} \quad 1 = \alpha - \beta\end{aligned} \tag{11.92}$$

takes out the distortion of an image. α and β are dependent on the scan range, the scan speed and on the scan history and have to be determined with exactly the same settings as for the measurement. $x_{\max}$ is the maximal scanning range. The condition for α and β guarantees that the image is transformed onto itself.

Similar equations as the empirical one shown above (11.92) can be derived by analyzing the movements of domain walls in piezo ceramics.

Passive Linearization: Measuring the Position. An alternative strategy is to measure the position of the piezo translators. Several possibilities exist.

1. The interferometers described above can be used to measure the elongation of the piezo elongation. The fiber optic interferometer is especially easy to implement. The coherence length of the laser only limits the measurement range. However the signal is of a periodic nature. Hence a direct use of the signal in a feedback circuit for the position is not possible. However as a measurement tool and, especially, as a calibration tool the interferometer is without competition. The wavelength of the light, for instance in a HeNe laser, is so well defined that the precision of the other components determines the error of the calibration or measurement.
2. The movement of the light spot on the quadrant detector can be used to measure the position of a piezo [11.164]. The output current changes by $0.5\,\mathrm{A/cm} \times P(\mathrm{W})/R(\mathrm{cm})$. Typical values ($P = 1\,\mathrm{mW}$, $R = 0.001\,\mathrm{cm}$) give $0.5\,\mathrm{A/cm}$. The noise limit is typically $0.15\,\mathrm{nm} \times \sqrt{\Delta f(\mathrm{Hz})/H(\mathrm{W/cm^2})}$. Again this means that the laser beam above would have a 0.1 nm noise limitation for a bandwidth of 21 Hz. The advantage of this method is that, in principle, one can linearize two axes with only one detector.
3. A knife-edge blocking part of a light beam incident on a photodiode can be used to measure the position of the piezo. This technique, commonly used in optical shear force detection [11.75, 165], has a sensitivity of better than 0.1 nm.
4. The capacitive detection [11.166, 167] of the cantilever deflection can be applied to the measurement of the piezo elongation. Equations (11.64) to (11.79) apply to the problem. This technique is used in some commercial instruments. The difficulties lie in the avoidance of fringe effects at the borders of the two plates. While conceptually simple, one needs the latest technology in surface preparation to get a decent linearity. The electronic circuits used for the readout are often proprietary.
5. Linear Variable Differential Transformers (LVDT) are a convenient means to measure positions down to 1 nm. They can be used together with a solid state joint setup, as often used for large scan range stages. Unlike the capacitive detection there are few difficulties in implementation. The sensors and the detection circuits LVDTs are available commercially.
6. A popular measurement technique is the use of strain gauges. They are especially sensitive when mounted on a solid state joint where the curvature is maximal. The resolution depends mainly on the induced curvature. A precision of 1 nm is attainable. The signals are low – a Wheatstone bridge is needed for the readout.

Active Linearization. Active linearization is done with feedback systems. Sensors need to be monotonic. Hence all the systems described above, with the exception of the interferometers are suitable. The most common solutions include the strain gauge approach, the capacitance measurement or the LVDT, which are all electronic solutions. Optical detection systems have the disadvantage that the intensity enters into the calibration.

Alternative Scanning Systems

The first STMs were based on piezo tripods [11.1]. The piezo tripod (Fig. 11.35) is an intuitive way to generate the three dimensional movement of a tip attached to its center. However, to get a suitable stability and scanning range, the tripod needs to be fairly large (about 50 mm). Some instruments use piezo stacks instead of monolithic piezoactuators. They are arranged in the tripod arrangement. Piezo stacks are thin layers of piezoactive materials glued together to form a device with up to 200 μm of actuation range. Preloading with a suitable metal casing reduces the nonlinearity.

If one tries to construct a homebuilt scanning system, the use of linearized scanning tables is recommended. They are built around solid state joints and actuated by piezo stacks. The joints guarantee that the movement is parallel with little deviation from the predefined scanning plane. Due to the construction it is easy to add measurement devices such as capacitive sensors, LVDTs

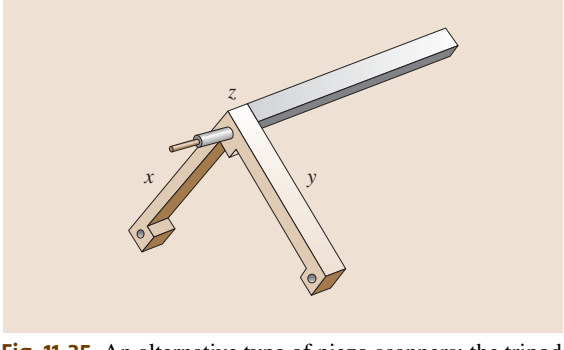


Fig. 11.35 An alternative type of piezo scanners: the tripod

or strain gauges which are essential for a closed loop linearization. Two-dimensional tables can be bought from several manufacturers. They have a linearity of better than 0.1% and a noise level of 10^{-4} to 10^{-5} of the maximal scanning range.

Control Systems

Basics. The electronics and software play an important role in the optimal performance of an SPM. Control electronics and software are supplied with commercial SPMs. Control electronic systems can use either analog or digital feedback. While digital feedback offers greater flexibility and the ease of configuration, analog feedback circuits might be better suited for ultralow noise operation. We will describe here the basic setups for AFMs.

Figure 11.36 shows a block schematic of a typical AFM feedback loop. The signal from the force transducer is fed into the feedback loop consisting mainly of a subtraction stage to get an error signal and an integrator. The gain of the integrator (high gain corresponds to short integration times) is set as high as possible without generating more than 1% overshoot. High gain minimizes the error margin of the current and forces the tip to follow the contours of constant density of states as well as possible. This operating mode is known as Constant Force Mode. A high voltage amplifier amplifies the outputs of the integrator. As AFMs using piezotubes usually require ±150 V at the output, the output of the integrator needs to be amplified by a high voltage amplifier.

In order to scan the sample, additional voltages at high tension are required to drive the piezo. For example, with a tube scanner, four scanning voltages are required, namely $+V_x$, $-V_x$, $+V_y$ and $-V_y$. The x- and y-scanning voltages are generated in a scan generator (analog or computer controlled). Both voltages are input to the two respective power amplifiers. Two inverting amplifiers generate the input voltages for the other two power amplifiers. The topography of the sample surface is determined by recording the input-voltage to the high voltage amplifier for the z-channel as a function of x and y (Constant Force Mode).

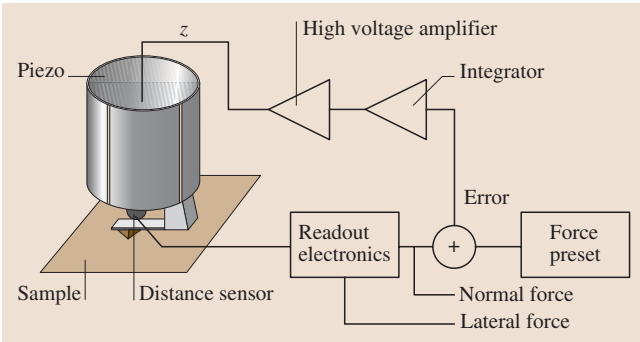


Fig. 11.36 Block schematic of the feedback control loop of an AFM

Another operating mode is the Variable Force Mode. The gain in the feedback loop is lowered and the scanning speed increased such that the force on the cantilever is no longer constant. Here the force is recorded as a function of x and y.

Force Spectroscopy. Four modes of spectroscopic imaging are in common use with force microscopes: measuring lateral forces, $\partial F/\partial z$, $\partial F/\partial x$ spatially resolved, and measuring force versus distance curves. Lateral forces can be measured by detecting the deflection of a cantilever in a direction orthogonal to the normal direction. The optical lever deflection method most easily does this. Lateral force measurements give indications of adhesion forces between the tip and the sample.

$\partial F/\partial z$ measurements probe the local elasticity of the sample surface. In many cases the measured quantity originates from a volume of a few cubic nanometers. The $\partial F/\partial z$ or local stiffness signal is proportional to Young's modulus, as far as one can define this quantity. Local stiffness is measured by vibrating the cantilever by a small amount in the z-direction. The expected signal for very stiff samples is zero: for very soft samples one gets, independent of the stiffness, also a constant signal. This signal is again zero for the optical lever deflection and equal to the driving amplitude for interferometeric measurements. The best sensitivity is obtained when the compliance of the cantilever matches the stiffness of the sample.

A third spectroscopic quantity is the lateral stiffness. It is measured by applying a small modulation in the x-direction on the cantilever. The signal is again optimal when the lateral compliance of the cantilever matches the lateral stiffness of the sample. The lateral stiffness is, in turn, related to the shear modulus of the sample.

Detailed information on the interaction of the tip and the sample can be gained by measuring force versus distance curves. It is necessary to have cantilevers with high enough compliance to avoid instabilities due to the attractive forces on the sample.

Using the Control Electronics as a Two-Dimensional Measurement Tool. Usually the control electronics of an AFM is used to control the x- and y-piezo signals while several data acquisition channels record the po-

sition dependent signals. The control electronics can be used in another way: it can be viewed as a two-dimensional function generator. What is normally the x- and y-signal can be used to control two independent variables of an experiment. The control logic of the AFM then ensures that the available parameter space is systematically probed at equally spaced points. An example is friction force curves measured along a line across a step on graphite.

Figure 11.37 shows the connections. The z-piezo is connected as usual, like the x-piezo. However the y-output is used to command the desired input parameter. The offset of the y-channel determines the position of the tip on the sample surface, together with the x-channel.

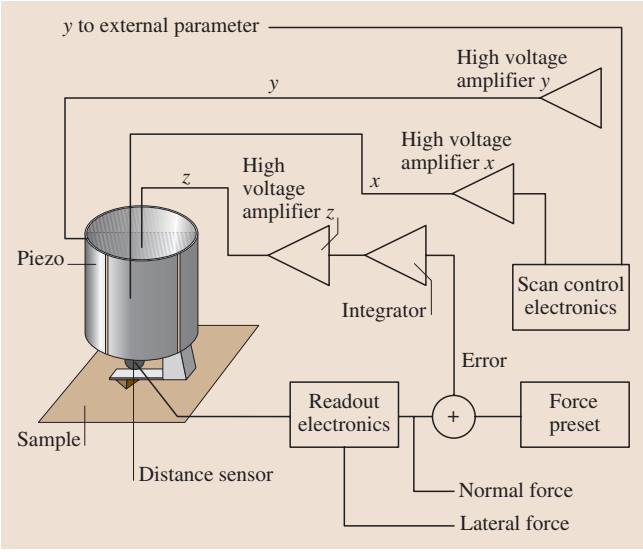


Fig. 11.37 Wiring of an AFM to measure friction force curves along a line

Some Imaging Processing Methods

The visualization and interpretation of images from AFMs is intimately connected to the processing of these images. An ideal AFM is a noise-free device that images a sample with perfect tips of known shape and has perfect linear scanning piezos. In reality, AFMs are not that ideal. The scanning device in AFMs is affected by distortions. The distortions are both linear and nonlinear. Linear distortions mainly result from imperfections in the machining of the piezotranslators causing crosstalk from the Z-piezo to the X- and Y-piezos, and vice versa. Among the linear distortions, there are two kinds which are very important. First, scanning piezos invariably have different sensitivities along the different scan axes due to the variation of the piezo material and uneven sizes of the electrode areas. Second, the same reasons might cause the scanning axes not to be orthogonal. Furthermore, the plane in which the piezoscanner moves for constant height z is hardly ever coincident with the sample plane. Hence, a linear ramp is added to the sample data. This ramp is especially bothersome when the height z is displayed as an intensity map.

The nonlinear distortions are harder to deal with. They can affect AFM data for a variety of reasons. First, piezoelectric ceramics do have a hysteresis loop, much like ferromagnetic materials. The deviations of piezoceramic materials from linearity increase with increasing amplitude of the driving voltage. The mechanical position for one voltage depends on the previously applied voltages to the piezo. Hence, to get the best positional accuracy, one should always approach a point on the sample from the same direction. Another type of nonlinear distortion of the images occurs when the scan frequency approaches the upper frequency limit of the X- and Y-drive amplifiers or the upper frequency limit of the feedback loop (z-component). This distortion, due to feedback loop, can only be minimized by reducing the scan frequency. On the other hand, there is a simple way to reduce distortions due to the X- and Y-piezo drive amplifiers. To keep the system as simple as possible, one normally uses a triangular waveform for driving the scanning piezos. However, triangular waves contain frequency components as multiples of the scan frequency. If the cutoff frequency of the X- and Y-drive electronics or of the feedback loop is too close to the scanning frequency (two or three times the scanning frequency), the triangular drive voltage is rounded off at the turning points. This rounding error causes, first, a distortion of the scan linearity and, second, through phase lags, the projection of part of the backward scan onto the forward scan. This type of distortion can be minimized by carefully selecting the scanning frequency and by using driving voltages for the X- and Y-piezos with waveforms like trapezoidal waves, which are closer to a sine wave. The values measured for X, Y, or Z piezos are affected by noise. The origin of this noise can be either electronic, some disturbances, or a property of the sample surface due to adsorbates. In addition to this incoherent noise, interference with main and other equipment nearby might be present. Depending on the type of noise, one can filter it in the real space or in Fourier space. The most important part of image processing is to visualize the measured data. Typical AFM data sets can consist of many thousands to over a million points per plane. There may be more than one image plane present.

The AFM data represents a topography in various data spaces.

Most commercial data acquisition systems use implicitly some kind of data processing. Since the original data is commonly subject to slopes on the surface, most programs use some kind of slope correction. The least disturbing way is to substrate a plane $z(x, y) = Ax + By + C$ from the data. The coefficients are determined by fitting $z(x, y)$ to the data. Another operation is to subtract a second order function such as $z(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey + F$. Again, the parameters are determined with a fit. This function is appropriate for almost planar data, where the nonlinearity of the piezos caused such a distortion.

In the image processing software from Digital Instruments, up to three operations are performed on the raw data. First, a zero-order flatten is applied. The flatten operation is used to eliminate image bow in the slow scan direction (caused by a physical bow in the instrument itself), slope in the slow scan direction, bands in the image (caused by differences in the scan height from one scan line to the next). The flattening operation takes each scan line and subtracts the average value of the height along each scan line from each point in that scan line. This brings each scan line to the same height. Next, a first-order plane-fit is applied in the fast scan direction. The plane-fit operation is used to eliminate bow and slope in the fast scan direction. The plane-fit operation calculated a best-fit plane for the image and subtracts it from the image. This plane has a constant non-zero slope in the fast scan direction. In some cases, a higher-order polynomial "plane" may be required. Depending upon the quality of the raw data, the flattening operation and/or the planefit operation may not be required at all.

References

11.1 G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel: Surface studies by scanning tunneling microscopy, Phys. Rev. Lett. **49** (1982) 57–61

11.2 G. Binnig, C. F. Quate, Ch. Gerber: Atomic force microscope, Phys. Rev. Lett. **56** (1986) 930–933

11.3 G. Binnig, Ch. Gerber, E. Stoll, T. R. Albrecht, C. F. Quate: Atomic resolution with atomic force microscope, Europhys. Lett. **3** (1987) 1281–1286

11.4 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

11.5 C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59** (1987) 1942–1945

11.6 R. Erlandsson, G. M. McClelland, C. M. Mate, S. Chiang: Atomic force microscopy using optical interferometry, J. Vac. Sci. Technol. A **6** (1988) 266–270

11.7 O. Marti, J. Colchero, J. Mlynek: Combined scanning force and friction microscopy of mica, Nanotechnol. **1** (1990) 141–144

11.8 G. Meyer, N. M. Amer: Simultaneous measurement of lateral and normal forces with an optical-beam-deflection atomic force microscope, Appl. Phys. Lett. **57** (1990) 2089–2091

11.9 B. Bhushan, J. Ruan: Atomic-scale friction measurements using friction force microscopy: Part II – Application to magnetic media, ASME J. Tribol. **116** (1994) 389–396

11.10 B. Bhushan, V. N. Koinkar, J. Ruan: Microtribology of magnetic media, Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. **208** (1994) 17–29

11.11 B. Bhushan, J. N. Israelachvili, U. Landman: Nanotribology: Friction, wear, and lubrication at the atomic scale, Nature **374** (1995) 607–616

11.12 S. Fujisawa, M. Ohta, T. Konishi, Y. Sugawara, S. Morita: Difference between the forces measured by an optical lever deflection and by an optical interferometer in an atomic force microscope, Rev. Sci. Instrum. **65** (1994) 644–647

11.13 S. Fujisawa, E. Kishi, Y. Sugawara, S. Morita: Fluctuation in 2-dimensional stick-slip phenomenon observed with 2-dimensional frictional force microscope, Jpn. J. Appl. Phys. **33** (1994) 3752–3755

11.14 S. Grafstrom, J. Ackermann, T. Hagen, R. Neumann, O. Probst: Analysis of lateral force effects on the topography in scanning force microscopy, J. Vac. Sci. Technol. B **12** (1994) 1559–1564

11.15 R. M. Overney, H. Takano, M. Fujihira, W. Paulus, H. Ringsdorf: Anisotropy in friction and molecular stick-slip motion, Phys. Rev. Lett. **72** (1994) 3546–3549

11.16 R. J. Warmack, X. Y. Zheng, T. Thundat, D. P. Allison: Friction effects in the deflection of atomic force microscope cantilevers, Rev. Sci. Instrum. **65** (1994) 394–399

11.17 N. A. Burnham, D. D. Domiguez, R. L. Mowery, R. J. Colton: Probing the surface forces of monolayer films with an atomic force microscope, Phys. Rev. Lett. **64** (1990) 1931–1934

11.18 N. A. Burham, R. J. Colton, H. M. Pollock: Interpretation issues in force microscopy, J. Vac. Sci. Technol. A **9** (1991) 2548–2556

11.19 C. D. Frisbie, L. F. Rozsnyai, A. Noy, M. S. Wrighton, C. M. Lieber: Functional group imaging by chemical force microscopy, Science **265** (1994) 2071–2074

11.20 V. N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A **14** (1996) 2378–2391

11.21 V. Scherer, B. Bhushan, U. Rabe, W. Arnold: Local elasticity and lubrication measurements using atomic force and friction force microscopy at ultrasonic frequencies, IEEE Trans. Mag. **33** (1997) 4077–4079

11.22 V. Scherer, W. Arnold, B. Bhushan: Lateral force microscopy using acoustic friction force microscopy, Surf. Interface Anal. **27** (1999) 578–587

11.23 B. Bhushan, S. Sundararajan: Micro/Nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46** (1998) 3793–3804

11.24 U. Krotil, T. Stifter, H. Waschipky, K. Weishaupt, S. Hild, O. Marti: Pulse force mode: A new method for the investigation of surface properties, Surf. Interface Anal. **27** (1999) 336–340

11.25 B. Bhushan, C. Dandavate: Thin-film friction and adhesion studies using atomic force microscopy, J. Appl. Phys. **87** (2000) 1201–1210

11.26 B. Bhushan: *Micro/Nanotribology and its Applications* (Kluwer, Dordrecht 1997)

11.27 B. Bhushan: *Principles and Applications of Tribology* (Wiley, New York 1999)

11.28 B. Bhushan: *Modern Tribology Handbook Vol. 1: Principles of Tribology* (CRC, Boca Raton 2001)

11.29 B. Bhushan: *Introduction to Tribology* (Wiley, New York 2002)

11.30 M. Reinstaedtler, U. Rabe, V. Scherer, U. Hartmann, A. Goldade, B. Bhushan, W. Arnold: On the nanoscale measurement of friction using atomic force microscope cantilever torsional resonances, Appl. Phys. Lett. **82** (2003) 2604–2606

11.31 N. A. Burnham, R. J. Colton: Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope, J. Vac. Sci. Technol. A **7** (1989) 2906–2913

11.32 P. Maivald, H. J. Butt, S. A. C. Gould, C. B. Prater, B. Drake, J. A. Gurley, V. B. Elings, P. K. Hansma: Using force modulation to image surface elasticities with the atomic force microscope, Nanotechnol. **2** (1991) 103–106

11.33 B. Bhushan, A. V. Kulkarni, W. Bonin, J. T. Wyrobek: Nano/Picoindentation measurements using capacitive transducer in atomic force microscopy, Philos. Mag. A **74** (1996) 1117–1128

11.34 B. Bhushan, V. N. Koinkar: Nanoindentation hardness measurements using atomic force microscopy, Appl. Phys. Lett. **75** (1994) 5741–5746

11.35 D. DeVecchio, B. Bhushan: Localized surface elasticity measurements using an atomic force microscope, Rev. Sci. Instrum. **68** (1997) 4498–4505

11.36 S. Amelio, A. V. Goldade, U. Rabe, V. Scherer, B. Bhushan, W. Arnold: Measurements of mechanical properties of ultra-thin diamond-like carbon coatings using atomic force acoustic microscopy, Thin Solid Films **392** (2001) 75–84

11.37 D. M. Eigler, E. K. Schweizer: Positioning single atoms with a scanning tunnelling microscope, Nature **344** (1990) 524–528

11.38 A. L. Weisenhorn, J. E. MacDougall, J. A. C. Gould, S. D. Cox, W. S. Wise, J. Massie, P. Maivald, V. B. Elings, G. D. Stucky, P. K. Hansma: Imaging and manipulating of molecules on a zeolite surface with an atomic force microscope, Science **247** (1990) 1330–1333

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

11.40 O. M. Leung, M. C. Goh: Orientation ordering of polymers by atomic force microscope tip-surface interactions, Science **225** (1992) 64–66

11.41 D. W. Abraham, H. J. Mamin, E. Ganz, J. Clark: Surface modification with the scanning tunneling microscope, IBM J. Res. Dev. **30** (1986) 492–499

11.42 R. M. Silver, E. E. Ehrichs, A. L. de Lozanne: Direct writing of submicron metallic features with a scanning tunnelling microscope, Appl. Phys. Lett. **51** (1987) 247–249

11.43 A. Kobayashi, F. Grey, R. S. Williams, M. Ano: Formation of nanometer-scale grooves in silicon with a scanning tunneling microscope, Science **259** (1993) 1724–1726

11.44 B. Parkinson: Layer-by-layer nanometer scale etching of two-dimensional substrates using the scanning tunneling microscopy, J. Am. Chem. Soc. **112** (1990) 7498–7502

11.45 A. Majumdar, P. I. Oden, J. P. Carrejo, L. A. Nagahara, J. J. Graham, J. Alexander: Nanometer-scale lithography using the atomic force microscope, Appl. Phys. Lett. **61** (1992) 2293–2295

11.46 B. Bhushan: Micro/Nanotribology and its applications to magnetic storage devices and MEMS, Tribol. Int. **28** (1995) 85–96

11.47 L. Tsau, D. Wang, K. L. Wang: Nanometer scale patterning of silicon(100) surface by an atomic force microscope operating in air, Appl. Phys. Lett. **64** (1994) 2133–2135

11.48 E. Delawski, B. A. Parkinson: Layer-by-layer etching of two-dimensional metal chalcogenides with the atomic force microscope, J. Am. Chem. Soc. **114** (1992) 1661–1667

11.49 B. Bhushan, G. S. Blackman: Atomic force microscopy of magnetic rigid disks and sliders and its applications to tribology, ASME J. Tribol. **113** (1991) 452–458

11.50 O. Marti, B. Drake, P. K. Hansma: Atomic force microscopy of liquid-covered surfaces: atomic resolution images, Appl. Phys. Lett. **51** (1987) 484–486

11.51 B. Drake, C. B. Prater, A. L. Weisenhorn, S. A. C. Gould, T. R. Albrecht, C. F. Quate, D. S. Cannell, H. G. Hansma, P. K. Hansma: Imaging crystals, polymers and processes in water with the atomic force microscope, Science **243** (1989) 1586–1589

11.52 M. Binggeli, R. Christoph, H. E. Hintermann, J. Colchero, O. Marti: Friction force measurements on potential controlled graphite in an electrolytic environment, Nanotechnol. **4** (1993) 59–63

11.53 G. Meyer, N. M. Amer: Novel optical approach to atomic force microscopy, Appl. Phys. Lett. **53** (1988) 1045–1047

11.54 J. H. Coombs, J. B. Pethica: Properties of vacuum tunneling currents: Anomalous barrier heights, IBM J. Res. Dev. **30** (1986) 455–459

11.55 M. D. Kirk, T. Albrecht, C. F. Quate: Low-temperature atomic force microscopy, Rev. Sci. Instrum. **59** (1988) 833–835

11.56 F. J. Giessibl, Ch. Gerber, G. Binnig: A low-temperature atomic force/scanning tunneling microscope for ultrahigh vacuum, J. Vac. Sci. Technol. B **9** (1991) 984–988

11.57 T. R. Albrecht, P. Grutter, D. Rugar, D. P. E. Smith: Low temperature force microscope with all-fiber interferometer, Ultramicroscopy **42–44** (1992) 1638–1646

11.58 H. J. Hug, A. Moser, Th. Jung, O. Fritz, A. Wadas, I. Parashikor, H. J. Güntherodt: Low temperature magnetic force microscopy, Rev. Sci. Instrum. **64** (1993) 2920–2925

11.59 C. Basire, D. A. Ivanov: Evolution of the lamellar structure during crystallization of a semicrystalline-amorphous polymer blend: Time-resolved hot-stage SPM study, Phys. Rev. Lett. **85** (2000) 5587–5590

11.60 H. Liu, B. Bhushan: Investigation of nanotribological properties of self-assembled monolayers with alkyl and biphenyl spacer chains, Ultramicroscopy **91** (2002) 185–202

11.61 J. Foster, J. Frommer: Imaging of liquid crystal using a tunneling microscope, Nature **333** (1988) 542–547

11.62 D. Smith, H. Horber, C. Gerber, G. Binnig: Smectic liquid crystal monolayers on graphite observed by scanning tunneling microscopy, Science **245** (1989) 43–45

11.63 D. Smith, J. Horber, G. Binnig, H. Nejoh: Structure, registry and imaging mechanism of alkylcyanobiphenyl molecules by tunnelling microscopy, Nature **344** (1990) 641–644

11.64 Y. Andoh, S. Oguchi, R. Kaneko, T. Miyamoto: Evaluation of very thin lubricant films, J. Phys. D **25** (1992) A71–A75

11.65 Y. Martin, C. C. Williams, H. K. Wickramasinghe: Atomic force microscope-force mapping and profiling on a sub 100-A scale, J. Appl. Phys. **61** (1987) 4723–4729

11.66 J. E. Stern, B. D. Terris, H. J. Mamin, D. Rugar: Deposition and imaging of localized charge on insulator surfaces using a force microscope, Appl. Phys. Lett. **53** (1988) 2717–2719

11.67 K. Yamanaka, H. Ogisco, O. Kolosov: Ultrasonic force microscopy for nanometer resolution subsurface imaging, Appl. Phys. Lett. **64** (1994) 178–180

11.68 K. Yamanaka, E. Tomita: Lateral force modulation atomic force microscope for selective imaging of friction forces, Jpn. J. Appl. Phys. **34** (1995) 2879–2882

11.69 U. Rabe, K. Janser, W. Arnold: Vibrations of free and surface-coupled atomic force microscope: Theory and experiment, Rev. Sci. Instrum. **67** (1996) 3281–3293

11.70 Y. Martin, H. K. Wickramasinghe: Magnetic imaging by force microscopy with 1000 Å resolution, Appl. Phys. Lett. **50** (1987) 1455–1457

11.71 D. Rugar, H. J. Mamin, P. Guethner, S. E. Lambert, J. E. Stern, I. McFadyen, T. Yogi: Magnetic force microscopy – General principles and application to longitudinal recording media, J. Appl. Phys. **63** (1990) 1169–1183

11.72 C. Schoenenberger, S. F. Alvarado: Understanding magnetic force microscopy, Z. Phys. B **80** (1990) 373–383

11.73 U. Hartmann: Magnetic force microscopy, Annu. Rev. Mater. Sci. **29** (1999) 53–87

11.74 D. W. Pohl, W. Denk, M. Lanz: Optical stethoscopy-image recording with resolution lambda/20, Appl. Phys. Lett. **44** (1984) 651–653

11.75 E. Betzig, J. K. Troutman, T. D. Harris, J. S. Weiner, R. L. Kostelak: Breaking the diffraction barrier – optical microscopy on a nanometric scale, Science **251** (1991) 1468–1470

11.76 E. Betzig, P. L. Finn, J. S. Weiner: Combined shear force and near-field scanning optical microscopy, Appl. Phys. Lett. **60** (1992) 2484

11.77 P. F. Barbara, D. M. Adams, D. B. O'Connor: Characterization of organic thin film materials with near-field scanning optical microscopy (NSOM), Annu. Rev. Mater. Sci. **29** (1999) 433–469

11.78 C. C. Williams, H. K. Wickramasinghe: Scanning thermal profiler, Appl. Phys. Lett. **49** (1986) 1587–1589

11.79 C. C. Williams, H. K. Wickramasinghe: Microscopy of chemical-potential variations on an atomic scale, Nature **344** (1990) 317–319

11.80 A. Majumdar: Scanning thermal microscopy, Annu. Rev. Mater. Sci. **29** (1999) 505–585

11.81 O. E. Husser, D. H. Craston, A. J. Bard: Scanning electrochemical microscopy – high resolution deposition and etching of materials, J. Electrochem. Soc. **136** (1989) 3222–3229

11.82 Y. Martin, D. W. Abraham, H. K. Wickramasinghe: High-resolution capacitance measurement and potentiometry by force microscopy, Appl. Phys. Lett. **52** (1988) 1103–1105

11.83 M. Nonnenmacher, M. P. O'Boyle, H. K. Wickramasinghe: Kelvin probe force microscopy, Appl. Phys. Lett. **58** (1991) 2921–2923

11.84 J. M. R. Weaver, D. W. Abraham: High resolution atomic force microscopy potentiometry, J. Vac. Sci. Technol. B **9** (1991) 1559–1561

11.85 D. DeVecchio, B. Bhushan: Use of a nanoscale Kelvin probe for detecting wear precursors, Rev. Sci. Instrum. **69** (1998) 3618–3624

11.86 B. Bhushan, A. V. Goldade: Measurements and analysis of surface potential change during wear of single-crystal silicon (100) at ultralow loads using Kelvin probe microscopy, Appl. Surf. Sci. **157** (2000) 373–381

11.87 P. K. Hansma, B. Drake, O. Marti, S. A. C. Gould, C. B. Prater: The scanning ion-conductance microscope, Science **243** (1989) 641–643

11.88 C. B. Prater, P. K. Hansma, M. Tortonese, C. F. Quate: Improved scanning ion-conductance microscope using microfabricated probes, Rev. Sci. Instrum. **62** (1991) 2634–2638

11.89 J. Matey, J. Blanc: Scanning capacitance microscopy, J. Appl. Phys. **57** (1985) 1437–1444

11.90 C. C. Williams: Two-dimensional dopant profiling by scanning capacitance microscopy, Annu. Rev. Mater. Sci. **29** (1999) 471–504

11.91 D. T. Lee, J. P. Pelz, B. Bhushan: Instrumentation for direct, low frequency scanning capacitance microscopy, and analysis of position dependent stray capacitance, Rev. Sci. Instrum. **73** (2002) 3523–3533

11.92 I. Giaever: Energy gap in superconductors measured by electron tunneling, Phys. Rev. Lett. **5** (1960) 147–148

11.93 P. K. Hansma, J. Tersoff: Scanning tunneling microscopy, J. Appl. Phys. **61** (1987) R1–R23

11.94 D. Sarid, V. Elings: Review of scanning force microscopy, J. Vac. Sci. Technol. B **9** (1991) 431–437

11.95 U. Durig, O. Zuger, A. Stalder: Interaction force detection in scanning probe microscopy: Methods and applications, J. Appl. Phys. **72** (1992) 1778–1797

11.96 J. Frommer: Scanning tunneling microscopy and atomic force microscopy in organic chemistry, Angew. Chem. Int. Ed. Engl. **31** (1992) 1298–1328

11.97 H. J. Güntherodt, R. Wiesendanger (Eds.): *Scanning Tunneling Microscopy I: General Principles and Applications to Clean and Adsorbate-Covered Surfaces* (Springer, Berlin, Heidelberg 1992)

11.98 R. Wiesendanger, H. J. Güntherodt (Eds.): *Scanning Tunneling Microscopy, II: Further Applications and Related Scanning Techniques* (Springer, Berlin, Heidelberg 1992)

11.99 D. A. Bonnell (Ed.): *Scanning Tunneling Microscopy and Spectroscopy – Theory, Techniques, and Applications* (VCH, New York 1993)

11.100 O. Marti, M. Amrein (Eds.): *STM and SFM in Biology* (Academic, San Diego 1993)

11.101 J. A. Stroscio, W. J. Kaiser (Eds.): *Scanning Tunneling Microscopy* (Academic, Boston 1993)

11.102 H. J. Güntherodt, D. Anselmetti, E. Meyer (Eds.): *Forces in Scanning Probe Methods* (Kluwer, Dordrecht 1995)

11.103 G. Binnig, H. Rohrer: Scanning tunnelling microscopy, Surf. Sci. **126** (1983) 236–244

11.104 B. Bhushan, J. Ruan, B. K. Gupta: A scanning tunnelling microscopy study of fullerene films, J. Phys. D **26** (1993) 1319–1322

11.105 R. L. Nicolaides, W. E. Yong, W. F. Packard, H. A. Zhou: Scanning tunneling microscope tip structures, J. Vac. Sci. Technol. A **6** (1988) 445–447

11.106 J. P. Ibe, P. P. Bey, S. L. Brandon, R. A. Brizzolara, N. A. Burnham, D. P. DiLella, K. P. Lee, C. R. K. Marrian, R. J. Colton: On the electrochemical etching of tips for scanning tunneling microscopy, J. Vac. Sci. Technol. A **8** (1990) 3570–3575

11.107 R. Kaneko, S. Oguchi: Ion-implanted diamond tip for a scanning tunneling microscope, Jpn. J. Appl. Phys. **28** (1990) 1854–1855

11.108 F. J. Giessibl: Atomic resolution of the silicon(111)–(7×7) surface by atomic force microscopy, Science **267** (1995) 68–71

11.109 B. Anczykowski, D. Krueger, K. L. Babcock, H. Fuchs: Basic properties of dynamic force spectroscopy with the scanning force microscope in experiment and simulation, Ultramicroscopy **66** (1996) 251–259

11.110 T. R. Albrecht and C. F. Quate: Atomic resolution imaging of a nonconductor by atomic force microscopy, J. Appl. Phys. **62** (1987) 2599–2602

11.111 S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P. K. Hansma: An atomic-resolution atomic-force microscope implemented using an optical lever, J. Appl. Phys. **65** (1989) 164–167

11.112 G. Meyer, N. M. Amer: Optical-beam-deflection atomic force microscopy: The NaCl(001) surface, Appl. Phys. Lett. **56** (1990) 2100–2101

11.113 A. L. Weisenhorn, M. Egger, F. Ohnesorge, S. A. C. Gould, S. P. Heyn, H. G. Hansma, R. L. Sinsheimer, H. E. Gaub, P. K. Hansma: Molecular resolution images of Langmuir–Blodgett films and DNA by atomic force microscopy, Langmuir **7** (1991) 8–12

11.114 J. Ruan, B. Bhushan: Atomic-scale and microscale friction of graphite and diamond using friction force microscopy, J. Appl. Phys. **76** (1994) 5022–5035

11.115 D. Rugar, P. K. Hansma: Atomic force microscopy, Phys. Today **43** (1990) 23–30

11.116 D. Sarid: *Scanning Force Microscopy* (Oxford Univ. Press, Oxford 1991)

11.117 G. Binnig: Force microscopy, Ultramicroscopy **42–44** (1992) 7–15

11.118 E. Meyer: Atomic force microscopy, Surf. Sci. **41** (1992) 3–49

11.119 H. K. Wickramasinghe: Progress in scanning probe microscopy, Acta Mater. **48** (2000) 347–358

11.120 A. J. den Boef: The influence of lateral forces in scanning force microscopy, Rev. Sci. Instrum. **62** (1991) 88–92

11.121 M. Radmacher, R. W. Tillman, M. Fritz, H. E. Gaub: From molecules to cells: Imaging soft samples with the atomic force microscope, Science **257** (1992) 1900–1905

11.122 F. Ohnesorge, G. Binnig: True atomic resolution by atomic force microscopy through repulsive and attractive forces, Science **260** (1993) 1451–1456

11.123 G. Neubauer, S. R. Coben, G. M. McClelland, D. Horne, C. M. Mate: Force microscopy with a bidirectional capacitance sensor, Rev. Sci. Instrum. **61** (1990) 2296–2308

11.124 T. Goddenhenrich, H. Lemke, U. Hartmann, C. Heiden: Force microscope with capacitive displacement detection, J. Vac. Sci. Technol. A **8** (1990) 383–387

11.125 U. Stahl, C. W. Yuan, A. L. Delozanne, M. Tortonese: Atomic force microscope using piezoresistive cantilevers and combined with a scanning electron microscope, Appl. Phys. Lett. **65** (1994) 2878–2880

11.126 R. Kassing, E. Oesterschulze: Sensors for scanning probe microscopy. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 35–54

11.127 C. M. Mate: Atomic-force-microscope study of polymer lubricants on silicon surfaces, Phys. Rev. Lett. **68** (1992) 3323–3326

11.128 S. P. Jarvis, A. Oral, T. P. Weihs, J. B. Pethica: A novel force microscope and point contact probe, Rev. Sci. Instrum. **64** (1993) 3515–3520

11.129 D. Rugar, H. J. Mamin, P. Guethner: Improved fiber-optical interferometer for atomic force microscopy, Appl. Phys. Lett. **55** (1989) 2588–2590

11.130 C. Schoenenberger, S. F. Alvarado: A differential interferometer for force microscopy, Rev. Sci. Instrum. **60** (1989) 3131–3135

11.131 D. Sarid, D. Iams, V. Weissenberger, L. S. Bell: Compact scanning-force microscope using laser diode, Opt. Lett. **13** (1988) 1057–1059

11.132 N. W. Ashcroft, N. D. Mermin: *Solid State Physics* (Holt Reinhart and Winston, New York 1976)

11.133 G. Binnig, D. P. E. Smith: Single-tube three-dimensional scanner for scanning tunneling microscopy, Rev. Sci. Instrum. **57** (1986) 1688

11.134 S. I. Park, C. F. Quate: Digital filtering of STM images, J. Appl. Phys. **62** (1987) 312

11.135 J. W. Cooley, J. W. Tukey: An algorithm for machine calculation of complex fourier series, Math. Computation **19** (1965) 297

11.136 J. Ruan, B. Bhushan: Atomic-scale friction measurements using friction force microscopy: Part I – General principles and new measurement techniques, ASME J. Tribol. **116** (1994) 378–388

11.137 T. R. Albrecht, S. Akamine, T. E. Carver, C. F. Quate: Microfabrication of cantilever styli for the atomic force microscope, J. Vac. Sci. Technol. A **8** (1990) 3386–3396

11.138 O. Marti, S. Gould, P. K. Hansma: Control electronics for atomic force microscopy, Rev. Sci. Instrum. **59** (1988) 836–839

11.139 O. Wolter, T. Bayer, J. Greschner: Micromachined silicon sensors for scanning force microscopy, J. Vac. Sci. Technol. B **9** (1991) 1353–1357

11.140 E. Meyer, R. Overney, R. Luthi, D. Brodbeck: Friction force microscopy of mixed Langmuir–Blodgett films, Thin Solid Films **220** (1992) 132–137

11.141 H. J. Dai, J. H. Hafner, A. G. Rinzler, D. T. Colbert, R. E. Smalley: Nanotubes as nanoprobes in scanning probe microscopy, Nature **384** (1996) 147–150

11.142 J. H. Hafner, C. L. Cheung, A. T. Woolley, C. M. Lieber: Structural and functional imaging with carbon nanotube AFM probes, Prog. Biophys. Mol. Biol. **77** (2001) 73–110

11.143 G. S. Blackman, C. M. Mate, M. R. Philpott: Interaction forces of a sharp tungsten tip with molecular films on silicon surface, Phys. Rev. Lett. **65** (1990) 2270–2273

11.144 S. J. O'Shea, M. E. Welland, T. Rayment: Atomic force microscope study of boundary layer lubrication, Appl. Phys. Lett. **61** (1992) 2240–2242

11.145 J. P. Cleveland, S. Manne, D. Bocek, P. K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64** (1993) 403–405

11.146 D. W. Pohl: Some design criteria in STM, IBM J. Res. Dev. **30** (1986) 417

11.147 W. T. Thomson, M. D. Dahleh: *Theory of Vibration with Applications*, 5th edn. (Prentice Hall, Upper Saddle River 1998)

11.148 J. Colchero: Reibungskraftmikroskopie. Ph.D. Thesis (University of Konstanz, Konstanz 1993)

11.149 G. M. McClelland, R. Erlandsson, S. Chiang: Atomic force microscopy: General principles and a new implementation. In: *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 6B, ed. by D. O. Thompson, D. E. Chimenti (Plenum, New York 1987) pp. 1307–1314

11.150 Y. R. Shen: *The Principles of Nonlinear Optics* (Wiley, New York 1984)

11.151 T. Baumeister, S. L. Marks: *Standard Handbook for Mechanical Engineers*, 7th edn. (McGraw-Hill, New York 1967)

11.152 J. Colchero, O. Marti, H. Bielefeldt, J. Mlynek: Scanning force and friction microscopy, Phys. Stat. Sol. **131** (1991) 73–75

11.153 R. Young, J. Ward, F. Scire: Observation of metal-vacuum-metal tunneling, field emission, and the transition region, Phys. Rev. Lett. **27** (1971) 922

11.154 R. Young, J. Ward, F. Scire: The topographiner: An instrument for measuring surface microtopography, Rev. Sci. Instrum. **43** (1972) 999

11.155 C. Gerber, O. Marti: Magnetostrictive positioner, IBM Tech. Disclosure Bull. **27** (1985) 6373

11.156 R. Garcìa Cantù, M. A. Huerta Garnica: Long-scan imaging by STM, J. Vac. Sci. Technol. A **8** (1990) 354
11.157 C. J. Chen: In situ testing and calibration of tube piezoelectric scanners, Ultramicroscopy **42–44** (1992) 1653–1658
11.158 R. G. Carr: J. Microscopy **152** (1988) 379
11.159 C. J. Chen: Electromechanical deflections of piezoelectric tubes with quartered electrodes, Appl. Phys. Lett. **60** (1992) 132
11.160 N. Libioulle, A. Ronda, M. Taborelli, J. M. Gilles: Deformations and nonlinearity in scanning tunneling microscope images, J. Vac. Sci. Technol. B **9** (1991) 655–658
11.161 E. P. Stoll: Restoration of STM images distorted by time-dependent piezo driver aftereffects, Ultramicroscopy **42–44** (1991) 1585–1589
11.162 R. Durselen, U. Grunewald, W. Preuss: Calibration and applications of a high precision piezo scanner for nanometrology, Scanning **17** (1995) 91–96
11.163 J. Fu: In situ testing and calibrating of Z-piezo of an atomic force microscope, Rev. Sci. Instrum. **66** (1995) 3785–3788
11.164 R. C. Barrett, C. F. Quate: Optical scan-correction system applied to atomic force microscopy, Rev. Sci. Instrum. **62** (1991) 1393
11.165 R. Toledo-Crow, P. C. Yang, Y. Chen, M. Vaez-Iravani: Near-field differential scanning optical microscope with atomic force regulation, Appl. Phys. Lett. **60** (1992) 2957–2959
11.166 J. E. Griffith, G. L. Miller, C. A. Green: A scanning tunneling microscope with a capacitance-based position monitor, J. Vac. Sci. Technol. B **8** (1990) 2023–2027
11.167 A. E. Holman, C. D. Laman, P. M. L. O. Scholte, W. C. Heerens, F. Tuinstra: A calibrated scanning tunneling microscope equipped with capacitive sensors, Rev. Sci. Instrum. **67** (1996) 2274–2280

12. Probes in Scanning Microscopies

Scanning probe microscopy (SPM) provides nanometer-scale mapping of numerous sample properties in essentially any environment. This unique combination of high resolution and broad applicability has lead to the application of SPM to many areas of science and technology, especially those interested in the structure and properties of materials at the nanometer scale. SPM images are generated through measurements of a tip-sample interaction. A well-characterized tip is the key element to data interpretation and is typically the limiting factor.

Commercially available atomic force microscopy (AFM) tips, integrated with force sensing cantilevers, are microfabricated from silicon and silicon nitride by lithographic and anisotropic etching techniques. The performance of these tips can be characterized by imaging nanometer-scale standards of known dimension, and the resolution is found to roughly correspond to the tip radius of curvature, the tip aspect ratio, and the sample height. Although silicon and silicon nitride tips have a somewhat large radius of curvature, low aspect ratio, and limited lifetime due to wear, the widespread use of AFM today is due in large part to the broad availability of these tips. In some special cases, small asperities on the tip can provide resolution much higher than the tip radius of curvature for low-Z samples such as crystal surfaces and ordered protein arrays.

Several strategies have been developed to improve AFM tip performance. Oxide sharpening improves tip sharpness and enhances tip asperities. For high-aspect-ratio samples such as integrated circuits, silicon AFM tips can be modified by focused ion beam (FIB) milling. FIB tips reach three-degree cone angles over lengths of several microns and can be fabricated at arbitrary angles. Other high resolution and high-aspect-ratio tips are produced by electron beam deposition (EBD) in which a carbon spike is deposited onto the tip apex from the background gases in an electron microscope. Finally, carbon nanotubes have been employed as AFM tips. Their nanometer-scale diameter, long length, high stiffness, and elastic buckling properties make carbon nanotubes possibly the ultimate tip material for AFM. Nanotubes can be manually attached to silicon or silicon nitride AFM tips or "grown" onto tips by chemical vapor deposition (CVD), which should soon make them widely available. In scanning tunneling microscopy (STM), the electron tunneling signal decays exponentially with tip-sample separation, so that in principle only the last few atoms contribute to the signal. STM tips are, therefore, not as sensitive to the nanoscale tip geometry and can be made by simple mechanical cutting or electrochemical etching of metal wires. In choosing tip materials, one prefers hard, stiff metals that will not oxidize or corrode in the imaging environment.

12.1 **Atomic Force Microscopy** 372
12.1.1 Principles of Operation 372
12.1.2 Standard Probe Tips 373
12.1.3 Probe Tip Performance 374
12.1.4 Oxide-Sharpened Tips 375
12.1.5 FIB tips 376
12.1.6 EBD tips 376
12.1.7 Carbon Nanotube Tips 376

12.2 **Scanning Tunneling Microscopy** 382
12.2.1 Mechanically Cut STM Tips 382
12.2.2 Electrochemically Etched STM Tips 383

References 383

In scanning probe microscopy (SPM), an image is created by raster scanning a sharp probe tip over a sample and measuring some highly localized tip-sample interaction as a function of position. SPMs are based on several interactions, the major types including scanning tunneling microscopy (STM), which measures an electronic tunneling current; atomic force microscopy (AFM), which measures force interactions; and near-field scan-

ning optical microscopy (NSOM), which measures local optical properties by exploiting near-field effects (Fig. 12.1). These methods allow the characterization of many properties (structural, mechanical, electronic, optical) on essentially any material (metals, semiconductors, insulators, biomolecules) and in essentially any environment (vacuum, liquid, or ambient air conditions). The unique combination of nanoscale resolution, previously the domain of electron microscopy, *and broad applicability* has led to the proliferation of SPM into virtually all areas of nanometer-scale science and technology.

Several enabling technologies have been developed for SPM, or borrowed from other techniques. Piezoelectric tube scanners allow accurate, sub-angstrom positioning of the tip or sample in three dimensions. Optical deflection systems and microfabricated cantilevers can detect forces in AFM down to the picoNewton range. Sensitive electronics can measure STM currents less than 1 picoamp. High transmission fiber optics and sensitive photodetectors can manipulate and detect small optical signals of NSOM. Environmental control has been developed to allow SPM imaging in UHV, cryogenic temperatures, at elevated temperatures, and in fluids. Vibration and drift have been controlled such that a probe tip can be held over a single molecule for hours of observation. Microfabrication techniques have been developed for the mass production of probe tips, making SPMs commercially available and allowing the development of many new SPM modes and combinations with other characterization methods. However, of all this SPM development over the past 20 years, what has received the least attention is perhaps the most important aspect: the probe tip.

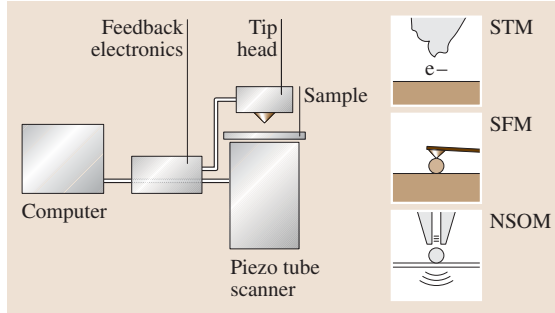


Fig. 12.1 A schematic of the components of a scanning probe microscope and the three types of signals observed: STM senses electron tunneling currents, AFM measures forces, and NSOM measures near-field optical properties via a sub-wavelength aperture

Interactions measured in SPMs occur at the tip-sample interface, which can range in size from a single atom to tens of nanometers. The size, shape, surface chemistry, electronic and mechanical properties of the tip apex will directly influence the data signal and the interpretation of the image. Clearly, the better characterized the tip the more useful the image information. In this chapter, the fabrication and performance of AFM and STM probes will be described.

12.1 Atomic Force Microscopy

AFM is the most widely used form of SPM, since it requires neither an electrically conductive sample, as in STM, nor an optically transparent sample or substrate, as in most NSOMs. Basic AFM modes measure the topography of a sample with the only requirement being that the sample is deposited on a flat surface and rigid enough to withstand imaging. Since AFM can measure a variety of forces, including van der Waals forces, electrostatic forces, magnetic forces, adhesion forces and friction forces, specialized modes of AFM can characterize the electrical, mechanical, and chemical properties of a sample in addition to its topography.

12.1.1 Principles of Operation

In AFM, a probe tip is integrated with a microfabricated force-sensing cantilever. A variety of silicon and silicon nitride cantilevers are commercially available with micron-scale dimensions, spring constants ranging from 0.01 to 100 N/m, and resonant frequencies ranging from 5 kHz to over 300 kHz. The cantilever deflection is detected by optical beam deflection, as illustrated in Fig. 12.2. A laser beam bounces off the back of the cantilever and is centered on a split photodiode. Cantilever deflections areproportional to the difference signal $V_A - V_B$. Sub-angstrom deflections can be deflected and, therefore, forces down to tens of picoNewtons can be measured. A more recently developed method of cantilever deflection measurement is through a piezoelectric layer on the cantilever that registers a voltage upon deflection [12.1].

A piezoelectric scanner rasters the sample under the tip while the forces are measured through deflections of the cantilever. To achieve more controlled imaging con-

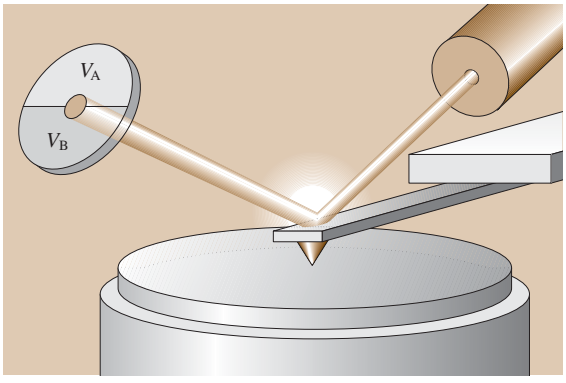


Fig. 12.2 An illustration of the optical beam deflection system that detects cantilever motion in the AFM. The voltage signal $V_A - V_B$ is proportional to the deflection

ditions, a feedback loop monitors the tip-sample force and adjusts the sample Z-position to hold the force constant. The topographic image of the sample is then taken from the sample Z-position data. The mode described is called contact mode, in which the tip is deflected by the sample due to repulsive forces, or "contact". It is generally only used for flat samples that can withstand lateral forces during scanning. To minimize lateral forces and sample damage, two AC modes have been developed. In these, the cantilever is driven into AC oscillation near its resonant frequency (tens to hundreds of kHz) with amplitudes of 5 to tens of s. When the tip approaches the sample, the oscillation is damped, and the reduced amplitude is the feedback signal, rather than the DC deflection. Again, topography is taken from the varying Z-position of the sample required to keep the tip oscillation amplitude constant. The two AC modes differ only in the nature of the interaction. In intermittent contact mode, also called tapping mode, the tip contacts the sample on each cycle, so the amplitude is reduced by ionic repulsion as in contact mode. In noncontact mode, long-range van der Waals forces reduce the amplitude by effectively shifting the spring constant experienced by the tip and changing its resonant frequency.

12.1.2 Standard Probe Tips

In early AFM work, cantilevers were made by hand from thin metal foils or small metal wires. Tips were created by gluing diamond fragments to the foil cantilevers or electrochemically etching the wires to a sharp point. Since these methods were labor intensive and not highly reproducible, they were not amenable to large-scale production. To address this problem, and the need for smaller cantilevers with higher resonant frequencies, batch fabrication techniques were developed (see Fig. 12.3). Building on existing methods to batch fabricate Si_3N_4 cantilevers, *Albrecht* et al. [12.2] etched an array of small square openings in an SiO_2 mask layer over a (100) silicon surface. The exposed square (100) regions were etched with KOH, an anisotropic etchant that terminates at the (111) planes, thus creating pyramidal etch pits in the silicon surface. The etch pit mask was then removed and another was applied to define the cantilever shapes with the pyramidal etch pits at the end. The Si wafer was then coated with a low stress Si_3N_4 layer by LPCVD. The Si_3N_4 fills the etch pit, using it as a mold to create a pyramidal tip. The silicon was later removed by etching to free the cantilevers and tips. Further steps resulting in the attachment of the cantilever to a macroscopic piece of glass are not described here. The resulting pyramidal tips were highly symmetric and had a tip radius of less than 30 nm, as determined by scanning electron microscopy (SEM). This procedure has likely not changed significantly, since commercially available Si_3N_4 tips are still specified to have a curvature radius of 30 nm.

Wolter et al. [12.3] developed methods to batch fabricate single-crystal Si cantilevers with integrated tips. Microfabricated Si cantilevers were first prepared using previously described methods, and a small mask was formed at the end of the cantilever. The Si around the mask was etched by KOH, so that the mask was under cut. This resulted in a pyramidal silicon tip under the mask, which was then removed. Again, this partial description of the full procedure only describes tip fabrication. With some refinements the silicon tips were made

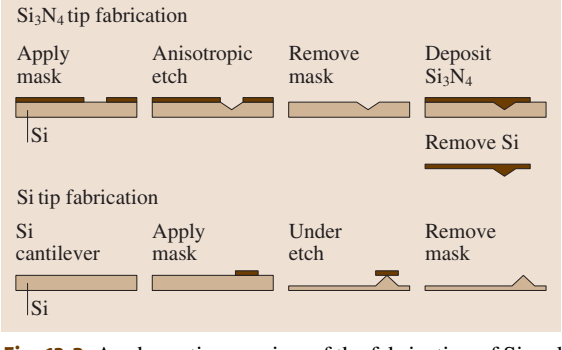


Fig. 12.3 A schematic overview of the fabrication of Si and Si_3N_4 tip fabrication as described in the text

in high yield with curvature radii of less than 10 nm. Si tips are sharper than Si_3N_4 tips, because they are directly formed by the anisotropic etch in single-crystal Si, rather than using an etch pit as a mask for deposited material. Commercially available silicon probes are made by similar refined techniques and provide a curvature typical radius of < 10 nm.

12.1.3 Probe Tip Performance

In atomic force microscopy the question of resolution can be a rather complicated issue. As an initial approximation, resolution is often considered strictly in geometrical terms that assume rigid tip-sample contact. The topographical image of a feature is broadened or narrowed by the size of the probe tip, so the resolution is approximately the width of the tip. Therefore, the resolution of AFM with standard commercially available tips is on the order of 5 to 10 nm. *Bustamante* and *Keller* [12.4] carried the geometrical model further by drawing an analogy to resolution in optical systems. Consider two sharp spikes separated by a distance d to be point objects imaged by AFM (see Fig. 12.4). Assume the tip has a parabolic shape with an end radius R. The tip-broadened image of these spikes will appear as inverted parabolas. There will be a small depression between the images of depth Δz. The two spikes are considered "resolved" if Δz is larger than the instrumental noise in the z direction. Defined in this manner, the resolution d, the minimum separation at which the spikes are resolved, is

$$d = 2\sqrt{2R\,(\Delta z)}\,, \tag{12.1}$$

where one must enter a minimal detectable depression for the instrument (Δz) to determine the resolution. So for a silicon tip with radius 5 nm and a minimum detectable Δz of 0.5 nm, the resolution is about 4.5 nm. However, the above model assumes the spikes are of equal height. *Bustamante* and *Keller* [12.4] went on to point out that if the height of the spikes is not equal, the resolution will be affected. Assuming a height difference of Δh, the resolution becomes:

$$d = \sqrt{2R}\left(\sqrt{\Delta z} + \sqrt{\Delta z + \Delta h}\right)\,. \tag{12.2}$$

For a pair of spikes with a 2 nm height difference, the resolution drops to 7.2 nm for a 5 nm tip and 0.5 nm minimum detectable Δz. While geometrical considerations are a good starting point for defining resolution, they ignore factors such as the possible compression and deformation of the tip and sample. *Vesenka* et al. [12.5] confirmed a similar geometrical resolution model by imaging monodisperse gold nanoparticles with tips characterized by transmission electron microscopy (TEM).

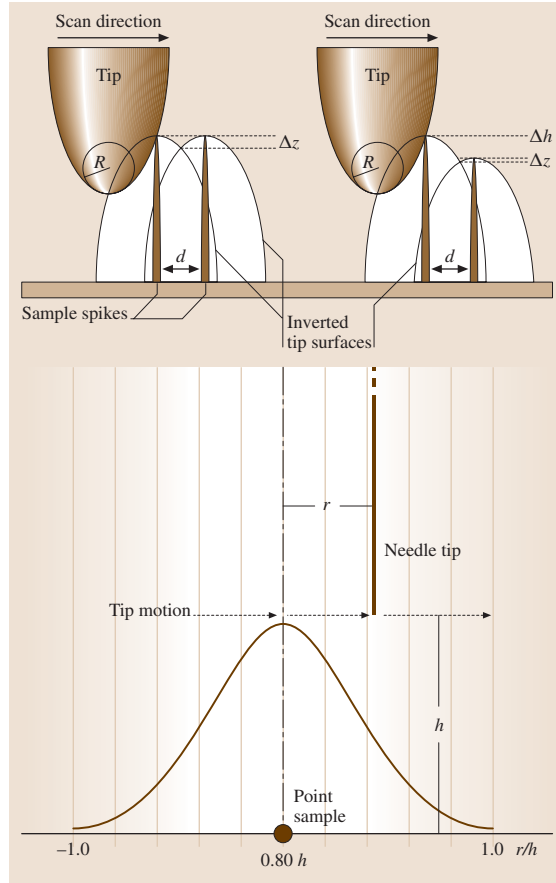


Fig. 12.4 The factors that determine AFM imaging resolution in contact mode (*top*) and noncontact mode (*bottom*), adapted from [12.4]

Noncontact AFM contrast is generated by long-range interactions such as van der Waals forces, so resolution will not simply be determined by geometry because the tip and sample are not in rigid contact. *Bustamante* and *Keller* [12.4] have derived an expression for the resolution in noncontact AFM for an idealized, infinitely thin "line" tip and a point particle as the sample (Fig. 12.4). Noncontact AFM is sensitive to the gradient of long-range forces, so the van der Waals force gradient was calculated as a function of position for the tip

at height h above the surface. If the resolution d is defined as the full width at half maximum of this curve, the resolution is:

$$d = 0.8h \; . \tag{12.3}$$

This shows that even for an ideal geometry, the resolution is fundamentally limited in noncontact mode by the tip-sample separation. Under UHV conditions, the tip-sample separation can be made very small, so atomic resolution is possible on flat, crystalline surfaces. Under ambient conditions, however, the separation must be larger to keep the tip from being trapped in the ambient water layer on the surface. This larger separation can lead to a point where further improvements in tip sharpness do not improve resolution. It has been found that imaging 5 nm gold nanoparticles in noncontact mode with carbon nanotube tips of 2 nm diameter leads to particle widths of 12 nm, larger than the 7 nm width one would expect assuming rigid contact [12.8]. However, in tapping mode operation, the geometrical definition of resolution is relevant, since the tip and sample come into rigid contact. When imaging 5 nm gold particles with 2 nm carbon nanotube tips in tapping mode, the expected 7 nm particle width is obtained [12.9].

The above descriptions of AFM resolution cannot explain the sub-nanometer resolution achieved on crystal surfaces [12.10] and ordered arrays of biomolecules [12.11] in contact mode with commercially available probe tips. Such tips have nominal radii of curvature ranging from 5 nm to 30 nm, an order of magnitude larger than the resolution achieved. A detailed model to explain the high resolution on ordered membrane proteins has been put forth by [12.6]. In this model, the larger part of the silicon nitride tip apex balances the tip-sample interaction through electrostatic forces, while a very small tip asperity interacts with the sample to provide contrast (see Fig. 12.5). This model is supported by measurements at varying salt concentrations to vary the electrostatic interaction strength and the observation of defects in the ordered samples. However, the existence of such asperities has never been confirmed by independent electron microscopy images of the tip. Another model, considered especially applicable to atomic resolution on crystal surfaces, assumes the tip is in contact with a region of the sample much larger than the resolution observed, and that force components matching the periodicity of the sample are transmitted to the tip, resulting in an "averaged" image of the periodic lattice. Regardless of the mechanism, the structures determined are accurate and make this a highly valuable method for membrane proteins. However, this level of resolution should not be expected for most biological systems.

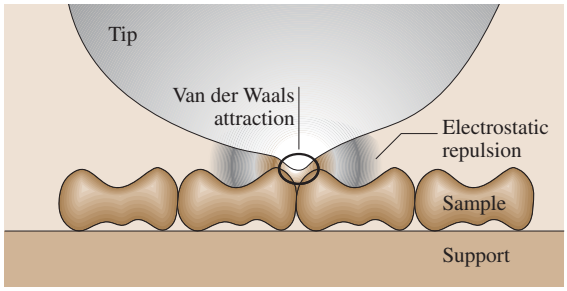


Fig. 12.5 A tip model to explain the high resolution obtained on ordered samples in contact mode, from [12.6]

12.1.4 Oxide-Sharpened Tips

Both Si and Si_3N_4 tips with increased aspect ratio and reduced tip radius can be fabricated through oxide sharpening of the tip. If a pyramidal or cone-shaped silicon tip is thermally oxidized to SiO_2 at low temperature ($< 1{,}050\,°C$), Si-SiO_2 stress formation reduces the oxidation rate at regions of high curvature. The result is a sharper, higher-aspect-ratio cone of silicon at the high curvature tip apex inside the outer pyramidal layer of SiO_2 (see Fig. 12.6). Etching the SiO_2 layer with HF then leaves tips with aspect ratios up to 10:1 and radii down to 1 nm [12.7], although 5–10 nm is the nominal specification for most commercially available tips. This oxide sharpening technique can also be applied to Si_3N_4 tips by oxidizing the silicon etch pits that are used as molds. As with tip fabrication, oxide sharpening is not quite as effective for Si_3N_4. Si_3N_4 tips were reported to have an

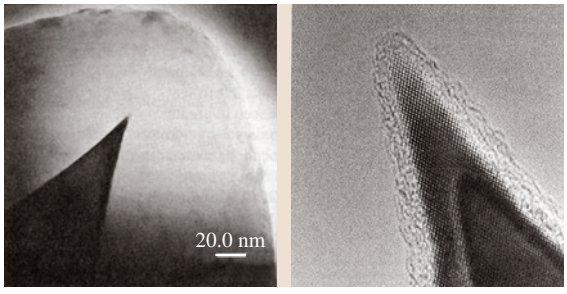


Fig. 12.6 Oxide sharpening of silicon tips. The *left image* shows a sharpened core of silicon in an outer layer of SiO_2. The *right image* is a higher magnification view of such a tip after the SiO_2 is removed. Adapted from [12.7]

11 nm radius of curvature [12.12], while commercially available oxide-sharpened Si_3N_4 tips have a nominal radius of < 20 nm.

12.1.5 FIB tips

A common AFM application in integrated circuit manufacture and MEMs is to image structures with very steep sidewalls such as trenches. To accurately image these features, one must consider the micron-scale tip structure, rather than the nanometer-scale structure of the tip apex. Since tip fabrication processes rely on anisotropic etchants, the cone half-angles of pyramidal tips are approximately 20 degrees. Images of deep trenches taken with such tips display slanted sidewalls and may not reach the bottom of the trench due to the tip broadening effects. To image such samples more faithfully, high-aspect-ratio tips are fabricated by focused ion beam (FIB) machining a Si tip to produce a sharp spike at the tip apex. Commercially available FIB tips have half cone angles of < 3 degrees over lengths of several microns, yielding aspect ratios of approximately 10:1. The radius of curvature at the tip end is similar to that of the tip before the FIB machining. Another consideration for high-aspect-ratio tips is the tip tilt. To ensure that the pyramidal tip is the lowest part of the tip-cantilever assembly, most AFM designs tilt the cantilever about 15 degrees from parallel. Therefore, even an ideal "line tip" will not give an accurate image of high steep sidewalls, but will produce an image that depends on the scan angle. Due to the versatility of the FIB machining, tips are available with the spikes at an angle to compensate for this effect.

12.1.6 EBD tips

Another method of producing high-aspect-ratio tips for AFM is called electron beam deposition (EBD). First developed for STM tips [12.13, 14], EBD tips were introduced for AFM by focusing an SEM onto the apex of a pyramidal tip arranged so that it pointed along the electron beam axis (see Fig. 12.7). Carbon material was deposited by the dissociation of background gases in the SEM vacuum chamber. *Schiffmann* [12.15] systematically studied the following parameters and how they affected EBD tip geometry:

Deposition time :	0.5 to 8 min
Beam current :	3–300 pA
Beam energy :	1–30 keV
Working distance :	8–48 mm .

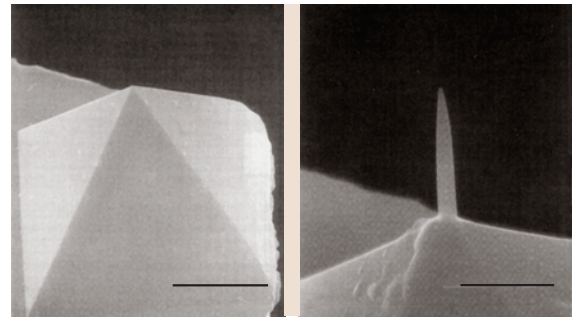


Fig. 12.7 A pyramidal tip before (*left*, 2-μm-scale bar) and after (*right*, 1-μm-scale bar) electron beam deposition, adapted from [12.13]

EBD tips were cylindrical with end radii of 20–40 nm, lengths of 1 to 5 μm, and diameters of 100 to 200 nm. Like FIB tips, EBD tips were found to achieve improved imaging of steep features. By controlling the position of the focused beam, the tip geometry can be further controlled. Tips were fabricated with lengths over 5 μm and aspect ratios greater than 100:1, yet these were too fragile to use as a tip in AFM [12.13].

12.1.7 Carbon Nanotube Tips

Carbon nanotubes are microscopic graphitic cylinders that are nanometers in diameter, yet many microns in length. Single-walled carbon nanotubes (SWNT) consist of single sp^2 hybridized carbon sheets rolled into seamless tubes and have diameters ranging from 0.7 to 3 nm.

Carbon Nanotube Structure

Larger structures called multiwalled carbon nanotubes (MWNT) consist of nested, concentrically arranged SWNT and have diameters ranging from 3 to 50 nm. Figure 12.8 shows a model of nanotube structure, as well as TEM images of a SWNT and a MWNT. The small diameter and high aspect ratio of carbon nanotubes suggests their application as high resolution, high-aspect-ratio AFM probes.

Carbon Nanotube Mechanical Properties

Carbon nanotubes possess exceptional mechanical properties that impact their use as probes. Their lateral stiffness can be approximated from that of a solid elastic rod:

$$k_{\text{lat}} = \frac{3\pi Y r^4}{4l^3} , \qquad (12.4)$$

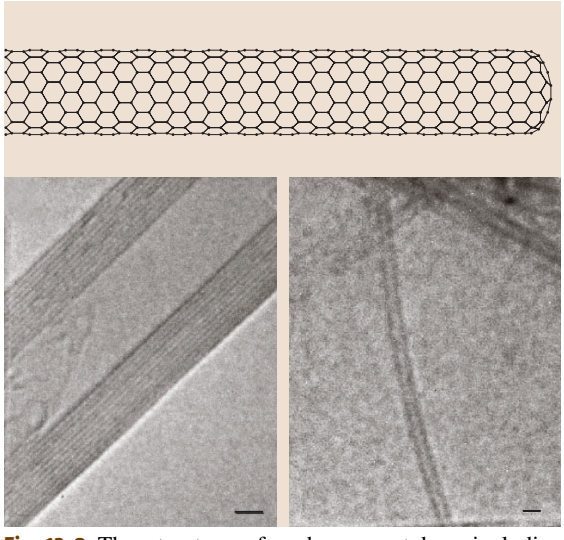


Fig. 12.8 The structure of carbon nanotubes, including TEM images of a MWNT (*left*) and a SWNT (*right*), from [12.16]

where the spring constant k_{lat} represents the restoring force per unit lateral displacement, r is the radius, l is the length, and Y is the Young's modulus (also called the elastic modulus) of the material. For the small diameters and extreme aspect ratios of carbon nanotube tips, the thermal vibrations of the probe tip at room temperature can become sufficient to degrade image resolution. These thermal vibrations can be approximated by equating $\frac{1}{2}k_{\mathrm{B}}T$ of thermal energy to the energy of an oscillating nanotube:

$$\frac{1}{2}k_{\mathrm{B}}T = \frac{1}{2}k_{\mathrm{lat}}a^2 \,, \tag{12.5}$$

where k_{B} is Boltzmann's constant, T is the temperature, and a is the vibration amplitude. Substituting for k_{lat} from (12.4) yields:

$$a = \sqrt{\frac{4k_{\mathrm{B}}Tl^3}{3\pi Yr^4}} \,. \tag{12.6}$$

The strong dependence on radius and length reveals that one must carefully control the tip geometry at this size scale. Equation (12.6) implies that the stiffer the material, i. e., the higher its Young's modulus, the smaller the thermal vibrations and the longer and thinner a tip can be. The Young's moduli of carbon nanotubes have been determined by measurements of the thermal vibration amplitude by TEM [12.18, 19] and by directly measuring the forces required to deflect a pinned carbon nanotube in an AFM [12.20]. These experiments revealed that the Young's modulus of carbon nanotubes is 1–2 TPa, in agreement with theoretical predictions [12.21]. This makes carbon nanotubes the stiffest known material and, therefore, the best for fabricating thin, high-aspect-ratio tips. A more detailed and accurate derivation of the thermal vibration amplitudes was derived for the Young's modulus measurements [12.18, 19].

Carbon nanotubes elastically buckle under large loads, rather than fracture or plastically deform like most materials. Nanotubes were first observed in the buckled state by transmission electron microscopy [12.17], as shown in Fig. 12.9. The first experimental evidence that nanotube buckling is elastic came from the application of nanotubes as probe tips [12.22], described in detail below. A more direct experimental observation of elastic buckling was obtained by deflecting nanotubes pinned to a low friction surface with an AFM tip [12.20]. Both reports found that the buckling force could be approximated with the macroscopic Euler buckling formula for an elastic column:

$$F_{\mathrm{Euler}} = \frac{\pi^3 Yr^4}{4l^2} \,. \tag{12.7}$$

The buckling force puts another constraint on the tip length: If the nanotube is too long the buckling force

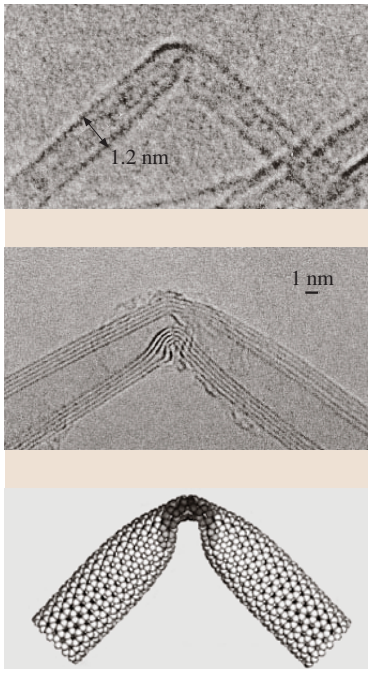


Fig. 12.9 TEM images and a model of a buckled nanotube, adapted from [12.17]

will be too low for stable imaging. The elastic buckling property of carbon nanotubes has significant implications for their use as AFM probes. If a large force is applied to the tip inadvertently, or if the tip encounters a large step in sample height, the nanotube can buckle to the side, then snap back without degraded imaging resolution when the force is removed, making these tips highly robust. No other tip material displays this buckling characteristic.

Manually Assembled Nanotube Probes

The first carbon nanotube AFM probes [12.22] were fabricated by techniques developed for assembling single-nanotube field emission tips [12.23]. This process, illustrated in Fig. 12.10, used purified MWNT material synthesized by the carbon arc procedure. The raw material, which must contain at least a few percent of long nanotubes ($> 10\,\mu m$) by weight, purified by oxidation to approximately 1% of its original mass. A torn edge of the purified material was attached to a micromanipulator by carbon tape and viewed under a high power optical microscope. Individual nanotubes and nanotube bundles were visible as filaments under dark field illumination. A commercially available AFM tip was attached to another micromanipulator opposing the nanotube material. Glue was applied to the tip apex from high vacuum carbon tape supporting the nanotube material. Nanotubes were then manually attached to the tip apex by micromanipulation. As assembled, MWNT tips were often too long for imaging due to thermal vibrations and low buckling forces described in Sect. 12.1.7. Nanotubes tips were shortened by applying 10 V pulses to the tip while it was near a sputtered niobium surface. This process etched ~ 100 nm lengths of nanotube per pulse.

The manually assembled MWNT tips demonstrated several important nanotube tip properties [12.22]. First, the high aspect ratio of the MWNT tips allowed the accurate imaging of trenches in silicon with steep sidewalls, similar to FIB and EBD tips. Second, elastic buckling was observed indirectly through force curves (see Fig. 12.11). Note that as the tip taps the sample, the amplitude drops to zero and a DC deflection is observed, because the nanotube is unbuckled and is essentially rigid. As the tip moves closer, the force on the nanotube eventually exceeds the buckling force. The nanotube buckles, allowing the vibration amplitude to partially recover, and the deflection remains constant. Numeric tip trajectory simulations could only reproduce these force curves if elastic buckling was included in the nanotube response. Finally, the nanotube tips were highly robust. Even after "tip crashes" or hundreds of controlled buckling cycles, the tip retained its resolution and high aspect ratio.

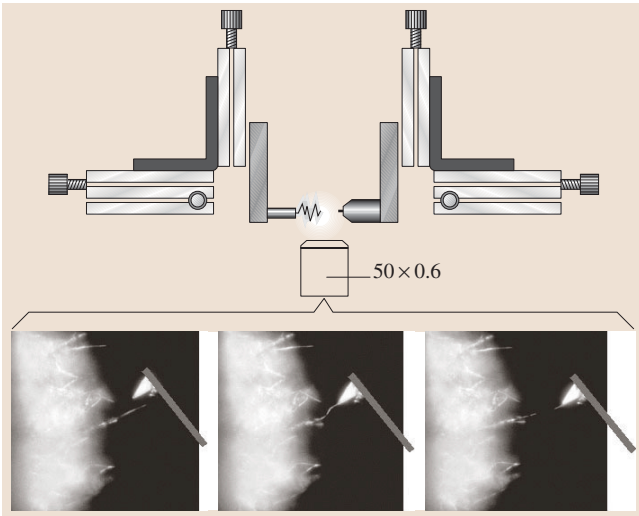


Fig. 12.10 A schematic drawing of the setup for manual assembly of carbon nanotube tips (*top*) and optical microscopy images of the assembly process (the cantilever was drawn in for clarity)

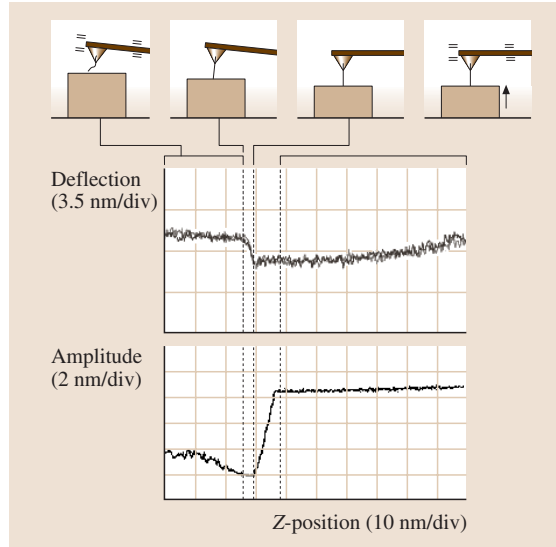


Fig. 12.11 Nanotube tip buckling. *Top* diagrams correspond to labeled regions of the force curves. As the nanotube tip buckles, the deflection remains constant and the amplitude increases, from [12.16]

Manual assembly of carbon nanotube probe tips is straightforward, but has several limitations. It is labor intensive and not amenable to mass production. Although MWNT tips have been made commercially available by this method, they are about ten times more expensive than silicon probes. The manual assembly method has also been carried out in an SEM, rather than an optical microscope [12.24]. This eliminates the need for pulse-etching, since short nanotubes can be attached to the tip, and the "glue" can be applied by EBD. But this is still not the key to mass production, since nanotube tips are made individually. MWNT tips provided a modest improvement in resolution on biological samples, but typical MWNT radii are similar to that of silicon tips, so they cannot provide the ultimate resolution possible with a SWNT tip. SWNT bundles can be attached to silicon probes by manual assembly. Pulse etching at times produces very high resolution tips that likely result from exposing a small number of nanotubes from the bundle, but this is not reproducible [12.25]. Even if a sample could be prepared that consisted of individual SWNT for manual assembly, such nanotubes would not be easily visible by optical microscopy or SEM.

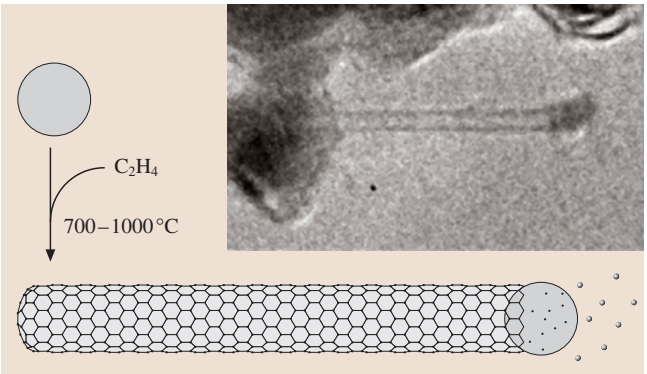


Fig. 12.12 CVD nanotube synthesis. Ethylene reacts with a nanometer-scale iron catalyst particle at high temperature to form a carbon nanotube. The *inset* in the upper right is a TEM image showing a catalyst particle at the end of a nanotube, from [12.16]

CVD Nanotube Probe Synthesis

The problems of manual assembly of nanotube probes discussed above can largely be solved by directly growing nanotubes onto AFM tips by metal-catalyzed chemical vapor deposition (CVD). The key features of the nanotube CVD process are illustrated in Fig. 12.12. Nanometer-scale metal catalyst particles are heated in a gas mixture containing hydrocarbon or CO. The gas molecules dissociate on the metal surface, and carbon is adsorbed into the catalyst particle. When this carbon precipitates, it nucleates a nanotube of similar diameter to the catalyst particle. Therefore, CVD allows control over nanotube size and structure, including the production of SWNTs [12.26] with radii as low as 3.5 Angstrom [12.27].

Several key issues must be addressed to grow nanotube AFM tips by CVD: (1) the alignment of the nanotubes at the tip, (2) the number of nanotubes that grow at the tip, and (3) the length of the nanotube tip. *Li* et al. [12.28] found that nanotubes grow perpendicular to a porous surface containing embedded catalyst. This approach was exploited to fabricate nanotube tips by CVD [12.29] with the proper alignment, as illustrated in Fig. 12.13. A flattened area of approximately 1–5 μm^2 was created on Si tips by scanning in contact mode at high load (1 μN) on a hard, synthetic diamond surface. The tip was then anodized in HF to create 100 nm-diameter pores in this flat surface [12.30]. It is important to only anodize the last 20–40 μm of the cantilever, which includes the tip, so that the rest

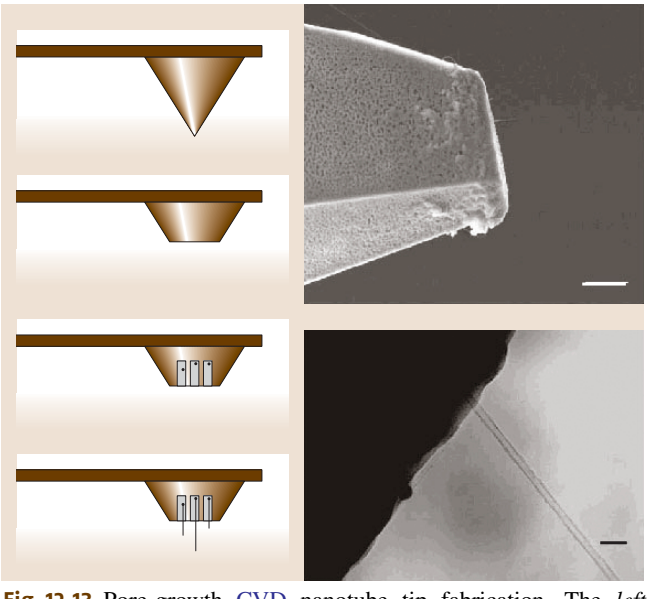


Fig. 12.13 Pore-growth CVD nanotube tip fabrication. The *left panel*, from top to bottom, shows the steps described in the text. The *upper right* is an SEM image of such a tip with a small nanotube protruding from the pores (scale bar is 1 μm). The *lower right* is a TEM of a nanotube protruding from the pores (scale bar is 20 nm), from [12.16]

of the cantilever is still reflective for use in the AFM. This was achieved by anodizing the tip in a small drop of HF under the view of an optical microscope. Next, iron was electrochemically deposited into the pores to form catalyst particles [12.31]. Tips prepared in this way were heated in low concentrations of ethylene at 800 °C, which is known to favor the growth of thin nanotubes [12.26]. When imaged by SEM, nanotubes were found to grow perpendicular to the surface from the pores as desired (Fig. 12.13). TEM revealed that the nanotubes were thin, individual, multiwalled nanotubes with typical radii ranging from 3–5 nm. If nanotubes did not grow in an acceptable orientation, the carbon could be removed by oxidation, and then CVD repeated to grow new nanotube tips.

These "pore-growth" CVD nanotube tips were typically several microns in length – too long for imaging – and were pulse-etched to a usable length of < 500 nm. The tips exhibited elastic buckling behavior and were very robust in imaging. In addition, the thin, individual nanotube tips enabled improved resolution [12.29] on isolated proteins. The pore-growth method demonstrated the potential of CVD to simplify the fabrication of nanotube tips, although there were still limitations. In particular, the porous layer was difficult to prepare and rather fragile.

An alternative approach for CVD fabrication of nanotube tips involves direct growth of SWNTs on the surface of a pyramidal AFM tip [12.32, 33]. In this "surface-growth" approach, an alumina/iron/molybdenum-powdered catalyst known to produce SWNT [12.26] was dispersed in ethanol at 1 mg/mL. Silicon tips were dipped in this solution and allowed to dry, leaving a sparse layer of ~ 100 nm catalyst clusters on the tip. When CVD conditions were applied, single-walled nanotubes grew along the silicon tip surface. At a pyramid edge, nanotubes can either bend to align with the edge, or protrude from the surface. If the energy required to bend the tube and follow the edge is less than the attractive nanotube-surface energy, then the nanotube will follow the pyramid edge to the apex. Therefore, nanotubes were effectively steered toward the tip apex by the pyramid edges. At the apex, the nanotube protruded from the tip, since the energetic cost of bending around the sharp silicon tip was too high. The high aspect ratio at the oxide-sharpened silicon tip apex was critical for good nanotube alignment. A schematic of this approach is shown in Fig. 12.14. Evidence for this model came from SEM investigations that show that a very high yield of tips contains nanotubes only at the apex, with very few protruding elsewhere

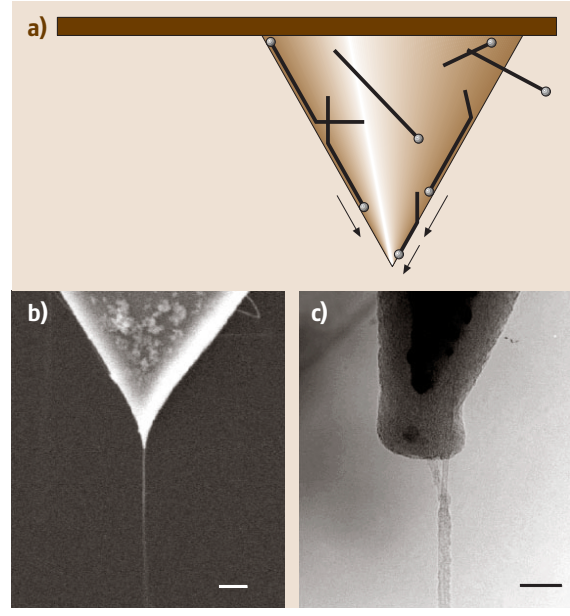


Fig. 12.14a–c Surface-growth nanotube tip fabrication. **(a)** Schematic represents the surface growth process in which nanotubes growing on the pyramidal tip are guided to the tip apex. **(b)**,**(c)** Images show **(b)** SEM (200-nm-scale bar) and **(c)** TEM (20-nm-scale bar) images of a surface growth tip, from [12.16]

from the pyramid. TEM analysis demonstrated that the tips typically consist of small SWNT bundles that are formed by nanotubes coming together from different edges of the pyramid to join at the apex, supporting the surface growth model described above (Fig. 12.14). The "surface growth" nanotube tips exhibit a high aspect ratio and high resolution imaging, as well as elastic buckling.

The surface growth method has been expanded to include wafer-scale production of nanotube tips with yields of over 90% [12.34], yet one obstacle remains to the mass production of nanotube probe tips. Nanotubes protruding from the tip are several microns long, and since they are so thin, they must be etched to less than 100 nm. While the pulse-etching step is fairly reproducible, it must be carried out on nanotube tips in a serial fashion, so surface growth does not yet represent a true method of batch nanotube tip fabrication.

Hybrid Nanotube Tip Fabrication: Pick-up Tips

Another method of creating nanotube tips is something of a hybrid between assembly and CVD. The motiva-

tion was to create AFM probes that have an *individual* SWNT at the tip to achieve the ultimate imaging resolution. In order to synthesize isolated SWNT, they must be nucleated at sites separated farther than their typical length. The alumina-supported catalyst contains a high density of catalyst particles per 100 nm cluster, so nanotube bundles cannot be avoided. To fabricate completely isolated nanotubes, isolated catalyst particles were formed by dipping a silicon wafer in an isopropyl alcohol solution of $Fe(NO_3)_3$. This effectively left a submonolayer of iron on the wafer, so that when it was heated in a CVD furnace, the iron became mobile and aggregated to form isolated iron particles. During CVD conditions, these particles nucleated and grew SWNTs. By controlling the reaction time, the SWNT lengths were kept shorter than their typical separation, so that the nanotubes never had a chance to form bundles. AFM analysis of these samples revealed 1–3 nm-diameter SWNT and un-nucleated particles on the surface (Fig. 12.15). However, there were tall objects that were difficult to image at a density of about 1 per 50 μm^2. SEM analysis at an oblique angle demonstrated that these were SWNTs that had grown perpendicular to the surface (Fig. 12.15).

In the "pick-up tip" method, these isolated SWNT substrates were imaged by AFM with silicon tips in air [12.9]. When the tip encountered a vertical SWNT, the oscillation amplitude was damped, so the AFM pulled the sample away from the tip. This pulled the SWNT into contact with the tip along its length, so that it became attached to the tip. This assembly process happened automatically when imaging in tapping mode – no special tip manipulation was required. When imaging a wafer with the density shown in Fig. 12.15, one nanotube was attached per 8 μm × 8 μm scan at 512×512 and 2 Hz, so a nanotube tip could be made in about 5 min. Since the as-formed SWNT tip continued to image, there was usually no evidence in the topographic image that a nanotube had been attached. However, the pick-up event was identified when the *Z*-voltage suddenly stepped to larger tip-

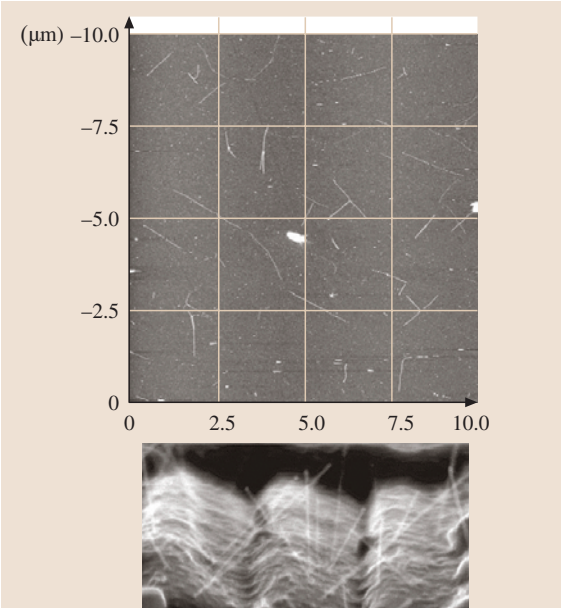


Fig. 12.15 Atomic force microscopy image (*top*) of a wafer with isolated nanotubes synthesized by CVD. The SEM view provides evidence that some of these nanotubes are arranged vertically

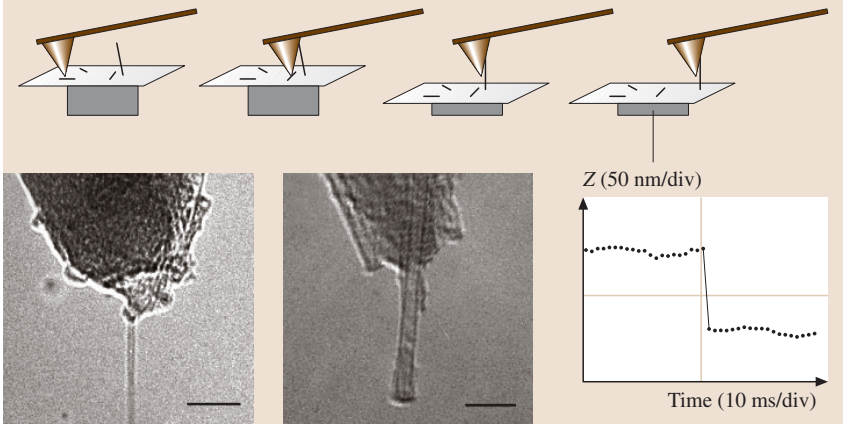


Fig. 12.16 Pick-up tip assembly of nanotube probes. The *top* illustrates the nanotube pick-up process that occurs while imaging vertical nanotubes in an AFM, including a trace of the *Z*-position during a pick-up event. The *lower left* TEM images show single nanotubes (diameters 0.9 nm and 2.8 nm) on an AFM tip fabricated by this method, adapted from [12.9]

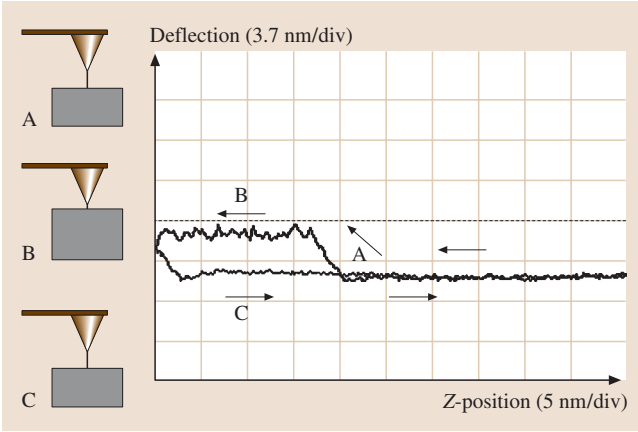


Fig. 12.17 The process by which nanotube tips can be shortened in AFM force curves. The hysteresis in the deflection trace (*bottom*) reveals that ∼ 17 nm were removed

sample separation due to the effective increase in tip length.

Individual SWNT tips must be quite short, typically less than 50 nm in length, for reasons outlined above. Pulse-etching, which removes 50–100 nm of nanotube length at a time, lacked the necessary precision for shortening pick-up tips, so the tips were shortened through force curves. A pick-up tip force curve is shown in Fig. 12.17. When the tip first interacted with the sample, the amplitude decreased to zero, and further approach generated a small deflection that ultimately saturated. However, this saturation was not due to buckling. Note that the amplitude did not recover, as in the buckling curves. This leveling was due to the nanotube sliding on the pyramidal tip, which was confirmed by the hysteresis in the amplitude and deflection curves. If the force curve was repeated, the tip showed no deflection or amplitude drop until further down, because the tip was essentially shorter.

Pick-up SWNT tips achieve the highest resolution of all nanotube tips, since they are always individuals rather than bundles. In tapping mode, they produce images of 5 nm gold particles that have a *full width* of ∼ 7 nm, the expected geometrical resolution for a 2 nm cylindrical probe [12.9]. Although the pick-up method is serial in nature, it may still be the key to the mass production of nanotube tips. Note that the original nanotube tip length can be measured electronically from the size of the Z-piezo step. The shortening can be electronically controlled through the hysteresis in the force curves. Therefore, the entire procedure (including tip exchange) can be automated by computer.

12.2 Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) was the original scanning probe microscopy and generally produces the highest resolution images, routinely achieving atomic resolution on flat, conductive surfaces. In STM, the probe tip consists of a sharpened metal wire that is held 0.3 to 1 nm from the sample. A potential difference of 0.1 V to 1 V between the tip and sample leads to tunneling currents on the order of 0.1 to 1 nA. As in AFM, a piezo-scanner rasters the sample under the tip, and the Z-position is adjusted to hold the tunneling current constant. The Z-position data represents the "topography", or in this case the surface, of constant electron density. As with other SPMs, the tip properties and performance greatly depend on the experiment being carried out. Although it is nearly impossible to prepare a tip with a known atomic structure, a number of factors are known to affect tip performance, and several preparation methods have been developed that produce good tips.

The nature of the sample being investigated and the scanning environment will affect the choice of the tip material and how the tip is fabricated. Factors to consider are mechanical properties – a hard material that will resist damage during tip-sample contact is desired. Chemical properties should also be considered – formation of oxides, or other insulating contaminants will affect tip performance. Tungsten is a common tip material because it is very hard and will resist damage, but its use is limited to ultrahigh vacuum (UHV) conditions, since it readily oxidizes. For imaging under ambient conditions an inert tip material such as platinum or gold is preferred. Platinum is typically alloyed with iridium to increase its stiffness.

12.2.1 Mechanically Cut STM Tips

STM tips can be fabricated by simple mechanical procedures such as grinding or cutting metal wires. Such

tips are not formed with highly reproducible shapes and have a large opening angle and a large radius of curvature in the range of 0.1 to 1 μm (see Fig. 12.18a). They are not useful for imaging samples with surface roughness above a few nanometers. However, on atomically flat samples, mechanically cut tips can achieve atomic resolution due to the nature of the tunneling signal, which drops exponentially with tip-sample separation. Since mechanically cut tips contain many small asperities on the larger tip structure, atomic resolution is easily achieved as long as one atom of the tip is just a few angstroms lower than all of the others.

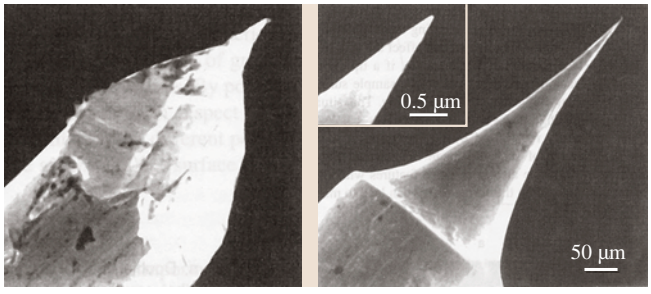


Fig. 12.18 A mechanically cut STM tip (*left*) and an electrochemically etched STM tip (*right*), from [12.35]

12.2.2 Electrochemically Etched STM Tips

For samples with more than a few nanometers of surface roughness, the tip structure in the nanometer-size range becomes an issue. Electrochemical etching can provide tips with reproducible and desirable shapes and sizes (Fig. 12.18), although the exact atomic structure of the tip apex is still not well controlled. The parameters of electrochemical etching depend greatly on the tip material and the desired tip shape. The following is an entirely general description. A fine metal wire (0.1–1 mm diameter) of the tip material is immersed in an appropriate electrochemical etchant solution. A voltage bias of 1–10 V is applied between the tip and a counterelectrode such that the tip is etched. Due to the enhanced etch rate at the electrolyte-air interface, a neck is formed in the wire. This neck is eventually etched thin enough so that it cannot support the weight of the part of the wire suspended in the solution, and it breaks to form a sharp tip. The widely varying parameters and methods will be not be covered in detail here, but many recipes are found in the literature for common tip materials [12.36–39].

References

12.1 R. Linnemann, T. Gotszalk, I. W. Rangelow, P. Dumania, E. Oesterschulze: Atomic force microscopy and lateral force microscopy using piezoresistive cantilevers, J. Vac. Sci. Technol. B **14**(2) (1996) 856–860

12.2 T. R. Albrecht, S. Akamine, T. E. Carver, C. F. Quate: Microfabrication of cantilever styli for the atomic force microscope, J. Vac. Sci. Technol. A **8**(4) (1990) 3386–3396

12.3 O. Wolter, T. Bayer, J. Greschner: Micromachined silicon sensors for scanning force microscopy, J. Vac. Sci. Technol. B **9**(2) (1991) 1353–1357

12.4 C. Bustamante, D. Keller: Scanning force microscopy in biology, Phys. Today **48**(12) (1995) 32–38

12.5 J. Vesenka, S. Manne, R. Giberson, T. Marsh, E. Henderson: Colloidal gold particles as an incompressible atomic force microscope imaging standard for assessing the compressibility of biomolecules, Biophys. J. **65** (1993) 992–997

12.6 D. J. Muller, D. Fotiadis, S. Scheuring, S. A. Muller, A. Engel: Electrostatically balanced subnanometer imaging of biological specimens by atomic force microscope, Biophys. J. **76**(2) (1999) 1101–1111

12.7 R. B. Marcus, T. S. Ravi, T. Gmitter, K. Chin, D. Liu, W. J. Orvis, D. R. Ciarlo, C. E. Hunt, J. Trujillo: Formation of silicon tips with <1 nm radius, Appl. Phys. Lett. **56**(3) (1990) 236–238

12.8 J. H. Hafner, C. L. Cheung, C. M. Lieber: unpublished results (2001)

12.9 J. H. Hafner, C. L. Cheung, T. H. Oosterkamp, C. M. Lieber: High-yield assembly of individual single-walled carbon nanotube tips for scanning probe microscopies, J. Phys. Chem. B **105**(4) (2001) 743–746

12.10 F. Ohnesorge, G. Binnig: True atomic resolution by atomic force microscopy through repulsive and attractive forces, Science **260** (1993) 1451–1456

12.11 D. J. Muller, D. Fotiadis, A. Engel: Mapping flexible protein domains at subnanometer resolution with the atomic force microscope, FEBS Lett. **430**(1–2 Special Issue SI) (1998) 105–111

12.12 S. Akamine, R. C. Barrett, C. F. Quate: Improved atomic force microscope images using microcantilevers with sharp tips, Appl. Phys. Lett. **57**(3) (1990) 316–318

12.13 D.J. Keller, C. Chih-Chung: Imaging steep, high structures by scanning force microscopy with electron beam deposited tips, Surf. Sci. **268** (1992) 333–339

12.14 T. Ichihashi, S. Matsui: In situ observation on electron beam induced chemical vapor deposition by transmission electron microscopy, J. Vac. Sci. Technol. B **6**(6) (1988) 1869–1872

12.15 K.I. Schiffmann: Investigation of fabrication parameters for the electron-beam-induced deposition of contamination tips used in atomic force microscopy, Nanotechnology **4** (1993) 163–169

12.16 J.H. Hafner, C.L. Cheung, A.T. Woolley, C.M. Lieber: Structural and functional imaging with carbon nanotube AFM probes, Prog. Biophys. Mol. Biol. **77**(1) (2001) 73–110

12.17 S. Iijima, C. Brabec, A. Maiti, J. Bernholc: Structural flexibility of carbon nanotubes, J. Chem. Phys. **104**(5) (1996) 2089–2092

12.18 M.M.J. Treacy, T.W. Ebbesen, J.M. Gibson: Exceptionally high Young's modulus observed for individual carbon nanotubes, Nature **381** (1996) 678–680

12.19 A. Krishnan, E. Dujardin, T.W. Ebbesen, P.N. Yianilos, M.M.J. Treacy: Young's modulus of single-walled nanotubes, Phys. Rev. B **58**(20) (1998) 14013–14019

12.20 E.W. Wong, P.E. Sheehan, C.M. Lieber: Nanobeam mechanics – elasticity, strength, and toughness of nanorods and nanotubes, Science **277**(5334) (1997) 1971–1975

12.21 J.P. Lu: Elastic properties of carbon nanotubes and nanoropes, Phys. Rev. Lett. **79**(7) (1997) 1297–1300

12.22 H.J. Dai, J.H. Hafner, A.G. Rinzler, D.T. Colbert, R.E. Smalley: Nanotubes as nanoprobes in scanning probe microscopy, Nature **384**(6605) (1996) 147–150

12.23 A.G. Rinzler, Y.H. Hafner, P. Nikolaev, L. Lou, S.G. Kim, D. Tomanek, D.T. Colbert, R.E. Smalley: Unraveling nanotubes: Field emission from atomic wire, Science **269** (1995) 1550

12.24 H. Nishijima, S. Kamo, S. Akita, Y. Nakayama, K.I. Hohmura, S.H. Yoshimura, K. Takeyasu: Carbon-nanotube tips for scanning probe microscopy: Preparation by a controlled process and observation of deoxyribonucleic acid, Appl. Phys. Lett. **74**(26) (1999) 4061–4063

12.25 S.S. Wong, A.T. Woolley, T.W. Odom, J.L. Huang, P. Kim, D.V. Vezenov , C.M. Lieber: Single-walled carbon nanotube probes for high-resolution nanostructure imaging, Appl. Phys. Lett. **73**(23) (1998) 3465–3467

12.26 J.H. Hafner, M.J. Bronikowski, B.R. Azamian, P. Nikolaev, A.G. Rinzler, D.T. Colbert, K.A. Smith, R.E. Smalley: Catalytic growth of single-wall carbon nanotubes from metal particles, Chem. Phys. Lett. **296**(1–2) (1998) 195–202

12.27 P. Nikolaev, M.J. Bronikowski, R.K. Bradley, F. Rohmund, D.T. Colbert, K.A. Smith, R.E. Smalley: Gas-phase catalytic growth of single-walled carbon nanotubes from carbon monoxide, Chem. Phys. Lett. **313**(1–2) (1999) 91–97

12.28 W.Z. Li, S.S. Xie, L.X. Qian, B.H. Chang, B.S. Zou, W.Y. Zhou, R.A. Zhao, G. Wang: Large-scale synthesis of aligned carbon nanotubes, Science **274**(5293) (1996) 1701–1703

12.29 J.H. Hafner, C.L. Cheung, C.M. Lieber: Growth of nanotubes for probe microscopy tips, Nature **398**(6730) (1999) 761–762

12.30 V. Lehmann: The physics of macroporous silicon formation, Thin Solid Films **255** (1995) 1–4

12.31 F. Ronkel, J.W. Schultze, R. Arensfischer: Electrical contact to porous silicon by electrodeposition of iron, Thin Solid Films **276**(1–2) (1996) 40–43

12.32 J.H. Hafner, C.L. Cheung, C.M. Lieber: Direct growth of single-walled carbon nanotube scanning probe microscopy tips, J. Am. Chem. Soc. **121**(41) (1999) 9750–9751

12.33 E.B. Cooper, S.R. Manalis, H. Fang, H. Dai, K. Matsumoto, S.C. Minne, T. Hunt, C.F. Quate: Terabit-per-square-inch data storage with the atomic force microscope, Appl. Phys. Lett. **75**(22) (1999) 3566–3568

12.34 E. Yenilmez, Q. Wang, R.J. Chen, D. Wang, H. Dai: Wafer scale production of carbon nanotube scanning probe tips for atomic force microscopy, Appl. Phys. Lett. **80**(12) (2002) 2225–2227

12.35 A. Stemmer, A. Hefti, U. Aebi, A. Engel: Scanning tunneling and transmission electron microscopy on identical areas of biological specimens, Ultramicroscopy **30**(3) (1989) 263

12.36 R. Nicolaides, L. Yong, W.E. Packard, W.F. Zhou, H.A. Blackstead, K.K. Chin, J.D. Dow, J.K. Furdyna, M.H. Wei, R.C. Jaklevic, W.J. Kaiser, A.R. Pelton, M.V. Zeller, J.J. Bellina: Scanning tunneling microscope tip structures, J. Vac. Sci. Technol. A **6**(2) (1988) 445–447

12.37 J.P. Ibe, P.P. Bey, S.L. Brandow, R.A. Brizzolara, N.A. Burnham, D.P. DiLella, K.P. Lee, C.R.K. Marrian, R.J. Colton: On the electrochemical etching of tips for scanning tunneling microscopy, J. Vac. Sci. Technol. A **8** (1990) 3570–3575

12.38 L. Libioulle, Y. Houbion, J.-M. Gilles: Very sharp platinum tips for scanning tunneling microscopy, Rev. Sci. Instrum. **66**(1) (1995) 97–100

12.39 A.J. Nam, A. Teren, T.A. Lusby, A.J. Melmed: Benign making of sharp tips for STM and FIM: Pt, Ir, Au, Pd, and Rh, J. Vac. Sci. Technol. B **13**(4) (1995) 1556–1559

13. Noncontact Atomic Force Microscopy and Its Related Topics

The scanning probe microscopy (SPM) such as the STM and the NC-AFM is the basic technology for the nanotechnology and also for the future bottom-up process. In Sect. 13.1, the principles of AFM such as operating modes and the frequency modulation method of the NC-AFM are fully explained. Then, in Sect. 13.2, applications of NC-AFM to semiconductors that make clear its potentials such as spatial resolution and functions are introduced. Next, in Sect. 13.3, applications of NC-AFM to insulators such as alkali halides, fluorides and transition metal oxides are introduced. At last, in Sect. 13.4, applications of NC-AFM to molecules such as carboxylate ($RCOO^-$) with R = H, CH_3, $C(CH_3)_3$ and CF_3 are introduced. Thus, the NC-AFM can observe atoms and molecules on various kinds of surfaces such as semiconductor, insulator and metal oxide with atomic/molecular resolutions. These sections are essential to understand the status of the art and the future possibility of NC-AFM that is the second generation of atom/molecule technology.

13.1 **Principles of Noncontact Atomic Force Microscope (NC-AFM)** 386
13.1.1 Imaging Signal in AFM 386
13.1.2 Experimental Measurement and Noise 387
13.1.3 Static AFM Operating Mode 387
13.1.4 Dynamic AFM Operating Mode 388
13.1.5 The Four Additional Challenges Faced by AFM 388
13.1.6 Frequency-Modulation AFM (FM-AFM) 389
13.1.7 Relation Between Frequency Shift and Forces 390
13.1.8 Noise in Frequency-Modulation AFM – Generic Calculation 391
13.1.9 Conclusion 391

13.2 **Applications to Semiconductors** 391
13.2.1 Si(111)7×7 Surface 392
13.2.2 Si(100)2×1 and Si(100)2×1:H Monohydride Surfaces 393
13.2.3 Metal-Deposited Si Surface 395

13.3 **Applications to Insulators** 397
13.3.1 Alkali Halides, Fluorides, and Metal Oxides 397
13.3.2 Atomically Resolved Imaging of a NiO(001) Surface 402
13.3.3 Atomically Resolved Imaging Using Noncoated and Fe-Coated Si Tips 402

13.4 **Applications to Molecules** 404
13.4.1 Why Molecules and What Molecules? 404
13.4.2 Mechanism of Molecular Imaging 404
13.4.3 Perspectives 407

References 407

The scanning tunneling microscope (STM) is an atomic tool based on an electric method that measures the tunneling current between a conductive tip and a conductive surface. It can electrically observe individual atoms/molecules. It can characterize or analyze the electronic nature around surface atoms/molecules. In addition, it can manipulate individual atoms/molecules. Hence, the STM is the first generation of atom/molecule technology. On the other hand, the atomic force microscope (AFM) is a unique atomic tool based on a mechanical method that can deal with even the insulator surface. Since the invention of noncontactAFM (NC-AFM) in 1995, the NC-AFM and the NC-AFM-based method have rapidly developed into a powerful surface tool on atomic/molecular scales, because the NC-AFM has the following characteristics: (1) it has true atomic resolution, (2) it can measure atomic force (so-called atomic force spectroscopy), (3) it can observe even insulators, and (4) it can measure mechanical responses such as elastic deformation. Thus, the NC-AFM is the second generation of atom/molecule technology. The scanning probe microscopy (SPM) such as the STM and the NC-AFM is the basic technology for nanotechnology and also for the future bottom-up process.

In Sect. 13.1, principles of NC-AFM will be fully introduced. Then, in Sect. 13.2, applications to semi-

conductors will be presented. Next, in Sect. 13.3, applications to insulators will be described. And, in Sect. 13.4, applications to molecules will be introduced. These sections are essential to understanding the status of the art and the future possibility of the NC-AFM.

13.1 Principles of Noncontact Atomic Force Microscope (NC-AFM)

The atomic force microscope (AFM), invented by *Binnig* [13.1] and introduced in 1986 by *Binnig*, *Quate*, and *Gerber* [13.2], is an offspring of the scanning tunneling microscope (STM) [13.3]. The STM is covered in books and review articles, e.g., [13.4–9]. Early on in the development of STM it became evident that relatively strong forces act between a tip in close proximity to a sample. It was found that these forces could be put to good use in the atomic force microscope (AFM). Detailed information about the noncontact AFM can be found in [13.10–12].

13.1.1 Imaging Signal in AFM

Figure 13.1 shows a sharp tip close to a sample. The potential energy between the tip and sample V_{ts} causes a z component of the tip-sample force $F_{ts} = -\partial V_{ts}/\partial z$. Depending on the mode of operation, the AFM uses F_{ts} or some entity derived from F_{ts} as the imaging signal.

Unlike the tunneling current, which has a very strong distance dependence, F_{ts} has long- and short-range contributions. We can classify the contributions by their range and strength. In vacuum, there are van der Waals, electrostatic, and magnetic forces with long-range (up to 100 nm) and short-range chemical forces (fractions of a nm).

The van der Waals interaction is caused by fluctuations in the electric dipole moment of atoms and their mutual polarization. For a spherical tip with radius R next to a flat surface (z is the distance between the plane connecting the centers of the surface atoms and the center of the closest tip atom), the van der Waals potential is given by [13.13]:

$$V_{\mathrm{vdW}} = -\frac{A_{\mathrm{H}}}{6z} \,. \tag{13.1}$$

The "Hamaker constant", A_{H}, depends on the type of materials (atomic polarizability and density) of the tip and sample and is on the order of 1 eV for most solids [13.13].

When the tip and sample are both conductive and have an electrostatic potential difference, $U \neq 0$, electrostatic forces are important. For a spherical tip with radius R, the force is given by [13.14]:

$$F_{\mathrm{electrostatic}} = -\frac{\pi\varepsilon_0 R U^2}{z} \,. \tag{13.2}$$

Chemical forces are more complicated. Empirical model potentials for chemical bonds are the Morse potential (see e.g., [13.13]).

$$V_{\mathrm{Morse}} = -E_{\mathrm{bond}}\left(2\,\mathrm{e}^{-\kappa(z-\sigma)} - \mathrm{e}^{-2\kappa(z-\sigma)}\right) \tag{13.3}$$

and the Lennard-Jones potential [13.13]:

$$V_{\mathrm{Lennard\text{-}Jones}} = -E_{\mathrm{bond}}\left(2\frac{\sigma^6}{z^6} - \frac{\sigma^{12}}{z^{12}}\right) \,. \tag{13.4}$$

These potentials describe a chemical bond with bonding energy E_{bond} and equilibrium distance σ. The Morse potential has an additional parameter – a decay length κ.

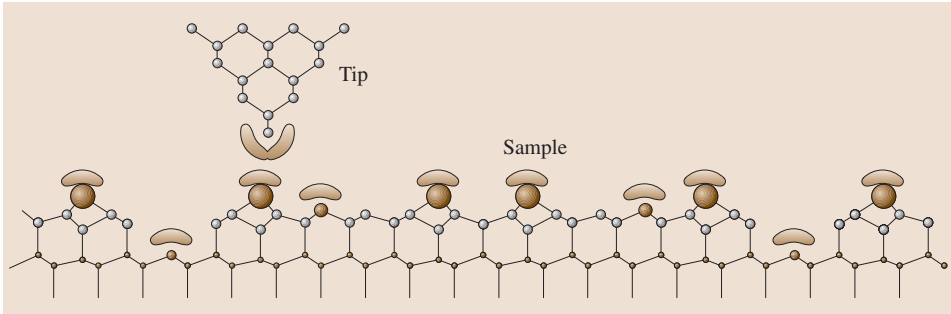


Fig. 13.1 Schematic view of an AFM tip close to a sample

13.1.2 Experimental Measurement and Noise

Forces between the tip and sample are typically measured by recording the deflection of a cantilever beam that has a tip mounted to its end (see Fig. 13.2). Today's microfabricated silicon cantilevers were first created by the *Calvin F. Quate* group [13.15–17] and *Wolter* et al. at IBM [13.18].

The cantilever is characterized by its spring constant k, eigenfrequency f_0, and quality factor Q.

For a rectangular cantilever with dimensions w, t, and L (see Fig. 13.2), the spring constant k is given by [13.6]:

$$k = \frac{E_Y w t^3}{4L^3} , \tag{13.5}$$

where E_Y is Young's modulus. The eigenfrequency f_0 is given by [13.6]:

$$f_0 = 0.162 \frac{t}{L^2} \sqrt{E_Y/\rho} , \tag{13.6}$$

where ρ is the mass density of the cantilever material. The Q-factor depends on the damping mechanisms present in the cantilever. For micromachined cantilevers operated in air, Q is typically a few hundred but can reach hundreds of thousands in vacuum.

In the first AFM, the deflection of the cantilever was measured with an STM – the backside of the cantilever was metalized, and a tunneling tip was brought close to it to measure the deflection [13.2]. Today's designs use optical (interferometer, beam-bounce) or electrical methods (piezoresistive, piezoelectric) for measuring the cantilever deflection. A discussion of the various techniques can be found in [13.19], descriptions of piezoresistive detection schemes are found in [13.17, 20], and piezoelectric methods are explained in [13.21–24].

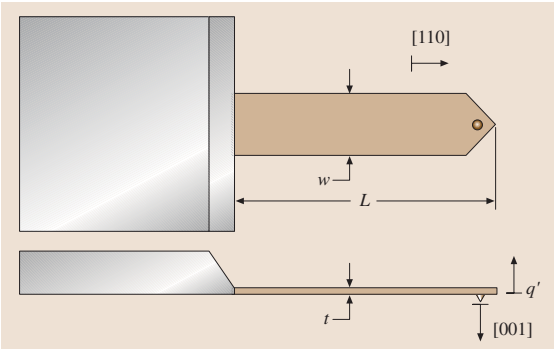


Fig. 13.2 Top view and side view of a microfabricated silicon cantilever (schematic)

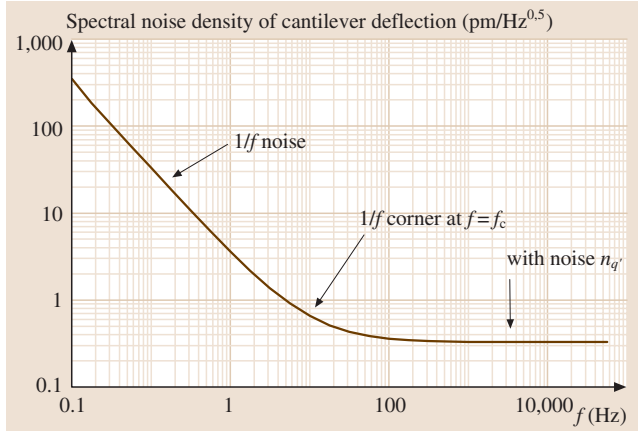


Fig. 13.3 Schematic view of $1/f$ noise apparent in force detectors. Static AFMs operate in a frequency range from 0.01 Hz to a few hundred Hz, while dynamic AFMs operate at frequencies around 10 kHz to a few hundred kHz. The noise of the cantilever deflection sensor is characterized by the $1/f$ corner frequency f_c and the constant deflection noise density $n_{q'}$ for the frequency range where white noise is dominating

The quality of the cantilever deflection measurement can be expressed in a schematic plot of the deflection noise density versus frequency, as in Fig. 13.3.

The noise density has a $1/f$ dependence for low frequency and merges into a constant noise density ("white noise") above the "$1/f$ corner frequency".

13.1.3 Static AFM Operating Mode

In the static mode of operation, the force translates into a deflection $q' = F_{ts}/k$ of the cantilever, yielding images as maps $z(x, y, F_{ts} = \text{const.})$. The noise level of the force measurement is then given by the cantilever's spring constant k times the noise level of the deflection measurement. In this respect, a small value for k increases force sensitivity. On the other hand, instabilities are more likely to occur with soft cantilevers (see Sect. 13.1.5). Because the deflection of the cantilever should be significantly larger than the deformation of the tip and sample, the cantilever should be much softer than the bonds between the bulk atoms in the tip and sample. Inter-atomic force constants in solids are in a range from 10 N/m to about 100 N/m – in biological samples, they can be as small as 0.1 N/m. Thus, typical values for k in the static mode are 0.01–5 N/m.

Even though it has been demonstrated that atomic resolution is possible with static AFM, the method can only be applied in certain cases. The detrimental effects of $1/f$ noise can be limited by working at low temperatures [13.25], where the coefficients of thermal expansion are very small, or by building the AFM of a material with a low thermal expansion coefficient [13.26]. The long-range attractive forces have to be cancelled by immersing the tip and sample in a liquid [13.26], or by partly compensating the attractive force by pulling at the cantilever after jump-to-contact has occurred [13.27]. *Jarvis* et al. have cancelled the long-range attractive force with an electromagnetic force applied to the cantilever [13.28]. Even with these restrictions, static AFM does not produce atomic resolution on reactive surfaces like silicon, as the chemical bonding of the AFM tip and sample pose an unsurmountable problem [13.29, 30].

13.1.4 Dynamic AFM Operating Mode

In the dynamic operation modes, the cantilever is deliberately vibrated. There are two basic methods of dynamic operation: amplitude modulation (AM) and frequency modulation (FM) operation. In AM-AFM [13.31], the actuator is driven by a fixed amplitude A_{drive} at a fixed frequency f_{drive}, where f_{drive} is close to f_0. When the tip approaches the sample, elastic and inelastic interactions cause a change in both the amplitude and the phase (relative to the driving signal) of the cantilever. These changes are used as the feedback signal. While the AM mode was initially used in a noncontact mode, it was later implemented very successfully at a closer distance range in ambient conditions involving repulsive tip-sample interactions.

The change in amplitude in AM mode does not occur instantaneously with a change in the tip-sample interaction, but on a time scale of $\tau_{\text{AM}} \approx 2Q/f_0$, and the AM mode is slow with high-Q cantilevers. However, the use of high Q-factors reduces noise. *Albrecht* et al. found a way to combine the benefits of high Q and high speed by introducing the frequency modulation (FM) mode [13.32], where the change in the eigenfrequency settles on a time scale of $\tau_{\text{FM}} \approx 1/f_0$.

Using the FM mode, the resolution was improved dramatically, and finally atomic resolution [13.33, 34] was obtained by reducing the tip-sample distance and working in vacuum. For atomic studies in vacuum, the FM mode (see Sect. 13.1.6) is now the preferred AFM technique. However, atomic resolution in vacuum can also be obtained with the AM mode, as demonstrated by *Erlandsson* et al. [13.35].

13.1.5 The Four Additional Challenges Faced by AFM

Some of the inherent AFM challenges are apparent by comparing the tunneling current and tip-sample force as a function of distance (Fig. 13.4).

The tunneling current is a monotonic function of the tip-sample distance and has a very sharp distance dependence. In contrast, the tip-sample force has long- and short-range components and is not monotonic.

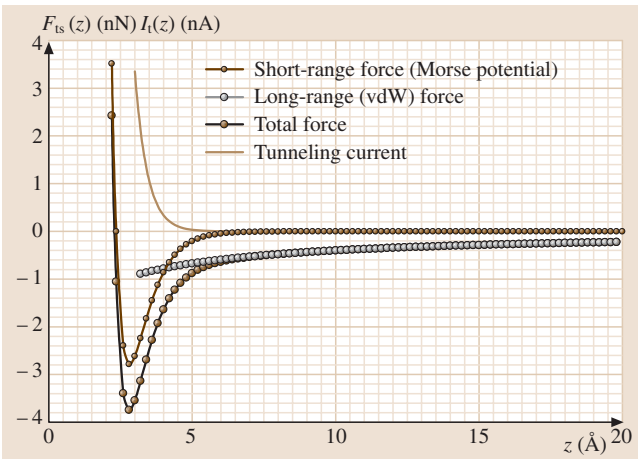


Fig. 13.4 Plot of tunneling current I_t and force F_{ts} (typical values) as a function of distance z between front atom and surface atom layer

Jump-to-Contact and Other Instabilities

If the tip is mounted on a soft cantilever, the initially attractive tip-sample forces can cause a sudden "jump-to-contact" when the tip approaches the sample. This instability occurs in the quasi-static mode if [13.36, 37]

$$k < \max\left(-\frac{\partial^2 V_{\text{ts}}}{\partial z^2}\right) = k_{\text{ts}}^{\max} \ . \tag{13.7}$$

Jump-to-contact can be avoided even for soft cantilevers by oscillating it at a large enough amplitude A [13.38]:

$$kA > \max\left(-F_{\text{ts}}\right) \ . \tag{13.8}$$

If hysteresis occurs in the $F_{\text{ts}}(z)$-relation, the energy ΔE_{ts} needs to be supplied to the cantilever for each oscillation cycle. If this energy loss is large compared to the intrinsic energy loss of the cantilever, amplitude control can become difficult. An additional approximate

criterion for k and A is then

$$\frac{kA^2}{2} \geq \frac{\Delta E_{\mathrm{ts}} Q}{2\pi} . \tag{13.9}$$

Contribution of Long-Range Forces

The force between tip and sample is composed of many contributions: electrostatic, magnetic, van der Waals, and chemical forces in vacuum. All of these force types except for the chemical forces have strong long-range components that conceal the atomic-force components. For imaging by AFM with atomic resolution, it is desirable to filter out the long-range force contributions and only measure the force components that vary at the atomic scale. While there is no way to discriminate between long- and short-range forces in static AFM, it is possible to enhance the short-range contributions in dynamic AFM through the proper choice of the oscillation amplitude A of the cantilever.

Noise in the Imaging Signal

Measuring the cantilever deflection is subject to noise, especially at low frequencies ($1/f$ noise). In static AFM, this noise is particularly problematic because of the approximate $1/f$ dependence. In dynamic AFM, the low-frequency noise is easily discriminated when using a bandpass filter with a center frequency around f_0.

Non-monotonic Imaging Signal

The tip-sample force is not monotonic. In general, the force is attractive for large distances, and upon decreasing the distance between tip and sample, the force turns repulsive (see Fig. 13.4). Stable feedback is only possible on a monotonic subbranch of the force curve.

Frequency-modulation AFM (FM-AFM) helps to overcome challenges. The non-monotonic imaging signal in AFM is a remaining complication for FM-AFM.

13.1.6 Frequency-Modulation AFM (FM-AFM)

In FM-AFM, a cantilever with eigenfrequency f_0 and spring constant k is subject to controlled positive feedback such that it oscillates with a constant amplitude A [13.32], as shown in Fig. 13.5.

Experimental Setup

The deflection signal is phase shifted, routed through an automatic gain control circuit and fed back to the actuator. The frequency f is a function of f_0, its quality factor Q, and the phase shift ϕ between the mechanical excitation generated at the actuator and the deflection of the cantilever. If $\phi = \pi/2$, the loop oscillates at $f = f_0$.

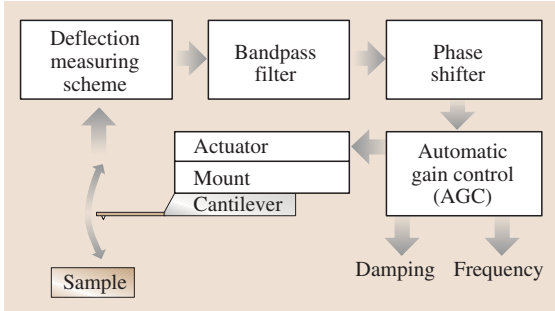


Fig. 13.5 Block diagram of a frequency-modulation force sensor

Three physical observables can be recorded: 1) a change in resonance frequency, Δf, 2) the control signal of the automatic gain control unit as a measure of the tip-sample energy dissipation and 3) an average tunneling current (for conducting cantilevers and tips).

Applications

FM-AFM was introduced by *Albrecht* and coworkers in magnetic force microscopy [13.32]. The noise level and imaging speed were enhanced significantly compared to amplitude modulation techniques. Achieving atomic resolution on the Si(111)-(7×7) surface has been an important step in the development of the STM [13.39], and in 1994, this surface was imaged by AFM with true atomic resolution for the first time [13.33] (see Fig. 13.6).

The initial parameters that provided *true atomic resolution* (see caption of Fig. 13.6) were found empirically. Surprisingly, the amplitude necessary for obtaining good

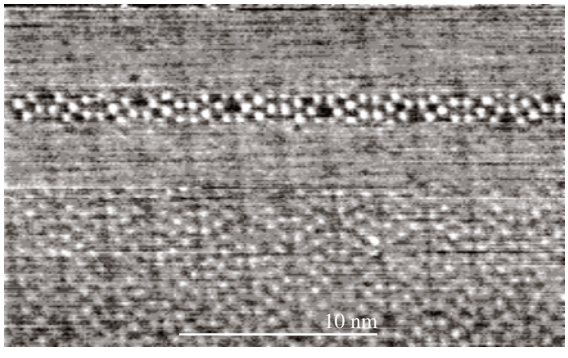


Fig. 13.6 First AFM image of the Si(111)-(7×7) surface. Parameters $k = 17\,\mathrm{Nm}$, $f_0 = 114\,\mathrm{kHz}$, $Q = 28{,}000$, $A = 34\,\mathrm{nm}$, $\Delta f = -70\,\mathrm{Hz}$, $V_t = 0\,\mathrm{V}$

results was very large compared to atomic dimensions. It turned out later that the amplitudes had to be that large in order to fulfill the stability criteria listed in Sect. 13.1.5. Cantilevers with $k \approx 2{,}000\,\mathrm{N/m}$ can be operated with amplitudes in the Å-range [13.24].

13.1.7 Relation Between Frequency Shift and Forces

The cantilever (spring constant k, effective mass m^*) is a macroscopic object and its motion can be described by classical mechanics. Figure 13.7 shows the deflection $q'(t)$ of the tip of the cantilever: It oscillates with an amplitude A at a distance $q(t)$ to a sample.

Generic Calculation

The Hamiltonian of the cantilever is:

$$H = \frac{p^2}{2m^*} + \frac{kq'^2}{2} + V_{\mathrm{ts}}(q)\,, \tag{13.10}$$

where $p = m^*\,\mathrm{d}q'/\mathrm{d}t$. The unperturbed motion is given by:

$$q'(t) = A\cos(2\square f_0 t)\,, \tag{13.11}$$

and the frequency is:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m^*}}\,. \tag{13.12}$$

If the force gradient $k_{\mathrm{ts}} = -\partial F_{\mathrm{ts}}/\partial z = \partial^2 V_{\mathrm{ts}}/\partial z^2$ is constant during the oscillation cycle, the calculation of the frequency shift is trivial:

$$\Delta f = \frac{f_0}{2k} k_{\mathrm{ts}}\,. \tag{13.13}$$

However, in classic FM-AFM, k_{ts} varies in orders of magnitude during one oscillation cycle, and a perturbation approach, as shown below, has to be employed for the calculation of the frequency shift.

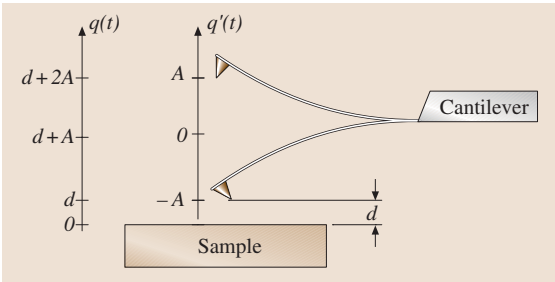


Fig. 13.7 Schematic view of an oscillating cantilever and definition of geometric terms

Hamilton–Jacobi Method

The first derivation of the frequency shift in FM-AFM was achieved in 1997 [13.38] using canonical perturbation theory [13.40]. The result of this calculation is:

$$\Delta f = -\frac{f_0}{kA^2}\left\langle F_{\mathrm{ts}} q' \right\rangle \tag{13.14}$$
$$= -\frac{f_0}{kA^2}\int_0^{1/f_0} F_{\mathrm{ts}}(d + A + q'(t))q'(t)\,\mathrm{d}t\,.$$

The applicability of first-order perturbation theory is justified, because in FM-AFM, E is typically in the range of several keVs, while V_{ts} is on the order of a few eVs. *Dürig* [13.41] has found a generalized algorithm that even allows the reconstruction of the tip-sample potential if not only the frequency shift, but the higher harmonics of the cantilever oscillation are known.

A Descriptive Expression for Frequency Shifts as a Function of the Tip–Sample Forces

With integration by parts, the complicated expression (13.14) transforms into a very simple expression that resembles (13.13) [13.42].

$$\Delta f = \frac{f_0}{2k}\int_{-A}^{A} k_{\mathrm{ts}}(z - q')\frac{\sqrt{A^2 - q'^2}}{\frac{\pi}{2}kA^2}\,\mathrm{d}q'\,. \tag{13.15}$$

This expression is closely related to (13.13): The constant k_{ts} is replaced by a weighted average, where the weight function $w(q', A)$ is a semicircle with radius A divided by the area of the semicircle $\pi A^2/2$ (see Fig. 13.8). For $A \to 0$, $w(q', A)$ is a representation of Dirac's delta function and the trivial zero-amplitude result of (13.13) is immediately recovered. The frequency shift results from a convolution between the tip-sample force gradient and the weight function. This convolution can easily be reversed with a linear transformation, and the tip-sample force can be recovered from the frequency shift versus distance curve [13.42].

The dependence of the frequency shift with amplitude confirms an empirical conjecture: Small amplitudes increase the sensitivity to short-range forces! Adjusting the amplitude in FM-AFM resembles tuning an optical spectrometer to a passing wavelength. When short-range interactions are to be probed, the amplitude should be in the range of the short-range forces. While using amplitudes in the Å-range has been elusive with conventional cantilevers, because of the instability problems described in Sect. 13.1.5, cantilevers with a stiffness on the order of 1,000 N/m like the ones introduced in [13.23] are well suited for small-amplitude operation.

13.1.8 Noise in Frequency-Modulation AFM – Generic Calculation

The vertical noise in FM-AFM is given by the ratio between the noise in the imaging signal and the slope of the imaging signal with respect to z:

$$\delta z = \frac{\delta \Delta f}{\left|\frac{\partial \Delta f}{\partial z}\right|} . \tag{13.16}$$

Figure 13.9 shows a typical frequency shift versus distance curve. Because the distance between the tip and sample is measured indirectly through the frequency shift, it is clearly evident from Fig. 13.9 that the noise in the frequency measurement $\delta\Delta f$ translates into vertical noise δz and is given by the ratio between $\delta\Delta f$ and the slope of the frequency shift curve $\Delta f(z)$ (13.16). Low vertical noise is obtained for a low-noise frequency measurement and a steep slope of the frequency shift curve.

The frequency noise $\delta\Delta f$ is typically inversely proportional to the cantilever amplitude A [13.32, 43]. The derivative of the frequency shift versus distance is constant for $A \ll \lambda$, where λ is the range of the tip-sample interaction and proportional to $A^{-1.5}$ for $A \gg \lambda$ [13.38]. Thus, minimal noise occurs if [13.44]

$$A_{\text{optimal}} \approx \lambda , \tag{13.17}$$

for chemical forces, $\lambda \approx 1$ Å. However, for stability reasons (Sect. 13.1.5), extremely stiff cantilevers are needed for small amplitude operation. The excellent noise performance of the stiff cantilever and small amplitude technique has been verified experimentally [13.24].

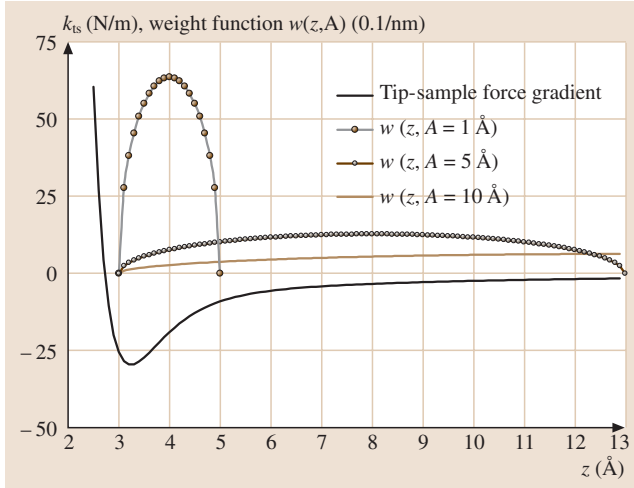


Fig. 13.8 Tip-sample force gradient k_{ts} and weight function for the calculation of the frequency shift

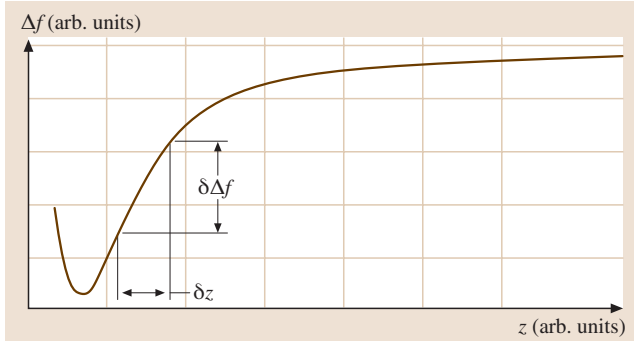


Fig. 13.9 Plot of frequency shift Δf as a function of tip-sample distance z. The noise in the tip-sample distance measurement is given by the noise of the frequency measurement $\delta\Delta f$ divided by the slope of the frequency shift curve

13.1.9 Conclusion

Dynamic force microscopy, in particular frequency-modulation atomic force microscopy, has matured into a viable technique that allows true atomic resolution of conducting and insulating surfaces and spectroscopic measurements on individual atoms [13.10, 45]. Even true atomic resolution in lateral force microscopy is now possible [13.46]. Challenges remain in the chemical composition and structural arrangement of the AFM tip.

13.2 Applications to Semiconductors

For the first time, corner holes and adatoms on Si(111)7×7 surface have been observed at a very local area by *Giessibl* using the pure noncontact AFM in ultrahigh vacuum (UHV) [13.33]. It became the breakthrough for the true atomic resolution imaging on a well-defined clean surface using noncontact AFM. Then, Si(111)7×7 [13.34, 35, 45, 47], InP(110) [13.48], and Si(100)2×1 [13.34] surfaces were successively

resolved with true atomic resolution. Furthermore, thermally induced motion of atoms or atomic-scale point defects on InP(110) surface have been observed at room temperature [13.48]. In this section, we will describe typical results of atomically resolved noncontact AFM imaging of semiconductor surfaces.

13.2.1 Si(111)7×7 Surface

Figure 13.10 shows the atomic resolution images of the Si(111)7×7 surface [13.49]. Here, Fig. 13.10a (TYPE I) was obtained using the Si tip without dangling, which is covered with an inert oxide layer. Fig. 13.10b (TYPE II) was obtained using the Si tip with dangling bond on which the Si atoms were deposited, due to the mechanical soft contact between the tip and the Si surface. The variable frequency shift mode was used. We can see not only adatoms and corner holes, but also missing adatoms described by the dimer-adatom-stacking fault (DAS) model. We can see that the image contrast in Fig. 13.10b is clearly stronger than that in Fig. 13.10a.

Interestingly, by using the Si tip with dangling bond, we observed contrast between inequivalent halves and between inequivalent adatoms of the 7×7 unit cell. Namely, as shown in Fig. 13.11a, the faulted halves surrounded with a solid line show brighter than the unfaulted halves surrounded with a broken line. Here, the positions of the faulted and unfaulted halves were decided from the step direction. From the cross-sectional profile along the long diagonal of the 7×7 unit cell in Fig. 13.11b, the heights of the corner adatoms are slightly higher than those of the adjacent center adatoms in the faulted and unfaulted halves of the unit cell. The measured corrugations are in the following decreasing order:

$$\text{Co-F} > \text{Ce-F} > \text{Co-U} > \text{Ce-U}\,,$$

where Co-F and Ce-F indicate the corner and center adatoms in faulted halves, and Co-U and Ce-U indicate the corner and center adatoms in unfaulted halves, respectively. As averaging over several units, the corrugation height differences are estimated to be 0.25 Å, 0.15 Å, and 0.05 Å for Co-F, Ce-F, and Co-U, respectively, referring to Ce-U. This tendency, that the heights of the corner adatoms are higher than those of the center adatoms, is consistent with the experimental results using a silicon tip [13.47], although they could not determine faulted and unfaulted halves of the unit cell in the measured AFM images. However, this tendency is completely contrary to the experimental results using a tungsten tip [13.35]. This difference may originate from the difference in the materials of the tip, which seems to affect the interaction between the tip and the reactive sample surface. Another possibility is that the tip is in contact with the surface during the small fraction of the oscillating cycle in their experiments [13.35].

We consider that contrast between inequivalent adatoms is not caused by tip artifacts for the following reasons: 1) Each adatom, corner hole and defect were clearly observed; 2) the apparent heights of adatoms are the same whether they are located adjacent to defects or not; 3) the same contrast in several images for the different tips has been observed.

It should be noted that the corrugation amplitude of adatoms ∼ 1.4 Å in Fig. 13.11b is higher than that of 0.8–1.0 Å obtained with the STM, although the depth of corner holes obtained with noncontact AFM is almost the same as that observed with STM. Besides, in noncontact mode AFM images, the corrugation ampli-

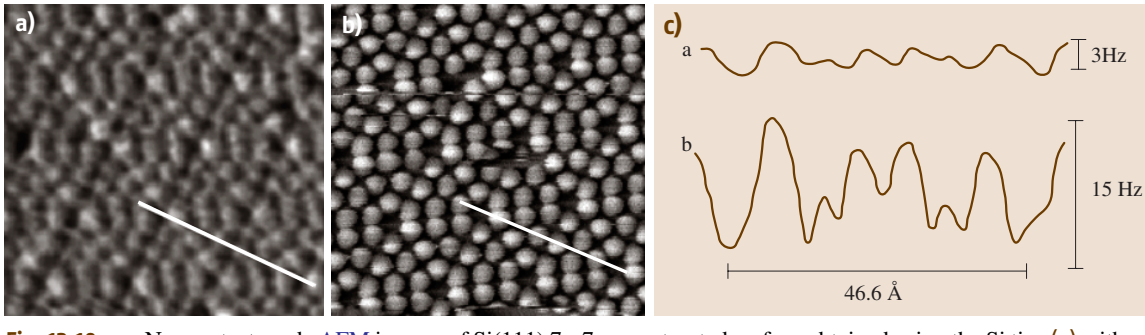


Fig. 13.10a–c Noncontact mode AFM images of Si(111) 7×7 reconstructed surface obtained using the Si tips **(a)** without and **(b)** with dangling bond. The scan area is 99 Å×99 Å. **(c)** The cross sectional profiles along long diagonal of the 7×7 unit cell indicated by *white lines* in **(a)** and **(b)**

tude of adatoms was frequently larger than the depth of the corner holes. The origin of such a large corrugation of adatoms may be due to the effect of the chemical interaction, but that is yet unclear.

The atom positions, surface energies, dynamic properties, and chemical reactivities on the Si(111)7×7 reconstructed surface have been extensively investigated theoretically and experimentally. From these investigations, the possible origins of the contrast between inequivalent adatoms in AFM images are the following: the true atomic heights that correspond to the adatom core positions, the stiffness (spring constant) of inter-atomic bonding with the adatoms that corresponds to the frequencies of the surface mode, the amount of charge of adatom, and the chemical reactivity of adatoms. Table 13.1 summarizes the decreasing orders of the inequivalent adatoms for individual property. From Table 13.1, we can see that the calculated adatom heights and the stiffness of inter-atomic bonding cannot explain the AFM data, while the amount of charge of adatom and the chemical reactivity of adatoms can explain our data. The contrast due to the amount of charge of adatom means that the AFM image originated from the difference of the vdW, or the electrostatic physical interactions between the tip and the valence electrons at the adatoms. The contrast due to the chemical reactivity of adatoms means that the AFM image originated from the difference of covalent bonding chemical interaction between the atoms at the tip apex and the dangling bond of adatoms. Thus, we can see there are two possible interactions that explain the strong contrast between inequivalent adatoms of the 7×7 unit cell observed using the Si tip with dangling bond.

A weak contrast image in Fig. 13.10a is due to vdW and/or electrostatic force interactions. On the other hand, strong contrast images in Figs. 13.10b and 13.11a are due to a covalent bonding formation between the AFM tip with Si atoms and Si adatoms. These results signify the capability of the noncontact mode AFM to image the variation in chemical reactivity of Si adatoms. In the future, by controlling an atomic species at the tip apex, the study of chemical reactivity on an atomic scale will be possible using the noncontact AFM.

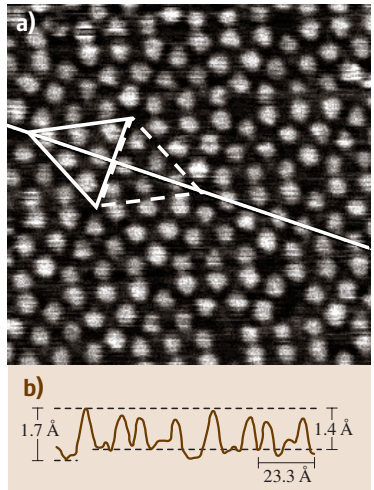


Fig. 13.11 **(a)** Noncontact mode AFM image with contrast of unequivalent adatoms and **(b)** a cross-sectional profile indicated by *white line*. The halves of the 7×7 unit cell surrounded with *solid line* and *broken line* correspond to faulted and unfaulted halves, respectively. The scan area is 89 Å×89 Å

13.2.2 Si(100)2×1 and Si(100)2×1:H Monohydride Surfaces

In order to investigate the imaging mechanism of the noncontact AFM, a comparative study between a reactive surface and an insensitive surface using the same tip is very useful. Si(100)2×1:H monohydride surface is a Si(100)2×1 reconstructed surface that is terminated by a hydrogen atom. It does not reconstruct as metal is deposited on the semiconductor surface. The surface structure hardly changes. Thus, Si(100)2×1:H monohydride surface is one of most useful surfaces for a model system to investigate the imaging mechanism, experimentally and theoretically. Furthermore, whether the interaction between a very small atom such as hy-

Table 13.1 Comparison between the adatom heights observed in an AFM image and the variety of properties for inequivalent adatoms

	Decreasing order	Agreement
AFM image	Co−F > Ce−F > Co−U > Ce−U	–
Calculated height	Co−F > Co−U > Ce−F > Ce−U	×
Stiffness of inter-atomic bonding	Ce−U > Co−U > Ce−F > Co−F	×
Amount of charge of adatom	Co−F > Ce−F > Co−U > Ce−U	○
Calculated chemical reactivity	Faulted > Unfaulted	○
Experimental chemical reactivity	Co−F > Ce−F > Co−U > Ce−U	○

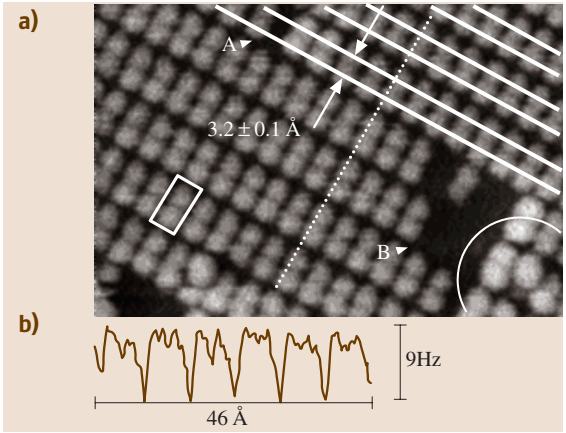


Fig. 13.12 **(a)** Noncontact AFM image of a Si(001)2×1 reconstructed surface. The scan area was 69×46 Å. One 2×1 unit cell is outlined with a *box*. *White rows* are superimposed to show the bright spots arrangement. The distance between bright spots on dimer row is 3.2 ± 0.1 Å. Into *white arc*, the alternative bright spots are shown. **(b)** Cross-sectional profile indicated by a *white dotted line*

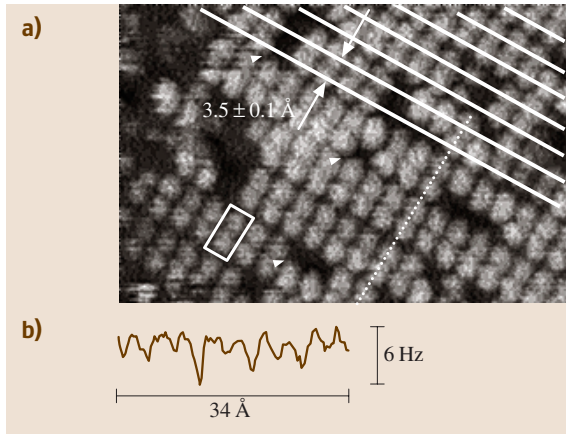


Fig. 13.13 **(a)** Noncontact AFM image of Si(001)2×1-H surface. The scan area was 69×46 Å. One 2×1 unit cell is outlined with a *box*. *White rows* are superimposed to show the bright spots arrangement. The distance between bright spots on dimer row is 3.5 ± 0.1 Å. **(b)** Cross-sectional profile indicated by a *white dotted line*

drogen and a tip apex is observable with noncontact AFM is interesting. Here, we show noncontact AFM images measured on a Si(100)2×1 reconstructed surface with a dangling bond and a Si(100)2×1:H monohydride surface with a dangling bond that is terminated by a hydrogen atom [13.50].

Figure 13.12a shows the atomic resolution image of a Si(100)2×1 reconstructed surface. The paired bright spots arranged like rows with a 2×1 symmetry was observed with clear contrast. Missing pairs of bright spots were observed, as indicated by arrows. Further, pairs of bright spots are shown by the white dashed arc and appear to be the stabilize-buckled asymmetric dimer structure. The distance between paired bright spots is 3.2 ± 0.1 Å.

Figure 13.13a shows the atomic resolution image of a Si(100)2×1:H monohydride surface. Paired bright spots arranged like rows were observed. Missing paired bright spots, as well as those paired in rows and single bright spots were observed, as indicated by arrows. The distance between paired bright spots is 3.5 ± 0.1 Å. This distance is 0.2 Å larger than that of a Si(100)2×1 reconstructed surface. It was found that the distance between bright spots increases in size due to the hydrogen termination.

The bright spots in Fig. 13.12 do not merely image the silicon atom site, because the distance between bright spots forming the dimer structure of Fig. 13.12a (3.2 ± 0.1 Å) is larger than the distance between silicon atoms of every dimer structure model (2.9 Å is the maximum distance between upper silicons on asymmetric dimer structures). This seems to be due to the contribution to imaging of the chemical bonding interaction between the dangling bond out of the silicon tip apex and the dangling bond on the Si(100)2×1 reconstructed surface. The chemical bonding interaction works with strong direction dependence between the dangling bond out of the silicon dimer structure on the Si(100)2×1 reconstructed surface and the dangling bond out of the silicon tip apex, and the dimer structure is obtained with a larger distance than between silicons on the surface.

The bright spots of Fig. 13.13 seem to be located at hydrogen atom sites on a Si(100)2×1:H monohydride surface, because the distance between bright spots forming the dimer structures (3.5 ± 0.1 Å) agrees approximately with the distance between the hydrogen, i. e., 3.52 Å. Thus, noncontact AFM atomically resolves the individual hydrogen atoms on the topmost layer. On this surface, the dangling bond is terminated by a hydrogen atom, and the hydrogen atom on the topmost layer does not have a chemical reactivity. Therefore, the interaction between the hydrogen atom on the top most layer and the silicon tip apex does not contribute to the chemical bonding interaction with strong direction de-

pendence as in a silicon surface, and the bright spots of a noncontact AFM image correspond to the hydrogen atom sites on the topmost layer.

13.2.3 Metal-Deposited Si Surface

In this section, we will introduce the comparative study of force interaction between a Si tip and a metal deposited Si surface and between a metal adsorbed Si tip and a metal-deposited Si surface [13.51, 52]. As for the metal-deposited Si surface, a Si(111)$\sqrt{3}\times\sqrt{3}$-Ag (hereafter referred to as $\sqrt{3}$-Ag) surface was used.

For the $\sqrt{3}$-Ag surface, the honeycomb-chained trimer (HCT) model has been accepted as the appropriate model. As shown in Fig. 13.14, this structure contains Si trimer at the second layer, 0.75 Å below the Ag trimer at the topmost layer. The topmost Ag atoms and the lower Si atoms form covalent bonds. The inter-atomic distances between the nearest neighbor Ag atoms forming Ag trimer and between lower Si atoms forming Si trimer are 3.43 Å and 2.31 Å, respectively. The apexes of Si trimers and Ag trimers face the $[11\bar{2}]$ direction and the direction tilted a little to the $[\bar{1}\bar{1}2]$ direction, respectively.

In Fig. 13.15, we show the noncontact AFM images measured using a normal Si tip at the frequency shift of (a) −37 Hz, (b) −43 Hz, and (c) −51 Hz, respectively. These frequency shifts correspond to the tip-sample distance of about 0–3 Å. We defined the zero position of tip-sample distance, i. e., the contact point, where the vibration amplitude began to decrease. The rhombus indicates the $\sqrt{3}\times\sqrt{3}$ unit cell. When the tip approached the surface, the contrast of the noncontact AFM images became strong and the pattern remarkably changed. That is, by moving the tip toward the sample surface, the hexagonal pattern, the trefoil-like pattern composed of three dark lines, and the triangle pattern can be observed sequentially. In Fig. 13.15a, the distance between the bright spots is 3.9 ± 0.2 Å. In Fig. 13.15c, the distance

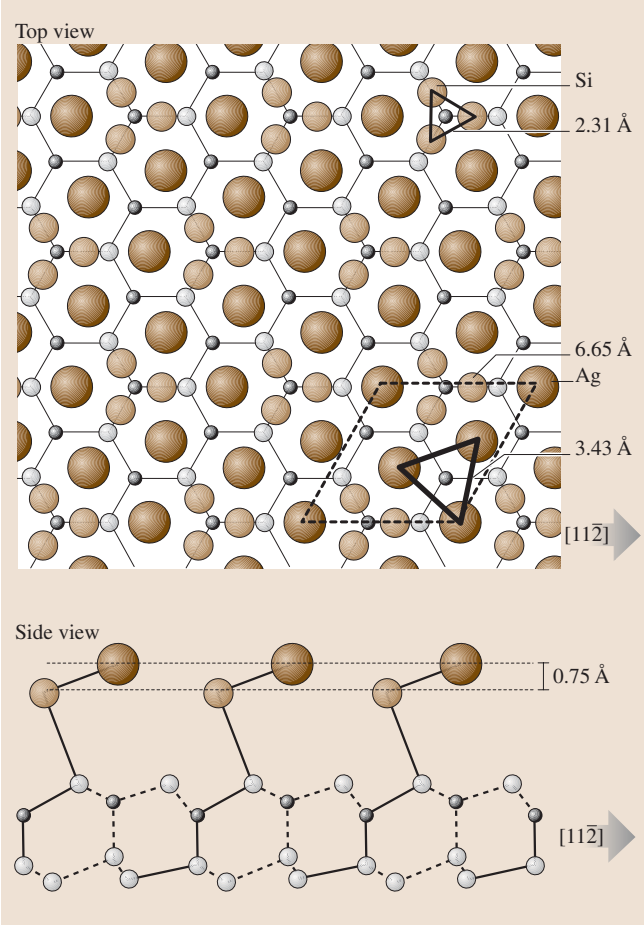


Fig. 13.14 HCT model for the structure of the Si(111)$\sqrt{3}\times\sqrt{3}$-Ag surface. *Black closed circle*, *gray closed circle*, *open circle*, and *closed circle* with *horizontal line* indicate Ag atom at the topmost layer, Si atom at the second layer, Si atom at the third layer, and Si atom at the fourth layer, respectively. *Rhombus* indicates $\sqrt{3}\times\sqrt{3}$ unit cell. *Thick*, *large solid triangle* indicates an Ag trimer. *Thin*, *small, solid triangle* indicates a Si trimer

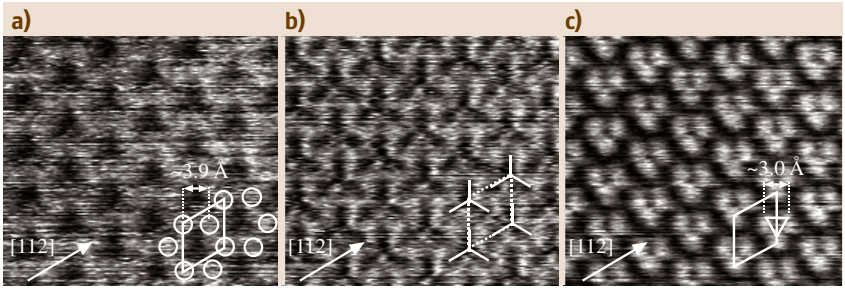


Fig. 13.15a–c Noncontact AFM images obtained at the frequency shifts of **(a)** −37 Hz, **(b)** −43 Hz, and **(c)** −51 Hz on a Si(111)$\sqrt{3}\times\sqrt{3}$-Ag surface. This distance dependence was obtained with the Si tip. Scan area is 38 Å × 34 Å. A *rhombus* indicates $\sqrt{3}\times\sqrt{3}$ unit cell

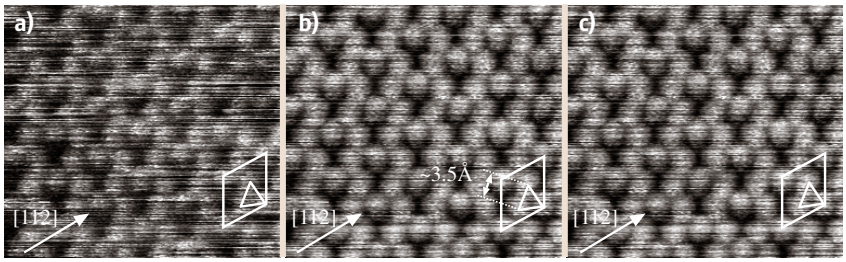


Fig. 13.16a–c Noncontact AFM images obtained at the frequency shifts of **(a)** −4.4 Hz, **(b)** −6.9 Hz, and **(c)** −9.4 Hz on a Si(111)$\sqrt{3}\times\sqrt{3}$-Ag surface. This distance dependence was obtained with the Ag adsorbed tip. Scan area is 38 Å × 34 Å

between the bright spots is 3.0 ± 0.2 Å, and the direction of the apex of all triangles composed of three bright spots is $[11\overline{2}]$.

In Fig. 13.16, we show the noncontact AFM images measured by using an Ag adsorbed tip at the frequency shift of (a) −4.4 Hz, (b) −6.9 Hz, and (c) −9.4 Hz, respectively. The tip-sample distances Z are roughly estimated to be $Z = 1.9$ Å, 0.6 Å, and ∼ 0 Å (in noncontact region), respectively. When the tip approached the surface, the pattern of the noncontact AFM images didn't change, although the contrasts became clearer. A triangle pattern can be observed. The distance between the bright spots is 3.5 ± 0.2 Å. The direction of the apex of all triangles composed of three bright spots is tilted a little from the $[\overline{11}2]$ direction.

Thus, noncontact AFM images measured on a Si(111)$\sqrt{3}\times\sqrt{3}$-Ag surface showed two types of distance dependence of the image patterns, depending on the atom species on the tip apex.

By using the normal Si tip with dangling bond, as in Fig. 13.15a, the measured distance between the bright spot of 3.9 ± 0.2 Å agrees with the distance of 3.84 Å between the center of the Ag trimer in the HCT model within the experimental error. Furthermore, the hexagonal pattern composed of six bright spots also agrees with the honeycomb structure of the Ag trimer in the HCT model. So the most appropriate site corresponding to the bright spots in Fig. 13.15a is the site of the center of the Ag trimers. In Fig. 13.15c, the measured distance of 3.0 ± 0.2 Å between the bright spots forming the triangle pattern does not agree with either the distance between the Si trimer of 2.31 Å or the distance between the Ag trimer of 3.43 Å in the HCT model, while the direction of the apex of triangles composed of three bright spots agrees with the $[11\overline{2}]$ direction of the apex of the Si trimer in the HCT model. So the most appropriate site corresponding to the bright spots in Fig. 13.15c is the intermediate site between Si atoms and Ag atoms. On the other hand, by using the Ag adsorbed tip, the measured distance of 3.5 ± 0.2 Å between the bright spots in Fig. 13.16 agrees with the distance of 3.43 Å between the nearest neighbor Ag atoms forming the Ag trimer at the topmost layer in the HCT model within the experimental error. In addition, the direction of the apex of the triangles composed of three bright spots also agrees with the direction of the apex of the Ag trimer, i. e., tilted $[\overline{11}2]$, in the HCT model. So the most appropriate site corresponding to the bright spots in Fig. 13.16 is the site of individual Ag atoms forming the Ag trimer at the topmost layer.

It should be noted that by using the noncontact AFM with an Ag adsorbed tip the individual Ag atom on the $\sqrt{3}$-Ag surface could be resolved in real space for the first time, although it could not be resolved by using the noncontact AFM with a Si tip. So far, the $\sqrt{3}$-Ag surface has been observed by the scanning tunneling microscope (STM) with atomic resolution. However, the STM can measure the local charge density of state near Fermi level on the surface. From the first principle calculation, it was proven that the unoccupied surface state is densely distributed around the center of the Ag trimer. As a result, bright contrast is obtained at the center of the Ag trimer with the STM.

Finally, we consider the origin of the atomic resolution imaging of the individual Ag atoms on the $\sqrt{3}$-Ag surface. Here, we discuss the difference of the force interactions between the Si and the Ag adsorbed tips. As shown in Fig. 13.17a, by using the Si tip, there is the dangling bond out of the topmost Si atom on the Si tip apex. As a result, the force interaction is dominated by the physical bonding interaction such as Coulomb force far from the surface and by chemical bonding interaction very close to the surface. In other words, if the reactive Si tip with dangling bond approaches the surface, at the distance far from the surface, the Coulomb force acts between the electron localized at the dangling bond out of the topmost Si atom on the tip apex and the positive charge distributed around the center of the Ag trimer. At the distance very close to the surface, the chemical bonding interaction will occur due to the onset of the or-

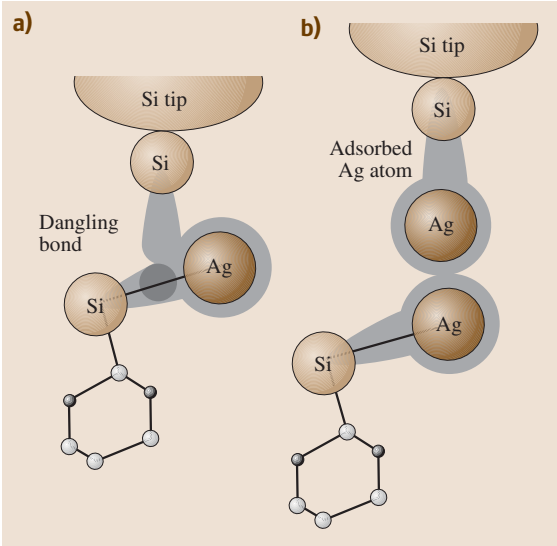


Fig. 13.17a,b Schematic illustration of **(a)** the Si atom with dangling bond and **(b)** the Ag adsorbed tip above the Si-Ag covalent bond on Si(111)$\sqrt{3}\times\sqrt{3}$-Ag surface

bital hybridization between the dangling bond out of the topmost Si atom on the Si tip apex and the Si-Ag covalent bond on the surface. Hence, the individual Ag atoms will not be resolved, the image pattern will change depending on the tip-sample distance. On the other hand, as shown in Fig. 13.17b, by using an Ag adsorbed tip, the dangling bond localized out of the topmost Si atom on the Si tip apex is terminated by the adsorbed Ag atom. As a result, even at very close tip-sample distance, the force interaction is dominated by the physical bonding interaction such as the vdW force interaction. If the Ag adsorbed tip approaches the surface, the vdW force acts between the Ag atom on the tip apex and Ag or Si atom on the surface. Ag atoms at the topmost layer on the $\sqrt{3}$-Ag surface are located higher than the Si atoms at the lower layer. Hence, the individual Ag atoms (or nearly true topography) will be resolved, and the image pattern will not change even at very close tip-sample distance. It should be emphasized that there is a possibility to identify or recognize atom species on a sample surface using the noncontact AFM if we can control an atomic species at the tip apex.

13.3 Applications to Insulators

Insulators such as alkali halides, fluorides, and metal oxides are key materials in many applications, including optics, microelectronics, catalysis, and so on. Surface properties are important in these technologies, but they are usually poorly understood. This is due to their low conductivity, which makes it difficult to investigate them using electron- and ion-based measurement techniques such as low-energy electron diffraction, ion scattering spectroscopy, and scanning tunneling microscopy (STM). Surface imaging by noncontact atomic force microscopy (NC-AFM) does not require the sample to exhibit high conductivity, because NC-AFM detects a force between the tip on the cantilever and the surface of the sample. Since the first report of atomically resolved NC-AFM on a Si(111) 7×7 surface [13.33], several groups have succeeded in obtaining "true" atomic resolution images of insulators, including defects, and it has been shown that NC-AFM is a powerful new tool for the atomic-scale surface investigation of insulators.

In this section, we will describe typical results of atomically resolved NC-AFM imaging of insulators such as alkali halides, fluorides, and metal oxides. For alkali halides and fluorides, we will focus on the contrast formation, which is the most important issue for interpreting atomically resolved images of binary compounds on the basis of experimental and theoretical results. For the metal oxides, typical examples of atomically resolved imaging will be shown. Also, simultaneous imaging using STM and NC-AFM will be described, and the difference between the STM and NC-AFM images will be demonstrated. Finally, we will describe the results obtained when we imaged an antiferromagnetic NiO(001) surface with a ferromagnetic Fe-coated tip in order to explore the possibility of detecting short-range magnetic interaction using the NC-AFM.

13.3.1 Alkali Halides, Fluorides, and Metal Oxides

The surfaces of alkali halides were the first insulating materials to be imaged by NC-AFM with true atomic resolution [13.53]. Up until now, there have been reports on atomically resolved images of (001) cleaved surfaces for single crystal NaF, RbBr, LiF, KI, NaCl [13.54], KBr [13.55], and thin films of NaCl(001) on Cu(111) [13.56]. In this section, we describe the contrast formation of alkali halides surface on the basis of experimental and theoretical results.

Alkali Halides

In experiments of alkali halides, the symmetry of the observed topographic images indicates that the protrusions exhibit only one type of ions, either positively or negatively charged ions. This leads to the conclusion that the atomic contrast is dominantly caused by electrostatic interactions between a charged atom at the tip apex and the surface ions, i. e., the long-range forces between the macroscopic tip and the sample such as van der Waals force is modulated by an alternating short-range electrostatic interaction of the surface ions. Theoretical work employing the atomistic simulation technique has revealed the mechanism for contrast formation on an ionic surface [13.57]. A significant part of the contrast is due to the displacement of ions in the force field, not only enhancing the atomic corrugations, but also contributing to the electrostatic potential by forming dipoles at the surface. The experimentally observed atomic corrugation height is determined by the interplay of the long- and short-range forces. In the case of NaCl, it has been experimentally demonstrated that a more blunt tip produces a larger corrugation when the tip-sample distance is shorter [13.54]. This result shows that the increased long-range forces induced by a blunt tip allow for more stable imaging closer to the surface. The stronger electrostatic short-range interaction and the larger ion displacement produce a more pronounced atomic corrugation.

At steps and kinks of an NaCl thin film on Cu(111), the amplitude of the atomic corrugations has been observed to increase by up to a factor of two, as large as that of atomically flat terraces [13.56]. This increase in the contrast is due to the low coordination of the step-edge and the kinked ions. The low coordination of the ions results in an enhancement of the electrostatic potential over the site and an increase in the displacement induced by the interaction with the tip. These results demonstrate that the short-range interaction dominates the atomic contrast and leads to a considerable displacement of the surface ions from the ideal position.

Theoretical study predicts that the image contrast depends on the chemical species of the tip apex. *Bennewitz* et al. [13.56] have performed the calculations using an MgO tip terminated by oxygen, an Mg ion, and an OH. The magnitude of the atomic contrast for the Mg-terminated tip shows a slight increase in comparison with an oxygen-terminated tip. The atomic contrast with the oxygen-terminated tip is dominated by the attractive electrostatic interaction between the oxygen on the tip apex and the Na ion, but the Mg-terminated tip attractively interacts with the Cl ion. Also, it has been demonstrated that the magnitude of the contrast for an OH-terminated tip is almost an order of magnitude less than for the oxygen- and Mg-terminated tips.

In order to explore the imaging mechanism for the alkali halides, the heterogeneous surface of NaCl crystal containing $(OCN)^-$ molecules as impurities has been observed by NC-AFM [13.58]. In this case, the atomic periodicity of the NaCl lattice is imaged. However, high-resolution investigations of impurities are limited by the convolution between the tip apex and the molecules at the surface. *Bennewitz* et al. [13.59] reported on the heterogeneous surfaces of alkali halide crystals. In this experiment, which was designed to observe the chemically nonhomogeneous alkali halide surface, a mixed crystal composing of 60% KCl and 40% KBr was used as the sample. In this crystal, which is a so-called solid solution, the Cl and Br ions are mixed randomly. The image of the cleaved $KCl_{0.6}Br_{0.4}(001)$ surface indicates that only one type of ion is imaged as protrusions as if it were a pure alkali halide crystal. However, there is a significant difference in comparison with the surface of a pure crystal. The amplitude of atomic corrugation varies strongly between the positions of the ions imaged as depressions. This variation in the corrugations corresponds to the constituents of the crystal, i. e., the Cl and Br ions, and it is concluded that the tip apex is negatively charged. Also, it is demonstrated that a comparison between the distribution of the deep depressions and the composition of the mixed crystal assigns the deep depressions to Br ions. Although an evaluation of the effect of the surface buckling induced by a random distribution of anions is needed in order to understand the contrast mechanism, this result demonstrates that NC-AFM measurement exhibits chemical sensitivity on an atomic scale. For applications to alkali halide surfaces, real space imaging with NC-AFM presents the possibility for chemical discrimination at the surface.

Fluorides

Fluorides are important materials for the progress of atomic-scale resolution NC-AFM imaging on insulators. There are reports in the literature of surface images for single crystal BaF_2, SrF_2 [13.60], and CaF_2 [13.61–63]. Surfaces of fluorite-type crystals are prepared by cleaving along the (111) planes. Their structure is more complex than the structure of alkali halides with the NaCl structure. The complexity is of great interest for atomic-resolution imaging using an NC-AFM and also for theoretical predictions of the interpretation of atomic-scale contrast information.

The first atomically resolved images of a $CaF_2(111)$ surface have been obtained in topographic mode [13.62] in which the tip-sample distance is regulated by maintaining a constant frequency shift. In this imaging mode, the surface ions mostly appear as spherical caps. Similar measurements have been performed on the (111) surfaces of SrF_2 and BaF_2.

Barth et al. [13.64] have found that the $CaF_2(111)$ surface images obtained using the constant height mode, in which the frequency shift is recorded with a very low loop gain, can be categorized into two contrast patterns. In the first, the ions appear as triangles, and in the other, they have the appearance of circles, similar to the contrast obtained in a topographic image. These differences are revealed by comparing the cross sections along the [121] direction. Theoretical studies demonstrated that these two different contrast patterns could be explained as a result of imaging with tips of different polarity [13.64–66]. When imaging with a positively charged (cation-terminated) tip, the triangular pattern appears. In this case, the contrast is dominated by the strong short-range electrostatic attraction between the positive tip and the negative F ions. The cross section along the [121] direction of the triangular image shows two maxima: one is a larger peak over the F ions located in the topmost layer, and the other is a smaller peak at the position of the F ions in the third layer. The minima appear at the position of the Ca ions in the second layer. When imaging with a negatively charged (anion-terminated) tip, the spherical image contrast appears, and the main periodicity is created by the Ca ions in the second layer between the topmost and the third F ion layers. In the cross section along the [121] direction, the large maxima correspond to the Ca sites, because of the strong attraction of the negative tip, and the minima appear at the sites of maximum repulsion over the F ions in the first layer. At a position between two F ions, there are smaller maxima. This reflects the weaker repulsion over the third layer F ion sites compared to the protruding topmost F ion sites and a slower decay in the contrast on one side of the Ca ions. These results demonstrate that the image contrast of a $CaF_2(111)$ surface depends on the state of the tip apex.

The triangular pattern obtained with a positively charged tip appears at relatively large tip-sample distance, as shown in Fig. 13.18a. On the cross section along the [121] direction, experimental results and theoretical studies both demonstrate the large peak and small shoulder characteristic for the triangular pattern image (Fig. 13.18d). When the tip approaches the surface more closely, the triangular pattern of the experimental images is more vivid (Fig. 13.18b), just as predicted in the theoretical works. On the cross section along the [121] direction, as the tip approaches, the amplitude of the shoulder increases until it is equal to that of the main peak, and this feature gives rise to the honeycomb pattern image, as shown Fig. 13.18c. These results demonstrate that there is a more complex process of image contrast formation. Also, from detailed theoretical analysis of the electrostatic potential [13.67], it is suggested that the change in displacement of the ions due to the proximity of the tip plays an important role in the formation of the image contrast. Such a drastic change in image contrast, depending on both the polarity of the terminated tip atom and the tip-sample distance, is inherent in

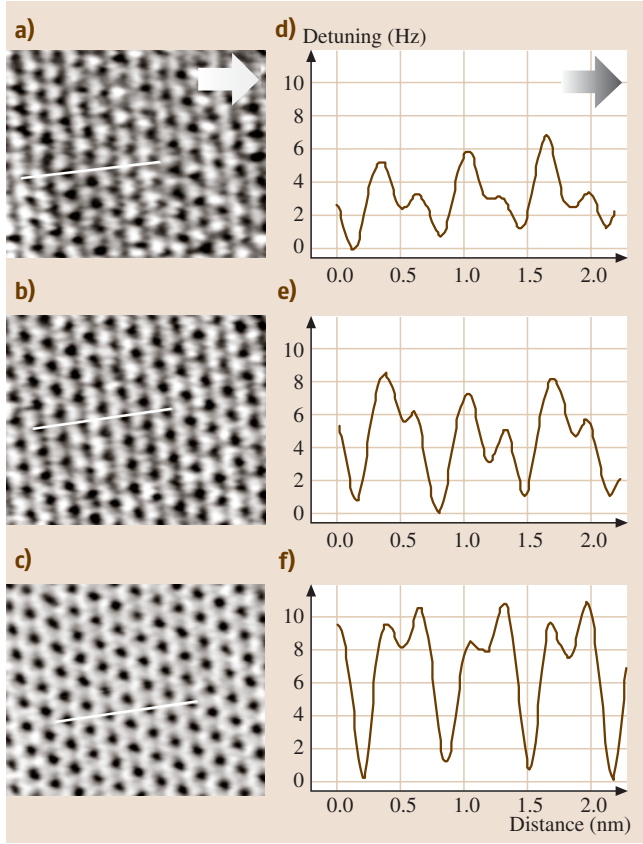


Fig. 13.18 **(a)–(c)** $CaF_2(111)$ surface images obtained by using the constant height mode. From **(a)** to **(c)** the frequency shift was lowered. The *white lines* represent the positions of the cross section. **(d)–(f)** The cross section extracted from the Fourier-filtered images of **(a)–(c)**. The *white* and *black arrows* represent the scanning direction. The images and the cross sections are from [13.64]

the fluoride (111) surface, and this image contrast feature cannot be seen on the (001) surface of alkali halides with a simple crystal structure.

These results have demonstrated that there is good agreement between experiment and theory. However, the results of careful experiments show another feature: The cross sections in the forward and backward scan directions do not yield identical results [13.64]. Also, it is shown that the structures of cross sections taken along the three equivalent [121] directions are different. It is thought that this can be attributed to the asymmetry of the nano-cluster at the tip apex, which leads to different interactions in the equivalent directions. A better understanding of the asymmetric image contrast may require a more complicated modeling of the tip structure. However, the most desirable solution would be the development of the suitable techniques for well-defined tip preparation. In fact, it should be mentioned that perfect tips on an atomic scale can occasionally be obtained. These tips do yield identical results in forward and backward scanning, and cross sections in the three equivalent directions taken with this tip are almost identical [13.68].

The fluoride (111) surface is an excellent standard surface for calibrating tips on an atomic scale. The polarity of the tip-terminated atom can be determined from the image contrast pattern (spherical or triangular pattern). The irregularities in the tip structure can be detected, since the surface structure is highly symmetric. Therefore, once such a tip has been prepared, it can be used as a calibrated tip for imaging unknown surfaces.

The states of the tip apex play an important role in interpreting NC-AFM images of alkali halide and fluoride surfaces. It is expected that the achievement of a good correlation between experimental and theoretical studies will help advance surface imaging for insulators by NC-AFM.

Metal Oxides

Most of the metal oxides that have attracted strong interest for their technological importance are insulative. Therefore, in the case of atomically resolved imaging of metal oxide surfaces by STM, efforts to increase the conductivity of the sample are needed, for example, the introduction of defects of anions or cations, doping with other atoms, and surface observations during heating of the sample. However, in principle, NC-AFM provides the possibility of observing nonconductive metal oxides without these efforts. In cases where the conductivity of the metal oxides is high enough for a tunneling current to flow, it should be noted that most of the surface images obtained by NC-AFM and STM are not identical.

Since the first report of atomically resolved images on a $TiO_2(110)$ surface with oxygen point defects [13.69], they have also been reported on $TiO_2(100)$ [13.70–72], $SnO_2(110)$ [13.73], NiO(001) [13.74, 75], $SrTiO_3(100)$ [13.76], and $CeO_2(111)$ [13.77] surfaces. Also, *Barth* et al. [13.78] have succeeded in obtaining atomically resolved NC-AFM images of a clean α-$Al_2O_3(0001)$ surface, which is impossible to investigate using STM. In this section, we describe typical results of the imaging of metal oxides by NC-AFM.

The α-$Al_2O_3(0001)$ surface exists in several ordered phases that can be reversibly transformed into each other by thermal treatments and oxygen exposure. It is known that the high temperature phase has a large $(\sqrt{31}\times\sqrt{31})R\pm 9°$ unit cell. However, the details of the atomic structure of this surface have not been revealed, and two models are proposed from the results of low-energy electron diffraction and X-ray diffraction studies. Recently, *Barth* et al. [13.78] have directly observed this reconstructed α-$Al_2O_3(0001)$ surface by NC-AFM. They confirmed that the dominant contrast of the low magnification image corresponds to a rhombic grid representing a unit cell of $(\sqrt{31}\times\sqrt{31})R+9°$, as shown in Fig. 13.19a. Also, more details of atomic structures are determined from the higher magnification image (Fig. 13.19b), which was taken at a reduced tip-sample distance. In this atomically resolved image, it is revealed that each side of the rhombus is intersected by 10 atomic rows, and that a hexagonal arrangement of atoms exists in the center of the rhombi (Fig. 13.19c). This feature agrees with the proposed surface structure that predicts order in the center of hexagonal surface domains and disorder at the domain boundaries. Their result is an excellent demonstration of the capabilities of the NC-AFM as a powerful tool for the atomic-scale surface investigation of insulators.

In addition, *Barth* et al. [13.78] have performed observations of the $Al_2O_3(0001)$ surface when it is exposed to water and hydrogen. It is suggested that the region of atomic disorder has an extremely high reactivity. Such investigations of the surface structure of adsorbed insulator substrates can only be provided by NC-AFM. The NC-AFM imaging of metal growth on oxides [13.79] has also been reported. A description of the adsorbed molecular arrangement on a TiO_2 surface can be found in the next section.

The atomic structure of $SrTiO_3(100)$-$(\sqrt{5}\times\sqrt{5})R26.6°$ surface, as well as that of $Al_2O_3(0001)$-$(\sqrt{31}\times\sqrt{31})R\pm 9°$ can be determined on the basis of

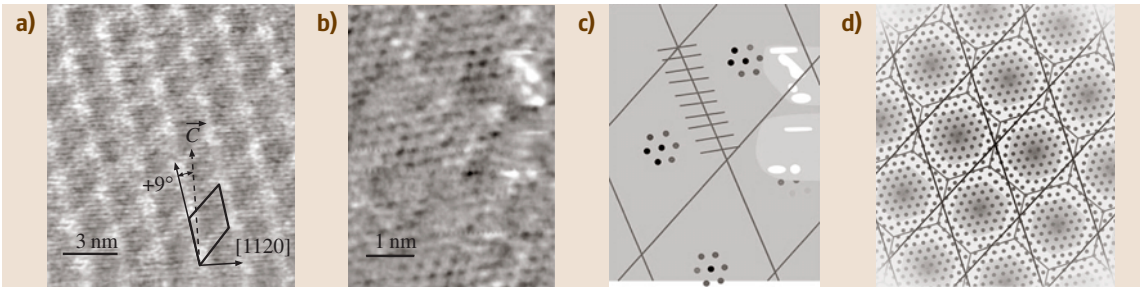


Fig. 13.19 **(a)** Image of the high temperature, reconstructed clean α-Al_2O_3 surface obtained by using the constant height mode. The *rhombus* represents the unit cell of the $(\sqrt{31}\times\sqrt{31})R+9°$ reconstructed surface. **(b)** Higher magnification image of **(a)**. Imaging was performed at a reduced tip-sample distance. **(c)** Schematic representation of the indicating regions of hexagonal order in the center of reconstructed rhombi. **(d)** Superposition of the hexagonal domain with reconstruction rhombi found by NC-AFM imaging. Atoms in *gray shaded regions* are well-ordered. The images and the schematic representations are from [13.78]

the results of NC-AFM imaging [13.76]. $SrTiO_3$ is one of the perovskite oxides and its (100) surface exhibits the many different kinds of reconstructed structures. In the case of the $(\sqrt{5}\times\sqrt{5})R26.6°$ reconstruction, the oxygen vacancy-Ti^{3+}-oxygen model (where the terminated surface is TiO_2 and the observed spots are related to oxygen vacancies) was proposed from the results of STM imaging. As shown in Fig. 13.20, *Kubo* et al. [13.76] have performed measurements using both STM and NC-AFM and have found that the size of the bright spots, as observed by NC-AFM, is always smaller than that for STM measurement and that the dark spots, which are not observed by STM, are arranged along the [001] and [010] directions in the NC-AFM image. A theoretical simulation of the NC-AFM image using first-principles calculations shows that the bright and dark spots correspond to Sr and oxygen atoms, respectively. It has been proposed that the structural model of the reconstructed surface consists of an ordered Sr adatom located at the oxygen fourfold site on the TiO_2-terminated layer (Fig. 13.20c).

Because STM images are related to the spatial distribution of the wave functions near the Fermi level, atoms, which do not have a local density of states near the Fermi level, are generally invisible even on the conductive materials. On the other hand, the NC-AFM image reflects the strength of the tip-sample interaction force originating from chemical, electrostatic interactions, and so on. Therefore, even STM and NC-AFM images obtained using an identical tip and sample will generally differ.

Simultaneous imaging of a $TiO_2(110)$ surface with STM and NC-AFM shows a typical result that differs in terms of the atoms that are visualized [13.72]. For simultaneous imaging, the tip-sample distance is controlled by keeping a constant frequency shift, and the tunneling current between the oscillating cantilever and the surface is measured. The two images show different atomic contrasts, i. e., the atomic corrugation of the

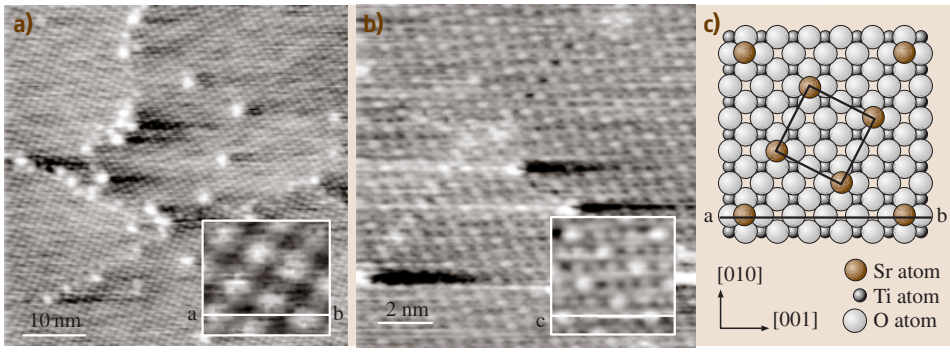


Fig. 13.20 **(a)** STM and **(b)** NC-AFM images of a $SrTiO_3(100)$ surface. **(c)** A proposed model of the $SrTiO_3(100)$-$(\sqrt{5}\times\sqrt{5})R26.6°$ surface reconstruction. The images and the schematic representations are from [13.76]

STM image is in antiphase in the (1×1) and in phase in the (1×2) with respect to the NC-AFM image. In the (1×1), the STM image shows that the dangling bond states at the tip apex overlap with the dangling bonds of the 3*d* states that are protruding from the Ti atom, while the NC-AFM primarily imaged the uppermost oxygen atom. The results of $SrTiO_3$ and TiO_2 surface imaging show that the information obtained by STM and NC-AFM differ, and that the simultaneous imaging of a metal oxide surface enables the investigation of a more detailed surface structure.

13.3.2 Atomically Resolved Imaging of a NiO(001) Surface

The transition metal oxides such as NiO, CoO, and FeO feature the simultaneous existence of an energy gap and unpaired electrons, which gives rise to magnetic properties. Such magnetic insulators are widely used for the exchange biasing in spin valve devices. The NC-AFM enables the direct surface imaging of magnetic insulators on an atomic scale. The forces detected by NC-AFM originate from several kinds of interaction between the surface and the tip, including magnetic interactions in some cases. Theoretical studies predict that short-range magnetic interactions such as the exchange interaction should enable the NC-AFM to image magnetic moments on an atomic scale. In this section, we will describe an imaging of the antiferromagnetic NiO(001) surface using a ferromagnetic Fe-coated tip in order to explore the detection of the short-range magnetic interaction. Also, theoretical studies of the exchange force interaction between a magnetic tip and a sample will be described.

Theoretical Studies of the Exchange Force

The NC-AFM measures various kinds of interaction between tip and sample such as van der Waals, and electrostatic and short-range chemical interactions. In the case of a magnetic tip and sample, the interaction detected by the NC-AFM includes the long-range magnetic dipole interaction and the short-range magnetic interaction. The short-range magnetic interaction between the atoms at the tip apex and the sample surface is governed by a combination of the Coulomb interaction and the Pauli exclusion principle and may be included among the short-range chemical interactions. The magnetic interaction energy depends on the electron spin state of the atoms, and the energy difference in spin alignments (parallel or anti-parallel) is referred to as the exchange interaction energy. Therefore, if the short-range magnetic interaction can be detected, an atomically resolved NC-AFM image, in which the short-range interaction contributes to the atomic protrusions, should reveal the local energy difference in the spin alignment.

In the past, extensive theoretical studies on the short-range magnetic interaction between a ferromagnetic tip and a ferromagnetic sample have been performed by a simple calculation [13.80], a tight-binding approximation [13.81], and first-principles calculations [13.82]. In the calculations performed by *Nakamura* et al. [13.82], three-atomic-layer Fe(001) films, which are separated by a vacuum gap, are used as a model for the tip and sample. The exchange force (F_{ex}) is defined as the difference of the forces (F_P, F_{AP}) in each spin configuration (parallel: P, anti-parallel: AP) of the tip and sample. The tip-sample distance dependency of the exchange force demonstrates that the amplitude of the exchange force is on the order of 10^{-10} N or above and is large enough to be detected by NC-AFM. Also, the site dependency of the exchange force indicates that the discrimination of the exchange force enables the direct imaging of the magnetic moments on an atomic scale. Recently, the interaction between a spin-polarized H atom and a Ni atom on a NiO(001) surface has been investigated theoretically [13.83]. *Foster* and *Shluger* demonstrated that the difference in magnitude of the exchange interaction between opposite spin Ni ions in a NiO surface could be sufficient to be measured in a low temperature NC-AFM experiment. Also, in order to investigate the effect of surface relaxation, they calculated the interaction of different metal tips with the MgO surface used as a model system. The results indicated that the interaction over the oxygen site is stronger than that over the Mg site, and it was pointed out that the tips should be coated in metals that interact weakly with the surface for the detection of the exchange interaction between the atom at the tip apex and the Ni atom on the NiO surface. However, we must take into account another possibility, that is, that a metal atom at the ferromagnetic tip apex may interact with a Ni atom on the second layer through a magnetic interaction mediated by the electrons in an oxygen atom on the surface.

13.3.3 Atomically Resolved Imaging Using Noncoated and Fe-Coated Si Tips

It is thought that the magnetic insulator NiO is one of the best candidates for detecting short-range magnetic interactions such as the exchange interaction, because an experimental configuration of a ferromagnetic tip

and an antiferromagnetic sample minimizes the long-range magnetic dipole interaction that might hinder the detection of the short-range interaction.

Below the Néel temperature of 525 K, NiO is antiferromagnetic and has a rhombohedral structure contracted along the ⟨111⟩ axes. The direction of the spin on nickel atoms in the same {111} plane is parallel, a so-called ferromagnetic sheet. The direction of spin in the adjacent {111} planes is anti-parallel. As a result, the spins at the Ni atoms on the (001) surface are arranged in a checkerboard pattern, i. e., the spin alignment along the ⟨111⟩ direction is antiferromagnetic in order and that along the ⟨1 −1 0⟩ direction is ferromagnetic in order.

Figure 13.21a shows an atomically resolved image of a NiO(001) surface with a ferromagnetic Fe-coated tip [13.84]. The bright protrusions correspond to atoms spaced about 0.42 nm apart, consistent with the expected periodic arrangement of the NiO(001) surface.

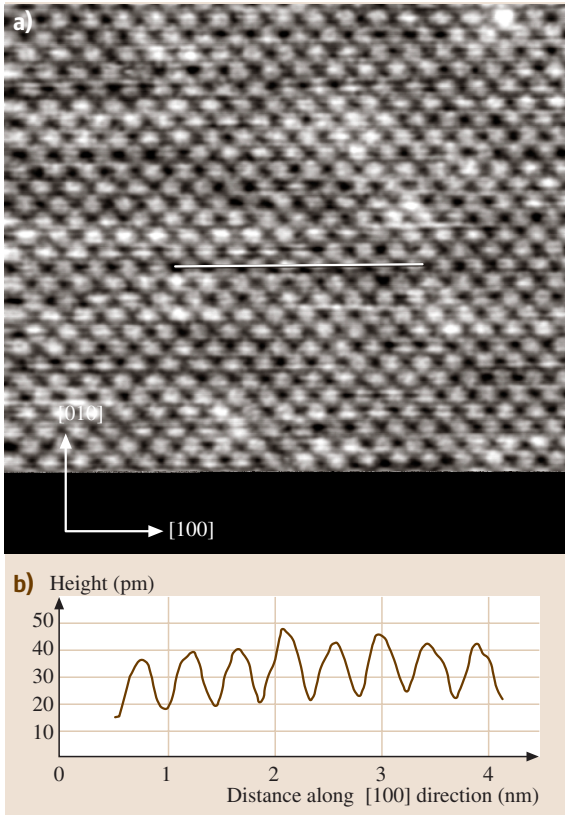


Fig. 13.21 **(a)** Atomically resolved image obtained with an Fe-coated tip. **(b)** shows the cross sections of the middle part in **(a)**. Their corrugations are about 30 pm

The corrugation amplitude is typically 30 pm, which is comparable to the value previously reported [13.74, 75], as shown in Fig. 13.21b. Also, because of the symmetry of the image, it was concluded that only one type of atom is visualized by NC-AFM. However, from this image, we think that it is difficult to distinguish which of the atoms are observed, oxygen or nickel. The theoretical works indicate that a metal tip interacts strongly with the oxygen atoms on the MgO(001) surface [13.83]. From this result, it is presumed that the bright protrusions correspond to the oxygen atoms. However, it is still questionable which of the atoms are visible with a Fe-coated tip.

If the short-range magnetic interaction is included in the interactions detected by NC-AFM, the corrugation amplitude of the atoms should depend on the direction of the spin at the atom-site. The results of a comparison of the corrugation amplitudes of the same species of atoms with opposite spin can reveal the difference in the short-range magnetic interaction, because the nonmagnetic interactions peculiar to a particular atomic species should be identical. From the results of first-principles calculations [13.82], the contribution of the short-range magnetic interaction to the measured corrugation amplitude is expected to be about a few percent of the total interaction. Discrimination of such small perturbations is therefore needed. In order to reduce the noise, the corrugation amplitude was added on the basis of the periodicity of the NC-AFM image. In addition, the topographical asymmetry, which is the index characterizing the difference in atomic corrugation amplitude, has been defined [13.85]. The result shows that the value of the topographical asymmetry calculated from the image obtained with an Fe-coated Si tip depends on the direction of summing of the corrugation amplitude, and that the dependency corresponds to the antiferromagnetic spin ordering of the NiO(001) surface. With a paramagnetic Si tip, no significant indication of a directional dependency of the topographical asymmetry can be seen. Therefore, it is believed that the results obtained using a Fe-coated tip support the conclusion that the dependency of the topographical asymmetry originates in the short-range magnetic interaction.

The measurements presented here demonstrate the possibility of imaging magnetic structures on an atomic scale by NC-AFM. However, the origin of the detected short-range magnetic interaction is under discussion. Therefore, further experiments such as the measurement of force-distance curves and a theoretical study of the interaction between the ferromagnetic tip and an antiferromagnetic sample are needed.

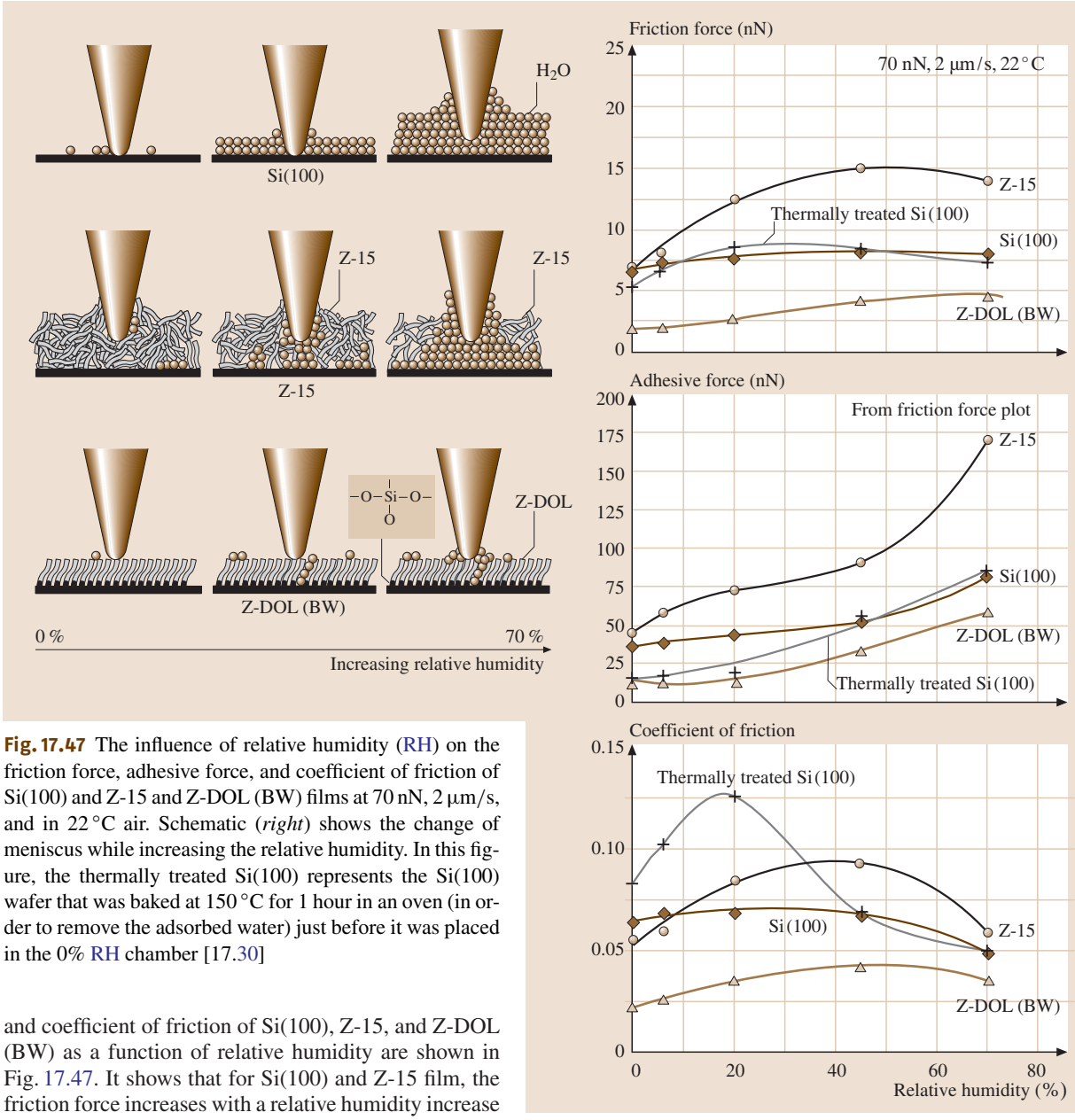


Fig. 17.47 The influence of relative humidity (RH) on the friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 and Z-DOL (BW) films at 70 nN, 2 μm/s, and in 22 °C air. Schematic (*right*) shows the change of meniscus while increasing the relative humidity. In this figure, the thermally treated Si(100) represents the Si(100) wafer that was baked at 150 °C for 1 hour in an oven (in order to remove the adsorbed water) just before it was placed in the 0% RH chamber [17.30]

and coefficient of friction of Si(100), Z-15, and Z-DOL (BW) as a function of relative humidity are shown in Fig. 17.47. It shows that for Si(100) and Z-15 film, the friction force increases with a relative humidity increase of up to 45%, and then slightly decreases with a further increase in the relative humidity. Z-DOL (BW) has a smaller friction force than Si(100) and Z-15 in the whole testing range. And its friction force shows a relative apparent increase when the relative humidity is higher than 45%. For Si(100), Z-15, and Z-DOL (BW), their adhesive forces increase with relative humidity, and their coefficients of friction increase with relative humidity up to 45%, after which they decrease with further increasing of the relative humidity. It is also observed that the humidity effect on Si(100) really depends on the history of the Si(100) sample. As the surface of Si(100) wafer readily adsorbs water in air, without any pre-treatment the Si(100) used in our study almost

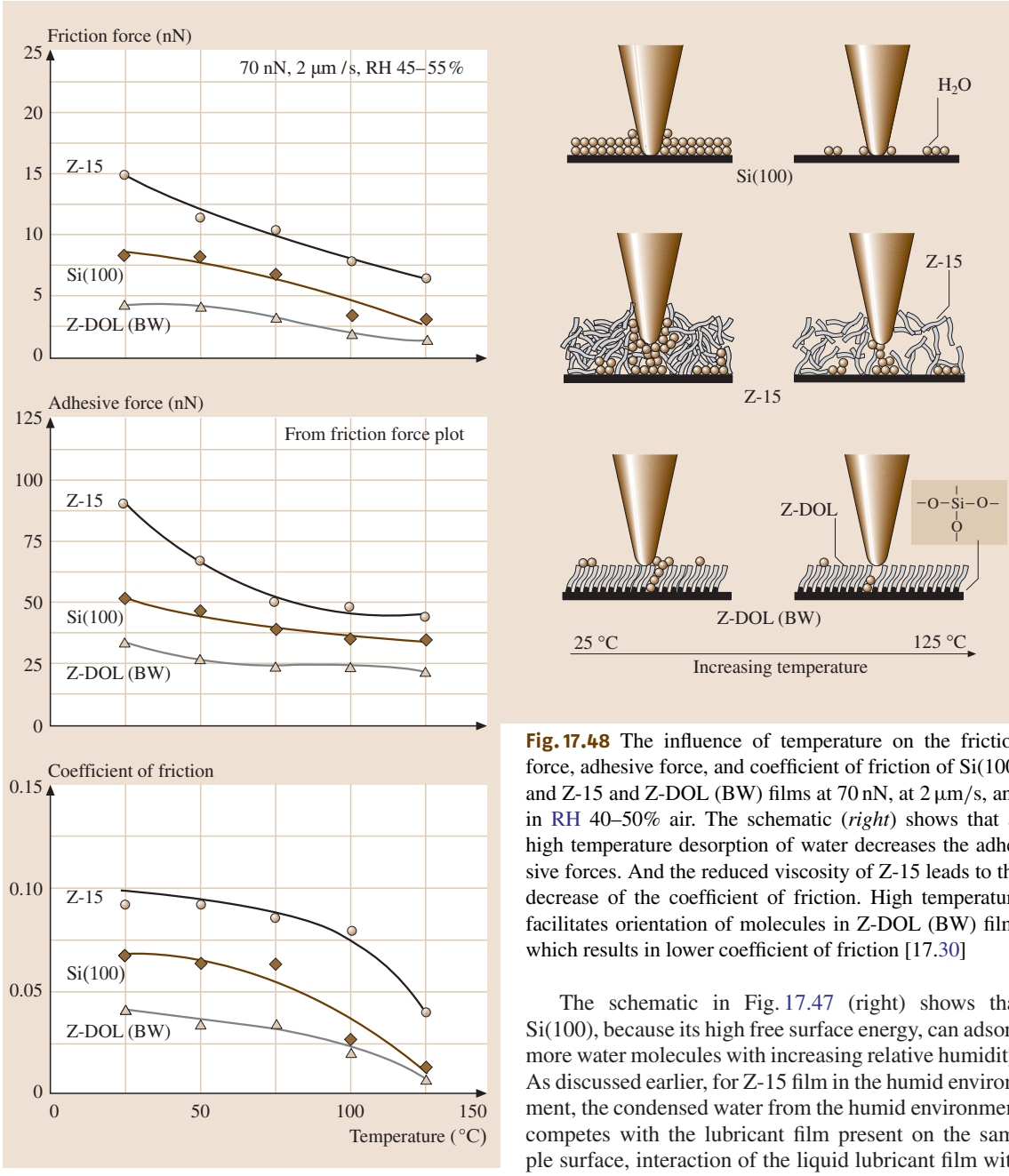


Fig. 17.48 The influence of temperature on the friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 and Z-DOL (BW) films at 70 nN, at 2 μm/s, and in RH 40–50% air. The schematic (*right*) shows that at high temperature desorption of water decreases the adhesive forces. And the reduced viscosity of Z-15 leads to the decrease of the coefficient of friction. High temperature facilitates orientation of molecules in Z-DOL (BW) film, which results in lower coefficient of friction [17.30]

reaches to its saturate stage of adsorbed water and is responsible for less effect during increasing relative humidity. However, once the Si(100) wafer was thermally treated by baking at 150 °C for 1 hour, a bigger effect was observed.

The schematic in Fig. 17.47 (right) shows that Si(100), because its high free surface energy, can adsorb more water molecules with increasing relative humidity. As discussed earlier, for Z-15 film in the humid environment, the condensed water from the humid environment competes with the lubricant film present on the sample surface, interaction of the liquid lubricant film with the silicon substrate is weakened, and a boundary layer of the liquid lubricant forms puddles. This dewetting allows water molecules to be adsorbed on the Si(100) substrate mixed with Z-15 molecules [17.27, 28]. Obviously, more water molecules can be adsorbed on Z-15 surface while increasing relative humidity. The more ad-

sorbed water molecules in the case of Si(100), along with lubricant molecules in Z-15 film case, form bigger water meniscus which leads to an increase of friction force, adhesive force, and coefficient of friction of Si(100) and Z-15 with humidity. But at a very high humidity of 70%, large quantities of adsorbed water can form a water layer that separates the tip and sample surface and act as a kind of lubricant, which causes a decrease in the friction force and coefficient of friction. For Z-DOL (BW) film, because of its hydrophobic surface properties, water molecules can be adsorbed at a humidity higher than 45% and causes an increase in the adhesive and friction forces.

To study the temperature effect on friction and adhesion, the variations of friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (BW) as a function of temperature are summarized in Fig. 17.48. It shows that the increasing temperature causes a decrease in friction force, adhesive force, and coefficient of friction of Si(100), Z-15, and Z-DOL (BW). The schematic (right) in Fig. 17.48 indicates that at high temperature, desorption of water leads to the decrease of friction force, adhesive forces, and coefficient of friction for all of the samples. For Z-15 film, the reduction of viscosity at high temperature also contributes to the decrease of friction force and coefficient of friction. In the case of Z-DOL (BW) film, molecules are more easily oriented at high temperature, which may be partly responsible for the low friction force and coefficient of friction.

To study the durability of lubricant films at the nanoscale, the friction of Si(100), Z-15, and Z-DOL (BW) as a function of the number of scanning cycles is shown in Fig. 17.49. As observed earlier, friction force and coefficient of friction of Z-15 is higher than that of Si(100) with the lowest values for Z-DOL (BW). During cycling, friction force and coefficient of friction of Si(100) show a slight decrease during the initial few cycles then remain constant. This is related to the removal of the top adsorbed layer. In the case of Z-15 film, the friction force and coefficient of friction show an increase during the initial few cycles, and then approach higher and stable values. This is believed to be caused by the attachment of the Z-15 molecules to the tip. The molecular interaction between these attached molecules to the tip and molecules on the film surface is responsible for an increase in friction. But after several scans, this molecular interaction reaches the equilibrium, and after that, friction force and coefficient of friction remain constant. In the case of Z-DOL (BW) film, the friction force and coefficient of friction start out low and remain

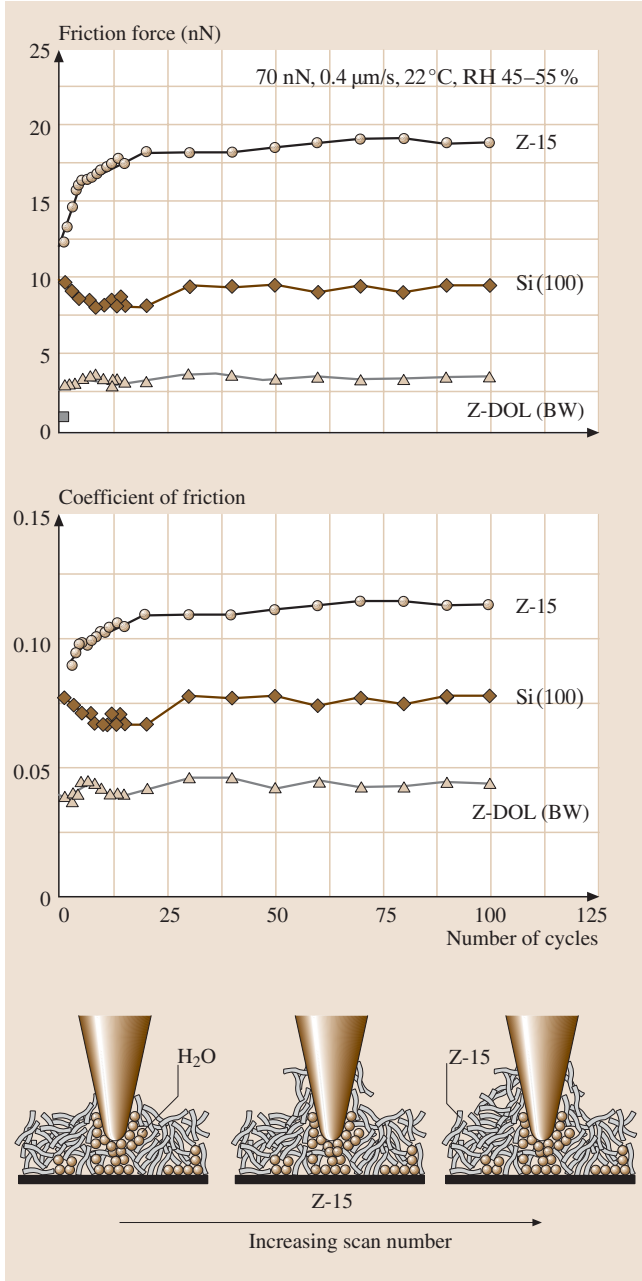


Fig. 17.49 Friction force and coefficient of friction versus number of sliding cycles for Si(100) and Z-15 and Z-DOL (BW) films at 70 nN, 0.8 μm/s, and in ambient air. Schematic (*bottom*) shows that some liquid Z-15 molecules can be attached to the tip. The molecular interaction between the attached molecules to the tip with the Z-15 molecules in the film results in an increase of the friction force with multiple scanning [17.30]

low during the entire test for 100 cycles, suggesting that Z-DOL (BW) molecules do not get attached or displaced as readily as Z-15 molecules.

As a brief summary, the influence of velocity, relative humidity, and temperature on the friction force of mobile Z-15 film is presented in Fig. 17.50. The changing trends are also addressed in this figure.

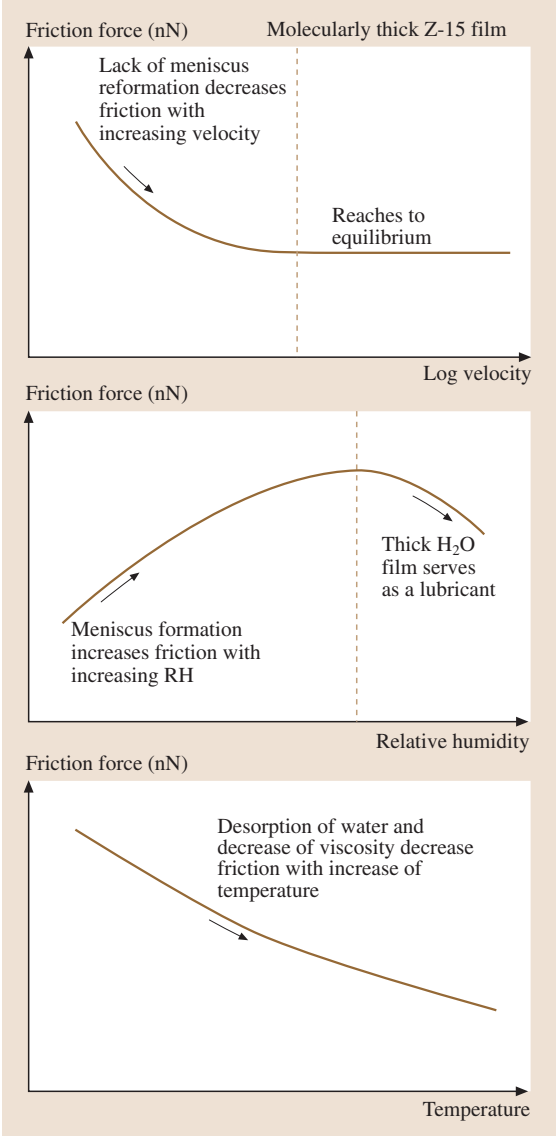


Fig. 17.50 Schematic shows the change of friction force of molecularly thick Z-15 films with log velocity, relative humidity, and temperature. The changing trends are also addressed in this figure [17.30]

17.5.2 Self-Assembled Monolayers

For lubrication of MEMS/NEMS, another effective approach involves the deposition of organized and dense molecular layers of long-chain molecules. Two common methods to produce monolayers and thin films are the Langmuir–Blodgett (L-B) deposition and self-assembled monolayers (SAMs) by chemical grafting of molecules. L-B films are physically bonded to the substrate by weak van der Waals attraction, while SAMs are chemically bonded via covalent bonds to the substrate. Because of the choice of chain length and terminal linking group that SAMs offer, they hold great promise for boundary lubrication of MEMS/NEMS. A number of studies have been conducted to study tribological properties of various SAMs [17.26, 29, 74–76]. It has been reported that SAMs with high-compliance long carbon chains exhibit low friction; chain compliance is desirable for low friction. Based on [17.29], the friction mechanism of SAMs is explained by a so-called "molecular spring" model (Fig. 17.51). According to this model, the chemically adsorbed self-assembled molecules on a substrate are just like assembled molecular springs anchored to the substrate. An asperity sliding on the surface of SAMs is like a tip sliding on the top of "molecular springs or brush". The molecular spring assembly has compliant features and can experience orientation and compression under load. The orientation of the molecular springs or brush under normal load reduces the shearing force at the interface, which, in turn, reduces the friction force.

The SAMs with high-compliance long carbon chains also exhibit the best wear resistance [17.29, 75]. In wear experiments, the wear depth as a function of normal load curves shows critical normal loads (Fig. 17.52). Below the critical normal load SAMs undergo orien-

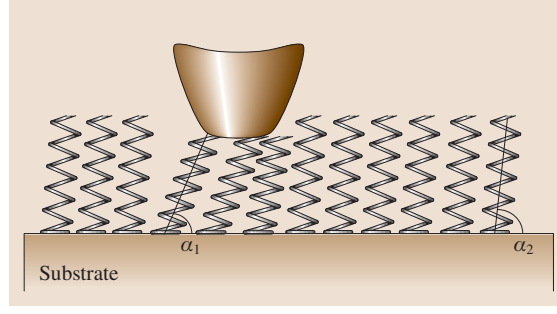


Fig. 17.51 Molecular spring model of SAMs. In this figure, $\alpha_1 < \alpha_2$, which is caused by the further orientation under the normal load applied by an asperity tip [17.29]

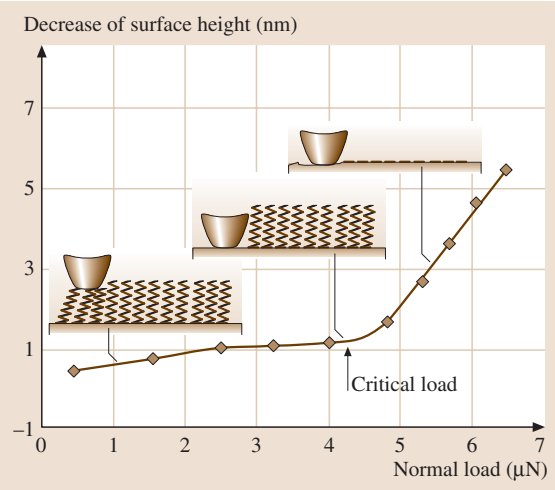


Fig. 17.52 Illustration of the wear mechanism of SAMs with increasing normal load [17.75]

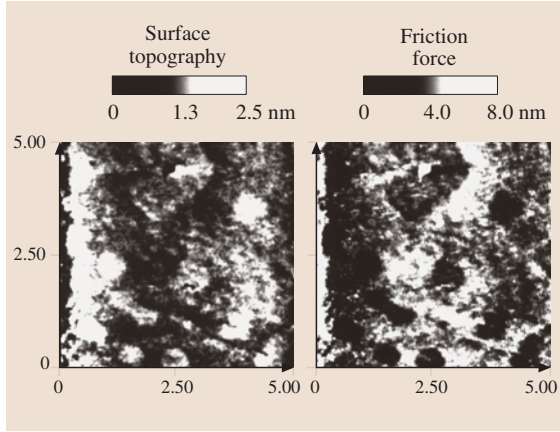


Fig. 17.53 Gray-scale plots of the surface topography and friction force obtained simultaneously for unbonded Demnum-type perfluoropolyether lubricant film on silicon [17.27]

tation, at the critical load SAMs wear away from the substrate due to weak interface bond strengths, while above the critical normal load severe wear takes place on the substrate.

17.5.3 Liquid Film Thickness Measurements

Liquid film thickness measurement of thin lubricant films (on the order of 10 nm or thicker) with nanometer lateral resolution can be made with the AFM [17.4, 59]. The lubricant thickness is obtained by measuring the force on the tip as it approaches, contacts, and pushes through the liquid film and ultimately contacts the substrate. The distance between the sharp snap-in (owing to the formation of a liquid meniscus between the film and the tip) at the liquid surface and the hard repulsion at the substrate surface is a measure of the liquid film thickness.

Lubricant film thickness mapping of ultrathin films (on the order of 2 nm) can be obtained using friction force microscopy [17.27] and adhesive force mapping [17.36]. Figure 17.53 shows gray-scale plots of the surface topography and friction force obtained simultaneously for unbonded Demnum-type PFPE lubricant film on silicon. The friction force plot shows well distinguished low and high friction regions roughly corresponding to high and low regions in surface topography (thick and thin lubricant regions). A uniformly lubricated sample does not show such a variation in the friction. Friction force imaging can thus be used to measure the lubricant uniformity on the sample surface, which cannot be identified by surface topography alone. Figure 17.54 shows the gray-scale plots of the adhesive force distribution for silicon samples coated uniformly and nonuniformly with Z-DOL type PFPE lubricant. It can be clearly seen that a region exists that has an adhesive force distinctly different from the other region for the nonuniformly coated sample. This implies that the liquid film thickness is nonuniform, giving rise to a difference in the meniscus forces.

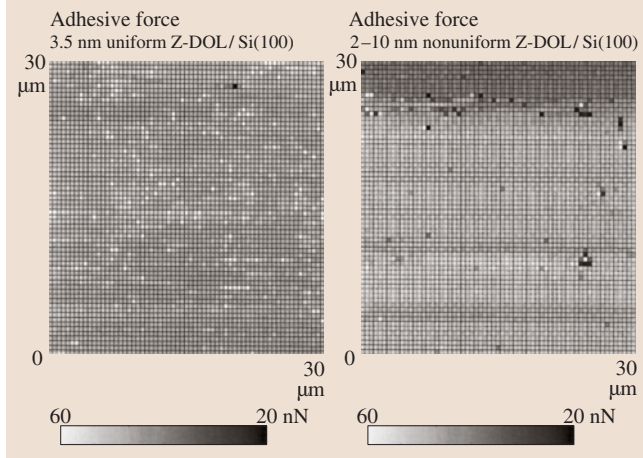


Fig. 17.54 Gray-scale plots of the adhesive force distribution of a uniformly coated, 3.5-nm thick unbonded Z-DOL film on silicon and 3- to 10-nm thick unbonded Z-DOL film on silicon that was deliberately coated nonuniformly by vibrating the sample during the coating process [17.36]

17.6 Closure

At most solid-solid interfaces of technological relevance, contact occurs at many asperities. A sharp AFM/FFM tip sliding on a surface simulates just one such contact. However, asperities come in all shapes and sizes. The effect of radius of a single asperity (tip) on the friction/adhesion performance can be studied using tips of different radii. AFM/FFM are used to study various tribological phenomena, including surface roughness, adhesion, friction, scratching, wear, indentation, detection of material transfer, and boundary lubrication. Measurement of atomic-scale friction of a freshly cleaved, highly oriented pyrolytic graphite exhibits the same periodicity as that of the corresponding topography. However, the peaks in friction and those in the corresponding topography are displaced relative to one another. Variations in atomic-scale friction and the observed displacement can be explained by the variation in interatomic forces in the normal and lateral directions. Local variations in microscale friction occur and are found to correspond to the local slopes, suggesting that a ratchet mechanism and collision effects are responsible for this variation. Directionality in the friction is observed on both micro- and macroscales, resulting from the surface roughness and surface preparation. Anisotropy in surface roughness accentuates this effect. Microscale friction is generally found to be smaller than the macrofriction, as there is less plowing contribution in microscale measurements. Microscale friction is load-dependent, and friction values increase with an increase in the normal load, approaching the macrofriction at contact stresses higher than the hardness of the softer material. The tip radius also affects adhesion and friction.

Mechanism of material removal on the microscale is studied. Wear rate for single-crystal silicon is negligible below 20 μN and is much higher and remains approximately constant at higher loads. Elastic deformation at low loads is responsible for negligible wear. Most of the wear debris is loose. SEM and TEM studies of the wear region suggest that the material on the microscale is removed by plastic deformation with a small contribution from elastic fracture; this observation corroborates the scratch data. Evolution of wear has also been studied using AFM. Wear is found to be initiated at nanoscratches. For a sliding interface requiring near-zero friction and wear, contact stresses should be below the hardness of the softer material to minimize plastic deformation, and surfaces should be free of nanoscratches. Furthermore, wear precursors can be detected at early stages of wear by using surface potential measurements. It is found that even in the case of zero wear (no measurable deformation of the surface using AFM), there can be a significant change in the surface potential inside the wear mark, which is useful for study of wear precursors. Detection of material transfer on a nanoscale is possible with AFM.

In situ surface characterization of local deformation of materials and thin coatings can be carried out using a tensile stage inside an AFM. An AFM can also be used for nanofabrication/nanomachining.

By using the force modulation technique, localized surface elasticity maps of composite materials with penetrating depths of less than 100 nm can be obtained. It is possible using phase contrast microscopy to get phase contrast maps or the contrast in viscoelastic properties of near surface regions. Scratching and indentation on nanoscales are powerful ways to screen for adhesion and resistance to deformation of ultrathin films. Modified AFM can be used to obtain load-displacement curves and for measurement of nanoindentation hardness and Young's modulus of elasticity, with depth of indentation as low as 1 nm. Hardness of ceramics on the nanoscale is found to be higher than that on the microscale. Ceramics exhibit significant plasticity and creep on a nanoscale.

Boundary lubrication studies and measurement of lubricant-film thickness with a lateral resolution on a nanoscale can be conducted using AFM. Chemically bonded lubricant films and self-assembled monolayers are superior in friction and wear resistance. For chemically bonded lubricant films, the adsorption of water, the formation of meniscus, and its change during sliding, viscosity, and surface properties play an important role on the friction, adhesion, and durability of these films. For SAMs, their friction mechanism is explained by a so-called "molecular spring" model. The films with high-compliance long carbon chains exhibit low friction and wear.

Investigations of wear, scratching, and indentation on the nanoscale using the AFM can provide insights into failure mechanisms of materials. Coefficients of friction, wear rates, and mechanical properties such as hardness have been found to be different on the nanoscale than on the macroscale; generally, coefficients of friction and wear rates on micro- and nanoscales are smaller, whereas hardness is greater. Therefore, micro/nanotribological studies may help define the regimes for ultralow friction and near-zero wear. These studies also provide insight into the atomic origins of adhesion, friction, wear, and lubrication mechanisms.

References

17.1 I. L. Singer, H. M. Pollock: *Fundamentals of Friction: Macroscopic and Microscopic Processes*, Nato Sci. Ser. E, Vol. 220 (Kluwer, Dordrecht 1992)

17.2 B. N. J. Persson, E. Tosatti: *Physics of Sliding Friction*, Nato Sci. Ser. E, Vol. 311 (Kluwer, Dordrecht 1996)

17.3 B. Bhushan: *Micro/Nanotribology and Its Applications*, Nato Sci. Ser. E, Vol. 330 (Kluwer, Dordrecht 1997)

17.4 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

17.5 B. Bhushan: Nanoscale tribophysics and tribomechanics, Wear **225–229** (1999) 465–492

17.6 B. Bhushan: Wear and mechanical characterisation on micro- to picoscales using AFM, Int. Mat. Rev. **44** (1999) 105–117

17.7 B. Bhushan: *Modern Tribology Handbook, Vol. 1: Principles of Tribology* (CRC, Boca Raton 2001)

17.8 B. Bhushan: *Fundamentals of tribology and bridging the gap between the macro- and micro/nanoscales*, NATO Sci. Ser. E, Vol. 10 (Kluwer, Dordrecht 2001)

17.9 B. Bhushan: Nano- to microscale wear and mechanical characterization studies using scanning probe microscopy, Wear **251** (2001) 1105–1123

17.10 B. Bhushan: *Introduction to Tribology* (Wiley, New York 2002)

17.11 B. Bhushan, J. N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374** (1995) 607–616

17.12 H. J. Guntherodt, D. Anselmetti, E. Meyer: *Forces in Scanning Probe Methods*, Nato Sci. Ser. E, Vol. 286 (Kluwer, Dordrecht 1995)

17.13 B. Bhushan: *Tribology and Mechanics of Magnetic Storage Devices*, 2nd edn. (Springer, New York 1996)

17.14 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

17.15 G. Binnig, C. F. Quate, Ch. Gerber: Atomic force microscopy, Phys. Rev. Lett. **56** (1986) 930–933

17.16 G. Binnig, Ch. Gerber, E. Stoll, T. R. Albrecht, C. F. Quate: Atomic resolution with atomic force microscope, Europhys. Lett. **3** (1987) 1281–1286

17.17 C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59** (1987) 1942–1945

17.18 B. Bhushan, J. Ruan: Atomic-scale friction measurements using friction force microscopy: part II – application to magnetic media, ASME J. Tribol. **116** (1994) 389–396

17.19 J. Ruan, B. Bhushan: Atomic-scale friction measurements using friction force microscopy: part I – general principles and new measurement techniques, ASME J. Tribol. **116** (1994) 378–388

17.20 J. Ruan, B. Bhushan: Atomic-scale and microscale friction of graphite and diamond using friction force microscopy, J. Appl. Phys. **76** (1994) 5022–5035

17.21 J. Ruan, B. Bhushan: Frictional behavior of highly oriented pyrolytic graphite, J. Appl. Phys. **76** (1994) 8117–8120

17.22 B. Bhushan, V. N. Koinkar, J. Ruan: Microtribology of magnetic media, Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. **208** (1994) 17–29

17.23 B. Bhushan, S. Sundararajan: Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46** (1998) 3793–3804

17.24 V. Scherer, W. Arnold, B. Bhushan: Active friction control using ultrasonic vibration. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 463–469

17.25 V. Scherer, W. Arnold, B. Bhushan: Lateral force microscopy using acoustic friction force microscopy, Surf. Interface Anal. **27** (1999) 578–587

17.26 B. Bhushan, A. V. Kulkarni, V. N. Koinkar, M. Boehm, L. Odoni, C. Martelet, M. Belin: Microtribological characterization of self-assembled and Langmuir–Blodgett monolayers by atomic and friction force microscopy, Langmuir **11** (1995) 3189–3198

17.27 V. N. Koinkar, B. Bhushan: Micro/nanoscale studies of boundary layers of liquid lubricants for magnetic disks, J. Appl. Phys. **79** (1996) 8071–8075

17.28 V. N. Koinkar, B. Bhushan: Microtribological studies of unlubricated and lubricated surfaces using atomic force/friction force microscopy, J. Vac. Sci. Technol. A. **14** (1996) 2378–2391

17.29 B. Bhushan, H. Liu: Nanotribological properties and mechanisms of alkylthiol and biphenyl thiol self-assembled monolayers studied by AFM, Phys. Rev. B **63** (2001) 245412–1–245412–11

17.30 H. Liu, B. Bhushan: Nanotribological characterization of molecularly-thick lubricant films for applications to MEMS/NEMS by AFM, Ultramicroscopy **97** (2003) 321–340

17.31 B. Bhushan, V. N. Koinkar: Tribological studies of silicon for magnetic recording applications, J. Appl. Phys. **75** (1994) 5741–5746

17.32 V. N. Koinkar, B. Bhushan: Microtribological properties of hard amorphous carbon protective coatings for thin film magnetic disks and heads, Proc. Inst. Mech. Eng. Part J: J. Eng. Tribol. **211** (1997) 365–372

17.33 S. Sundararajan, B. Bhushan: Development of a continuous microscratch technique in an atomic force microscope and its application to study scratch resistance of ultra-thin hard amorphous carbon coatings, J. Mater. Res. **16** (2001) 75–84

17.34 B. Bhushan, V. N. Koinkar: Nanoindentation hardness measurements using atomic force microscopy, Appl. Phys. Lett. **64** (1994) 1653–1655

17.35 B. Bhushan, A. V. Kulkarni, W. Bonin, J. T. Wyrobek: Nano/picoindentation measurement using a capacitance transducer system in atomic force microscopy, Philos. Mag. **74** (1996) 1117–1128

17.36 B. Bhushan, C. Dandavate: Thin-film friction and adhesion studies using atomic force microscopy, J. Appl. Phys. **87** (2000) 1201–1210

17.37 B. Bhushan: Micro/nanotribology and its applications to magnetic storage devices and MEMS, Tribol. Int. **28** (1995) 85–95

17.38 D. DeVecchio, B. Bhushan: Use of a nanoscale Kelvin probe for detecting wear precursors, Rev. Sci. Instrum. **69** (1998) 3618–3624

17.39 B. Bhushan, A. V. Goldade: Measurements and analysis of surface potential change during wear of single crystal silicon (100) at ultralow loads using Kelvin probe microscopy, Appl. Surf. Sci **157** (2000) 373–381

17.40 B. Bhushan, A. V. Goldade: Kelvin probe microscopy measurements of surface potential change under wear at low loads, Wear **244** (2000) 104–117

17.41 M. S. Bobji, B. Bhushan: Atomic force microscopic study of the micro-cracking of magnetic thin films under tension, Scripta Mater. **44** (2001) 37–42

17.42 M. S. Bobji, B. Bhushan: In-situ microscopic surface characterization studies of polymeric thin films during tensile deformation using atomic force microscopy, J. Mater. Res. **16** (2001) 844–855

17.43 P. Maivald, H. J. Butt, S. A. C. Gould, C. B. Prater, B. Drake, J. A. Gurley, V. B. Elings, P. K. Hansma: Using force modulation to image surface elasticities with the atomic force microscope, Nanotechnol. **2** (1991) 103–106

17.44 D. DeVecchio, B. Bhushan: Localized surface elasticity measurements using an atomic force microscope, Rev. Sci. Instrum. **68** (1997) 4498–4505

17.45 V. Scherer, B. Bhushan, U. Rabe, W. Arnold: Local elasticity and lubrication measurements using atomic force and friction force microscopy at ultrasonic frequencies, IEEE Trans. Mag. **33** (1997) 4077–4079

17.46 H. U. Krotil, T. Stifter, H. Waschipky, K. Weishaupt, S. Hild, O. Marti: Pulse force mode: A new method for the investigation of surface properties, Surf. Interface Anal. **27** (1999) 336–340

17.47 S. Amelio, A. V. Goldade, U. Rabe, V. Scherer, B. Bhushan, W. Arnold: Measurements of elastic properties of ultra-thin diamond-like carbon coatings using atomic force acoustic microscopy, Thin Solid Films **392** (2001) 75–84

17.48 B. Anczykowski, D. Kruger, K. L. Babcock, H. Fuchs: Basic properties of dynamic force microscopy with the scanning force microscope in experiment and simulation, Ultramicroscopy **66** (1996) 251–259

17.49 J. Tamayo, R. Garcia: Deformation, contact time, and phase contrast in tapping mode scanning force microscopy, Longmuir **12** (1996) 4430–4435

17.50 R. Garcia, J. Tamayo, M. Calleja, F. Garcia: Phase contrast in tapping-mode scanning force microscopy, Appl. Phys. A **66** (1998) 309–312

17.51 W. W. Scott, B. Bhushan: Use of phase imaging in atomic force microscopy for measurement of viscoelastic contrast in polymer nanocomposites and molecularly-thick lubricant films, Ultramicroscopy **97** (2003) 151–169

17.52 B. Bhushan, J. Qi: Phase contrast imaging of nanocomposites and molecularly- thick lubricant films in magnetic media, Nanotechnol. **14** (2003) 886–895

17.53 V. N. Koinkar, B. Bhushan: Microtribological studies of Al_2O_3-TiC, polycrystalline and single-crystal Mn-Zn ferrite and SiC head slider materials, Wear **202** (1996) 110–122

17.54 E. Meyer, R. Overney, R. Luthi, D. Brodbeck, L. Howald, J. Frommer, H. J. Guntherodt, O. Wolter, M. Fujihira, T. Takano, Y. Gotoh: Friction force microscopy of mixed Langmuir–Blodgett films, Thin Solid Films **220** (1992) 132–137

17.55 C. D. Frisbie, L. F. Rozsnyai, A. Noy, M. S. Wrighton, C. M. Lieber: Functional group imaging by chemical force microscopy, Science **265** (1994) 2071–2074

17.56 V. N. Koinkar, B. Bhushan: Effect of scan size and surface roughness on microscale friction measurements, J. Appl. Phys. **81** (1997) 2472–2479

17.57 S. Sundararajan, B. Bhushan: Topography-induced contributions to friction forces measured using an atomic force/friction force microscope, J. Appl. Phys. **88** (2000) 4825–4831

17.58 B. Bhushan: *Principles and Applications of Tribology* (Wiley, New York 1999)

17.59 B. Bhushan, G. S. Blackman: Atomic force microscopy of magnetic rigid disks and sliders and its applications to tribology, ASME J. Tribol. **113** (1991) 452–458

17.60 M. Reinstaedtler, U. Rabe, V. Scherer, U. Hartmann, A. Goldade, B. Bhushan, W. Arnold: On the nanoscale measurement of friction using atomic-force microscope cantilever torsional resonances, Appl. Phys. Lett. **82** (2003) 2604–2606

17.61 T. Stifter, O. Marti, B. Bhushan: Theoretical investigation of the distance dependence of capillary and van der Waals forces in scanning probe microscopy, Phys. Rev. B **62** (2000) 13667–13673

17.62 U. D. Schwarz, O. Zwoerner, P. Koester, R. Wiesendanger: Friction force spectroscopy in the low-load regime with well-defined tips. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 233–238

17.63 B. Bhushan, A. V. Kulkarni: Effect of normal load on microscale friction measurements, Thin Solid Films **278** (1996) 49–56; Errata: **293** 333

17.64 V. N. Koinkar, B. Bhushan: Scanning and transmission electron microscopies of single-crystal silicon microworn/machined using atomic force microscopy, J. Mater. Res. **12** (1997) 3219–3224

17.65 X. Zhao, B. Bhushan: Material removal mechanism of single-crystal silicon on nanoscale and at ultralow loads, Wear **223** (1998) 66–78

17.66 B. Bhushan, P. S. Mokashi, T. Ma: A new technique to measure Poisson's ratio of ultrathin polymeric films using atomic force microscopy, Rev. Sci. Instrumen. **74** (2003) 1043–1047

17.67 A. V. Kulkarni, B. Bhushan: Nanoscale mechanical property measurements using modified atomic force microscopy, Thin Solid Films **290–291** (1996) 206–210

17.68 A. V. Kulkarni, B. Bhushan: Nano/picoindentation measurements on single-crystal aluminum using modified atomic force microscopy, Mater. Lett. **29** (1996) 221–227

17.69 A. V. Kulkarni, B. Bhushan: Nanoindentation measurement of amorphous carbon coatings, J. Mater. Res. **12** (1997) 2707–2714

17.70 J. Ruan, B. Bhushan: Nanoindentation studies of fullerene films using atomic force microscopy, J. Mater. Res. **8** (1993) 3019–3022

17.71 N. A. Fleck, G. M. Muller, M. F. Ashby, J. W. Hutchinson: Strain gradient plasticity: Theory and experiment, Acta Metall. Mater. **42** (1994) 475–487

17.72 W. B. Li, J. L. Henshall, R. M. Hooper, K. E. Easterling: The mechanism of indentation creep, Acta Metall. Mater. **39** (1991) 3099–3110

17.73 F. P. Bowden, D. Tabor: *The Friction and Lubrication of Solids*, Vol. 1 (Clarendon, Oxford 1950)

17.74 H. Liu, B. Bhushan: Investigation of the adhesion, friction, and wear properties of biphenyl thiol self-assembled monolayers by atomic force microscopy, J. Vac. Sci. Technol. A **19** (2001) 1234–1240

17.75 H. Liu, B. Bhushan: Investigation of nanotribological properties of self-assembled monolayers with alkyl and biphenyl spacer chains, Ultramicroscopy **91** (2002) 185–202

17.76 B. Bhushan: Self-assembled monolayers for controlling hydrophobicity and/or friction and wear. In: *Modern Tribology Handbook, Vol. 2: Materials, Coatings, and Industrial Applications*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 909–929

18. Surface Forces and Nanorheology of Molecularly Thin Films

In this chapter, we describe the static and dynamic normal forces that occur between surfaces in vacuum or liquids and the different modes of friction that can be observed between (i) bare surfaces in contact (dry or interfacial friction), (ii) surfaces separated by a thin liquid film (lubricated friction), and (iii) surfaces coated with organic monolayers (boundary friction).

Experimental methods suitable for measuring normal surface forces, adhesion and friction (lateral or shear) forces of different magnitude at the molecular level are described. We explain the molecular origin of van der Waals, electrostatic, solvation and polymer mediated interactions, and basic models for the contact mechanics of adhesive and nonadhesive elastically deforming bodies. The effects of interaction forces, molecular shape, surface structure and roughness on adhesion and friction are discussed.

Simple models for the contributions of the adhesion force and external load to interfacial friction are illustrated with experimental data on both unlubricated and lubricated systems, as measured with the surface forces apparatus. We discuss rate-dependent adhesion (adhesion hysteresis) and how this is related to friction. Some examples of the transition from wearless friction to friction with wear are shown.

Lubrication in different lubricant thickness regimes is described together with explanations of nanorheological concepts. The occurrence of and transitions between smooth and stick–slip sliding in various types of dry (unlubricated and solid boundary lubricated) and liquid lubricated systems are discussed based on recent experimental results and models for stick–slip involving memory distance and dilatency.

18.1 **Introduction: Types of Surface Forces** 544

18.2 **Methods Used to Study Surface Forces** 546
- 18.2.1 Force Laws 546
- 18.2.2 Adhesion Forces 547
- 18.2.3 The SFA and AFM 547
- 18.2.4 Some Other Force-Measuring Techniques 549

18.3 **Normal Forces Between Dry (Unlubricated) Surfaces** 550
- 18.3.1 Van der Waals Forces in Vacuum and Inert Vapors 550
- 18.3.2 Charge Exchange Interactions 552
- 18.3.3 Sintering and Cold Welding 553

18.4 **Normal Forces Between Surfaces in Liquids** 554
- 18.4.1 Van der Waals Forces in Liquids 554
- 18.4.2 Electrostatic and Ion Correlation Forces 554
- 18.4.3 Solvation and Structural Forces 557
- 18.4.4 Hydration and Hydrophobic Forces 559
- 18.4.5 Polymer-Mediated Forces 561
- 18.4.6 Thermal Fluctuation Forces 563

18.5 **Adhesion and Capillary Forces** 564
- 18.5.1 Capillary Forces 564
- 18.5.2 Adhesion Mechanics 566
- 18.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch 566
- 18.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis 567

18.6 **Introduction: Different Modes of Friction and the Limits of Continuum Models** 569

18.7 **Relationship Between Adhesion and Friction Between Dry (Unlubricated and Solid Boundary Lubricated) Surfaces** 571
- 18.7.1 Amontons' Law and Deviations from It Due to Adhesion: The Cobblestone Model 571
- 18.7.2 Adhesion Force and Load Contribution to Interfacial Friction 572
- 18.7.3 Examples of Experimentally Observed Friction of Dry Surfaces... 576
- 18.7.4 Transition from Interfacial to Normal Friction with Wear 579

18.8 **Liquid Lubricated Surfaces** 580
18.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime 580
18.8.2 Friction of Intermediate Thickness Films 582
18.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology 584
18.9 **Role of Molecular Shape and Surface Structure in Friction** 591
References 594

18.1 Introduction: Types of Surface Forces

In this chapter, we discuss the most important types of surface forces and the relevant equations for the force and friction laws. Several different attractive and repulsive forces operate between surfaces and particles. Some forces occur in vacuum, for example, attractive van der Waals and repulsive hard-core interactions. Other types of forces can arise only when the interacting surfaces are separated by another condensed phase, which is usually a liquid. The most common types of surface forces and their main characteristics are listed in Table 18.1.

In *vacuum*, the two main long-range interactions are the attractive van der Waals and electrostatic (Coulomb) forces. At smaller surface separations (corresponding to molecular contact at surface separations of $D \approx 0.2$ nm), additional attractive interactions can be found such as covalent or metallic bonding forces. These attractive forces are stabilized by the hard-core repulsion. Together they determine the surface and interfacial energies of planar surfaces, as well as the strengths of materials and adhesive junctions. Adhesion forces are often strong enough to elastically or plastically deform bodies or particles when they come into contact.

In *vapors* (e.g., atmospheric air containing water and organic molecules), solid surfaces in or close to contact will generally have a surface layer of chemisorbed or physisorbed molecules, or a capillary condensed liquid bridge between them. A surface layer usually causes the adhesion to decrease, but in the case of capillary condensation, the additional Laplace pressure or attractive "capillary" force may make the adhesion between the surfaces stronger than in inert gas or vacuum.

When totally immersed in a *liquid*, the force between particles or surfaces is completely modified from that in vacuum or air (vapor). The van der Waals attraction is generally reduced, but other forces can now arise that can qualitatively change both the range and even the sign of the interaction. The attractive force in such a system can be either stronger or weaker than in the absence of the intervening liquid. For example, the overall attraction can be stronger in the case of two hydrophobic surfaces separated by water, but weaker for two hydrophilic surfaces. Depending on the different forces that may be operating simultaneously in solution, the overall force law is not generally monotonically attractive even at long range; it can be repulsive, or the force can change sign at some finite surface separation. In such cases, the potential energy minimum, which determines the adhesion force or energy, occurs not at true molecular contact between the surfaces, but at some small distance farther out.

The forces between surfaces in a liquid medium can be particularly complex at *short range*, i. e., at surface separations below a few nanometers or 4–10 molecular diameters. This is partly because with increasing confinement, a liquid ceases to behave as a structureless continuum with bulk properties; instead, the size and shape of its molecules begin to determine the overall interaction. In addition, the surfaces themselves can no longer be treated as inert and structureless walls (i. e., mathematically flat) and their physical and chemical properties at the atomic scale must now be taken into account. The force laws will then depend on whether the surfaces are amorphous or crystalline (and whether the lattices of crystalline surfaces are matched or not), rough or smooth, rigid or soft (fluid-like), and hydrophobic or hydrophilic.

It is also important to distinguish between *static* (i. e., equilibrium) interactions and *dynamic* (i. e., nonequilibrium) forces such as viscous and friction forces. For example, certain liquid films confined between two contacting surfaces may take a surprisingly long time to equilibrate, as may the surfaces themselves, so that the short-range and adhesion forces appear to be time-dependent, resulting in "aging" effects.

Table 18.1 Types of surface forces in vacuum vs. in liquid (colloidal forces)

Type of force	Subclasses or alternative names	Main characteristics
	Attractive forces	
van der Waals	Debye induced dipole force (v & s) London dispersion force (v & s) Casimir force (v & s)	Ubiquitous, occurs both in vacuum and in liquids
Electrostatic	Ionic bond (v) Coulombic force (v & s) Hydrogen bond (v) Charge-exchange interaction (v & s) Acid–base interaction (s) "Harpooning" interaction (v)	Strong, long-range, arises in polar solvents; requires surface charging or charge-separation mechanism
Ion correlation	van der Waals force of polarizable ions (s)	Requires mobile charges on surfaces in a polar solvent
Quantum mechanical	Covalent bond (v) Metallic bond (v) Exchange interaction (v)	Strong, short-range, responsible for contact binding of crystalline surfaces
Solvation	Oscillatory force (s) Depletion force (s)	Mainly entropic in origin, the oscillatory force alternates between attraction and repulsion
Hydrophobic	Attractive hydration force (s)	Strong, apparently long-range; origin not yet understood
Specific binding	"Lock-and-key" or complementary binding (v & s) Receptor–ligand interaction (s) Antibody–antigen interaction (s)	Subtle combination of different non-covalent forces giving rise to highly specific binding; main recognition mechanism of biological systems
	Repulsive forces	
van der Waals	van der Waals disjoining pressure (s)	Arises only between dissimilar bodies interacting in a medium
Electrostatic	Coulombic force (v & s)	Arises only for certain constrained surface charge distributions
Quantum mechanical	Hard-core or steric repulsion (v) Born repulsion (v)	Short-range, stabilizing attractive covalent and ionic binding forces, effectively determine molecular size and shape
Solvation	Oscillatory solvation force (s) Structural force (s) Hydration force (s)	Monotonically repulsive forces, believed to arise when solvent molecules bind strongly to surfaces
Entropic	Osmotic repulsion (s) Double-layer force (s) Thermal fluctuation force (s) Steric polymer repulsion (s) Undulation force (s) Protrusion force (s)	Due to confinement of molecular or ionic species; requires mechanism that keeps trapped species between the surfaces
	Dynamic interactions	
Non-equilibrium	Hydrodynamic forces (s) Viscous forces (s) Friction forces (v & s) Lubrication forces (s)	Energy-dissipating forces occurring during relative motion of surfaces or bodies

Note: (v) applies only to interactions in vacuum, (s) applies only to interactions in solution (or to surfaces separated by a liquid), and (v & s) applies to interactions occurring both in vacuum and in solution.

18.2 Methods Used to Study Surface Forces

18.2.1 Force Laws

The full force law $F(D)$ between two surfaces, i. e., the force F as a function of surface separation D, can be measured in a number of ways [18.2–6]. The simplest is to move the base of a spring by a known amount, ΔD_0. Figure 18.1 illustrates this method when applied to the interaction of two magnets. However, the method is applicable also at the microscopic or molecular level, and it forms the basis of all direct force-measuring apparatuses such as the surface forces apparatus (SFA; [18.3, 7]) and the atomic force microscope (AFM; [18.8–10]). If there is a detectable force between the surfaces, this will cause the force-measuring spring to deflect by ΔD_s, while the surface separation changes by ΔD. These three displacements are related by

$$\Delta D_s = \Delta D_0 - \Delta D \,. \tag{18.1}$$

The difference in force, ΔF, between the initial and final separations is given by

$$\Delta F = k_s \Delta D_s \,, \tag{18.2}$$

where k_s is the spring constant. The equations above provide the basis for measurements of the force difference between any two surface separations. For example, if a force-measuring apparatus with a known k_s can measure D (and thus ΔD), ΔD_0, and ΔD_s, the force difference ΔF can be measured between a large initial or reference separation D, where the force is zero ($F = 0$), and another separation $D - \Delta D$. By working one's way in increasing increments of $\Delta D = \Delta D_0 - \Delta D_s$, the full force law $F(D)$ can be constructed over any desired distance regime.

In order to measure an equilibrium force law, it is essential to establish that the two surfaces have stopped moving before the displacements are measured. When displacements are measured while two surfaces are still in relative motion, one also measures a viscous or frictional contribution to the total force. Such dynamic force measurements have enabled the viscosities of liquids near surfaces and in thin films to be accurately determined [18.11–13].

In practice, it is difficult to measure the forces between two perfectly flat surfaces, because of the stringent requirement of perfect alignment for making reliable measurements at distances of a few tenths of a nanometer. It is far easier to measure the forces between curved surfaces, e.g., two spheres, a sphere and a flat surface, or two crossed cylinders. Furthermore, the force $F(D)$ measured between two curved surfaces can be directly related to the energy per unit area $E(D)$ between two flat surfaces at the same separation, D, by the so-called Derjaguin approximation [18.14]:

$$E(D) = \frac{F(D)}{2\pi R} \,, \tag{18.3}$$

where R is the radius of the sphere (for a sphere and a flat surface) or the radii of the cylinders (for two crossed cylinders).

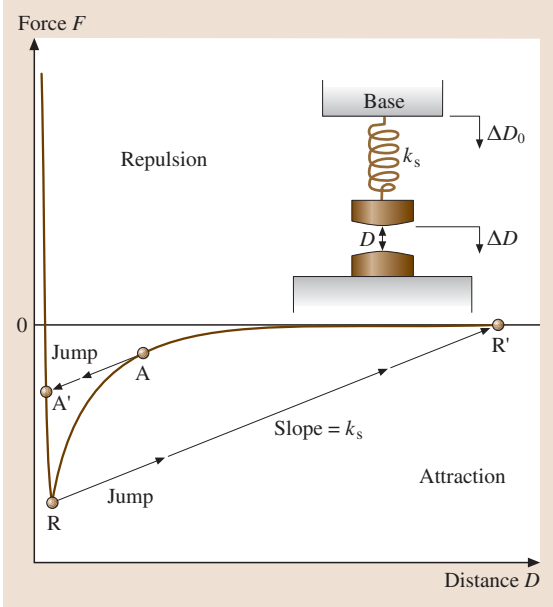


Fig. 18.1 Schematic attractive force law between two macroscopic objects such as two magnets, or between two microscopic objects such as the van der Waals force between a metal tip and a surface. On lowering the base supporting the spring, the latter will expand or contract such that at any equilibrium separation D the attractive force balances the elastic spring restoring force. If the gradient of the attractive force $\mathrm{d}F/\mathrm{d}D$ exceeds the gradient of the spring's restoring force (defined by the spring constant k_s), the upper surface will jump from A into contact at A′ (A for "advancing"). On separating the surfaces by raising the base, the two surfaces will jump apart from R to R′ (R for "receding"). The distance R−R′ multiplied by k_s gives the adhesion force, i. e., the value of F at the point R (after [18.1] with permission)

18.2.2 Adhesion Forces

The most direct way to measure the adhesion of two solid surfaces (such as two spheres or a sphere on a flat) is to suspend one of them on a spring and measure the adhesion or "pull-off" force needed to separate the two bodies from its deflection. If k_s is the stiffness of the force-measuring spring and ΔD the distance the two surfaces jump apart when they separate, then the adhesion force F_s is given by

$$F_s = F_{max} = k_s \Delta D \,, \tag{18.4}$$

where we note that in liquids, the maximum or minimum in the force may occur at some nonzero surface separation (see Fig. 18.7). From F_s and a known surface geometry, and assuming that the surfaces were everywhere in molecular contact, one may also calculate the surface or interfacial energy γ. For an elastically deformable sphere of radius R on a flat surface, or for two crossed cylinders of radius R, we have [18.3, 15]

$$\gamma = \frac{F_s}{3\pi R} \,, \tag{18.5}$$

while for two spheres of radii R_1 and R_2

$$\gamma = \frac{F_s}{3\pi}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \,, \tag{18.6}$$

where γ is in units of $\mathrm{J\,m^{-2}}$ (see Sect. 18.5.2).

18.2.3 The SFA and AFM

In a typical force-measuring experiment, at least two of the above displacement parameters – ΔD_0, ΔD, and ΔD_s – are directly or indirectly measured, and from these the third displacement and the resulting force law $F(D)$ are deduced using (18.1) and (18.2) together with a measured value of k_s. For example, in SFA experiments, ΔD_0 is changed by expanding or contracting a piezoelectric crystal by a known amount or by moving the base of the spring with sensitive motor-driven mechanical stages. The resulting change in surface separation ΔD is measured optically, and the spring deflection ΔD_s can then be obtained according to (18.1). In AFM experiments, ΔD_0 and ΔD_s are measured using a combination of piezoelectric, optical, capacitance or magnetic techniques, from which the change in surface separation ΔD is deduced. Once a force law is established, the geometry of the two surfaces (e.g., their radii) must also be known before the results can be compared with theory or with other experiments.

The SFA (Fig. 18.2) is used for measurements of adhesion and force laws between two curved molecularly smooth surfaces immersed in liquids or controlled vapors [18.3, 7, 16]. The surface separation is measured by multiple beam interferometry with an accuracy of ± 0.1 nm. From the shape of the interference fringes one also obtains the radius of the surfaces, R, and any surface deformation that arises during an interaction [18.17–19]. The resolution in the lateral direction is about 1 μm. The surface separation can be independently controlled to within 0.1 nm, and the force sensitivity is about 10^{-8} N. For a typical surface radius of $R \approx 1$ cm, γ values can be measured to an accuracy of about $10^{-3}\,\mathrm{mJ\,m^{-2}}$.

Several different materials have been used to form the surfaces in the SFA, including mica [18.20, 21], silica [18.22], sapphire [18.23], and polymer sheets [18.24]. These materials can also be used as supporting substrates in experiments on the forces between adsorbed or chemically bound polymer layers [18.13, 25–30], surfactant and lipid monolayers and bilayers [18.31–34], and metal and metal oxide layers [18.35–42]. The range of liquids and vapors that can be used is almost endless, and have thus far included aqueous solutions, organic liquids and solvents, polymer melts, various petroleum oils and lubricant liquids, and liquid crystals.

Friction attachments for the SFA [18.43–47] allow for the two surfaces to be sheared laterally past each other at varying sliding speeds or oscillating frequencies, while simultaneously measuring both the transverse (frictional or shear) force and the normal force (load) between them. The ranges of friction forces and sliding speeds that can be studied with such methods are currently 10^{-7}–10^{-1} N and 10^{-13}–$10^{-2}\,\mathrm{m\,s^{-1}}$, respectively [18.48]. The externally applied load, L, can be varied continuously, and both positive and negative loads can be applied. The distance between the surfaces, D, their true molecular contact area, their elastic (or viscoelastic or elastohydrodynamic) deformation, and their lateral motion can all be monitored simultaneously by recording the moving interference fringe pattern.

In the atomic force microscope (Fig. 18.3), the force is measured by monitoring the deflection of a soft cantilever supporting a sub-microscopic tip ($R \approx 10$–200 nm) as this is interacting with a flat, macroscopic surface [18.8, 49, 50]. The measurements can be done in a vapor or liquid. The normal (bending) spring stiffness of the cantilever can be as small as $0.01\,\mathrm{N\,m^{-1}}$, allowing measurements of normal forces as small as 1 pN (10^{-12} N), which corresponds to the bond strength

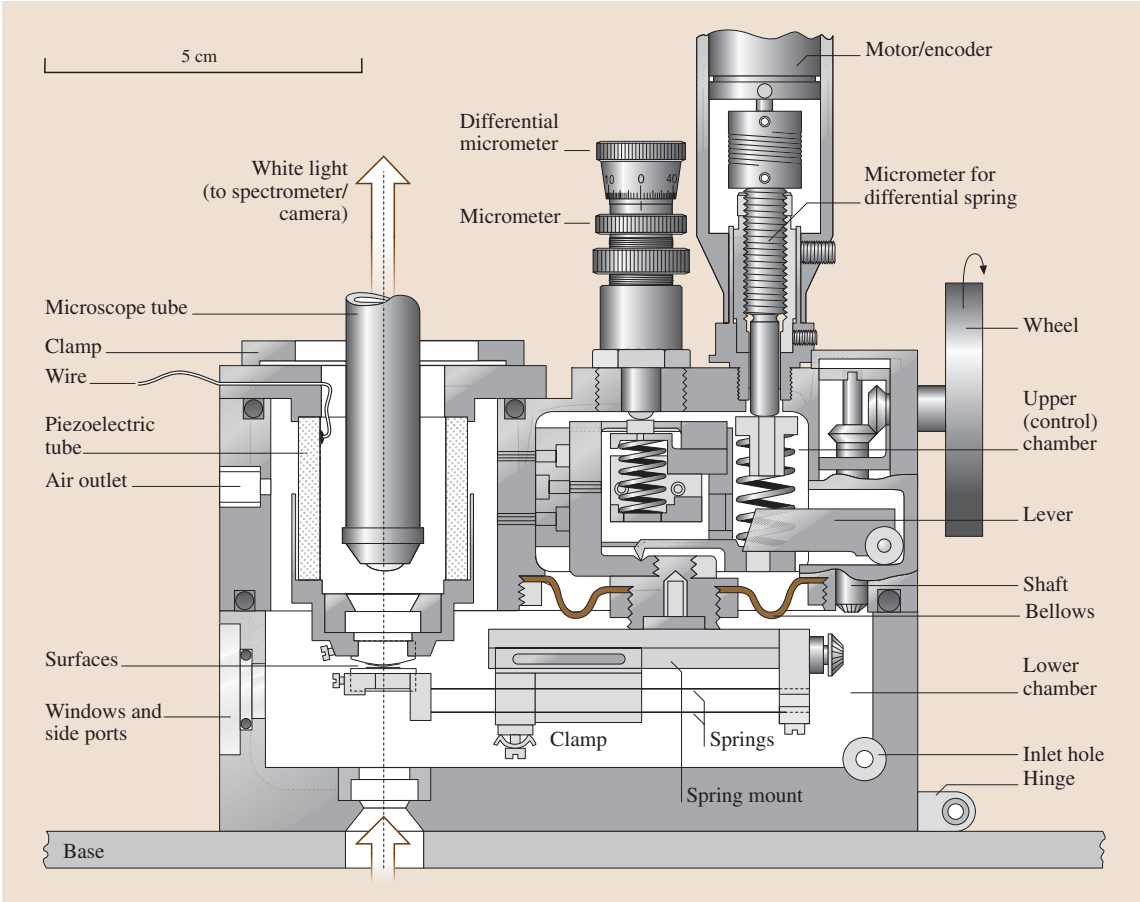


Fig. 18.2 A surface forces apparatus (SFA) where the intermolecular forces between two macroscopic, cylindrical surfaces of local radius R can be directly measured as a function of surface separation over a large distance regime from tenths of a nanometer to micrometers. Local or transient surface deformations can be detected optically. Various attachments for moving one surface laterally with respect to the other have been developed for friction measurements in different regimes of sliding velocity and sliding distance (after [18.16] with permission)

of single molecules [18.51, 52]. Distances can be inferred with an accuracy of about 1 nm, and changes in distance can be measured to about 0.1 nm. Since the contact area is small, different interaction regimes can be resolved on samples with a heterogeneous composition on lateral scales of about ten nanometers. Height differences and the roughness of the sample can be measured directly from the cantilever deflection or, alternatively, by using a feedback system to raise or lower the sample so that the deflection (the normal force) is kept constant during a scan over the area of interest. Weak interaction forces and larger (microscopic) interaction areas can be investigated by replacing the tip with a micrometer-sized sphere to form a "colloidal probe" [18.9].

The atomic force microscope can also be used for friction measurements (lateral force microscopy, LFM) by monitoring the torsion of the cantilever as the sample is scanned in the direction perpendicular to the long axis of the cantilever [18.10, 50, 53, 54]. Typically, the stiffness of the cantilever to lateral bending is much larger than to bending in the normal direction and to torsion, so that these signals are decoupled and height and friction can be detected simultaneously. The torsional spring constant can be as low as $0.1\,\mathrm{N\,m^{-1}}$, giving a lateral (friction) force sensitivity of 10^{-11} N.

Rapid technical developments have facilitated the calibrations of the normal [18.55, 56] and lateral spring constants [18.54, 57], as well as in situ measurements of

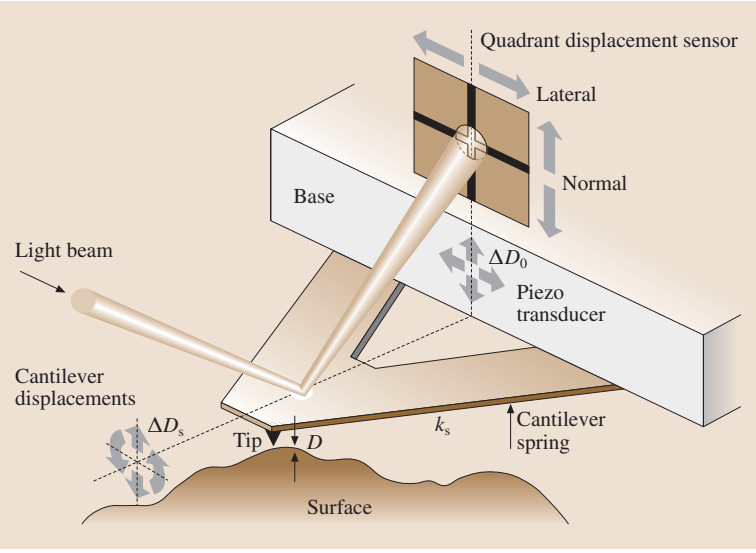


Fig. 18.3 Schematic drawing of an atomic force microscope (AFM) tip supported on a triangular cantilever interacting with an arbitrary solid surface. The normal force and topology are measured by monitoring the calibrated deflection of the cantilever as the tip is moved across the surface by means of a piezoelectric transducer. Various designs have been developed that move either the sample or the cantilever during the scan. Friction forces can be measured from the torsion of the cantilever when the scanning is in the direction perpendicular to its long axis (after [18.62] with permission)

the macroscopic tip radius [18.58]. Cantilevers of different shapes with a large range of spring constants, tip radii, and surface treatments (inorganic or organic coatings) are commercially available. The flat surface, and also the particle in the colloidal probe technique, can be any material of interest. However, remaining difficulties with this technique are that the distance between the tip and the substrate, D, and the deformations of the tip and sample are not directly measurable. Another important difference between the AFM/LFM and SFA techniques is the different size of the contact area, and the related observation that even when a cantilever with a very low spring constant is used in the AFM, the pressure in the contact zone is typically much higher than in the SFA. Hydrodynamic effects in liquids also affect the measurements of normal forces differently on certain time scales [18.59–61].

18.2.4 Some Other Force-Measuring Techniques

A large number of other techniques are available for the measurements of the normal forces between solid or fluid surfaces (see [18.5, 62]). The techniques discussed in this section are not used for lateral (friction) force measurements, but are commonly used to study normal forces, particularly in biological systems.

Micropipette aspiration is used to measure the forces between cells or vesicles, or between a cell or vesicle and another surface [18.63–65]. The cell or vesicle is held by suction at the tip of a glass micropipette and deforms elastically in response to the net interactions with another surface and to the applied suction. The shape of the deformed surface (cell membrane) is measured and used to deduce the force between the surfaces and the membrane tension [18.64]. The membrane tension, and thus the stiffness of the cell or vesicle, is regulated by applying different hydrostatic pressures. Forces can be measured in the range of 0.1 pN to 1 nN, and the distance resolution is a few nanometers. The interactions between a colloidal particle and another surface can be studied by attaching the particle to the cell membrane [18.66].

In the osmotic stress technique, pressures are measured between colloidal particles in aqueous solution, membranes or bilayers, or other ordered colloidal structures (viruses, DNA). The separation between the particle surfaces and the magnitude of membrane undulations are measured by X-ray or neutron scattering techniques. This is combined with a measurement of the osmotic pressure of the solution [18.67–70]. The technique has been used to measure repulsive forces, such as DLVO interactions, steric forces, and hydration forces [18.71]. The sensitivity in pressure is 0.1 mN m^{-2}, and distances can be resolved to 0.1 nm.

The optical tweezers technique is based on the trapping of dielectric particles at the center of a focused laser beam by restoring forces arising from radiation pressure and light intensity gradients [18.72, 73]. The forces experienced by particles as they are moved toward or away from one another can be measured with a sensitivity in the pN range. Small biological molecules are typically attached to a larger bead of a material

with suitable refractive properties. Recent development allows determinations of position with nanometer resolution [18.74], which makes this technique useful for studying the forces during extension of single molecules.

In total internal reflection microscopy (TIRM), the potential energy between a micrometer-sized colloidal particle and flat surface in aqueous solution is deduced from the average equilibrium height of the particle above the surface, measured from the intensity of scattered light. The average height ($D \approx 10$–100 nm) results from a balance of gravitational force, radiation pressure from a laser beam focused at the particle from below, and intermolecular forces [18.75]. The technique is particularly suitable for measuring weak forces (sensitivity ca. 10^{-14} N), but is more difficult to use for systems with strong interactions. A related technique is reflection interference contrast microscopy (RICM), where optical interference is used to also monitor changes in the shape of the approaching colloidal particle or vesicle [18.76].

An estimate of bond strengths can be obtained from the hydrodynamic shear force exerted by a fluid on particles or cells attached to a substrate [18.77, 78]. At a critical force, the bonds are broken and the particle or cell will be detached and move with the velocity of the fluid. This method requires knowledge of the contact area and the flow velocity profile of the fluid. Furthermore, a uniform stress distribution in the contact area is generally assumed. At low bond density, this technique can be used to determine the strength of single bonds (1 pN).

18.3 Normal Forces Between Dry (Unlubricated) Surfaces

18.3.1 Van der Waals Forces in Vacuum and Inert Vapors

Forces between macroscopic bodies (such as colloidal particles) across vacuum arise from interactions between the constituent atoms or molecules of each body across the gap separating them. These intermolecular interactions are electromagnetic forces between permanent or induced dipoles (van der Waals forces), and between ions (electrostatic forces). In this section, we describe the van der Waals forces, which occur between all atoms and molecules and between all macroscopic bodies (see [18.3]).

The interaction between two permanent dipoles with a fixed relative orientation can be attractive or repulsive. For the specific case of two freely rotating permanent dipoles in a liquid or vapor (orientational or Keesom interaction), and for a permanent dipole and an induced dipole in an atom or polar or nonpolar molecule (induction or Debye interaction), the interaction is on average always attractive. The third type of van der Waals interaction, the fluctuation or London dispersion interaction, arises from instantaneous polarization of one nonpolar or polar molecule due to fluctuations in the charge distribution of a neighboring nonpolar or polar molecule (Fig. 18.4a). Correlation between these fluctuating induced dipole moments gives an attraction that is present between any two molecules or surfaces across vacuum. At very small separations, the interaction will ultimately be repulsive as the electron clouds of atoms and molecules begin to overlap. The total interaction is thus a combination of a short-range repulsion and a relatively long-range attraction.

Except for in highly polar materials such as water, London dispersion interactions give the largest contribution (70–100 %) to the van der Waals attraction. The interaction energy of the van der Waals force between atoms or molecules depends on the separation r as

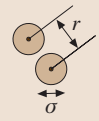
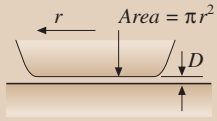
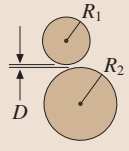
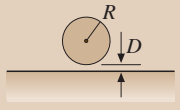
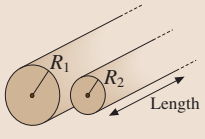
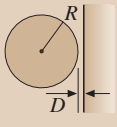
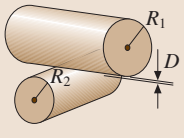
$$E(D) = \frac{-C_{\mathrm{vdW}}}{r^6} , \tag{18.7}$$

where the constant C_{vdW} depends on the dipole moments and polarizabilities of the molecules. At large separations (> 10 nm), the London interaction is weakened by a randomizing effect caused by the rapid fluctuations. That is, the induced temporary dipole moment of one molecule may have changed during the time needed for the transmission of the electromagnetic wave (photon) generated by its fluctuating charge density to another molecule and the return of the photon generated by the induced fluctuation in this second molecule. This phenomenon is called retardation and causes the interaction energy to decay as r^{-7} at large separations [18.79].

Dispersion interactions are to a first approximation additive, and their contribution to the interaction energy between two macroscopic bodies (such as colloidal particles) across vacuum can be found by summing the

pair-wise interactions [18.80]. The interaction is generally described in terms of the Hamaker constant, A_H. Another approach is to treat the interacting bodies and an intervening medium as continuous phases and determine the strength of the interaction from bulk dielectric properties of the materials [18.81, 82]. Unlike the pair-wise summation, this method takes into account the screening of the interactions between molecules inside the bodies by the molecules closer to the surfaces and the effects of the intervening medium. For the interaction between material 1 and material 3 across material 2, the non-retarded Hamaker constant given by the Lifshitz theory

Table 18.2 Van der Waals interaction energy and force between macroscopic bodies of different geometries

Geometry of bodies with surfaces D apart ($D \ll R$)			van der Waals interaction Energy, E	van der Waals interaction Force, F
Two atoms or small molecules		$r \geqslant \sigma$	$\frac{-C_{vdW}}{r^6}$	$\frac{-6C_{vdW}}{r^7}$
Two flat surfaces (per unit area)		$r \gg D$	$\frac{-A_H}{12\pi D^2}$	$\frac{-A_H}{6\pi D^3}$
Two spheres or macromolecules of radii R_1 and R_2		$R_1, R_2 \gg D$	$\frac{-A_H}{6D}\left(\frac{R_1 R_2}{R_1+R_2}\right)$	$\frac{-A_H}{6D^2}\left(\frac{R_1 R_2}{R_1+R_2}\right)$
Sphere or macromolecule of radius R near a flat surface		$R \gg D$	$\frac{-A_H R}{6D}$	$\frac{-A_H R}{6D^2}$
Two parallel cylinders or rods of radii R_1 and R_2 (per unit length)		$R_1, R_2 \gg D$	$\frac{-A_H}{12\sqrt{2}\,D^{3/2}}\left(\frac{R_1 R_2}{R_1+R_2}\right)^{1/2}$	$\frac{-A_H}{8\sqrt{2}\,D^{5/2}}\left(\frac{R_1 R_2}{R_1+R_2}\right)^{1/2}$
Cylinder of radius R near a flat surface (per unit length)		$R \gg D$	$\frac{-A_H\sqrt{R}}{12\sqrt{2}\,D^{3/2}}$	$\frac{-A_H\sqrt{R}}{8\sqrt{2}\,D^{5/2}}$
Two cylinders or filaments of radii R_1 and R_2 crossed at 90°		$R_1, R_2 \gg D$	$\frac{-A_H\sqrt{R_1 R_2}}{6D}$	$\frac{-A_H\sqrt{R_1 R_2}}{6D^2}$

A negative force (A_H positive) implies attraction, a positive force means repulsion (A_H negative) (after [18.62] with permission)

is approximately [18.3]:

$$A_{\mathrm{H},123} = A_{\mathrm{H},\nu=0} + A_{\mathrm{H},\nu>0}$$
$$\approx \tfrac{3}{4} k_{\mathrm{B}} T \left(\tfrac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}\right)\left(\tfrac{\varepsilon_3-\varepsilon_2}{\varepsilon_3+\varepsilon_2}\right)$$
$$+ \frac{3h\nu_{\mathrm{e}}}{8\sqrt{2}} \frac{(n_1^2-n_2^2)(n_3^2-n_2^2)}{\sqrt{(n_1^2+n_2^2)}\sqrt{(n_3^2+n_2^2)}\left(\sqrt{(n_1^2+n_2^2)}+\sqrt{(n_3^2+n_2^2)}\right)}, \quad (18.8)$$

where the first term ($\nu = 0$) represents the permanent dipole and dipole–induced dipole interactions and the second ($\nu > 0$) the London (dispersion) interaction. ε_i and n_i are the static dielectric constants and refractive indexes of the materials, respectively. ν_{e} is the frequency of the lowest electron transition (around $3\times10^{15}\,\mathrm{s}^{-1}$). Either one of the materials 1, 2, or 3 in (18.8) can be vacuum or air ($\varepsilon = n = 1$). A_{H} is typically 10^{-20}–10^{-19} J (the higher values are found for metals) for interactions between solids and liquids across vacuum or air.

The interaction energy between two macroscopic bodies is dependent on the geometry and is always attractive between two bodies of the same material [A_{H} positive, see (18.8)]. The van der Waals interaction energy and force laws ($F = -\mathrm{d}E(D)/\mathrm{d}D$) for some common geometries are given in Table 18.2. Because of the retardation effect, the equations in Table 18.2 will lead to an overestimation of the dispersion force at large separations. It is, however, apparent that the interaction energy between macroscopic bodies decays more slowly with separation (i. e., has a longer range) than between two molecules.

For inert nonpolar surfaces, e.g., consisting of hydrocarbons or van der Waals solids and liquids, the Lifshitz theory has been found to apply even at molecular contact, where it can be used to predict the surface energies (surface tensions) of such solids and liquids. For example, for hydrocarbon surfaces, $A_{\mathrm{H}} = 5\times10^{-20}$ J. Inserting this value into the equation for two flat surfaces (Table 18.2) and using a "cut-off" distance of $D_0 \approx 0.165$ nm as an effective separation when the surfaces are in contact [18.3], we obtain for the surface energy γ (which is defined as half the interaction energy)

$$\gamma = \frac{E}{2} = \frac{A_{\mathrm{H}}}{24\pi D_0^2} \approx 24\,\mathrm{mJ\,m^{-2}}\,, \quad (18.9)$$

a value that is typical for hydrocarbon solids and liquids [18.83].

If the adhesion force is measured between a spherical surface of radius $R = 1$ cm and a flat surface using an SFA, we expect the adhesion force to be (see Table 18.2) $F = A_{\mathrm{H}}R/(6D_0^2) = 4\pi R\gamma \approx 3.0$ mN. Using a spring constant of $k_{\mathrm{s}} = 100\,\mathrm{N\,m^{-1}}$, such an adhesive force will cause the two surfaces to jump apart by $\Delta D = F/k_{\mathrm{s}} = 30\,\mu$m, which can be accurately measured. (For elastic bodies that deform in adhesive contact, R changes during the interaction and the measured adhesion force is 25% lower, see Sect. 18.5.2). Surface energies of solids can thus be directly measured with the SFA and, in principle, with the AFM if the contact geometry can be quantified. The measured values are in good agreement with calculated values based on the known surface energies γ of the materials, and for nonpolar low-energy solids they are well accounted for by the Lifshitz theory [18.3].

18.3.2 Charge Exchange Interactions

Electrostatic interactions are present between ions (Coulomb interactions), between ions and permanent dipoles, and between ions and nonpolar molecules in which a charge induces a dipole moment. The interaction energy between ions or between a charge and a fixed permanent dipole can be attractive or repulsive. For an induced dipole or a freely rotating permanent dipole in vacuum or air, the interaction energy with a charge is always attractive.

Spontaneous charge transfer may occur between two dissimilar materials in contact [18.84]. The phenomenon is especially prominent in contacts between a metal, for example, mercury, and a material with low conductivity, but is also observed, for example, between two different polymer layers. During separation, rolling or sliding of one body over the other, the surfaces experience both charge transition from one surface to the other and charge transfer (conductance) along each surface (Fig. 18.4b). The latter process is typically slower, and, as a result, charges remain on the surfaces as they are separated in vacuum or dry nitrogen gas. The charging gives rise to a strong adhesion with adhesion energies of over $1{,}000\,\mathrm{mJ\,m^{-2}}$, similar to fracture or cohesion energies of the solid bodies themselves [18.84, 85]. Upon separating the surfaces farther apart, a strong, long-range electrostatic attraction is observed. The charging can be decreased through discharges across the gap between the surfaces (which requires a high charging) or through conducting in the solids. It has been suggested that charge exchange interactions are particularly important in rolling friction between dry surfaces (which can simplistically be thought of as an adhesion–separation process), where the distance dependence of forces acting normally to the surfaces plays a larger role than

in sliding friction. Recent experiments on the sliding friction between metal–insulator surfaces indicate that stick–slip would be accompanied by charge transfer events [18.86].

Photo-induced charge transfer, or harpooning, involves the transfer of an electron between an atom in a molecular beam or at a solid surface (typically alkali or transition metal) to an atom or molecule in a gas (typically halides) to form a negatively charged molecular ion in a highly excited vibrational state. This transfer process can occur at atomic distances of 0.5–0.7 nm, which is far from molecular contact. The formed molecular ion is attracted to the surface and chemisorbs onto it. Photo-induced charge transfer processes also occur in the photosynthesis in green plants and in photoelectrochemical cells (solar cells) at the junction between two semiconductors or between a semiconductor and an electrolyte solution [18.87].

18.3.3 Sintering and Cold Welding

When macroscopic particles in a powder or in a suspension come into molecular contact, they can bond together to form a network or solid body with very different density and shear strength compared to the powder (a typical example is porcelain). The rate of bonding is dependent on the surface energy (causing a stress at the edge of the contact) and the atomic mobility (diffusion rate) of the contacting materials. To increase the diffusion rate, objects formed from powders are heated to about one-half of the melting temperature of the components in a process called sintering, which can be done in different atmospheres or in a liquid.

In the sintering process, the surface energy of the system is lowered due to the reduction of total surface area (Fig. 18.4c). In metal and ceramic systems, the most important mechanism is solid-state diffusion, initially surface diffusion. As the surface area decreases and the grain boundaries increase at the contacts, grain boundary diffusion and diffusion through the crystal lattice become more important. The grain boundaries will eventually migrate, so that larger particles are formed (coarsening). Mass can also be transferred through evaporation and condensation, and through viscous and plastic flow. In liquid-phase sintering, the materials can melt, which increases the mass transport. Amorphous materials like polymers and glasses do not have real grain boundaries and sinter by viscous flow [18.88].

Some of these mechanisms (surface diffusion and evaporation–condensation) reduce the surface area and increase the grain size (coarsening) without densification, in contrast to bulk transport mechanisms like grain boundary diffusion and plastic and viscous flow. As the material becomes denser, elongated pores collapse to form smaller, spherical pores with a lower surface energy. Models for sintering typically consider the size and growth rate of the grain boundary (the "neck") formed between two spherical particles. At a high stage of densification, the sintering stress σ at the curved neck between two particles is given by [18.88]

$$\sigma = \frac{2\gamma_{SS}}{G} + \frac{2\gamma_{SV}}{r_p} , \tag{18.10}$$

where γ_{SS} is the solid–solid grain boundary energy, γ_{SV} is the solid–vapor surface energy, G is the grain size, and r_p is the radius of the pore.

A related phenomenon is cold welding, which is the spontaneous formation of strong junctions between clean (unoxidized) metal surfaces with a mutual solubility when they are brought in contact with or without an applied pressure. The plastic deformations accompanying the formation and breaking of such contacts on a molecular scale during motion of one surface normally (see Fig. 18.10c,d) or laterally (shearing) with respect to the other have been studied both experimentally [18.89] and theoretically [18.90–95]. The breaking of a cold welded contact is generally associated with damage or deformation of the surface structure.

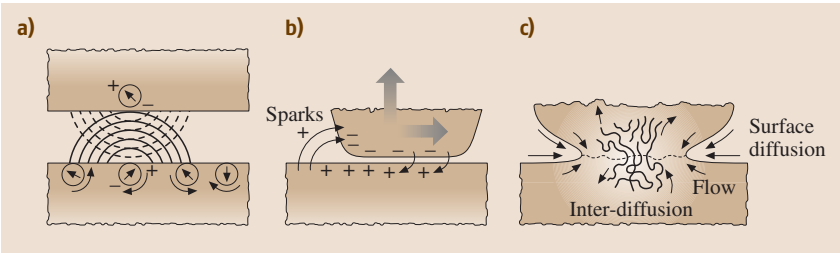


Fig. 18.4a–c Schematic representation of **(a)** van der Waals interaction (dipole–induced dipole interaction), **(b)** charge exchange, which acts to increase adhesion and friction forces, and **(c)** sintering between two surfaces

18.4 Normal Forces Between Surfaces in Liquids

18.4.1 Van der Waals Forces in Liquids

The dispersion interaction in a medium will be significantly lower than in vacuum, since the attractive interaction between two solute molecules in a medium (solvent) involves displacement and reorientation of the nearest-neighbor solvent molecules. Even though the surrounding medium may change the dipole moment and polarizability from that in vacuum, the interaction between two identical molecules remains attractive in a binary mixture. The extension of the interactions to the case of two macroscopic bodies is the same as described in Sect. 18.3.1. Typically, the Hamaker constants for interactions in a medium are an order of magnitude lower than in vacuum. Between macroscopic surfaces in liquids, van der Waals forces become important at distances below 10–15 nm and may at these distances start to dominate interactions of different origin that have been observed at larger separations.

Figure 18.5 shows the measured van der Waals forces between two crossed cylindrical mica surfaces in water and various salt solutions. Good agreement is obtained between experiment and theory. At larger surface separations, above about 5 nm, the measured forces fall off more rapidly than as D^{-2}. This retardation effect (see Sect. 18.3.1) is also predicted by Lifshitz theory and is due to the time needed for propagation of the induced dipole moments over large distances.

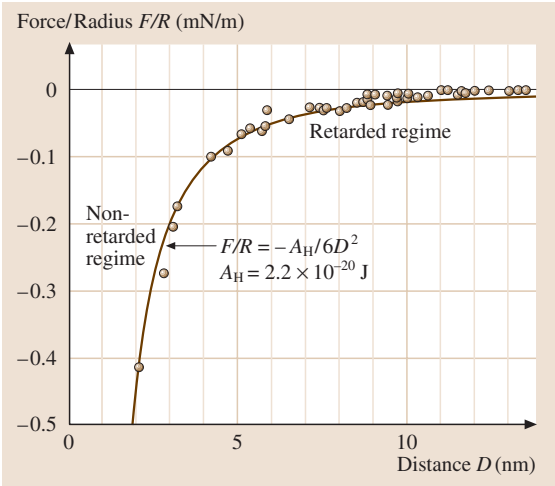


Fig. 18.5 Attractive van der Waals force F between two curved mica surfaces of radius $R \approx 1$ cm measured in water and various aqueous electrolyte solutions. The electrostatic interaction has been subtracted from the total measured force. The measured non-retarded Hamaker constant is $A_H = 2.2 \times 10^{-20}$ J. Retardation effects are apparent at distances larger than 5 nm, as expected theoretically. (After [18.3]. Copyright 1991, with permission from Elsevier Science)

From Fig. 18.5, we may conclude that at separations above about 2 nm, or 8 molecular diameters of water, the *continuum* Lifshitz theory is valid. This would mean that water films as thin as 2 nm may be expected to have bulk-like properties, at least as far as their interaction forces are concerned. Similar results have been obtained with other liquids, where in general continuum properties are manifested, both as regards their interactions and other properties such as viscosity, at a film thickness larger than 5 or 10 molecular diameters. In the absence of a solvent (in vacuum), the agreement of measured van der Waals forces with the continuum Lifshitz theory is generally good at all separations down to molecular contact ($D = D_0$).

Van der Waals interactions in a system of three or more different materials (see (18.8)) can be attractive or repulsive, depending on their dielectric properties. Numerous experimental studies show the attractive van der Waals forces in various systems [18.3], and also repulsive van der Waals forces have been measured directly [18.96]. A practical consequence of the repulsive interaction obtained across a medium with intermediate dielectric properties is that the van der Waals forces will give rise to preferential, nonspecific adsorption of molecules with an intermediate dielectric constant. This is commonly seen as adsorption of vapors or solutes to a solid surface. It is also possible to diminish the attractive interaction between dispersed colloidal particles by adsorption of a thin layer of material with dielectric properties close to those of the surrounding medium (matching of refractive index), or by adsorption of a polymer that gives a steric repulsive force that keeps the particles separated at a distance where the magnitude of the van der Waals attraction is negligible. Thermal motion will then keep the particles dispersed.

18.4.2 Electrostatic and Ion Correlation Forces

Most surfaces in contact with a highly polar liquid (such as water) acquire a surface charge, either by dissociation of ions from the surface into the solution or by preferential adsorption of certain ions from the solution.

The surface charge is balanced by a layer of oppositely charged ions (counterions) in the solution at some small distance from the surface (see [18.3]). In dilute solution, this distance is the Debye length, κ^{-1}, which is purely a property of the electrolyte *solution*. The Debye length falls with increasing ionic strength (i. e., with the molar concentration M_i and valency z_i) of the ions in solution:

$$\kappa^{-1} = \left(\frac{\varepsilon\varepsilon_0 k_B T}{e^2 N_A \sum_i z_i^2 M_i} \right)^{1/2} , \tag{18.11}$$

where e is the electronic charge. For example, for 1:1 electrolytes at 25 °C, $\kappa^{-1} = 0.304\,\text{nm}/\sqrt{M_{1:1}}$, where M_i is given in M (mol dm^{-3}). κ^{-1} is thus ca. 10 nm in a 1 mM NaCl solution and 0.3 nm in a 1 M solution. In totally pure water at pH 7, where $M_i = 10^{-7}$ M, κ^{-1} is 960 nm, or about 1 µm. The Debye length also relates the surface charge density σ of a surface to the electrostatic surface potential ψ_0 via the Grahame equation, which for 1:1 electrolytes can be expressed as:

$$\sigma = \sqrt{8\varepsilon\varepsilon_0 k_B T} \sinh\left(e\psi_0/2k_B T\right) \times \sqrt{M_{1:1}} . \tag{18.12}$$

Since the Debye length is a measure of the thickness of the diffuse atmosphere of counterions near a charged surface, it also determines the range of the electrostatic "double-layer" interaction between two charged surfaces. The electrostatic double-layer interaction is an entropic effect that arises upon decreasing the thickness of the liquid film containing the dissolved ions. Because of the attractive force between the dissolved ions and opposite charges on the surfaces, the ions stay between the surfaces, but an osmotic repulsion arises as their concentration increases. The long-range electrostatic interaction energy at large separations (weak overlap) between two similarly charged molecules or surfaces is typically repulsive and is roughly an exponentially decaying function of D:

$$E(D) \approx +C_{ES}\, e^{-\kappa D} , \tag{18.13}$$

where C_{ES} is a constant that depends on the geometry of the interacting surfaces, on their surface charge density, and the solution conditions (Table 18.3). We see that the Debye length is the decay length of the interaction energy between two surfaces (and of the mean potential away from one surface). C_{ES} can be determined by solving the so-called Poisson–Boltzmann equation or by using other theories [18.97, 98]. The equations in Table 18.3 are expressed in terms of a constant, Z, defined as

$$Z = 64\pi\varepsilon\varepsilon_0 (k_B T/e)^2 \tanh^2\left[ze\psi_0/(4k_B T)\right] , \tag{18.14}$$

which depends only on the properties of the *surfaces*.

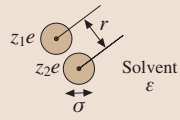
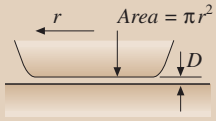
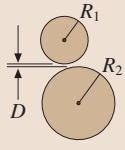
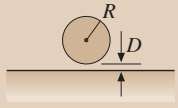
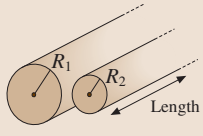
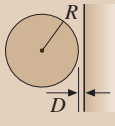
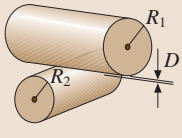
The above approximate expressions are accurate only for surface separations larger than about one Debye length. At smaller separations one must use numerical solutions of the Poisson–Boltzmann equation to obtain the exact interaction potential, for which there are no simple expressions. In the limit of small D, it can be shown that the interaction energy depends on whether the surfaces remain at constant potential ψ_0 (as assumed in the above equations) or at constant charge σ (when the repulsion exceeds that predicted by the above equations), or somewhere between these limits. In the "constant charge limit" the total *number* of counterions in the compressed film does not change as D is decreased, whereas at constant potential, the *concentration* of counterions is constant.The limiting pressure (or force per unit area) at constant charge is the osmotic pressure of the confined ions:

$$\begin{aligned} F &= k_B T \times \text{ion number density} \\ &= 2\sigma k_B T/(zeD), \quad \text{for } D \ll \kappa^{-1} . \end{aligned} \tag{18.15}$$

That is, as $D \to 0$ the double-layer pressure at constant surface charge becomes infinitely repulsive and independent of the salt concentration (at constant potential the force instead becomes a constant at small D). However, at small separations, the van der Waals attraction (which goes as D^{-2} between two spheres or as D^{-3} between two planar surfaces, see Table 18.2) wins out over the double-layer repulsion, unless some other short-range interaction becomes dominant (see Sect. 18.4.4). This is the theoretical prediction that forms the basis of the so-called Derjaguin–Landau–Verwey–Overbeek (DLVO) theory [18.97, 99], illustrated in Fig. 18.6.

Because of the different distance dependence of the van der Waals and electrostatic interactions, the total force law, as described by the DLVO theory, can show several minima and maxima. Typically, the depth of the outer (secondary) minimum is a few $k_B T$, enough to cause reversible flocculation of particles from an aqueous dispersion. If the force barrier between the secondary and primary minimum is lowered, for example, by increasing the electrolyte concentration, particles can be irreversibly coagulated in the primary minimum. In practice, other forces (described

Table 18.3 Electrical "double-layer" interaction energy $E(D)$ and force ($F = -\mathrm{d}E/\mathrm{d}D$) between macroscopic bodies

Geometry of bodies with surfaces D apart ($D \ll R$)			Electric "double-layer" interaction Energy, E	Force, F
Two ions or small molecules		$r \geqslant \sigma$	$\dfrac{+z_1z_2e^2}{4\pi\varepsilon\varepsilon_0 r}\dfrac{\mathrm{e}^{-\kappa(r-\sigma)}}{(1+\kappa\sigma)}$	$\dfrac{+z_1z_2e^2}{4\pi\varepsilon\varepsilon_0 r^2}\dfrac{(1+\kappa r)}{(1+\kappa\sigma)}\mathrm{e}^{-\kappa(r-\sigma)}$
Two flat surfaces (per unit area)		$r \gg D$	$(\kappa/2\pi)\,Z\mathrm{e}^{-\kappa D}$	$\left(\kappa^2/2\pi\right)Z\mathrm{e}^{-\kappa D}$
Two spheres or macromolecules of radii R_1 and R_2		$R_1, R_2 \gg D$	$\left(\dfrac{R_1R_2}{R_1+R_2}\right)Z\mathrm{e}^{-\kappa D}$	$\kappa\left(\dfrac{R_1R_2}{R_1+R_2}\right)Z\mathrm{e}^{-\kappa D}$
Sphere or macromolecule of radius R near a flat surface		$R \gg D$	$RZ\mathrm{e}^{-\kappa D}$	$\kappa RZ\mathrm{e}^{-\kappa D}$
Two parallel cylinders or rods of radii R_1 and R_2 (per unit length)		$R_1, R_2 \gg D$	$\dfrac{\kappa^{1/2}}{\sqrt{2\pi}}\left(\dfrac{R_1R_2}{R_1+R_2}\right)^{1/2}Z\mathrm{e}^{-\kappa D}$	$\dfrac{\kappa^{3/2}}{\sqrt{2\pi}}\left(\dfrac{R_1R_2}{R_1+R_2}\right)^{1/2}Z\mathrm{e}^{-\kappa D}$
Cylinder of radius R near a flat surface (per unit length)		$R \gg D$	$\kappa^{1/2}\sqrt{\dfrac{R}{2\pi}}Z\mathrm{e}^{-\kappa D}$	$\kappa^{3/2}\sqrt{\dfrac{R}{2\pi}}Z\mathrm{e}^{-\kappa D}$
Two cylinders or filaments of radii R_1 and R_2 crossed at 90°		$R_1, R_2 \gg D$	$\sqrt{R_1R_2}Z\mathrm{e}^{-\kappa D}$	$\kappa\sqrt{R_1R_2}Z\mathrm{e}^{-\kappa D}$

The interaction energy and force for bodies of different geometries is based on the Poisson–Boltzmann equation (a continuum, mean-field theory). Equation (18.14) gives the interaction constant Z (in terms of surface potential ψ_0) for the interaction between similarly charged (ionized) surfaces in aqueous solutions of monovalent electrolyte. It can also be expressed in terms of the surface charge density σ by applying the Grahame equation (18.12) (after [18.62] with permission)

in the following sections) often appear at very small separations, so that the full force law between two surfaces or colloidal particles in solution can be more complex than might be expected from the DLVO theory.

There are situations when the double-layer interaction can be attractive at short range even between surfaces of similar charge, especially in systems with charge regulation due to dissociation of chargeable groups on the surfaces [18.100]; ion condensa-

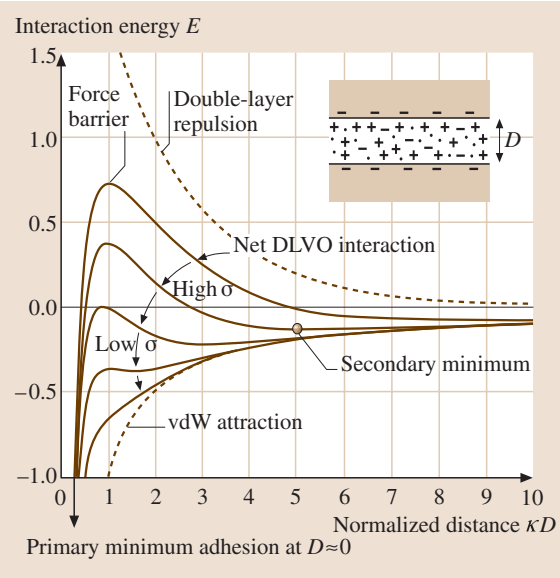


Fig. 18.6 Schematic plots of the DLVO interaction potential energy E between two flat, charged surfaces (or, according to the Derjaguin approximation, (18.3), the force F between two curved surfaces) as a function of the surface separation normalized by the Debye length, κ^{-1}. The van der Waals attraction (inverse power-law dependence on D) together with the repulsive electrostatic "double-layer" force (roughly exponential) at different surface charge σ (or potential, see (18.12)) determine the net interaction potential in aqueous electrolyte solution (after [18.62] with permission)

tion [18.101], which may lower the effective surface charge density in systems containing di-and trivalent counterions; or ion correlation, which is an additional van der Waals-like attraction due to mobile and therefore highly polarizable counterions located at the surface [18.102]. The ion correlation (or charge fluctuation) force becomes significant at separations below 4 nm and increases with the surface charge density σ and the valency z of the counterions. Computer simulations have shown that at high charge density and monovalent counterions, the ion correlation force can reduce the effective double-layer repulsion by 10–15 %. With divalent counterions, the ion correlation force was found to exceed the double-layer repulsion and the total force then became attractive at a separation below 2 nm even in dilute electrolyte solution [18.103]. Experimentally, such short-range attractive forces have been found in charged bilayer systems [18.104, 105].

18.4.3 Solvation and Structural Forces

When a liquid is confined within a restricted space, for example, a very thin film between two surfaces, it ceases to behave as a structureless continuum. At small surface separations (below about 10 molecular diameters), the van der Waals force between two surfaces or even two solute molecules in a liquid (solvent) is no longer a smoothly varying attraction. Instead, there arises an additional "solvation" force that generally oscillates between attraction and repulsion with distance, with a periodicity equal to some mean dimension σ of the liquid molecules [18.106]. Figure 18.7a shows the force law between two smooth mica surfaces across the hydrocarbon liquid tetradecane, whose inert, chain-like molecules have a width of $\sigma \approx 0.4$ nm.

The short-range oscillatory force law is related to the "density distribution function" and "potential of mean force" characteristic of intermolecular interactions in liquids. These forces arise from the confining effects two surfaces have on liquid molecules, forcing them to order into quasi-discrete layers. Such layers are energetically or entropically favored and correspond to the minima in the free energy, whereas fractional layers are disfavored (energy maxima). This effect is quite general and arises in all simple liquids when they are confined between two smooth, rigid surfaces, both flat and curved. Oscillatory forces do not require any attractive liquid–liquid or liquid–wall interaction, only two hard walls confining molecules whose shape is not too irregular and that are free to exchange with molecules in a bulk liquid reservoir. In the absence of any attractive pressure between the molecules, the bulk liquid density could be maintained by an external hydrostatic pressure – in real liquids attractive van der Waals forces play the role of such an external pressure.

Oscillatory forces are now well understood theoretically, at least for simple liquids, and a number of theoretical studies and computer simulations of various confined liquids (including water) that interact via some form of Lennard-Jones potential have invariably led to an oscillatory solvation force at surface separations below a few molecular diameters [18.107–114]. In a first approximation, the oscillatory force law may be described by an exponentially decaying cosine function of the form

$$E \approx E_0 \cos(2\pi D/\sigma)\, \mathrm{e}^{-D/\sigma} \,, \qquad (18.16)$$

where both theory and experiments show that the oscillatory period and the characteristic decay length of the envelope are close to σ.

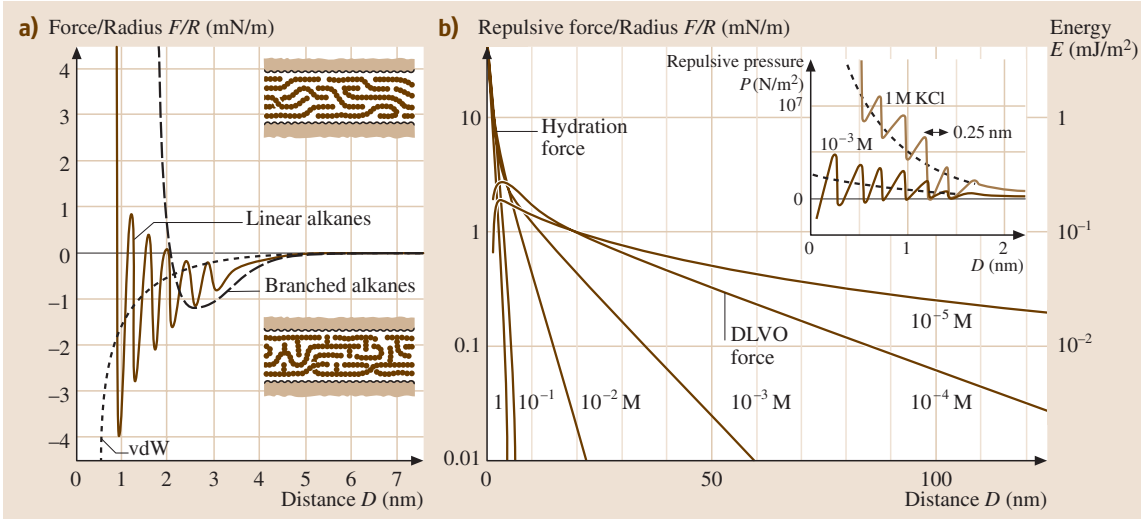


Fig. 18.7 **(a)** *Solid curve:* Forces measured between two mica surfaces across saturated linear chain alkanes such as *n*-tetradecane and *n*-hexadecane [18.115, 116]. The 0.4 nm periodicity of the oscillations indicates that the molecules are preferentially oriented parallel to the surfaces, as shown schematically in the *upper insert*. The theoretical continuum van der Waals attraction is shown as a *dotted curve*. *Dashed curve:* Smooth, non-oscillatory force law exhibited by irregularly shaped alkanes (such as 2-methyloctadecane) that cannot order into well-defined layers (*lower insert*) [18.116, 117]. Similar non-oscillatory forces are also observed between "rough" surfaces, even when these interact across a saturated linear chain liquid. This is because the irregularly shaped surfaces (rather than the liquid) now prevent the liquid molecules from ordering in the gap. **(b)** Forces measured between charged mica surfaces in KCl solutions of varying concentrations [18.20]. In dilute solutions (10^{-5} and 10^{-4} M), the measured forces are excellently described by the DLVO theory, based on exact solutions to the nonlinear Poisson–Boltzmann equation for the electrostatic forces and the Lifshitz theory for the van der Waals forces (using a Hamaker constant of $A_H = 2.2 \times 10^{-20}$ J). At higher concentrations, as more hydrated K^+ cations adsorb onto the negatively charged surfaces, an additional hydration force appears superimposed on the DLVO interaction at distances below 3–4 nm. This force has both an oscillatory and a monotonic component. *Insert:* Short-range hydration forces between mica surfaces shown as pressure versus distance. The lower and upper curves show surfaces 40% and 95% saturated with K^+ ions. At larger separations, the forces are in good agreement with the DLVO theory. (After [18.3]. Copyright 1991, with permission from Elsevier Science)

Once the solvation zones of the two surfaces overlap, the mean liquid density in the gap is no longer the same as in the bulk liquid. Since the van der Waals interaction depends on the optical properties of the liquid, which in turn depends on the density, the van der Waals and the oscillatory solvation forces are not strictly additive. It is more correct to think of the solvation force as *the* van der Waals force at small separations with the molecular properties and density variations of the medium taken into account. It is also important to appreciate that solvation forces do not arise simply because liquid molecules tend to structure into semi-ordered layers at surfaces. They arise because of the disruption or *change* of this ordering during the approach of a second surface. The two effects are related; the greater the tendency toward structuring at an isolated surface the greater the solvation force between two such surfaces, but there is a real distinction between the two phenomena that should be borne in mind.

Oscillatory forces lead to different adhesion values depending on the energy minimum from which two surfaces are being separated. For an interaction energy described by (18.16), "quantized" adhesion energies will be E_0 at $D = 0$ (primary minimum), E_0/e at $D = \sigma$, E_0/e^2 at $D = 2\sigma$, etc. E_0 can be thought of as a depletion force (see Sect. 18.4.5) that is approximately given by the osmotic limit $E_0 \approx -k_B T/\sigma^2$, which can exceed the contribution to the adhesion energy in contact from the van der Waals forces (at $D_0 \approx 0.15$–0.20 nm, as discussed in Sect. 18.3.1, keeping in mind that the

Lifshitz theory fails to describe the force law at *intermediate* distances). Such multivalued adhesion forces have been observed in a number of systems, including the interactions of fibers.

Measurements of oscillatory forces between different surfaces across both aqueous and nonaqueous liquids have revealed their richness of properties [18.118–121], for example, their great sensitivity to the shape and rigidity of the solvent molecules, to the presence of other components, and to the structure of the confining surfaces (see Sects. 18.5.3 and 18.9). In particular, the oscillations can be smeared out if the molecules are irregularly shaped (e.g., branched) and therefore unable to pack into ordered layers, or when the interacting surfaces are rough or fluid-like (see Sect. 18.4.6).

It is easy to understand how oscillatory forces arise between two flat, plane parallel surfaces. Between two curved surfaces, e.g., two spheres, one might imagine the molecular ordering and oscillatory forces to be smeared out in the same way that they are smeared out between two randomly rough surfaces (see Sect. 18.5.3); however, this is not the case. Ordering can occur so long as the curvature or roughness is itself regular or uniform, i. e., not random. This is due to the Derjaguin approximation (18.3). If the energy between two flat surfaces is given by a decaying oscillatory function (for example, a cosine function as in (18.16)), then the force (and energy) between two curved surfaces will also be an oscillatory function of distance with some phase shift. Likewise, two surfaces with regularly curved regions will also retain their oscillatory force profile, albeit modified, as long as the corrugations are truly regular, i. e., periodic. On the other hand, surface roughness, even on the nanometer scale, can smear out oscillations if the roughness is random and the confined molecules are smaller than the size of the surface asperities [18.122, 123]. If an organic liquid contains small amounts of water, the expected oscillatory force can be replaced by a strongly attractive capillary force (see Sect. 18.5.1).

18.4.4 Hydration and Hydrophobic Forces

The forces occurring in water and electrolyte solutions are more complex than those occurring in nonpolar liquids. According to continuum theories, the attractive van der Waals force is always expected to win over the repulsive electrostatic "double-layer" force at small surface separations (Fig. 18.6). However, certain surfaces (usually oxide or hydroxide surfaces such as clays or silica) swell spontaneously or repel each other in aqueous solution, even at high salt concentrations. Yet in all these systems one would expect the surfaces or particles to remain in strong adhesive contact or coagulate in a primary minimum if the only forces operating were DLVO forces.

There are many other aqueous systems in which the DLVO theory fails and where there is an additional short-range force that is not oscillatory but monotonic. Between hydrophilic surfaces this force is exponentially repulsive and is commonly referred to as the *hydration*, or *structural*, force. The origin and nature of this force has long been controversial, especially in the colloidal and biological literature. Repulsive hydration forces are believed to arise from strongly hydrogen-bonding surface groups, such as hydrated ions or hydroxyl (−OH) groups, which modify the hydrogen-bonding network of liquid water adjacent to them. Because this network is quite extensive in range [18.124], the resulting interaction force is also of relatively long range.

Repulsive hydration forces were first extensively studied between clay surfaces [18.125]. More recently, they have been measured in detail between mica and silica surfaces [18.20–22], where they have been found to decay exponentially with decay lengths of about 1 nm. Their effective range is about 3 to 5 nm, which is about twice the range of the oscillatory solvation force in water. Empirically, the hydration repulsion between two hydrophilic surfaces appears to follow the simple equation

$$E = E_0 \mathrm{e}^{-D/\lambda_0} , \tag{18.17}$$

where $\lambda_0 \approx 0.6–1.1$ nm for 1:1 electrolytes and $E_0 = 3–30\,\mathrm{mJ\,m^{-2}}$ depending on the hydration (hydrophilicity) of the surfaces, higher E_0 values generally being associated with lower λ_0 values.

The interactions between molecularly smooth mica surfaces in dilute electrolyte solutions obey the DLVO theory (Fig. 18.7b). However, at higher salt concentrations, specific to each electrolyte, hydrated cations bind to the negatively charged surfaces and give rise to a repulsive hydration force [18.20, 21]. This is believed to be due to the energy needed to dehydrate the bound cations, which presumably retain some of their water of hydration on binding. This conclusion was arrived at after noting that the strength and range of the hydration forces increase with the known hydration numbers of the electrolyte cations in the order: $Mg^{2+} > Ca^{2+} > Li^{+} \sim Na^{+} > K^{+} > Cs^{+}$. Similar trends are observed with other negatively charged colloidal surfaces.

While the hydration force between two mica surfaces is overall repulsive below a distance of 4 nm, it is not always monotonic below about 1.5 nm but exhibits oscillations of mean periodicity of 0.25 ± 0.03 nm, roughly equal to the diameter of the water molecule. This is shown in the insert in Fig. 18.7b, where we may note that the first three minima at $D = 0$, 0.28, and 0.56 nm occur at negative energies, a result that rationalizes observations on certain colloidal systems. For example, clay platelets such as montmorillonite often repel each other increasingly strongly as they come closer together, but they are also known to stack into stable aggregates with water interlayers of typical thickness 0.25 and 0.55 nm between them [18.126, 127], suggestive of a turnabout in the force law from a monotonic repulsion to discretized attraction. In chemistry we would refer to such structures as stable hydrates of fixed stoichiometry, whereas in physics we may think of them as experiencing an oscillatory force.

Both surface force and clay swelling experiments have shown that hydration forces can be modified or "regulated" by exchanging ions of different hydration on surfaces, an effect that has important practical applications in controlling the stability of colloidal dispersions. It has long been known that colloidal particles can be precipitated (coagulated or flocculated) by increasing the electrolyte concentration, an effect that was traditionally attributed to the reduced screening of the electrostatic double-layer repulsion between the particles due to the reduced Debye length. However, there are many examples where colloids are stabilized at high salt concentrations, not at low concentrations. This effect is now recognized as being due to the increased hydration repulsion experienced by certain surfaces when they bind highly hydrated ions at higher salt concentrations. Hydration regulation of adhesion and interparticle forces is an important practical method for controlling various processes such as clay swelling [18.126, 127], ceramic processing and rheology [18.128, 129], material fracture [18.128], and colloidal particle and bubble coalescence [18.130].

Water appears to be unique in having a solvation (hydration) force that exhibits both a monotonic and an oscillatory component. Between hydrophilic surfaces the monotonic component is repulsive (Fig. 18.7b), but between hydrophobic surfaces it is attractive and the final adhesion is much greater than expected from the Lifshitz theory.

A hydrophobic surface is one that is inert to water in the sense that it cannot bind to water molecules via ionic or hydrogen bonds. Hydrocarbons and fluorocarbons are hydrophobic, as is air, and the strongly attractive hydrophobic force has many important manifestations and consequences such as the low solubility or miscibility of water and oil molecules, micellization, protein folding, strong adhesion and rapid coagulation of hydrophobic surfaces, non-wetting of water on hydrophobic surfaces, and hydrophobic particle attachment to rising air bubbles (the basic principle of froth flotation).

In recent years, there has been a steady accumulation of experimental data on the force laws between various hydrophobic surfaces in aqueous solution [18.131–143]. These studies have found that the hydrophobic force law between two macroscopic surfaces is of surprisingly long range, decaying exponentially with a characteristic decay length of 1 to 2 nm in the separation range of 0 to 10 nm, and then more gradually further out. The hydrophobic force can be far stronger than the van der Waals attraction, especially between hydrocarbon surfaces in water, for which the Hamaker constant is quite small. The magnitude of the hydrophobic attraction has been found to decrease with the decreasing hydrophobicity (increasing hydrophilicity) of lecithin lipid bilayer surfaces [18.31] and silanated surfaces [18.139], whereas examples of the opposite trend have been shown for some Langmuir–Blodgett-deposited monolayers [18.144].

For two surfaces in water the purely hydrophobic interaction energy (ignoring DLVO and oscillatory forces) in the range of 0 to 10 nm is given by

$$E = -2\gamma\, \mathrm{e}^{-D/\lambda_0} \,, \tag{18.18}$$

where typically $\lambda_0 = 1$–2 nm, and $\gamma = 10$–$50\ \mathrm{mJ\,m^{-2}}$. The higher value corresponds to the interfacial energy of a pure hydrocarbon–water interface.

At a separation below 10 nm, the hydrophobic force appears to be insensitive or only weakly sensitive to changes in the type and concentration of electrolyte ions in the solution. The absence of a "screening" effect by ions attests to the non-electrostatic origin of this interaction. In contrast, some experiments have shown that at separations greater than 10 nm, the attraction does depend on the intervening electrolyte, and that in dilute solutions, or solutions containing divalent ions, it can continue to exceed the van der Waals attraction out to separations of 80 nm [18.136, 145].

The long-range nature of the hydrophobic interaction has a number of important consequences. It accounts for the rapid coagulation of hydrophobic particles in water and may also account for the rapid folding of proteins. It also explains the ease with which water films rupture on hydrophobic surfaces. In this case, the

van der Waals force across the water film is repulsive and therefore favors wetting, but this is more than offset by the attractive hydrophobic interaction acting between the two hydrophobic phases across water. Hydrophobic forces are increasingly being implicated in the adhesion and fusion of biological membranes and cells. It is known that both osmotic and electric field stresses enhance membrane fusion, an effect that may be due to the concomitant increase in the hydrophobic area exposed between two adjacent surfaces.

From the previous discussion we can infer that hydration and hydrophobic forces are not of a simple nature. These interactions are probably the most important, yet the least understood of all the forces in aqueous solutions. The unusual properties of water and the nature of the surfaces (including their homogeneity and stability) appear to be equally important. Some particle surfaces can have their hydration forces regulated, for example, by ion exchange. Others appear to be intrinsically hydrophilic (e.g., silica) and cannot be coagulated by changing the ionic condition, but can be rendered hydrophobic by chemically modifying their surface groups. For example, on heating silica to above 600 °C, two adjacent surface silanol (−OH) groups release a water molecule and form a hydrophobic siloxane (−O−) group, whence the repulsive hydration force changes into an attractive hydrophobic force.

How do these exponentially decaying repulsive or attractive forces arise? Theoretical work and computer simulations [18.109, 111, 146, 147] suggest that the solvation forces in water should be purely oscillatory, whereas other theoretical studies [18.148–153] suggest a monotonically exponential repulsion or attraction, possibly superimposed on an oscillatory force. The latter is consistent with experimental findings, as shown in the inset to Fig. 18.7b, where it appears that the oscillatory force is simply additive with the monotonic hydration and DLVO forces, suggesting that these arise from essentially different mechanisms.

It is probable that the short-range hydration force between all smooth, rigid, or crystalline surfaces (e.g., mineral surfaces such as mica) has an oscillatory component. This may or may not be superimposed on a monotonic force due to image interactions [18.150] and/or structural or hydrogen bonding interactions [18.148, 149].

Like the repulsive hydration force, the origin of the hydrophobic force is still unknown. *Luzar* et al. [18.152] carried out a Monte Carlo simulation of the interaction between two hydrophobic surfaces across water at separations below 1.5 nm. They obtained a decaying oscillatory force superimposed on a monotonically attractive curve. In more recent work [18.154, 155], it has been suggested that hydrophobic surfaces generate a depleted region of water around them, and that a long-range attractive force due to depletion arises between two such surfaces.

It is questionable whether the hydration or hydrophobic force should be viewed as an ordinary type of solvation or structural force that is reflecting the packing of water molecules. The energy (or entropy) associated with the hydrogen bonding network, which extends over a much larger region of space than the molecular correlations, is probably at the root of the long-range interactions of water. The situation in water appears to be governed by much more than the molecular packing effects that dominate the interactions in simpler liquids.

18.4.5 Polymer-Mediated Forces

Polymers or macromolecules are chain-like molecules consisting of many identical segments (monomers or repeating units) held together by covalent bonds. The size of a polymer coil in solution or in the melt is determined by a balance between van der Waals attraction (and hydrogen bonding, if present) between polymer segments, and the entropy of mixing, which causes the polymer coil to expand. In polymer melts above the glass transition temperature, and at certain conditions in solution, the attraction between polymer segments is exactly balanced by the entropy effect. The polymer solution will then behave virtually ideally, and the density distribution of segments in the coil is Gaussian. This is called the theta (θ) condition, and it occurs at the theta or Flory temperature for a particular combination of polymer and solvent or solvent mixture. At lower temperatures (in a poor or bad solvent), the polymer–polymer interactions dominate over the entropic, and the coil will shrink or precipitate. At higher temperatures (good solvent conditions), the polymer coil will be expanded.

High molecular weight polymers form large coils, which significantly affect the properties of a solution even when the total mass of polymer is very low. The radius of the polymer coil is proportional to the segment length, a, and the number of segments, n. At theta conditions, the hydrodynamic radius of the polymer coil (the root-mean-square separation of the ends of one polymer chain) is theoretically given by $R_h = a\,n^{1/2}$, and the unperturbed radius of gyration (the average root-mean-square distance of a segment from the center of mass of the molecule) is $R_g = a\,(n/6)^{1/2}$. In a good sol-

vent the perturbed size of the polymer coil, the Flory radius R_F, is proportional to $n^{3/5}$.

Polymers interact with surfaces mainly through van der Waals and electrostatic interactions. The physisorption of polymers containing only one type of segment is reversible and highly dynamic, but the rate of exchange of adsorbed chains with free chains in the solution is low, since the polymer remains bound to the surface as long as one segment along the chain is adsorbed. The adsorption energy per segment is on the order of $k_B T$. In a good solvent, the conformation of a polymer on a surface is very different from the coil conformation in bulk solution. Polymers adsorb in "trains", separated by "loops" extending into solution and dangling "tails" (the ends of the chain). Compared to adsorption at lower temperatures, good solvent conditions favor more of the polymer chain being in the solvent, where it can attain its optimum conformation. As a result, the extension of the polymer is longer, even though the total amount of adsorbed polymer is lower. In a good solvent, the polymer chains can also be effectively repelled from a surface, if the loss in conformational entropy close to the surface is not compensated for by a gain in enthalpy from adsorption of segments. In this case, there will be a layer of solution (thickness $\approx R_g$) close to the surfaces that is depleted of polymer.

The interaction forces between two surfaces across a polymer solution will depend on whether the polymer is adsorbing onto the surfaces or is repelled from them, and also on whether the interaction occurs at "true" or "restricted" thermodynamic equilibrium. At true or full equilibrium, the polymer between the surfaces can equilibrate (exchange) with polymer in the bulk solution at all surface separations. Some theories [18.156, 157] predict that at full equilibrium, the polymer chains would move from the confined gap into the bulk solution where they could attain entropically more favorable conformations, and that a monotonic attraction at all distances would result from bridging and depletion interactions (which will be discussed below). Other theories suggest that the interaction at small separations would be ultimately repulsive, since some polymer chains would remain in the gap due to their attractive interactions with many sites on the surface (enthalpic) – more sites would be available to the remaining polymer chains if some others desorbed and diffused out from the gap [18.64, 158, 159].

At restricted equilibrium, the polymer is kinetically trapped, and the adsorbed amount is thus constant as the surfaces are brought toward each other, but the chains can still rearrange on the surfaces and in the gap. Experimentally, the true equilibrium situation is very difficult to attain, and most experiments are done at restricted equilibrium conditions. Even the equilibration of conformations assumed in theoretical models for restricted equilibrium conditions can be so slow that this condition is difficult to reach.

In systems of adsorbing polymer, bridging of chains from one surface to the other can give rise to a long-range attraction, since the bridging chains would gain conformational entropy if the surfaces were closer together. In poor solvents, both bridging and intersegment interactions contribute to an attraction [18.26]. However, regardless of solvent and equilibrium conditions, a strong repulsion due to the osmotic interactions is seen as small surface separations in systems of adsorbing polymers at restricted equilibrium.

In systems containing high concentrations of non-adsorbing polymer, the difference in solute concentration in the bulk and between the surfaces at separations smaller than the approximate polymer coil diameter ($2R_g$, i. e., when the polymer has been squeezed out from the gap between the surfaces) may give rise to an attractive osmotic force ("depletion attraction") [18.160–164]. In addition, if the polymer coils become initially compressed as the surfaces approach each other, this can give rise to a repulsion ("depletion stabilization") at large separations [18.163]. For a system of two cylindrical surfaces or radius R, the maximum depletion force, F_{dep}, is expected to occur when the surfaces are in contact and is given by multiplying the depletion (osmotic) pressure, $P_{dep} = \rho k_B T$, by the contact area πr^2, where r is given by the chord theorem: $r^2 = (2R - R_g)R_g \approx 2RR_g$ [18.3]:

$$F_{dep}/R = -2\pi R_g \rho k_B T \ , \tag{18.19}$$

where ρ is the number density of the polymer in the bulk solution.

If a part of the polymer (typically an end group) is different from the rest of the chain, this part may preferentially adsorb to the surface. End-adsorbed polymer is attached to the surface at only one point, and the extension of the chain is dependent on the grafting density, i. e., the average distance, s, between adsorbed end groups on the surface (Fig. 18.8). One distinguishes between different regions of increasing overlap of the chains (stretching) called pancake, mushroom, and brush regimes [18.165]. In the mushroom regime, where the coverage is sufficiently low so that there is no overlap between neighboring chains, the thickness of the adsorbed layer is proportional to $n^{1/2}$ (i. e., to R_g) at theta conditions and to $n^{3/5}$ in a good solvent.

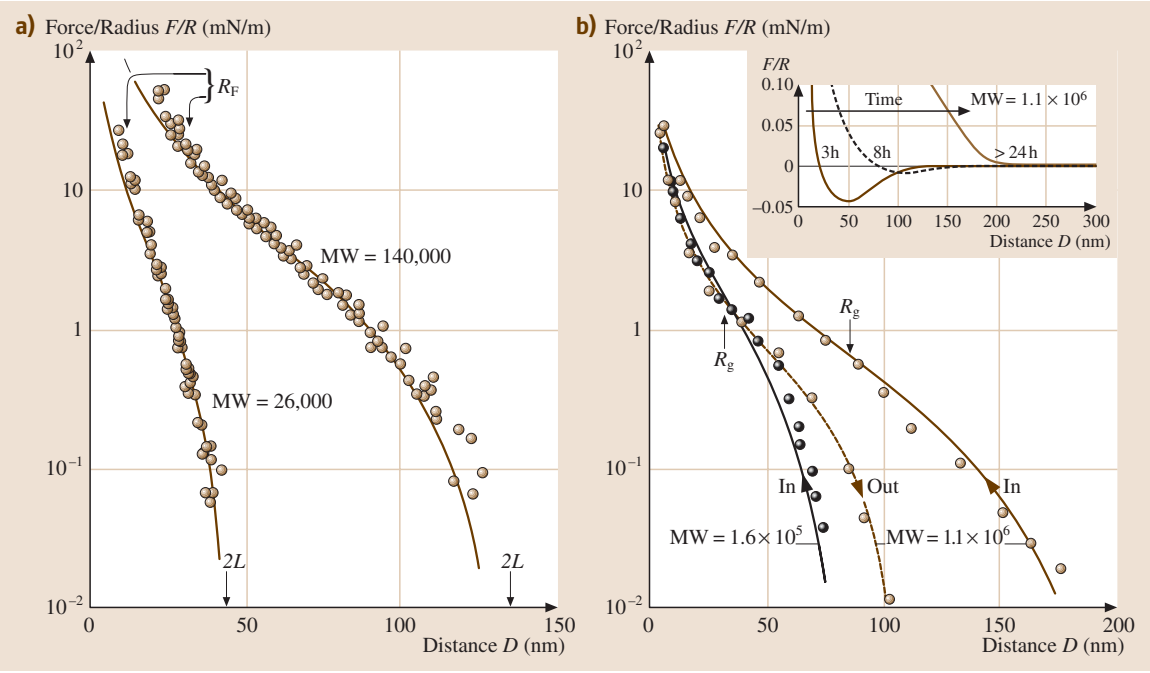


Fig. 18.8a,b Experimentally determined forces in systems of two interacting polymer brushes: **(a)** Polystyrene brush layers grafted via an adsorbing chain-end group onto mica surfaces in toluene (a good solvent for polystyrene). *Left curve:* MW = 26,000 g/mol, $R_F = 12$ nm. *Right curve*: MW = 140,000 g/mol, $R_F = 32$ nm. Both force curves were reversible on approach and separation. The *solid curves* are theoretical fits using the Alexander–de Gennes theory with the following measured parameters: spacing between attachments sites: $s = 8.5$ nm, brush thickness: $L = 22.5$ nm and 65 nm, respectively. (Adapted from [18.171]). **(b)** Polyethylene oxide layers physisorbed onto mica from 150 µg/ml solution in aqueous 0.1 M KNO_3 (a good solvent for polyethylene oxide). *Main figure:* Equilibrium forces at full coverage after ∼ 16 h adsorption time. *Left curve:* MW = 160,000 g/mol, $R_g = 32$ nm. *Right curve:* MW = 1,100,000 g/mol, $R_g = 86$ nm. Note the hysteresis (irreversibility) on approach and separation for this *physisorbed* layer, in contrast to the absence of hysteresis with *grafted* chains in case **(a)**. The *solid curves* are based on a modified form of the Alexander–de Gennes theory. *Inset* in **(b)**: Evolution of the forces with the time allowed for the higher MW polymer to adsorb from solution. Note the gradual reduction in the attractive bridging component. (Adapted from [18.172–174]. After [18.3]. Copyright 1991, with permission from Elsevier Science)

Several models [18.165–170] have been developed for the extension and interactions between two brushes (strongly stretched grafted chains). They are based on a balance between osmotic pressure within the brush layers (uncompressed and compressed) and elastic energy of the chains and differ mainly in the assumptions of the segment density profile, which can be a step function or parabolic. At high coverage (in the brush regime),where the chains will avoid overlapping each other, the thickness of the layer is proportional to n.

Experimental work on both monodisperse [18.27,28, 171, 175] and polydisperse [18.30] systems at different solvent conditions has confirmed the expected range and magnitude of the repulsive interactions resulting from compression of densely packed grafted layers.

18.4.6 Thermal Fluctuation Forces

If a surface is not rigid but very soft or even fluid-like, this can act to smear out any oscillatory solvation force. This is because the thermal fluctuations of such interfaces make them dynamically "rough" at any instant, even though they may be perfectly smooth on a time average. The types of surfaces that fall into this category are fluid-like amphiphilic surfaces of micelles, bilayers, emulsions, soap films, etc., but also solid colloidal par-

ticle surfaces that are coated with surfactant monolayers, as occurs in lubricating oils, paints, toners, etc.

These thermal fluctuation forces (also called entropic or steric forces) are usually short-range and repulsive and are very effective at stabilizing the attractive van der Waals forces at some small but finite separation. This can reduce the adhesion energy or force by up to three orders of magnitude. It is mainly for this reason that fluid-like micelles and bilayers, biological membranes, emulsion droplets, or gas bubbles adhere to each other only very weakly.

Because of their short range it was, and still is, commonly believed that these forces arise from water ordering or "structuring" effects at surfaces, and that they reflect some unique or characteristic property of water. However, it is now known that these repulsive forces also exist in other liquids. Moreover, they appear to become stronger with increasing temperature, which is unlikely if the force would originate from molecular ordering effects at surfaces. Recent experiments, theory, and computer simulations [18.176–178] have shown that these repulsive forces have an entropic origin arising from the osmotic repulsion between exposed thermally mobile surface groups once these overlap in a liquid. These phenomena include undulating and peristaltic forces between membranes or bilayers, and, on the molecular scale, protrusion and head group overlap forces where the interactions also are influenced by hydration forces.

18.5 Adhesion and Capillary Forces

18.5.1 Capillary Forces

When considering the adhesion of two solid surfaces or particles in air or in a liquid, it is easy to overlook or underestimate the important role of capillary forces, i. e., forces arising from the Laplace pressure of curved menisci formed by condensation of a liquid between and around two adhering surfaces (Fig. 18.9).

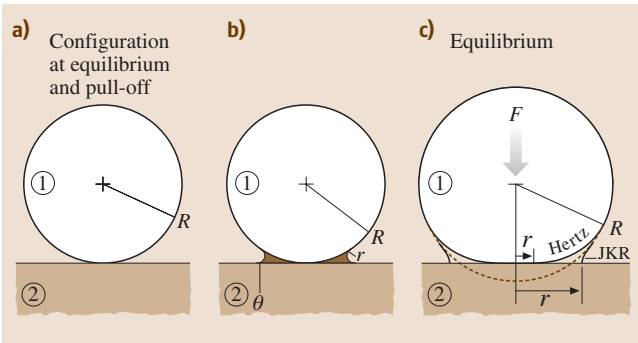


Fig. 18.9a–c Adhesion and capillary forces: **(a)** a non-deforming sphere on a rigid, flat surface in an inert atmosphere and **(b)** in a vapor that can "capillary condense" around the contact zone. At equilibrium, the concave radius, r, of the liquid meniscus is given by the Kelvin equation. For a concave meniscus to form, the contact angle θ has to be less than 90°. In the case of hydrophobic surfaces surrounded by water, a vapor cavity can form between the surfaces. As long as the surfaces are perfectly smooth, the contribution of the meniscus to the adhesion force is independent of r. (After [18.1] with permission.) **(c)** Elastically deformable sphere on a rigid flat surface in the absence (Hertz) and presence (JKR) of adhesion [**(a)** and **(c)** after [18.3]. Copyright 1991, with permission from Elsevier Science]

The adhesion force between a non-deformable spherical particle of radius R and a flat surface in an inert atmosphere (Fig. 18.9a) is

$$F_s = 4\pi R \gamma_{SV} \,. \tag{18.20}$$

But in an atmosphere containing a condensable vapor, the expression above is replaced by

$$F_s = 4\pi R(\gamma_{LV}\cos\theta + \gamma_{SL}) \,, \tag{18.21}$$

where the first term is due to the Laplace pressure of the meniscus and the second is due to the direct adhesion of the two contacting solids within the liquid. Note that the above equation does not contain the radius of curvature, r, of the liquid meniscus (see Fig. 18.9b). This is because for smaller r the Laplace pressure γ_{LV}/r increases, but the area over which it acts decreases by the same amount, so the two effects cancel out. Experiments with inert liquids, such as hydrocarbons, condensing between two mica surfaces indicate that (18.21) is valid for values of r as small as 1–2 nm, corresponding to vapor pressures as low as 40% of saturation [18.121, 179, 180]. Capillary condensation also occurs in binary liquid systems, e.g., when water dissolved in hydrocarbon liquids condense around two contacting hydrophilic surfaces or when a vapor cavity forms in water around two hydrophobic surfaces. In the case of water condensing from vapor or from oil, it also appears that the bulk value of γ_{LV} is applicable for meniscus radii as small as 2 nm.

The capillary condensation of liquids, especially water, from vapor can have additional effects on the

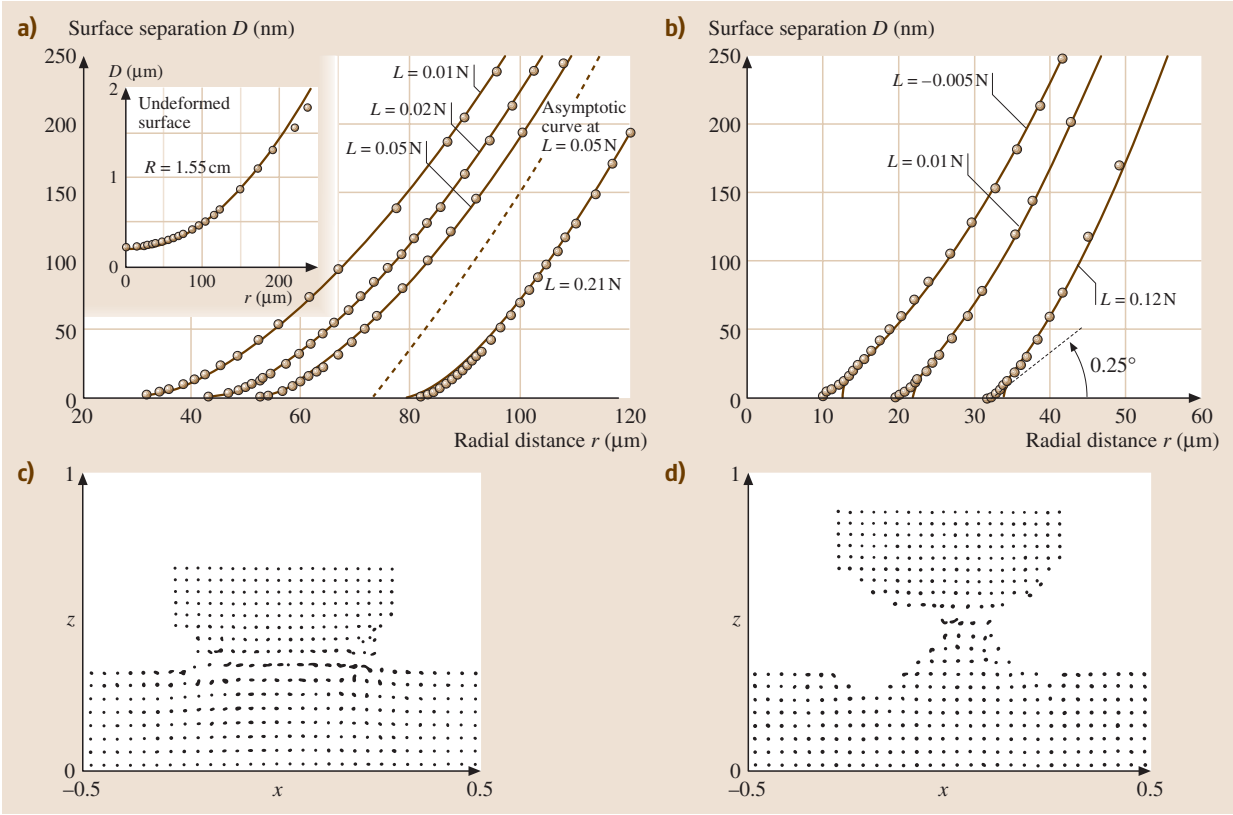


Fig. 18.10a–d Experimental and computer simulation data on contact mechanics for ideal Hertz and JKR contacts. **(a)** Measured profiles of surfaces in nonadhesive contact (*circles*) compared with Hertz profiles (*continuous curves*). The system was mica surfaces in a concentrated KCl solution in which they do not adhere. When not in contact, the surface shape is accurately described by a sphere of radius $R = 1.55$ cm (*inset*). The applied loads were 0.01, 0.02, 0.05, and 0.21 N. The last profile was measured in a different region of the surfaces where the local radius of curvature was 1.45 cm. The Hertz profiles correspond to central displacements of $\delta = 66.5$, 124, 173, and 441 nm. The *dashed line* shows the shape of the undeformed sphere corresponding to the curve at a load of 0.05 N; it fits the experimental points at larger distances (not shown). **(b)** Surface profiles measured with adhesive contact (mica surfaces adhering in dry nitrogen gas) at applied loads of -0.005, 0.01, and 0.12 N. The continuous lines are JKR profiles obtained by adjusting the central displacement in each case to get the best fit to points at larger distances. The values are $\delta = -4.2$, 75.6, and 256 nm. Note that the scales of this figure exaggerate the apparent angle at the junction of the surfaces. This angle, which is insensitive to load, is only about 0.25°. **(c)** and **(d)** Molecular dynamics simulation illustrating the formation of a connective neck between an Ni tip (topmost eight layers) and an Au substrate. The figures show the atomic configuration in a slice through the system at indentation **(c)** and during separation **(d)**. Note the crystalline structure of the neck. Distances are given in units of x and z, where $x = 1$ and $z = 1$ correspond to 6.12 nm. [**(a)** and **(b)** after [18.181]. Copyright 1987, with permission from Elsevier Science. **(c)** and **(d)** after [18.91], with kind permission from Kluwer Academic Publishers]

physical state of the contact zone. For example, if the surfaces contain ions, these will diffuse and build up within the liquid bridge, thereby changing the chemical composition of the contact zone, as well as influencing the adhesion. In the case of surfaces covered with surfactant or polymer molecules (amphiphilic surfaces), the molecules can overturn on exposure to humid air, so that the surface nonpolar groups become replaced by polar groups, which renders the surfaces hydrophilic. When two such surfaces come into contact, water will condense around the contact zone and the adhesion force will also be affected – generally increasing well above

the value expected for inert hydrophobic surfaces. It is apparent that the adhesion in vapor or a solvent is often largely determined by capillary forces arising from the condensation of liquid that may be present only in very small quantities, e.g., 10–20 % of saturation in the vapor, or 20 ppm in the solvent.

18.5.2 Adhesion Mechanics

Two bodies in contact deform as a result of surface forces and/or applied normal forces. For the simplest case of two interacting elastic spheres (a model that is easily extended to an elastic sphere interacting with an undeformable surface (or vice versa)) and in the absence of attractive surface forces, the vertical central displacement (compression) was derived by *Hertz* [18.182]. In this model, the displacement and the contact area are equal to zero when no external force (load) is applied, i. e., at the points of contact and of separation.

In systems where attractive surface forces are present between the surfaces, the deformations are more complicated. Modern theories of the adhesion mechanics of two contacting solid surfaces are based on the Johnson–Kendall–Roberts (JKR) theory [18.15, 183], or on the Derjaguin–Muller–Toporov (DMT) theory [18.184–186]. The JKR theory is applicable to easily deformable, large bodies with high surface energy, whereas the DMT theory better describes very small and hard bodies with low surface energy [18.187].

In the JKR theory, two spheres of radii R_1 and R_2, bulk elastic modulus K, and surface energy γ will flatten due to attractive surface forces when in contact at no external load. The contact area will increase under an external load or normal force F, such that at mechanical equilibrium the radius of the contact area, r, is given by

$$r^3 = \frac{R}{K}\left[F + 6\pi R\gamma + \sqrt{12\pi R\gamma F + (6\pi R\gamma)^2}\right] , \tag{18.22}$$

where $R = R_1 R_2/(R_1 + R_2)$. In the absence of surface energy, γ, equation (18.22) is reduced to the expression for the radius of the contact area in the Hertz model. Another important result of the JKR theory gives the adhesion force or "pull-off" force:

$$F_S = -3\pi R\gamma_S , \tag{18.23}$$

where the surface energy, γ_S, is defined as $W = 2\gamma_S$, where W is the reversible work of adhesion. Note that according to the JKR theory a finite elastic modulus, K, while having an effect on the load–area curve, has no effect on the adhesion force, an interesting and unexpected result that has nevertheless been verified experimentally [18.15, 181, 188, 189].

Equation (18.22) and (18.23) provide the framework for analyzing results of adhesion measurements of contacting solids, known as contact mechanics [18.183, 190], and for studying the effects of surface conditions and time on adhesion energy hysteresis (see Sect. 18.5.4).

18.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch

In a contact between two rough surfaces, the real area of contact varies with the applied load in a different manner than between smooth surfaces [18.191, 192]. For a Hertzian contact, it has been shown that the contact area for rough surfaces increases approximately linearly with the applied normal force (load), F, instead of as $F^{2/3}$ for smooth surfaces. In systems with attractive surface forces, there is a competition between this attraction and repulsive forces arising from compression of high asperities. As a result, the adhesion in such systems can be very low, especially if the surfaces are not easily deformed [18.193–195]. The opposite is possible for soft (viscoelastic) surfaces where the real (molecular) contact area might be larger than for two perfectly smooth surfaces.

Adhesion forces may also vary depending on the commensurability of the crystallographic lattices of the interacting surfaces. *McGuiggan* and *Israelachvili* [18.196] measured the adhesion between two mica surfaces as a function of the orientation (twist angle) of their surface lattices. The forces were measured in air, water, and an aqueous salt solution where oscillatory structural forces were present. In humid air, the adhesion was found to be relatively independent of the twist angle θ due to the adsorption of a 0.4-nm-thick amorphous layer of organics and water at the interface. In contrast, in water, sharp adhesion peaks (energy minima) occurred at $\theta = 0°, \pm 60°, \pm 120°$ and 180°, corresponding to the "coincidence" angles of the surface lattices (Fig. 18.11). As little as $\pm 1°$ away from these peaks, the energy decreased by 50%. In aqueous KCl solution, due to potassium ion adsorption the water between the surfaces becomes ordered, resulting in an oscillatory force profile where the adhesive minima occur at discrete separations of about 0.25 nm, corresponding to integral numbers of water layers. The whole interaction potential was now found to depend on the orientation of

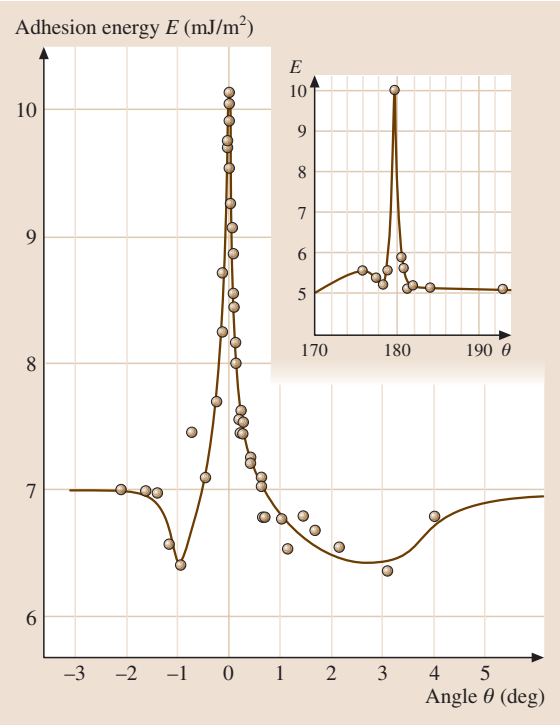


Fig. 18.11 Adhesion energy for two mica surfaces in contact in water (in the primary minimum of an oscillatory force curve) as a function of the mismatch angle θ about $\theta = 0°$ and 180° between the mica surface lattices. (After [18.197] with permission)

the surface lattices, and the effect extended at least four molecular layers.

It has also been appreciated that the structure of the confining surfaces is just as important as the nature of the liquid for determining the solvation forces [18.90, 122, 123, 198–202]. Between two surfaces that are completely flat but "unstructured", the liquid molecules will order into layers, but there will be no lateral ordering within the layers. In other words, there will be positional ordering normal but not parallel to the surfaces. If the surfaces have a crystalline (periodic) lattice, this may induce ordering parallel to the surfaces, as well, and the oscillatory force then also depends on the structure of the surface lattices. Further, if the two lattices have different dimensions ("mismatched" or "incommensurate" lattices), or if the lattices are similar but are not in register relative to each other, the oscillatory force law is further modified [18.196, 203] and the tribological properties of the film are also influenced, as discussed in Sect. 18.9 [18.203, 204].

As shown by the experiments, these effects can alter the magnitude of the adhesive minima found at a given separation within the last one or two nanometers of a thin film by a factor of two. The force barriers (maxima) may also depend on orientation. This could be even more important than the effects on the minima. A high barrier could prevent two surfaces from coming closer together into a much deeper adhesive well. Thus the maxima can effectively contribute to determining not only the final separation of two surfaces, but also their final adhesion. Such considerations should be particularly important for determining the thickness and strength of intergranular spaces in ceramics, the adhesion forces between colloidal particles in concentrated electrolyte solution, and the forces between two surfaces in a crack containing capillary condensed water.

For surfaces that are *randomly* rough, oscillatory forces become smoothed out and disappear altogether, to be replaced by a purely monotonic solvation force [18.116, 122, 123]. This occurs even if the liquid molecules themselves are perfectly capable of ordering into layers. The situation of *symmetric* liquid molecules confined between *rough* surfaces is therefore not unlike that of *asymmetric* molecules between *smooth* surfaces (see Sect. 18.4.3 and Fig. 18.7a). To summarize, for there to be an oscillatory solvation force, the liquid molecules must be able to be correlated over a reasonably long range. This requires that both the liquid molecules and the surfaces have a high degree of order or symmetry. If either is missing, so will the oscillations. A roughness of only a few tenths of a nanometer is often sufficient to eliminate any oscillatory component of the force law.

18.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis

Under ideal conditions the adhesion energy is a well-defined thermodynamic quantity. It is normally denoted by E or W (the work of adhesion) or γ (the surface tension, where $W = 2\gamma$) and gives the reversible work done on bringing two surfaces together or the work needed to separate two surfaces from contact. Under ideal, equilibrium conditions these two quantities are the same, but under most realistic conditions they are not: The work needed to separate two surfaces is always greater than that originally gained by bringing them together. An understanding of the molecular mechanisms underlying this phenomenon is essential for understanding many adhesion phenomena, energy dissipation during loading–unloading cycles, contact angle hysteresis, and

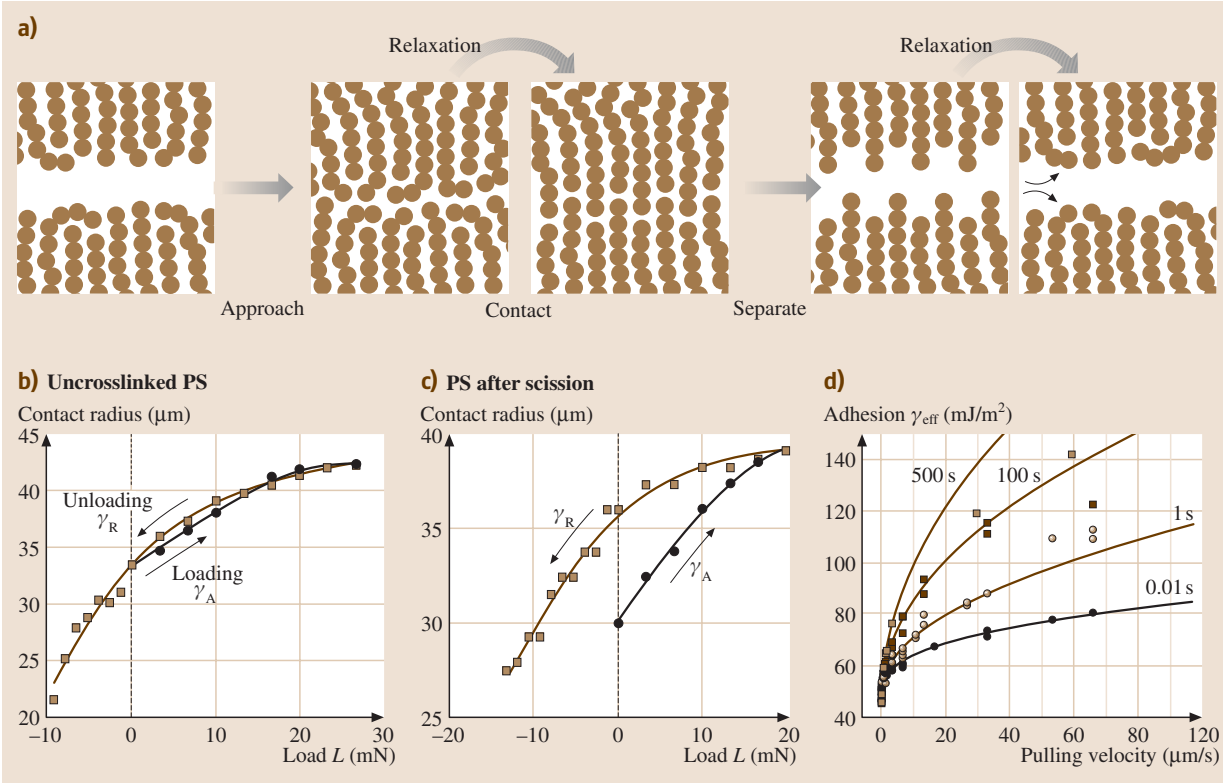


Fig. 18.12 **(a)** Schematic representation of interpenetrating chains. **(b)** and **(c)**: JKR plots (contact radius r as a function of applied load L) showing small adhesion hysteresis for uncrosslinked polystyrene and larger adhesion hysteresis after chain scission at the surfaces after 18 h irradiation with ultraviolet light in an oxygen atmosphere. The adhesion hysteresis continues to increase with the irradiation time. **(d)** Rate-dependent adhesion of CTAB surfactant monolayers. The *solid curves* [18.205] are fits to experimental data on CTAB adhesion after different contact times [18.206] using an approximate analytical solution for a JKR model, including crack tip dissipation. Due to the limited range of validity of the approximation, the fits rely on the low effective adhesion energy part of the experimental data only. From the fits one can determine the thermodynamic adhesion energy, the characteristic dissipation velocity, and the intrinsic dissipation exponent of the model. [**(a)** after [18.207]. Copyright 1993 American Chemical Society. **(b)** and **(c)** after [18.208]. Copyright 2002 American Association for the Advancement of Science. **(d)** after [18.205]. Copyright 2000 American Chemical Society]

the molecular mechanisms associated with many frictional processes.

It is wrong to think that hysteresis arises because of some imperfection in the system such as rough or chemically heterogeneous surfaces, or because the supporting material is viscoelastic. Adhesion hysteresis can arise even between perfectly smooth and chemically homogenous surfaces supported by perfectly elastic materials. It can be responsible for such phenomena as rolling friction and elastoplastic adhesive contacts [18.190, 209–212] during loading–unloading and adhesion–decohesion cycles.

Adhesion hysteresis may be thought of as being due to mechanical effects such as instabilities, or chemical effects such as interdiffusion, interdigitation, molecular reorientations and exchange processes occurring at an interface after contact, as illustrated in Fig. 18.12. Such processes induce roughness and chemical heterogeneity even though initially (and after separation and re-equilibration) both surfaces are perfectly smooth and chemically homogeneous. In general, if the energy change, or work done, on separating two surfaces from adhesive contact is not fully recoverable on bringing the two surfaces back into contact again, the adhesion

hysteresis may be expressed as

$$\underset{\text{Receding}}{W_R} > \underset{\text{Advancing}}{W_A}$$

or

$$\Delta W = (W_R - W_A) > 0 \,, \tag{18.24}$$

where W_R and W_A are the adhesion or surface energies for receding (separating) and advancing (approaching) two solid surfaces, respectively.

Hysteresis effects are also commonly observed in wetting/dewetting phenomena [18.213]. For example, when a liquid spreads and then retracts from a surface the advancing contact angle θ_A is generally larger than the receding angle θ_R. Since the contact angle, θ, is related to the liquid–vapor surface tension, γ_L, and the solid–liquid adhesion energy, W, by the Dupré equation,

$$(1 + \cos\theta)\gamma_L = W \,, \tag{18.25}$$

we see that *wetting hysteresis* or *contact angle hysteresis* ($\theta_A > \theta_R$) actually implies adhesion hysteresis, $W_R > W_A$, as given by (18.24).

Energy dissipating processes such as adhesion and contact angle hysteresis arise because of practical constraints of the *finite time* of measurements and the *finite elasticity* of materials. This prevents many loading–unloading or approach–separation cycles from being thermodynamically reversible, even though if carried out infinitely slowly they would be. By thermodynamically irreversible one simply means that one cannot go through the approach–separation cycle via a continuous series of equilibrium states, because some of these are connected via spontaneous – and therefore thermodynamically irreversible – instabilities or transitions where energy is liberated and therefore "lost" via heat or phonon release [18.214]. This is an area of much current interest and activity, especially regarding the fundamental molecular origins of adhesion and friction in polymer and surfactant systems, and the relationships between them [18.190, 205, 206, 208, 210, 215–218].

18.6 Introduction: Different Modes of Friction and the Limits of Continuum Models

Most frictional processes occur with the sliding surfaces becoming damaged in one form or another [18.209]. This may be referred to as "normal" friction. In the case of brittle materials, the damaged surfaces slide past each other while separated by relatively large, micron-sized wear particles. With more ductile surfaces, the damage remains localized to nanometer-sized, plastically deformed asperities. Some features of the friction between damaged surfaces will be described in Sect. 18.7.4.

There are also situations in which sliding can occur between two perfectly smooth, undamaged surfaces. This may be referred to as "interfacial" sliding or

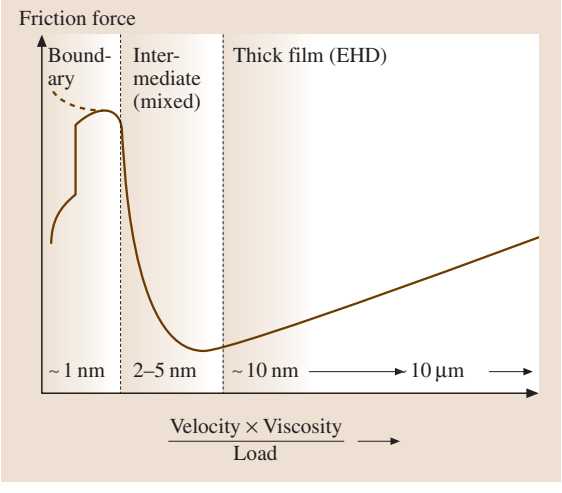


Fig. 18.13 Stribeck curve: an empirical curve giving the trend generally observed in the friction forces or friction coefficients as a function of sliding velocity, the bulk viscosity of the lubricating fluid, and the applied load. The three friction/lubrication regimes are known as the boundary lubrication regime (see Sect. 18.7), the intermediate or mixed lubrication regime (Sect. 18.8.2), and thick film or elastohydrodynamic (EHD) lubrication regime (Sect. 18.8.1). The film thicknesses believed to correspond to each of these regimes are also shown. For thick films, the "friction" force is purely viscous, e.g., Couette flow at low shear rates, but may become complicated at higher shear rates where EHD deformations of surfaces can occur during sliding. (After [18.1] with permission)

Table 18.4 The three main tribological regimes characterizing the changing properties of liquids subjected to an increasing confinement between two solid surfaces[a]

Regime	Conditions for getting into this regime	Static/equilibrium properties[b]	Dynamic properties[c]
Bulk	• Thick films (> 10 molecular diameters, $\gg R_g$ for polymers) • Low or zero loads • High shear rates	Bulk (continuum) properties: • Bulk liquid density • No long-range order	Bulk (continuum) properties: • Newtonian viscosity • Fast relaxation times • No glass temperature • No yield point • Elastohydrodynamic lubrication
Intermediate mixed	• Intermediately thick films (4–10 molecular diameters, $\sim R_g$ for polymers) • Low loads or pressure	Modified fluid properties include: • Modified positional and orientational order[a] • Medium to long-range molecular correlations • Highly entangled states	Modified rheological properties include: • Non-Newtonian flow • Glassy states • Long relaxation times • Mixed lubrication
Boundary	• Molecularly thin films (< 4 molecular diameters) • High loads or pressure • Low shear rates • Smooth surfaces or asperities	Onset of non-fluidlike properties: • Liquid-like to solid-like phase transitions • Appearance of new liquid-crystalline states • Epitaxially induced long-range ordering	Onset of tribological properties: • No flow until yield point or critical shear stress reached • Solid-like film behavior characterized by defect diffusion, dislocation motion, shear melting • Boundary lubrication

Based on work by *Granick* [18.219], *Hu* and *Granick* [18.220], and others [18.38, 207, 221] on the dynamic properties of short chain molecules such as alkanes and polymer melts confined between surfaces

[a] Confinement can lead to an increased or decreased order in a film, depending both on the surface lattice structure and the geometry of the confining cavity

[b] In each regime both the static and dynamic properties change. The static properties include the film density, the density distribution function, the potential of mean force, and various positional and orientational order parameters

[c] Dynamic properties include viscosity, viscoelastic constants, and tribological yield points such as the friction coefficient and critical shear stress

"boundary" friction and is the focus of the following sections. The term "boundary lubrication" is more commonly used to denote the friction of surfaces that contain a thin protective lubricating layer such as a surfactant monolayer, but here we shall use the term more broadly to include any molecularly thin solid, liquid, surfactant, or polymer film.

Experiments have shown that as a liquid film becomes progressively thinner, its physical properties change, at first quantitatively and then qualitatively [18.44, 47, 219, 220, 222, 223]. The quantitative changes are manifested by an increased viscosity, non-Newtonian flow behavior, and the replacement of normal melting by a glass transition, but the film remains recognizable as a liquid (Fig. 18.13). In tribology, this regime is commonly known as the "mixed lubrication" regime, where the rheological properties of a film are intermediate between the bulk and boundary properties. One may also refer to it as the "intermediate" regime (Table 18.4).

For even thinner films, the changes in behavior are more dramatic, resulting in a qualitative change in properties. Thus first-order phase transitions can now occur to solid or liquid-crystalline phases [18.46,201,207,221, 224–227], whose properties can no longer be characterized even qualitatively in terms of bulk or continuum liquid properties such as viscosity. These films now exhibit yield points (characteristic of fracture in solids) and their molecular diffusion and relaxation times can be ten orders of magnitude longer than in the bulk liquid or even in films that are just slightly thicker. The three friction regimes are summarized in Table 18.4.

18.7 Relationship Between Adhesion and Friction Between Dry (Unlubricated and Solid Boundary Lubricated) Surfaces

18.7.1 Amontons' Law and Deviations from It Due to Adhesion: The Cobblestone Model

Early theories and mechanisms for the dependence of friction on the applied normal force or load, L, were developed by *da Vinci*, *Amontons*, *Coulomb* and *Euler* [18.228]. For the macroscopic objects investigated, the friction was found to be directly proportional to the load, with no dependence on the contact area. This is described by the so-called Amontons' law:

$$F = \mu L \,, \tag{18.26}$$

where F is the shear or friction force and μ is a constant defined as the coefficient of friction. This friction law has a broad range of applicability and is still the principal means of quantitatively describing the friction between surfaces. However, particularly in the case of adhering surfaces, Amontons' law does not adequately describe the friction behavior with load, because of the finite friction force measured at zero and even negative applied loads.

When a lateral force, or shear stress, is applied to two surfaces in adhesive contact, the surfaces initially remain "pinned" to each other until some critical shear force is reached. At this point, the surfaces begin to slide past each other either smoothly or in jerks. The frictional force needed to initiate sliding from rest is known as the *static* friction force, denoted by F_s, while the force needed to maintain smooth sliding is referred to as the *kinetic* or *dynamic* friction force, denoted by F_k. In general, $F_s > F_k$. Two sliding surfaces may also move in regular jerks, known as stick–slip sliding, which is discussed in more detail in Sect. 18.8.3. Such friction forces cannot be described by models used for thick films that are viscous (see Sect. 18.8.1) and, therefore, shear as soon as the smallest shear force is applied.

Experimentally, it has been found that during both smooth and stick–slip sliding the local geometry of the contact zone remains largely unchanged from the static geometry. In an adhesive contact, the contact area as a function of load is thus generally well described by the JKR equation, (18.22). The friction force between two molecularly smooth surfaces sliding in adhesive contact is not simply proportional to the applied load, L, as might be expected from Amontons' law. There is an additional adhesion contribution that is proportional to the area of contact, A. Thus, in general, the interfacial friction force of dry, unlubricated surfaces sliding smoothly past each other in adhesive contact is given by

$$F = F_k = S_c A + \mu L \,, \tag{18.27}$$

where S_c is the "critical shear stress" (assumed to be constant), $A = \pi r^2$ is the contact area of radius r given by (18.22), and μ is the coefficient of friction. For low loads we have:

$$\begin{aligned} F &= S_c A = S_c \pi r^2 \\ &= S_c \pi \left[\frac{R}{K} \left(L + 6\pi R\gamma + \sqrt{12\pi R\gamma L + (6\pi R\gamma)^2} \right) \right]^{2/3} ; \end{aligned} \tag{18.28}$$

whereas for high loads (or high μ), (18.27) reduces to Amontons' law: $F = \mu L$. Depending on whether the friction force in (18.27) is dominated by the first or second term, one may refer to the friction as *adhesion-controlled* or *load-controlled*, respectively.

The following friction model, first proposed by *Tabor* [18.229] and developed further by *Sutcliffe* et al. [18.230], *McClelland* [18.231], and *Homola* et al. [18.45], has been quite successful at explaining the interfacial and boundary friction of two solid crystalline surfaces sliding past each other in the absence of wear. The surfaces may be unlubricated, or they may be

separated by a monolayer or more of some boundary lubricant or liquid molecules. In this model, the values of the critical shear stress S_c, and coefficient of friction μ, of (18.27) are calculated in terms of the energy needed to overcome the attractive intermolecular forces and compressive externally applied load as one surface is raised and then slid across the molecular-sized asperities of the other.

This model (variously referred to as the *interlocking asperity model*, *Coulomb friction*, or the *cobblestone model*) is similar to pushing a cart over a road of cobblestones where the cartwheels (which represent the molecules of the upper surface or film) must be made to roll over the cobblestones (representing the molecules of the lower surface) before the cart can move. In the case of the cart, the downward force of gravity replaces the attractive intermolecular forces between two material surfaces. When at rest, the cartwheels find grooves between the cobblestones where they sit in potential energy minima, and so the cart is at some stable mechanical equilibrium. A certain lateral force (the "push") is required to raise the cartwheels against the force of gravity in order to initiate motion. Motion will continue as long as the cart is pushed, and rapidly stops once it is no longer pushed. Energy is dissipated by the liberation of heat (phonons, acoustic waves, etc.) every time a wheel hits the next cobblestone. The cobblestone model is not unlike the *Coulomb* and *interlocking asperity* models of friction [18.228] except that it is being applied at the molecular level and for a situation where the external load is augmented by attractive intermolecular forces.

There are thus two contributions to the force pulling two surfaces together: the externally applied load or pressure, and the (internal) attractive intermolecular forces that determine the adhesion between the two surfaces. Each of these contributions affects the friction force in a different way, which we will discuss in more detail below.

18.7.2 Adhesion Force and Load Contribution to Interfacial Friction

Adhesion Force Contribution

Consider the case of two surfaces sliding past each other, as shown in Fig. 18.14. When the two surfaces are initially in adhesive contact, the surface molecules will adjust themselves to fit snugly together [18.233], in an analogous manner to the self-positioning of the cartwheels on the cobblestone road. A small tangential force applied to one surface will therefore not result in

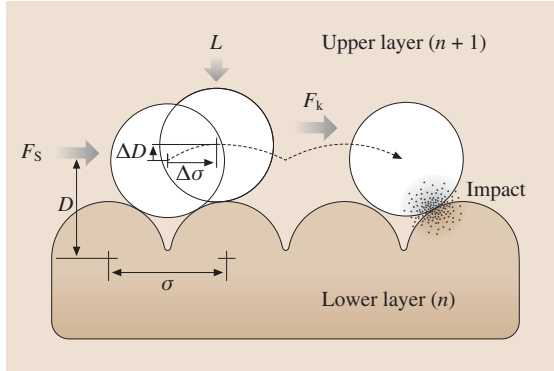


Fig. 18.14 Schematic illustration of how one molecularly smooth surface moves over another when a lateral force F is applied (the "cobblestone model"). As the upper surface moves laterally by some fraction of the lattice dimension $\Delta\sigma$, it must also move up by some fraction of an atomic or molecular dimension ΔD before it can slide across the lower surface. On impact, some fraction ε of the kinetic energy is "transmitted" to the lower surface, the rest being "reflected" back to the colliding molecule (upper surface). (After [18.232] with permission)

the sliding of that surface relative to the other. The attractive van der Waals forces between the surfaces must first be overcome by having the surfaces separate by a small amount. To initiate motion, let the separation between the two surfaces increase by a small amount ΔD, while the lateral distance moved is $\Delta\sigma$. These two values will be related via the geometry of the two surface lattices. The energy put into the system by the force F acting over a lateral distance $\Delta\sigma$ is

$$\text{Input energy: } F \times \Delta\sigma \ . \tag{18.29}$$

This energy may be equated with the change in interfacial or surface energy associated with separating the surfaces by ΔD, i. e., from the equilibrium separation $D = D_0$ to $D = (D_0 + \Delta D)$. Since $\gamma \propto D^{-2}$ for two flat surfaces, the energy cost may be approximated by:

$$\text{Surface energy change} \times \text{area:}$$
$$2\gamma A\left[1 - D_0^2/(D_0+\Delta D)^2\right] \approx 4\gamma A(\Delta D/D_0) \ , \tag{18.30}$$

where γ is the surface energy, A the contact area, and D_0 the surface separation at equilibrium. During steady state sliding (kinetic friction), not all of this energy will be "lost" or absorbed by the lattice every time the surface molecules move by one lattice spacing: Some fraction will be reflected during each impact of the "cartwheel"

molecules [18.231]. Assuming that a fraction ε of the above surface energy is "lost" every time the surfaces move across the characteristic length $\Delta\sigma$ (Fig. 18.14), we obtain after equating (18.29) and (18.30)

$$S_c = \frac{F}{A} = \frac{4\gamma\varepsilon\Delta D}{D_0\Delta\sigma} . \tag{18.31}$$

For a typical hydrocarbon or a van der Waals surface, $\gamma \approx 25\,\mathrm{mJ\,m^{-2}}$. Other typical values would be: $\Delta D \approx 0.05\,\mathrm{nm}$, $D_0 \approx 0.2\,\mathrm{nm}$, $\Delta\sigma \approx 0.1\,\mathrm{nm}$, and $\varepsilon \approx 0.1$–0.5. Using the above parameters, (18.31) predicts $S_c \approx (2.5$–$12.5)\times 10^7\,\mathrm{N\,m^{-2}}$ for van der Waals surfaces. This range of values compares very well with typical experimental values of $2\times 10^7\,\mathrm{N\,m^{-2}}$ for hydrocarbon or mica surfaces sliding in air (see Fig. 18.16) or separated by one molecular layer of cyclohexane [18.45].

The above model suggests that all interfaces, whether dry or lubricated, dilate just before they shear or slip. This is a small but important effect: The dilation provides the crucial extra space needed for the molecules to slide across each other or flow. This dilation has been computed by *Thompson* and *Robbins* [18.201] and *Zaloj* et al. [18.234] and measured by *Dhinojwala* et al. [18.235].

This model may be extended, at least semi-quantitatively, to lubricated sliding, where a thin liquid film is present between the surfaces. With an increase in the number of liquid layers between the surfaces, D_0 increases while ΔD decreases, hence the lower the friction force. This is precisely what is observed, but with more than one liquid layer between two surfaces the situation becomes too complex to analyze analytically (actually, even with one or no interfacial layers, the calculation of the fraction of energy dissipated per molecular collision ε is not a simple matter). Furthermore, even in systems as simple as linear alkanes, interdigitation and interdiffusion have been found to contribute strongly to the properties of the system [18.114, 236]. Sophisticated modeling based on computer simulations is now required, as described in the following section.

Relation Between Boundary Friction and Adhesion Energy Hysteresis

While the above equations suggest that there is a direct correlation between friction and adhesion, this is not the case. The correlation is really between friction and adhesion hysteresis, described in Sect. 18.5.4. In the case of friction, this subtle point is hidden in the factor ε, which is a measure of the amount of energy absorbed (dissipated, transferred, or "lost") by the lower surface when it is impacted by a molecule from the upper surface. If $\varepsilon = 0$, all the energy is reflected, and there will be no kinetic friction force nor any adhesion hysteresis, but the absolute magnitude of the adhesion force or energy will remain finite and unchanged. This is illustrated in Figs. 18.17 and 18.19.

The following simple model shows how adhesion hysteresis and friction may be quantitatively related. Let $\Delta\gamma = \gamma_R - \gamma_A$ be the adhesion energy hysteresis per unit area, as measured during a typical loading–unloading cycle (see Figs. 18.17a and 18.19c,d). Now consider the same two surfaces sliding past each other and assume that frictional energy dissipation occurs through the same mechanism as adhesion energy dissipation, and that both occur over the same characteristic molecular length scale σ. Thus, when the two surfaces (of contact area $A = \pi r^2$) move a distance σ, equating the frictional energy ($F\times\sigma$) to the dissipated adhesion energy ($A\times\Delta\gamma$), we obtain

$$\text{Friction force: } F = \frac{A\times\Delta\gamma}{\sigma} = \frac{\pi r^2}{\sigma}(\gamma_R - \gamma_A) , \tag{18.32}$$

$$\text{or} \quad \text{Critical shear stress: } S_c = F/A = \Delta\gamma/\sigma , \tag{18.33}$$

which is the desired expression and has been found to give order of magnitude agreement between measured friction forces and adhesion energy hysteresis [18.207]. If we equate (18.33) with (18.31), since $4\Delta D/(D_0\Delta\sigma) \approx 1/\sigma$, we obtain the intuitive relation

$$\varepsilon = \frac{\Delta\gamma}{\gamma} . \tag{18.34}$$

External Load Contribution to Interfacial Friction

When there is no interfacial adhesion, S_c is zero. Thus, in the absence of any adhesive forces between two surfaces, the only "attractive" force that needs to be overcome for sliding to occur is the externally applied load or pressure.

For a preliminary discussion of this question, it is instructive to compare the magnitudes of the *externally* applied pressure to the *internal* van der Waals pressure between two smooth surfaces. The internal van der Waals pressure between two flat surfaces is given (see Table 18.2) by $P = A_H/6\pi D_0^3 \approx 1\,\mathrm{GPa}$ (10^4 atm), using a typical Hamaker constant of $A_H = 10^{-19}\,\mathrm{J}$, and assuming $D_0 \approx 2\,\mathrm{nm}$ for the equilibrium interatomic spacing. This implies that we should not expect the externally

applied load to affect the interfacial friction force F, as defined by (18.27), until the externally applied pressure L/A begins to exceed ~ 100 MPa (10^3 atm). This is in agreement with experimental data [18.237] where the effect of load became dominant at pressures in excess of 10^3 atm.

For a more general semiquantitative analysis, again consider the cobblestone model used to derive (18.31), but now include an additional contribution to the surface energy change of (18.30) due to the work done against the external load or pressure, $L\Delta D = P_{\text{ext}} A \Delta D$ (this is equivalent to the work done against gravity in the case of a cart being pushed over cobblestones). Thus:

$$S_c = \frac{F}{A} = \frac{4\gamma\varepsilon\Delta D}{D_0 \Delta\sigma} + \frac{P_{\text{ext}}\varepsilon\Delta D}{\Delta\sigma} , \tag{18.35}$$

which gives the more general relation

$$S_c = F/A = C_1 + C_2 P_{\text{ext}} , \tag{18.36}$$

where $P_{\text{ext}} = L/A$ and C_1 and C_2 are constants characteristic of the surfaces and sliding conditions. The constant $C_1 = 4\gamma\varepsilon\Delta D/(D_0\Delta\sigma)$ depends on the mutual adhesion of the two surfaces, while both C_1 and $C_2 = \varepsilon\Delta D/\Delta\sigma$ depend on the topography or atomic bumpiness of the surface groups (Fig. 18.14). The smoother the surface groups the smaller the ratio $\Delta D/\Delta\sigma$ and hence the lower the value of C_2. In addition, both C_1 and C_2 depend on ε (the fraction of energy dissipated per collision), which depends on the relative masses of the shearing molecules, the sliding velocity, the temperature, and the characteristic molecular relaxation processes of the surfaces. This is by far the most difficult parameter to compute, and yet it is the most important since it represents the energy transfer mechanism in any friction process, and since ε can vary between 1 and 0, it determines whether a particular friction force will be large or close to zero. Molecular simulations offer the best way to understand and predict the magnitude of ε, but the complex multi-body nature of the problem makes simple conclusions difficult to draw [18.239–241]. Some of the basic physics of the energy transfer and dissipation of the molecular collisions can be drawn from simplified models such as a 1-D three-body system [18.214].

Finally, the above equation may also be expressed in terms of the friction force F:

$$F = S_c A = C_1 A + C_2 L . \tag{18.37}$$

Equations similar to (18.36) and (18.37) were previously derived by *Derjaguin* [18.242,243] and by *Briscoe*

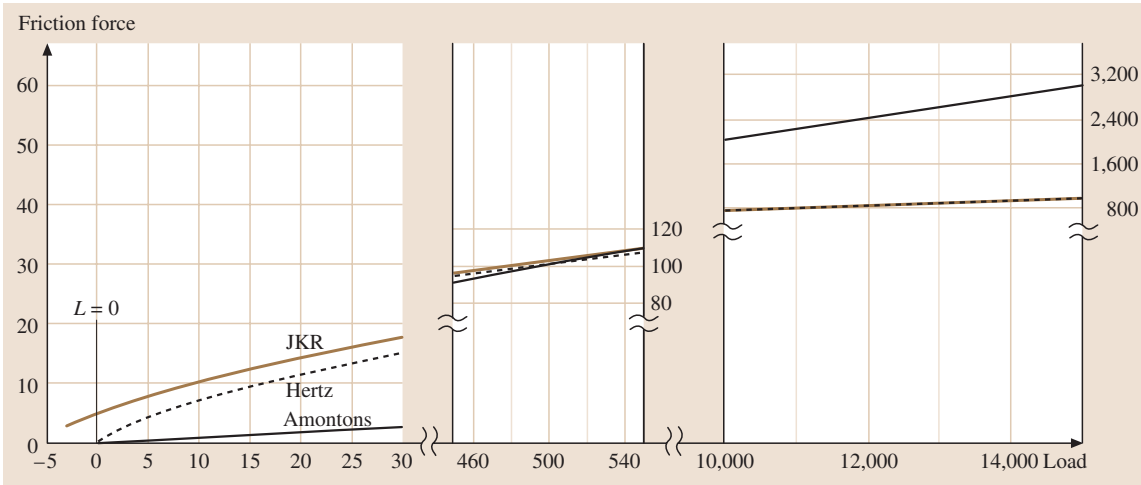


Fig. 18.15 Friction as a function of load for smooth surfaces. At low loads, the friction is dominated by the $C_1 A$ term of (18.37). The adhesion contribution (JKR curve) is most prominent near zero load where the Hertzian and Amontons' contributions to the friction are minimal. As the load increases, the adhesion contribution becomes smaller as the JKR and Hertz curves converge. In this range of loads, the linear $C_2 L$ contribution surpasses the area contribution to the friction. At much higher loads the explicit load dependence of the friction dominates the interactions, and the observed behavior approaches Amontons' law. It is interesting to note that for smooth surfaces the pressure over the contact area does not increase as rapidly as the load. This is because as the load is increased, the surfaces deform to increase the surface area and thus moderate the contact pressure. (After [18.238] with permission of Kluwer Academic Publishers)

and *Evans* [18.244], where the constants C_1 and C_2 were interpreted somewhat differently than in this model.

In the absence of any attractive interfacial force, we have $C_1 \approx 0$, and the second term in (18.36) and (18.37) should dominate (Fig. 18.15). Such situations typically arise when surfaces repel each other across the lubricating liquid film. In such cases, the total frictional force should be low and should increase *linearly* with the external load according to

$$F = C_2 L \,. \tag{18.38}$$

An example of such lubricated sliding occurs when two mica surfaces slide in water or in salt solution (see Fig. 18.20), where the short-range "hydration" forces between the surfaces are repulsive. Thus, for sliding in 0.5 M KCl it was found that $C_2 = 0.015$ [18.245]. Another case where repulsive surfaces eliminate the adhesive contribution to friction is for polymer chains attached to surfaces at one end and swollen by a good solvent [18.175]. For this class of systems, $C_2 < 0.001$ for a finite range of polymer layer compressions (normal loads, L). The low friction between the surfaces in this regime is attributed to the entropic repulsion between the opposing brush layers with a minimum of entanglement between the two layers. However, with higher normal loads, the brush layers become compressed and begin to entangle, which results in higher friction (see [18.246]).

It is important to note that (18.38) has exactly the same form as Amontons' Law

$$F = \mu L \,, \tag{18.39}$$

where μ is the coefficient of friction.

At the molecular level a thermodynamic analog of the Coulomb or cobblestone models (see Sect. 18.7.1) based on the "contact value theorem" [18.3, 238, 245] can explain why $F \propto L$ also holds at the microscopic or molecular level. In this analysis we consider the surface molecular groups as being momentarily compressed and decompressed as the surfaces move along. Under irreversible conditions, which always occur when a cycle is completed in a finite amount of time, the energy "lost" in the compression–decompression cycle is dissipated as heat. For two non-adhering surfaces, the stabilizing pressure P_i acting locally between any two elemental contact points i of the surfaces may be expressed by the contact value theorem [18.3]:

$$P_i = \rho_i k_B T = k_B T / V_i \,, \tag{18.40}$$

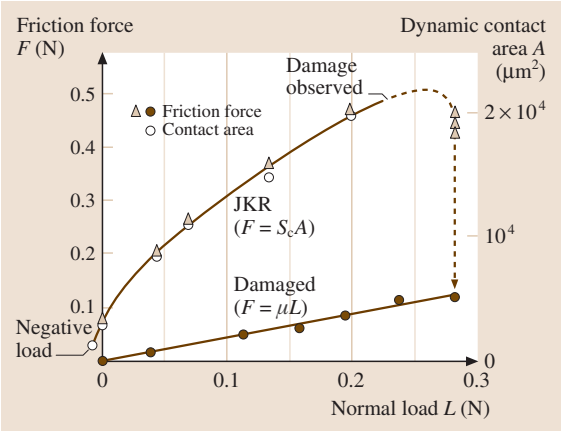


Fig. 18.16 Friction force F and contact area A vs. load L for two mica surfaces sliding in adhesive contact in dry air. The contact area is well described by the JKR theory, (18.22), even during sliding, and the friction force is found to be directly proportional to this area, (18.28). The *vertical dashed line* and *arrow* show the transition from interfacial to normal friction with the onset of wear (*lower curve*). Sliding velocity is 0.2 μm s^{-1}. (After [18.45] with permission, copyright 1989 American Society of Mechanical Engineers)

where $\rho_i = V_i^{-1}$ is the local number density (per unit volume) or activity of the interacting entities, be they molecules, atoms, ions or the electron clouds of atoms. This equation is essentially the osmotic or entropic pressure of a gas of confined molecules. As one surface moves across the other, local regions become compressed and decompressed by a volume ΔV_i. The work done per cycle can be written as $\varepsilon P_i \Delta V_i$, where ε ($\varepsilon \leq 1$) is the fraction of energy per cycle "lost" as heat, as defined earlier. The energy balance shows that for each compression–decompression cycle, the dissipated energy is related to the friction force by

$$F_i x_i = \varepsilon P_i \Delta V_i \,, \tag{18.41}$$

where x_i is the lateral distance moved per cycle, which can be the distance between asperities or the distance between surface lattice sites. The pressure at each contact junction can be expressed in terms of the local normal load L_i and local area of contact A_i as $P_i = L_i / A_i$. The volume change over a cycle can thus be expressed as $\Delta V_i = A_i z_i$, where z_i is the vertical distance of confinement. Inserting these into (18.41), we get

$$F_i = \varepsilon L_i (z_i / x_i) \,, \tag{18.42}$$

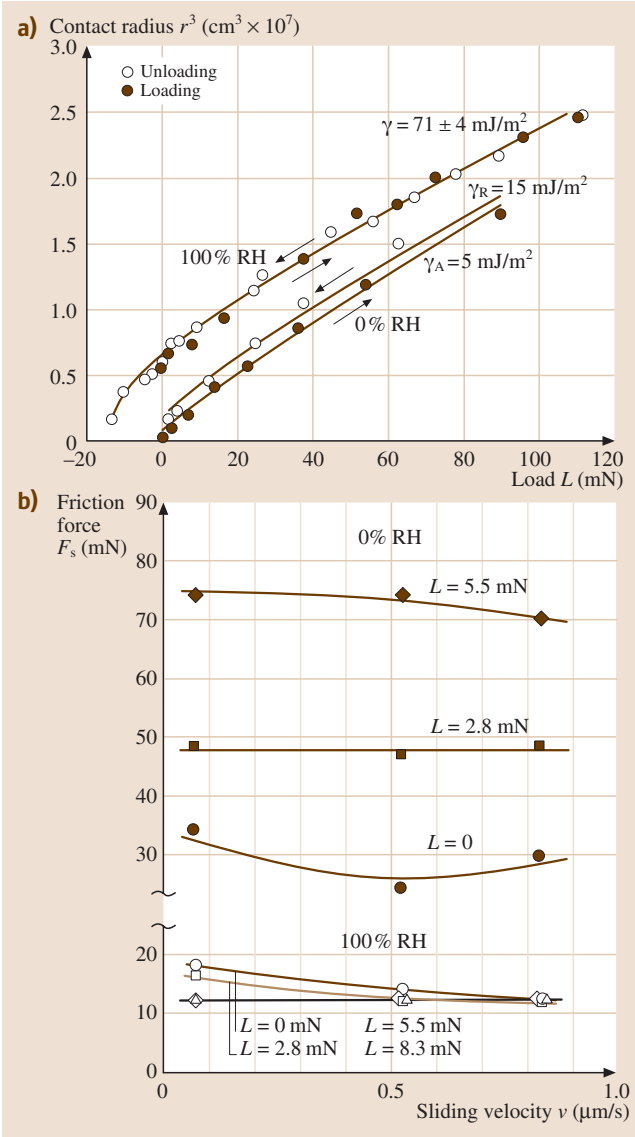


Fig. 18.17 **(a)** Contact radius r versus externally applied load L for loading and unloading of two hydrophilic silica surfaces exposed to dry and humid atmospheres. Note that while the adhesion is higher in humid air, the *hysteresis* in the adhesion is higher in dry air. **(b)** Effect of velocity on the static friction force F_s for hydrophobic (heat-treated electron beam evaporated) silica in dry and humid air. The effects of humidity, load, and sliding velocity on the friction forces, as well as the stick–slip friction of the hydrophobic surfaces, are qualitatively consistent with a "friction" phase diagram representation as in Fig. 18.28. (After [18.41]. Copyright 1994, with permission from Elsevier Science)

which is independent of the local contact area A_i. The total friction force is thus

$$F = \sum F_i = \sum \varepsilon L_i (z_i/x_i)$$
$$= \varepsilon \langle z_i/x_i \rangle \sum L_i = \mu L \,, \qquad (18.43)$$

where it is assumed that on average the local values of L_i and P_i are independent of the local "slope" z_i/x_i. Therefore, the friction coefficient μ is a function only of the average surface topography and the sliding velocity, but is independent of the local (real) or macroscopic (apparent) contact areas.

While this analysis explains non-adhering surfaces, there is still an additional explicit contact area contribution for the case of adhering surfaces, as in (18.37). The distinction between the two cases arises because the initial assumption of the contact value theorem, (18.40), is incomplete for adhering systems. A more appropriate starting equation would reflect the full intermolecular interaction potential, including the attractive interactions, in addition to the purely repulsive contributions of (18.40), much as the van der Waals equation of state modifies the ideal gas law.

18.7.3 Examples of Experimentally Observed Friction of Dry Surfaces

Numerous model systems have been studied with a surface forces apparatus (SFA) modified for friction experiments (see Sect. 18.2.3). The apparatus allows for control of load (normal force) and sliding speed, and simultaneous measurement of surface separation, surface shape, true (molecular) area of contact between smooth surfaces, and friction forces. A variety of both unlubricated and solid and liquid lubricated surfaces have been studied both as smooth single-asperity contacts and after they have been roughened by shear-induced damage.

Figure 18.16 shows the contact area, A, and friction force, F, both plotted against the applied load, L, in an experiment in which two molecularly smooth surfaces of mica in adhesive contact were slid past each other in an atmosphere of dry nitrogen gas. This is an example of the low load, adhesion controlled limit, which is excellently described by (18.28). In a number of different experiments, S_c was measured to be 2.5×10^7 N m^{-2} and to be independent of the sliding velocity [18.45,247]. Note that there is a friction force even at negative loads, where the surfaces are still sliding in adhesive contact.

Figure 18.17 shows the correlation between adhesion hysteresis and friction for two surfaces consisting of silica films deposited on mica substrates [18.41].

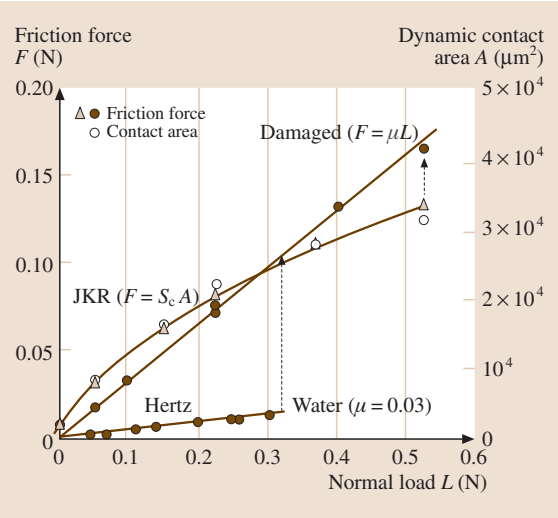


Fig. 18.18 Sliding of mica surfaces, each coated with a 2.5-nm-thick monolayer of calcium stearate surfactant, in the absence of damage (obeying JKR type boundary friction) and in the presence of damage (obeying Amontons' type normal friction). Note that both for this system and for the bare mica in Figs. 18.16 and 18.20, the friction force obeys Amontons' law with a friction coefficient of $\mu \approx 0.3$ after damage occurs. At much higher applied loads, the undamaged surfaces also follow Amontons' type sliding, but for a different reason: The dependence on adhesion becomes smaller. *Lower line:* Interfacial sliding with a monolayer of water between the mica surfaces, shown for comparison. (After [18.247]. Copyright 1990, with permission from Elsevier Science)

The friction between undamaged hydrophobic silica surfaces showed stick–slip both at dry conditions and at 100% relative humidity. Similar to the mica surfaces in Figs. 18.16, 18.18, and 18.20, the friction of damaged silica surfaces obeyed Amontons' law with a friction coefficient of 0.25 both at dry conditions and at 55% relative humidity.

The high friction force of unlubricated sliding can often be reduced by treating the solid surface with a boundary layer of some other solid material that exhibits lower friction such as a surfactant monolayer, or by ensuring that during sliding, a thin liquid film remains between the surfaces (as will be discussed in Sect. 18.8). The effectiveness of a solid boundary lubricant layer on reducing the forces of friction is illustrated in Fig. 18.18. Comparing this with the friction of the unlubricated/untreated surfaces (Fig. 18.16) shows that the critical shear stress has been reduced by a factor of about ten: from 2.5×10^7 to $3.5 \times 10^6\,\mathrm{N\,m^{-2}}$. At much higher applied loads or pressures, the friction force is proportional to the load, rather than the area of contact [18.237], as expected from (18.27).

Yamada and *Israelachvili* [18.248] studied the adhesion and friction of fluorocarbon surfaces (surfactant-coated boundary lubricant layers), which were compared to those of hydrocarbon surfaces. They concluded that well-ordered fluorocarbon surfaces have a high friction, in spite of their lower adhesion energy (in agreement with previous findings). The low friction coefficient of PTFE (Teflon) must, therefore, be due to some effect other than a low adhesion. For example, the softness of PTFE, which allows material to flow at the interface, which thus behaves like a fluid lubricant. On a related issue, *Luengo* et al. [18.249] found that C_{60} surfaces also exhibited low adhesion but high friction. In both cases the high friction appears to arise from the bulky surface groups – fluorocarbon compared to hydrocarbon groups in the former, large fullerene spheres in the latter. Apparently, the fact that C_{60} molecules rotate *in their lattice* does not make them a good lubricant: The molecules of the opposing surface must still climb over them in order to slide, and this requires energy that is independent of whether the surface molecules are fixed or freely rotating. Larger particles such as ~ 25 nm-sized nanoparticles (also known as "inorganic fullerenes") do appear to produce low friction by behaving like molecular ball bearings, but the potential of this promising new class of solid lubricant has still to be explored [18.250].

Figure 18.19 illustrates the relationship between adhesion hysteresis and friction for surfactant-coated surfaces under different conditions. This effect, however, is much more general and has been shown to hold for other surfaces as well [18.41, 232, 251].

Direct comparisons between absolute adhesion energies and friction forces show little correlation. In some cases, higher adhesion energies for the same system under different conditions correspond with lower friction forces. For example, for hydrophilic silica surfaces (Fig. 18.17) it was found that with increasing relative humidity the adhesion energy *increases*, but the adhesion energy hysteresis measured in a loading–unloading cycle *decreases*, as does the friction force [18.41]. For hydrophobic silica surfaces under dry conditions, the friction at load $L = 5.5\,\mathrm{mN}$ was $F = 75\,\mathrm{mN}$. For the same sample, the adhesion energy hysteresis was $\Delta\gamma = 10\,\mathrm{mJ\,m^{-2}}$, with a contact area of $A \approx 10^{-8}\,\mathrm{m^2}$ at the same load. Assuming a value for the characteristic distance σ on the order of one lattice spacing, $\sigma \approx 1$ nm, and inserting these values into (18.32), the friction force

is predicted to be $F \approx 100$ mN for the kinetic friction force, which is close to the measured value of 75 mN. Alternatively, we may conclude that the dissipation factor is $\varepsilon = 0.75$, i. e., that almost all the energy is dissipated as heat at each molecular collision.

A liquid lubricant film (Sect. 18.8.3) is usually much more effective at lowering the friction of two surfaces than a solid boundary lubricant layer. However, to successfully use a liquid lubricant, it must "wet" the surfaces, that is, it should have a high affinity for the surfaces, so that not all the liquid molecules become squeezed out when the surfaces come close together, even under a large compressive load. Another important requirement is that the liquid film remains a liquid under tribological conditions, i. e., that it does not epitaxially solidify between the surfaces.

Effective lubrication usually requires that the lubricant be injected between the surfaces, but in some cases the liquid can be made to condense from the vapor. This is illustrated in Fig. 18.20 for two untreated mica surfaces sliding with a thin layer of water between them. A monomolecular film of water (of thickness 0.25 nm per surface) has reduced S_c from its value for dry surfaces (Fig. 18.16) by a factor of more than 30, which

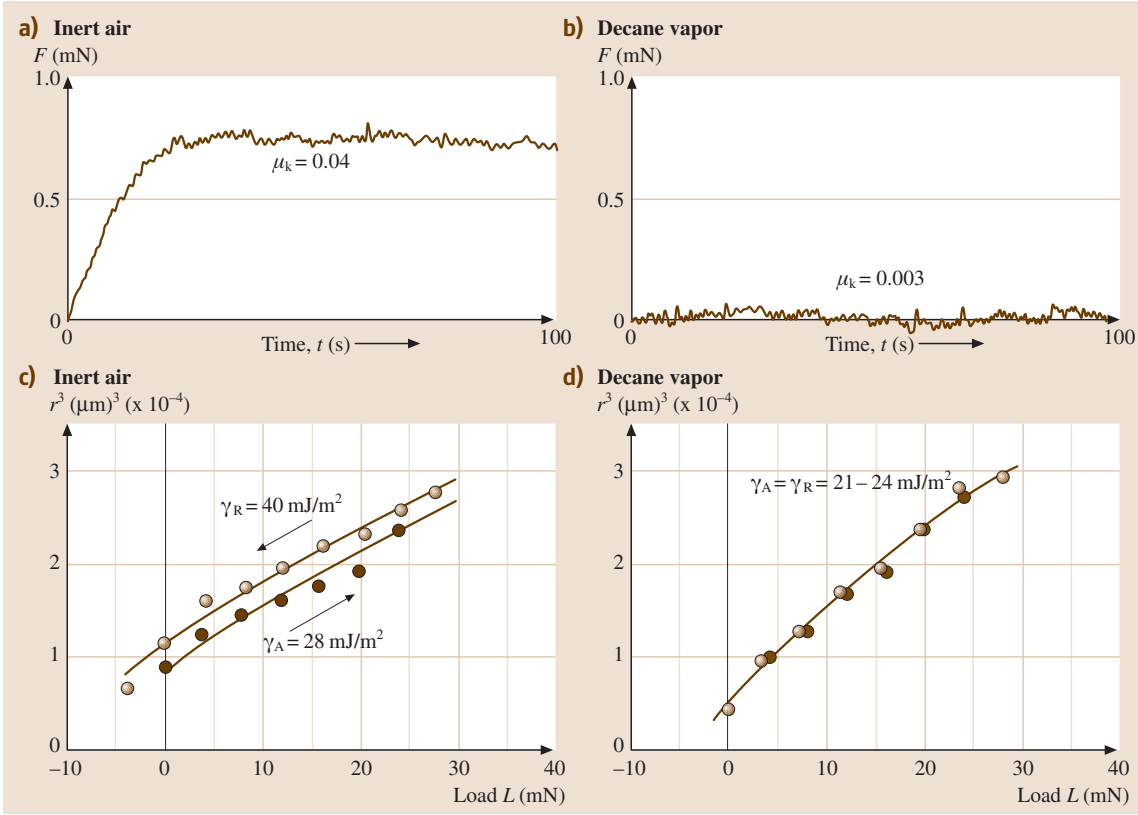


Fig. 18.19a–d *Top:* Friction traces for two fluid-like calcium alkylbenzene sulfonate monolayer-coated surfaces at 25 °C showing that the friction force is much higher between dry monolayers **(a)** than between monolayers whose fluidity has been enhanced by hydrocarbon penetration from vapor **(b)**. *Bottom:* Contact radius vs. load (r^3 vs. L) data measured for the same two surfaces as above and fitted to the JKR equation (18.22), shown by the *solid curves*. For dry monolayers **(c)** the adhesion energy on unloading ($\gamma_R = 40\,\mathrm{mJ\,m^{-2}}$) is greater than that on loading ($\gamma_R = 28\,\mathrm{mJ\,m^{-2}}$), indicative of an adhesion energy hysteresis of $\Delta\gamma = \gamma_R - \gamma_A = 12\,\mathrm{mJ\,m^{-2}}$. For monolayers exposed to saturated decane vapor **(d)** their adhesion hysteresis is zero ($\gamma_A = \gamma_R$), and both the loading and unloading data are well fitted by the thermodynamic value of the surface energy of fluid hydrocarbon chains, $\gamma = 24\,\mathrm{mJ\,m^{-2}}$. (After [18.207] with permission. Copyright 1993 American Chemical Society)

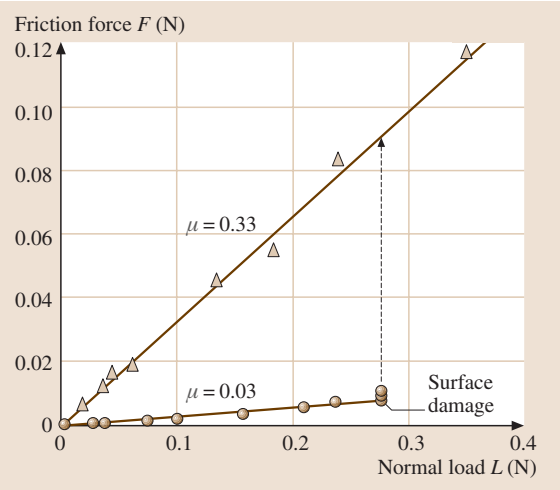


Fig. 18.20 Two mica surfaces sliding past each other while immersed in a 0.01 M KCl salt solution (nonadhesive conditions). The water film is molecularly thin, 0.25 to 0.5 nm thick, and the interfacial friction force is very low: $S_c \approx 5 \times 10^5\ \mathrm{N\,m^{-2}}$, $\mu \approx 0.03$ (before damage occurs). After the surfaces have become damaged, the friction coefficient is ca. 0.3. (After [18.247]. Copyright 1990, with permission from Elsevier Science)

may be compared with the factor of ten attained with the boundary lubricant layer (of thickness 2.5 nm per surface) in Fig. 18.18. Water appears to have unusual lubricating properties and usually gives wearless friction with no stick–slip.

The effectiveness of a water film only 0.25 nm thick to lower the friction force by more than an order of magnitude is attributed to the "hydrophilicity" of the mica surface (mica is "wetted" by water) and to the existence of a strongly repulsive short-range hydration force between such surfaces in aqueous solutions, which effectively removes the adhesion-controlled contribution to the friction force [18.245]. It is also interesting that a 0.25-nm-thick water film between two mica surfaces is sufficient to bring the coefficient of friction down to 0.01–0.02, a value that corresponds to the unusually low friction of ice. Clearly, a single monolayer of water can be a very good lubricant – much better than most other monomolecular liquid films – for reasons that will be discussed in Sect. 18.9.

Dry polymer layers (Fig. 18.21) typically show a high initial static friction ("stiction") as sliding commences from rest in adhesive contact. The development of the friction force after a change in sliding direction, a gradual transition from stick–slip to smooth sliding, is shown in Fig. 18.21. A correlation between adhesion hysteresis and friction similar to the one observed for silica surfaces in Fig. 18.17 can be seen for dry polymer layers below their glass transition temperature. As shown in Fig. 18.12b,c, the adhesion hysteresis for polystyrene surfaces can be increased by irradiation to induce scission of chains, and it has been found that the steady state friction force (kinetic friction) shows a similar increase with irradiation time [18.208].

Figure 18.22 shows an example of a computer simulation of the sliding of two unlubricated silicon surfaces (modeled as a tip sliding over a planar surface) [18.91]. The sliding proceeds through a series of stick–slip events, and information on the friction force and the local order of the initially crystalline surfaces can be obtained. Similar studies for cold welding systems [18.91] have demonstrated the occurrence of shear or friction damage within the sliding surface (tip) as the lowest layer of it adheres to the bottom surface.

18.7.4 Transition from Interfacial to Normal Friction with Wear

Frictional damage can have many causes such as adhesive tearing at high loads or overheating at high sliding speeds. Once damage occurs, there is a transition from "interfacial" to "normal" or load-controlled friction as the surfaces become forced apart by the torn-out as-

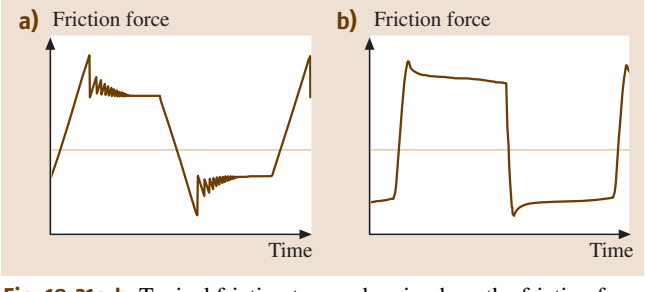


Fig. 18.21a,b Typical friction traces showing how the friction force varies with the sliding time for two symmetric, glassy polymer films at dry conditions. Qualitative features that are common to both polystyrene and polyvinyl benzyl chloride: **(a)** Decaying stick–slip motion is observed until smooth sliding is attained if the motion continues for a sufficiently long distance. **(b)** Smooth sliding observed at sufficiently high speeds. Similar observations have been made by *Berthoud* et al. [18.252] in measurements on polymethyl methacrylate. (After [18.208] with permission. Copyright 2002 American Association for the Advancement of Science)

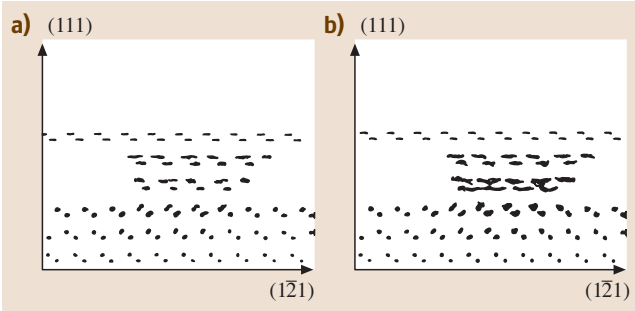


Fig. 18.22a,b Computer simulation of the sliding of two contacting Si surfaces (a tip and a flat surface). Shown are particle trajectories in a constant-force simulation, $F_{z,\text{external}} = -2.15 \times 10^{-8}$ N, viewed along the (10$\bar{1}$) direction just before **(a)** and after **(b)** a stick–slip event for a large, initally ordered, dynamic tip. (After [18.91] with permission of Kluwer Academic Publishers)

perities (wear particles). For low loads, the friction changes from obeying $F = S_c A$ to obeying Amontons' law, $F = \mu L$, as shown in Figs. 18.16 and 18.18, and sliding now proceeds smoothly with the surfaces separated by a 10–100 nm forest of wear debris (in this case, mica flakes). The wear particles keep the surfaces apart over an area that is much greater than their size, so that even one submicroscopic particle or asperity can cause a significant reduction in the area of contact and, therefore, in the friction [18.247]. For this type of frictional sliding, one can no longer talk of the molecular contact area of the two surfaces, although the macroscopic or "apparent" area is still a useful parameter.

One remarkable feature of the transition from interfacial to normal friction of brittle surfaces is that while the strength of interfacial friction, as reflected in the values of S_c, is very dependent on the type of surface and on the liquid film between the surfaces, this is not the case once the transition to normal friction has occurred. At the onset of damage, the material properties of the underlying substrates control the friction. In Figs. 18.16, 18.18, and 18.20 the friction for the damaged surfaces is that of any damaged mica–mica system, $\mu \approx 0.3$, *independent of the initial surface coatings or liquid films between the surfaces*. A similar friction coefficient was found for damaged silica surfaces [18.41].

In order to practically modify the frictional behavior of such brittle materials, it is important to use coatings that will both alter the interfacial tribological character and remain intact and protect the surfaces from damage during sliding. *Berman* et al. [18.253] found that the friction of a strongly bound octadecyl phosphonic acid monolayer on alumina surfaces was higher than for untreated, undamaged α-alumina surfaces, but the bare surfaces easily became damaged upon sliding, resulting in an ultimately higher friction system with greater wear rates than the more robust monolayer-coated surfaces.

Clearly, the mechanism and factors that determine *normal* friction are quite different from those that govern *interfacial* friction (Sects. 18.7.1–18.7.2). This effect is not general and may only apply to brittle materials. For example, the friction of ductile surfaces is totally different and involves the continuous plastic deformation of contacting surface asperities during sliding, rather than the rolling of two surfaces on hard wear particles [18.209]. Furthermore, in the case of ductile surfaces, water and other surface-active components do have an effect on the friction coefficients under "normal" sliding conditions.

18.8 Liquid Lubricated Surfaces

18.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime

Experimentally, it is usually difficult to unambiguously establish which type of sliding mode is occurring, but an empirical criterion, based on the Stribeck curve (Fig. 18.13), is often used as an indicator. This curve shows how the friction force or the coefficient of friction is expected to vary with sliding speed, depending on which type of friction regime is operating. For thick liquid lubricant films whose behavior can be described by bulk (continuum) properties, the friction forces are essentially the hydrodynamic or viscous drag forces. For example, for two plane parallel surfaces of area A separated by a distance D and moving laterally relative to each other with velocity v, if the intervening liquid is *Newtonian*, i. e., if its viscosity η is independent of the shear rate, the frictional force experienced by the surfaces is given by

$$F = \frac{\eta A v}{D} , \tag{18.44}$$

where the shear rate $\dot{\gamma}$ is defined by

$$\dot{\gamma} = \frac{v}{D} \,. \tag{18.45}$$

At higher shear rates, two additional effects often come into play. First, certain properties of liquids may change at high $\dot{\gamma}$ values. In particular, the effective viscosity may become non-Newtonian, one form given by

$$\eta = \dot{\gamma}^n \,, \tag{18.46}$$

where $n = 0$ (i. e., $\eta_{\text{eff}} = \text{constant}$) for Newtonian fluids, $n > 0$ for shear thickening (dilatent) fluids, and $n < 0$ for shear thinning (pseudoplastic) fluids (the latter become less viscous, i. e., flow more easily, with increasing shear rate). An additional effect on η can arise from the higher local stresses (pressures) experienced by the liquid film as $\dot{\gamma}$ increases. Since the viscosity is generally also sensitive to the pressure (usually increasing with P), this effect also acts to increase η_{eff} and thus the friction force.

A second effect that occurs at high shear rates is surface deformation, arising from the large hydrodynamic forces acting on the sliding surfaces. For example, Fig. 18.23 shows how two surfaces deform elastically when the sliding speed increases to a high value. These deformations alter the hydrodynamic friction forces, and this type of friction is often referred to as *elastohydrodynamic lubrication* (EHD or EHL), as mentioned in Table 18.4.

How thin can a liquid film be before its dynamic, e.g., viscous flow, behavior ceases to be described by bulk properties and continuum models? Concerning the static properties, we have already seen in Sect. 18.4.3 that films composed of simple liquids display continuum behavior down to thicknesses of 4–10 molecular diameters. Similar effects have been found to apply to the dynamic properties, such as the viscosity, of simple liquids in thin films. Concerning viscosity measurements, a number of dynamic techniques were recently developed [18.11–13, 43, 254, 255] for directly measuring the viscosity as a function of film thickness and shear rate across very thin liquid films between two surfaces. By comparing the results with theoretical predictions of fluid flow in thin films, one can determine the effective positions of the shear planes and the onset of non-Newtonian behavior in very thin films.

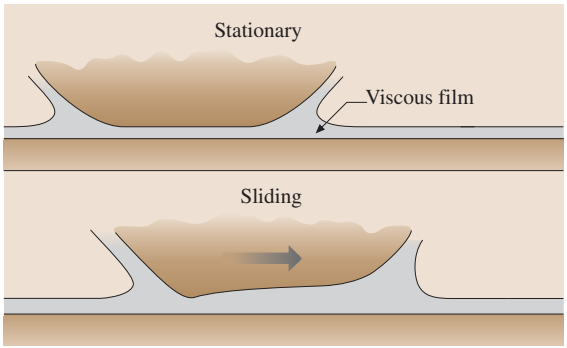
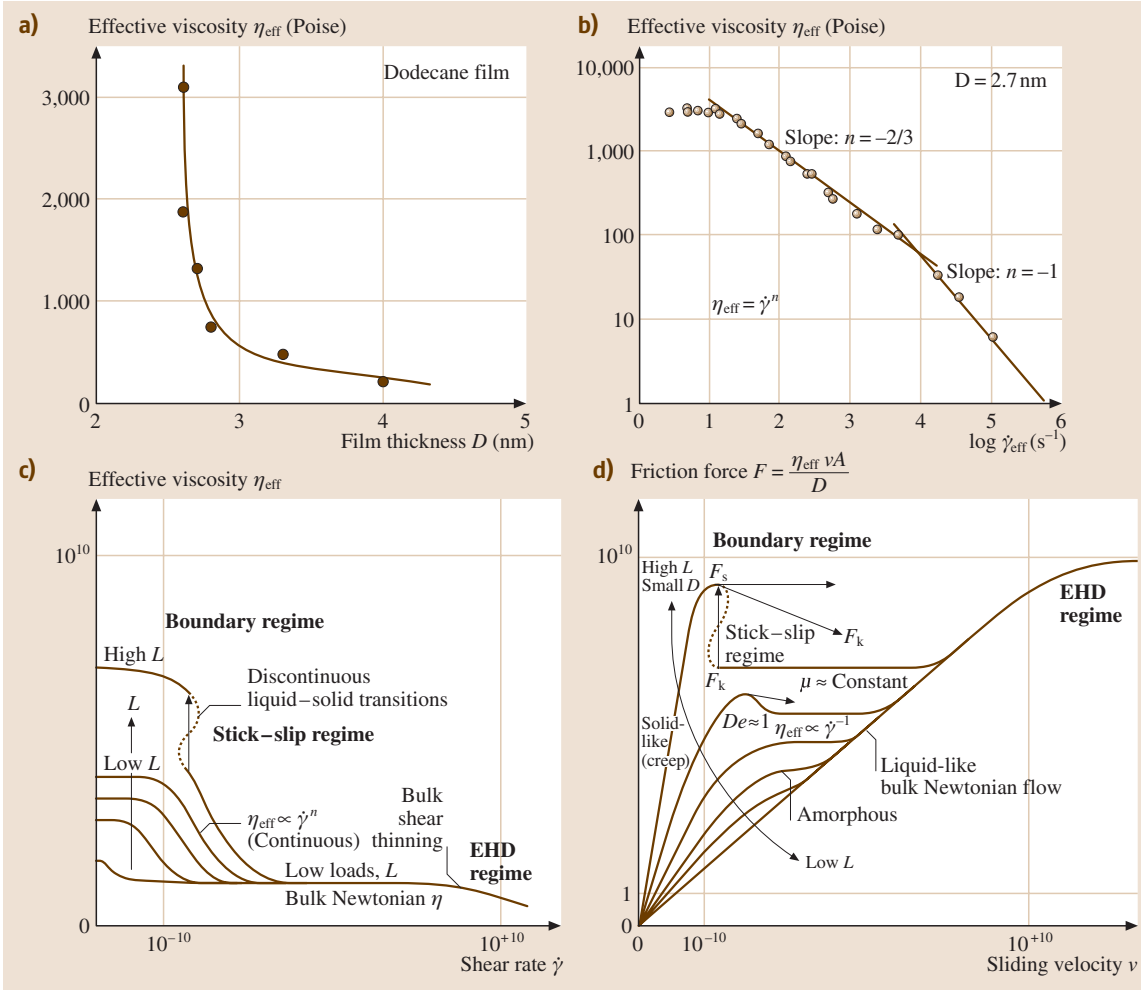


Fig. 18.23 *Top:* Stationary surfaces (one more deformable and one rigid) separated by a thick liquid film. *Bottom:* Elastohydrodynamic deformation of the upper surface during sliding. (After [18.1] with permission)

The results show that for simple liquids, including linear chain molecules such as alkanes, the viscosity in thin films is the same, within 10%, as the bulk even for films as thin as 10 molecular diameters (or segment widths) [18.11–13, 254, 255]. This implies that the shear plane is effectively located within one molecular diameter of the solid–liquid interface, and these conclusions were found to remain valid even at the highest shear rates studied (of $\sim 2 \times 10^5\ \text{s}^{-1}$). With water between two mica or silica surfaces [18.22, 254–256] this has been found to be the case (to within $\pm 10\%$) down to surface separations as small as 2 nm, implying that the shear planes must also be within a few tenths of a nanometer of the solid–liquid interfaces. These results appear to be independent of the existence of electrostatic "double-layer" or "hydration" forces. For the case of the simple liquid toluene confined between surfaces with adsorbed layers of C_{60} molecules, this type of viscosity measurement has shown that the traditional no-slip assumption for flow at a solid interface does not always hold [18.257]. The C_{60} layer at the mica–toluene interface results in a "full slip" boundary, which dramatically lowers the viscous drag or effective viscosity for regular Couette or Poiseuille flow.

With polymeric liquids (polymer melts) such as polydimethylsiloxanes (PDMS) and polybutadienes (PBD), or with polystyrene (PS) adsorbed onto surfaces from solution, the far-field viscosity is again equal to the bulk value, but with the non-slip plane (hydrodynamic layer thickness) being located at $D = 1\text{–}2R_g$ away from each surface [18.11, 47], or at $D = L$ or less for polymer brush layers of thickness L per surface [18.13, 258]. In contrast, the same technique was used to show that for non-adsorbing polymers in solution, there is actually a depletion layer of nearly pure solvent that exists at the surfaces that affects the confined solution flow properties [18.256]. These effects are observed from near contact to surface separations in excess of 200 nm.



Further experiments with surfaces closer than a few molecular diameters ($D < 2$–4 nm for simple liquids, or $D < 2$–$4R_g$ for polymer fluids) indicate that large deviations occur for thinner films, described below. One important conclusion from these studies is, therefore, that the dynamic properties of simple liquids, including water, near an *isolated* surface are similar to those of the bulk liquid *already within the first layer of molecules adjacent to the surface*, only changing when another surface approaches the first. In other words, the viscosity and position of the shear plane near a surface are not simply a property of that surface, but of how far that surface is from another surface. The reason for this is because when two surfaces are close together, the constraining effects on the liquid molecules between them are much more severe than when there is only one surface. Another obvious consequence of the above is that one should not make measurements on a single, isolated solid–liquid interface and then draw conclusions about the state of the liquid or its interactions in a thin film *between* two surfaces.

18.8.2 Friction of Intermediate Thickness Films

For liquid films in the thickness range between 4 and 10 molecular diameters, the properties can be significantly different from those of bulk films. Still, the fluids remain recognizable as fluids; in other words, they do not undergo a phase transition into a solid

◀ **Fig. 18.24a–d** Typical rheological behavior of liquid films in the mixed lubrication regime. **(a)** Increase in effective viscosity of dodecane film between two mica surfaces with decreasing film thickness. At distances larger than 4–5 nm, the effective viscosity η_{eff} approaches the bulk value η_{bulk} and does not depend on the shear rate $\dot{\gamma}$. (After [18.219]. Copyright 1991 American Association for the Advancement of Science.) **(b)** Non-Newtonian variation of η_{eff} with shear rate of a 2.7-nm-thick dodecane film at a net normal pressure of 0.12 MPa and at 28 °C. The effective viscosity decays as a power law, as in (18.46). In this example, $n = 0$ at the lowest $\dot{\gamma}$ and changes to $n = -2/3$ and -1 at higher $\dot{\gamma}$. For films of bulk thickness, dodecane is a low viscosity Newtonian fluid ($n = 0$). **(c)** Proposed general friction map of effective viscosity η_{eff} (arbitrary units) as a function of effective shear rate $\dot{\gamma}$ (arbitrary units) on logarithmic scales. Three main classes of behavior emerge: (i) Thick films: elastohydrodynamic sliding. At $L = 0$, approximating bulk conditions, η_{eff} is independent of shear rate except when shear thinning might occur at sufficiently large $\dot{\gamma}$. (ii) Boundary layer films, intermediate regime. A Newtonian regime is again observed [$\eta_{\text{eff}} = \text{constant}$, $n = 0$ in (18.46)] at low loads and low shear rates, but η_{eff} is much higher than the bulk value. As the shear rate $\dot{\gamma}$ increases beyond $\dot{\gamma}_{\text{min}}$, the effective viscosity starts to drop with a power-law dependence on the shear rate [see panel **(b)**], with n in the range $-1/2$ to -1 most commonly observed. As the shear rate $\dot{\gamma}$ increases still more, beyond $\dot{\gamma}_{\text{max}}$, a second Newtonian plateau is encountered. (iii) Boundary layer films, high load. The η_{eff} continues to grow with load and to be Newtonian provided that the shear rate is sufficiently low. Transition to sliding at high velocity is discontinuous ($n < -1$) and usually of the stick–slip variety. **(d)** Proposed friction map of friction force as a function of sliding velocity in various tribological regimes. With increasing load, Newtonian flow in the elastohydrodynamic (EHD) regimes crosses into the boundary regime of lubrication. Note that even EHD lubrication changes, at the highest velocities, to limiting shear stress response. At the highest loads (L) and smallest film thickness (D), the friction force goes through a maximum (the static friction, F_s), followed by a regime where the friction coefficient (μ) is roughly constant with increasing velocity (meaning that the kinetic friction, F_k, is roughly constant). Non-Newtonian shear thinning is observed at somewhat smaller load and larger film thickness; the friction force passes through a maximum at the point where $De = 1$. De, the Deborah number, is the point at which the applied shear rate exceeds the natural relaxation time of the boundary layer film. The velocity axis from 10^{-10} to 10^{10} (arbitrary units) indicates a large span. (Panels **(b)–(d)** after [18.259]. Copyright 1996, with permission from Elsevier Science)

or liquid-crystalline phase. This regime has recently been studied by *Granick* et al. [18.44, 219, 220, 222, 223], who used a friction attachment [18.43, 44] to the SFA where a sinusoidal input signal to a piezoelectric device makes the two surfaces slide back and forth laterally past each other at small amplitudes. This method provides information on the real and imaginary parts (elastic and dissipative components, respectively) of the shear modulus of thin films at different shear rates and film thickness. *Granick* [18.219] and *Hu* et al. [18.223] found that films of simple liquids become non-Newtonian in the 2.5–5 nm regime (about 10 molecular diameters, see Fig. 18.24). Polymer melts become non-Newtonian at much larger film thicknesses, depending on their molecular weight [18.47].

Klein and *Kumacheva* [18.46, 226, 260] studied the interaction forces and friction of small quasi-spherical liquid molecules such as cyclohexane between molecularly smooth mica surfaces. They concluded that surface epitaxial effects can cause the liquid film to already solidify at 6 molecular diameters, resulting in a sudden (discontinuous) jump to high friction at low shear rates. Such dynamic first-order transitions, however, may depend on the shear rate.

A generalized friction map (Fig. 18.24c,d) has been proposed by *Luengo* et al. [18.259] that illustrates the changes in η_{eff} from bulk Newtonian behavior ($n = 0$, $\eta_{\text{eff}} = \eta_{\text{bulk}}$) through the transition regime where n reaches a minimum of -1 with decreasing shear rate to the solid-like creep regime at very low $\dot{\gamma}$ where n returns to 0. A number of results from experimental, theoretical, and computer simulation work have shown values of n from $-1/2$ to -1 for this transition regime for a variety of systems and assumptions [18.219, 220, 239, 261–267].

The intermediate regime appears to extend over a narrow range of film thickness, from about 4 to 10 molecular diameters or polymer radii of gyration. Thinner films begin to adopt "boundary" or "interfacial" friction properties (described below, see also Table 18.5). Note that the intermediate regime is actually a very narrow one when defined in terms of film thickness, for example, varying from about $D = 2$ to 4 nm for hexadecane films [18.219].

A fluid's effective viscosity η_{eff} in the intermediate regime is usually higher than in the bulk, but η_{eff} usu-

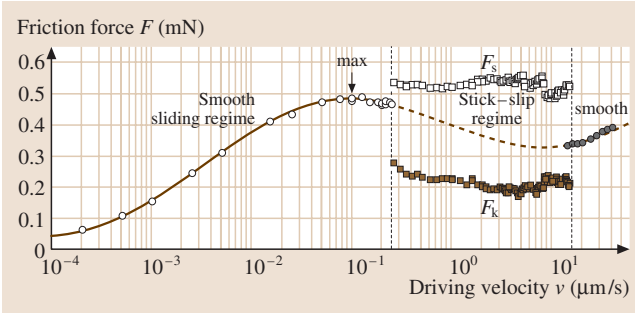


Fig. 18.25 Transition from smooth sliding to "inverted" stick–slip and to a second smooth sliding regime with increasing driving velocity during shear of two adsorbed surfactant monolayers in aqueous solution at a load of $L = 4.5$ mN and $T = 20\,°C$. The smooth sliding (*open circles*) to inverted stick–slip (*squares*) transition occurs at $v_c \sim 0.3\,\mu m/s$. Prior to the transition, the kinetic stress levels off at after a logarithmic dependence on velocity. The quasi-smooth regime persists up to the transition at v_c. At high driving velocities (*filled circles*), a new transition to a smooth sliding regime is observed between 14 and 17 μm/s. (After [18.268] with permission)

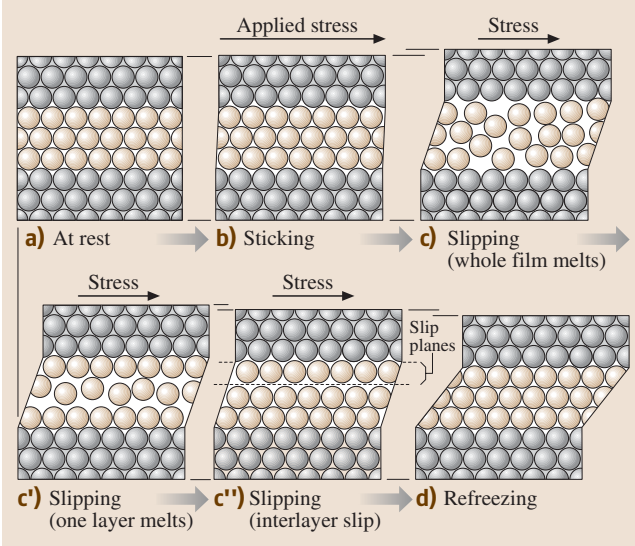


Fig. 18.26a–d Idealized schematic illustration of molecular rearrangements occurring in a molecularly thin film of spherical or simple chain molecules between two solid surfaces during shear. Depending on the system, a number of different molecular configurations within the film are possible during slipping and sliding, shown here as stages (c): total disorder as the whole film melts; (c′): partial disorder; and (c″): order persists even during sliding with slip occurring at a single slip plane either within the film or at the walls. A dilation is predicted in the direction normal to the surfaces. (After [18.224] with permission)

ally *decreases* with increasing sliding velocity, v (known as *shear thinning*). When two surfaces slide in the intermediate regime, the motion tends to thicken the film (dilatancy). This sends the system into the bulk EHL regime where, as indicated by (18.44), the friction force now *increases* with velocity. This initial decrease, followed by an increase, in the frictional forces of many lubricant systems is the basis for the empirical Stribeck curve of Fig. 18.13. In the transition from bulk to boundary behavior there is first a quantitative change in the material properties (viscosity and elasticity), which can be continuous, to discontinuous qualitative changes that result in yield stresses and non-liquidlike behavior.

The rest of this section is devoted to friction in the boundary regime. Boundary friction may be thought of as applying to the case where a lubricant film is present, but where this film is of molecular dimensions – a few molecular layers or less.

18.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology

When a liquid is confined between two surfaces or within any narrow space whose dimensions are less than 4 to 10 molecular diameters, both the static (equilibrium) and dynamic properties of the liquid, such as its compressibility and viscosity, can no longer be described even qualitatively in terms of the bulk properties. The molecules confined within such molecularly thin films become ordered into layers ("out-of-plane" ordering), and within each layer they can also have lateral order ("in-plane" ordering). Such films may be thought of as behaving more like a liquid crystal or a solid than a liquid.

As described in Sect. 18.4.3, the measured *normal* forces between two solid surfaces across molecularly thin films exhibit exponentially decaying oscillations, varying between attraction and repulsion with a periodicity equal to some molecular dimension of the solvent molecules. Thus most liquid films can sustain a finite normal stress, and the adhesion force between two surfaces across such films is "quantized", depending on the thickness (or number of liquid layers) between the surfaces. The structuring of molecules in thin films and the oscillatory forces it gives rise to are now reasonably well understood, both experimentally and theoretically, at least for simple liquids.

Work has also recently been done on the dynamic, e.g., viscous or shear, forces associated with molecularly thin films. Both experiments [18.38, 46, 203, 221,

226,227,269,270] and theory [18.200,201,261,271] indicate that even when two surfaces are in steady state sliding they still prefer to remain in one of their stable potential energy minima, i. e., a sheared film of liquid can retain its basic layered structure. Thus even during motion the film does not become totally liquid-like. Indeed, if there is some "in-plane" ordering within a film, it will exhibit a yield-point before it begins to flow. Such films can therefore sustain a finite shear stress, in addition to a finite normal stress. The value of the yield stress depends on the number of layers comprising the film and represents another "quantized" property of molecularly thin films [18.200].

The dynamic properties of a liquid film undergoing shear are very complex. Depending on whether the film is more liquid-like or solid-like, the motion will be smooth or of the stick–slip type. During sliding, transitions can occur between n layers and $(n-1)$ or $(n+1)$ layers (see Fig. 18.27). The details of the motion depend critically on the externally applied load, the temperature, the sliding velocity, the twist angle between the two surface lattices, and the sliding direction relative to the lattices.

Smooth and Stick–Slip Sliding

Recent advances in friction-measuring techniques have enabled the interfacial friction of molecularly thin films to be measured with great accuracy. Some of these advances have involved the surface forces apparatus technique [18.38, 45, 47, 221, 247, 269, 270], while others have involved the atomic force microscope [18.10, 53, 231, 272]. In addition, computer simulations [18.90,200,201,271,273,274] have become sufficiently sophisticated to enable fairly complex tribological systems to be studied. All these advances are necessary if one is to probe such subtle effects as smooth or stick–slip friction, transient and memory effects, and ultralow friction mechanisms at the molecular level.

The theoretical models presented in this section will be concerned with a situation commonly observed experimentally: Stick–slip occurs between a static state with high friction and a low-friction kinetic state, and a transition from this sliding regime to smooth sliding can be induced by an increase in velocity. Experimental data on various systems showing this behavior are shown in Figs. 18.27, 18.30b, and 18.31a. Recent studies on adhesive systems have revealed the possibility of other dynamic responses such as "inverted" stick–slip between two kinetic states of higher and lower friction and with a transition from smooth sliding to stick–slip with increasing velocity, as shown in Fig. 18.25 [18.268].

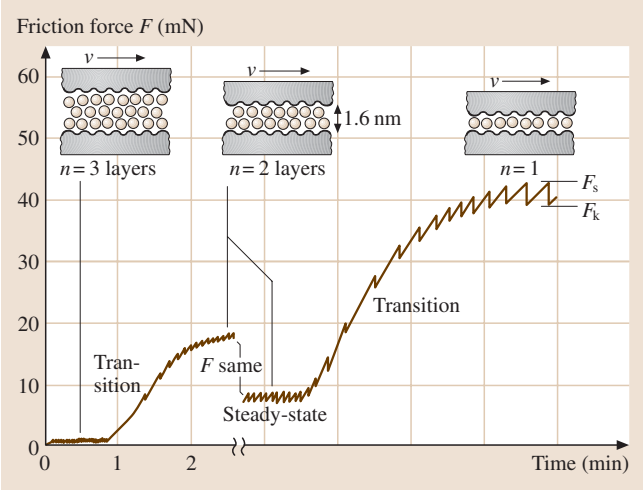


Fig. 18.27 Measured change in friction during interlayer transitions of the silicone liquid octamethylcyclotetrasiloxane (OMCTS), an inert liquid whose quasispherical molecules have a diameter of 0.8 nm. In this system, the shear stress $S_c = F/A$ was found to be constant as long as the number of layers, n, remained constant. Qualitatively similar results have been obtained with other quasi-spherical molecules such as cyclohexane [18.269]. The shear stresses are only weakly dependent on the sliding velocity v. However, for sliding velocities above some critical value, v_c, the stick–slip disappears and sliding proceeds smoothly at the kinetic value. (After [18.221] with permission)

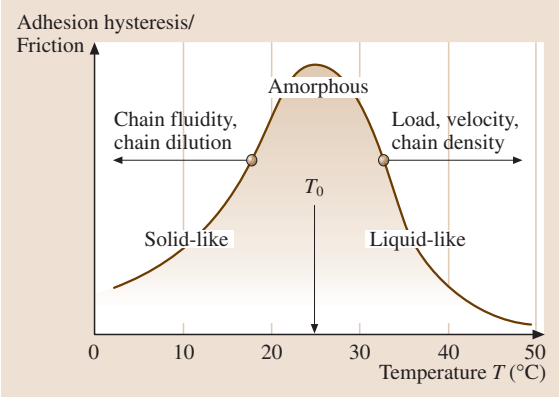


Fig. 18.28 Schematic friction phase diagram representing the trends observed in the boundary friction of a variety of different surfactant monolayers. The characteristic bell-shaped curve also correlates with the adhesion energy hysteresis of the monolayers. The *arrows* indicate the direction in which the whole curve is dragged when the load, velocity, etc., is increased. (After [18.232] with permission)

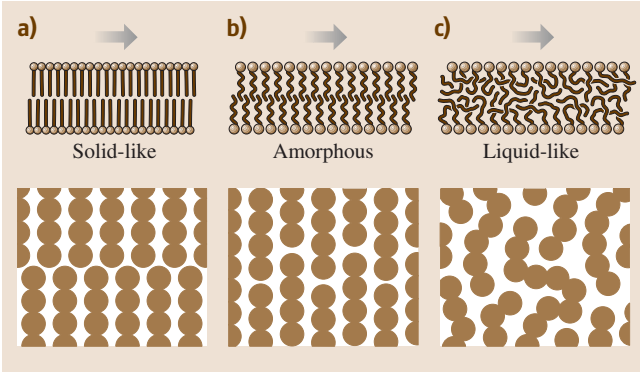


Fig. 18.29a–c Different dynamic phase states of boundary monolayers during adhesive contact and/or frictional sliding. Solid-like **(a)** and liquid-like monolayers **(c)** exhibit low adhesion hysteresis and friction. Increasing the temperature generally shifts a system from the left to the right. Changing the load, sliding velocity, or other experimental conditions can also change the dynamic phase state of surface layers, as shown in Fig. 18.28. (After [18.232] with permission)

With the added insights provided by computer simulations, a number of distinct molecular processes have been identified during smooth and stick–slip sliding in model systems for the more familiar static-to-kinetic stick–slip and transition from stick–slip to smooth sliding. These are shown schematically in Fig. 18.26 for the case of spherical liquid molecules between two solid crystalline surfaces. The following regimes may be identified:

Surfaces at rest (Fig. 18.26a): Even with no externally applied load, solvent–surface epitaxial interactions can cause the liquid molecules in the film to attain a solid-like state. Thus at rest the surfaces are stuck to each other through the film.

Sticking regime (frozen, solid-like film) (Fig. 18.26b): A progressively increasing lateral shear stress is applied. The solid-like film responds elastically with a small lateral displacement and a small increase or "dilatancy" in film thickness (less than a lattice spacing or molecular dimension, σ). In this regime the film retains its frozen, solid-like state: All the strains are elastic and reversible, and the surfaces remain effectively stuck to each other. However, slow creep may occur over long time periods.

Slipping and sliding regimes (molten, liquid-like film) (Fig. 18.26c,c′,c″): When the applied shear stress or force has reached a certain critical value, the *static* friction force, F_s, the film suddenly melts (known as "shear melting") or rearranges to allow for wall-slip or slip within the film to occur at which point the two surfaces begin to slip rapidly past each other. If the applied stress is kept at a high value, the upper surface will continue to slide indefinitely.

Refreezing regime (resolidification of film) (Fig. 18.26d): In many practical cases, the rapid slip of the upper surface relieves some of the applied force, which eventually falls below another critical value, the *kinetic* friction force F_k, at which point the film resolidifies and the whole stick–slip cycle is repeated. On the other hand, if the slip rate is smaller than the rate at which the external stress is applied, the surfaces will continue to slide smoothly in the kinetic state and there will be no more stick–slip. The critical velocity at which stick–slip disappears is discussed in more detail in Sect. 18.8.3.

Experiments with linear chain (alkane) molecules show that the film thickness remains quantized during sliding, so that the structure of such films is probably more like that of a nematic liquid crystal where the liquid molecules have become shear aligned in some direction, enabling shear motion to occur while retaining some order within the film. Experiments on the friction of two molecularly smooth mica surfaces separated by three molecular layers of the liquid OMCTS (Fig. 18.27) show how the friction increases to higher values in a quantized way when the number of layers falls from $n = 3$ to $n = 2$ and then to $n = 1$.

Computer simulations for simple spherical molecules [18.201] further indicate that during the slip the film thickness is roughly 15% higher than at rest (i. e., the film density falls), and that the order parameter within the film drops from 0.85 to about 0.25. Such dilatancy has been investigated both experimentally [18.235] and in further computer simulations [18.234]. The changes in thickness and in order parameter are consistent with a disorganized state for the whole film during the slip [18.271], as illustrated schematically in Fig. 18.26c. At this stage, we can only speculate on other possible configurations of molecules in the slipping and sliding regimes. This probably depends on the shapes of the molecules (e.g., whether spherical or linear or branched), on the atomic structure of the surfaces, on the sliding velocity, etc. [18.275]. Figure 18.26c,c′,c″ shows three possible sliding modes wherein the shearing film either totally melts, or where the molecules retain their layered structure and where slip occurs between two or more layers. Other sliding modes, for example, involving the movement of dislocations or disclinations are also possible, and it is unlikely that one single mechanism applies in all cases.

Both friction and adhesion hysteresis vary nonlinearly with temperature, often peaking at some

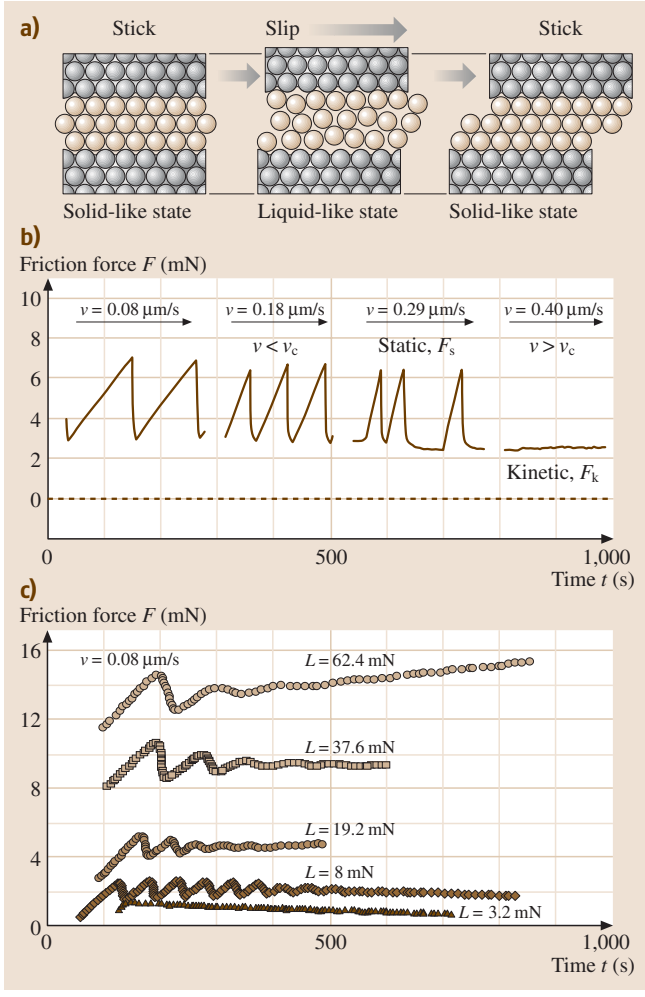


Fig. 18.30 **(a)** "Phase transitions" model of stick–slip where a thin liquid film alternately freezes and melts as it shears, shown here for 22 spherical molecules confined between two solid crystalline surfaces. In contrast to the velocity-dependent friction model, the intrinsic friction force is assumed to change abruptly (at the transitions), rather than smooth or continuously. The resulting stick–slip is also different, for example, the peaks are sharper and the stick–slip disappears above some critical velocity v_c. Note that while the slip displacement is here shown to be only two lattice spacings, in most practical situations it is much larger, and that freezing and melting transitions at surfaces or in thin films may not be the same as freezing or melting transitions between the bulk solid and liquid phases. **(b)** Exact reproduction of a chart-recorder trace of the friction force for hexadecane between two untreated mica surfaces at increasing sliding velocity v, plotted as a function of time. In general, with increasing sliding speed, the stick–slip spikes increase in frequency and decrease in magnitude. As the critical sliding velocity v_c is approached, the spikes become erratic, eventually disappearing altogether at $v = v_c$. At higher sliding velocities the sliding continues smoothly in the kinetic state. Such friction traces are fairly typical for simple liquid lubricants and dry boundary lubricant systems (see Fig. 18.31a) and may be referred to as the "conventional" type of static–kinetic friction (in contrast to Fig. 18.25). Experimental conditions: contact area $A = 4\times10^{-9}\,\mathrm{m}^2$, load $L = 10\,\mathrm{mN}$, film thickness $D = 0.4–0.8\,\mathrm{nm}$, $v = 0.08–0.4\,\mu\mathrm{m\,s}^{-1}$, $v_c \approx 0.3\,\mu\mathrm{m\,s}^{-1}$, atmosphere: dry N_2 gas, $T = 18\,°\mathrm{C}$. [**(a)** and **(b)** after [18.276] with permission. Copyright 1993 American Chemical Society.] **(c)** Friction response of a thin squalane (a branched hydrocarbon) film at different loads and a constant sliding velocity $v = 0.08\,\mu\mathrm{m\,s}^{-1}$, slightly above the critical velocity for this system at low loads. Initally, with increasing load, the stick–slip amplitude and the mean friction force decrease with sliding time or sliding distance. However, at high loads or pressures, the mean friction force increases with time, and the stick–slip takes on a more symmetrical, sinusoidal shape. At all loads investigated, the stick–slip component gradually decayed as the friction proceeded towards smooth sliding. (*Gourdon* and *Israelachvili*, unpublished data)

particular temperature, T_0. The temperature dependence of these forces can, therefore, be represented on a friction phase diagram such as the one shown in Fig. 18.28. Experiments have shown that T_0, and the whole bell-shaped curve, are shifted along the temperature-axis (as well as in the vertical direction) in a systematic way when the load, sliding velocity, etc., are varied. These shifts also appear to be highly correlated with one another, for example, an increase in temperature producing effects that are similar to *decreasing* the sliding speed or load.

Such effects are also commonly observed in other energy dissipating phenomena such as polymer viscoelasticity [18.277], and it is likely that a similar physical mechanism is at the heart all of such phenomena. A possible molecular process underlying the energy dissipation of chain molecules during boundary layer sliding is illustrated in Fig. 18.29, which shows the three main dynamic phase states of boundary monolayers.

In contrast to the characteristic relaxation time associated with fluid lubricants, it has been established

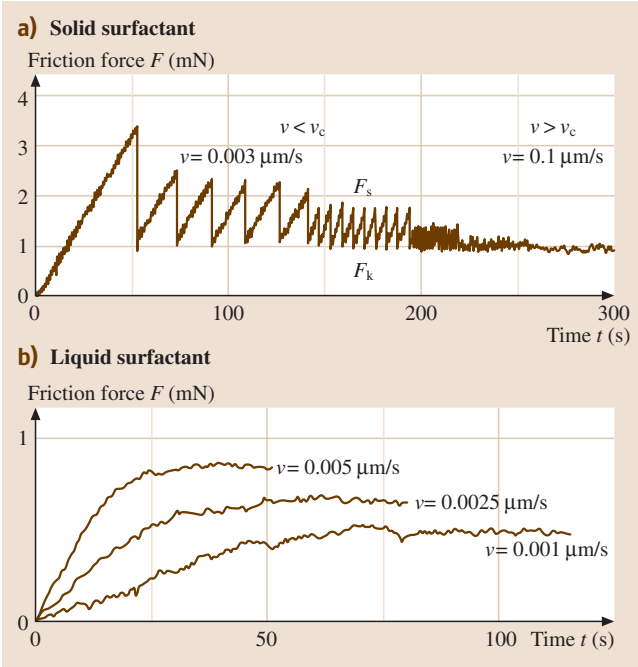


Fig. 18.31 **(a)** Exact reproduction of chart-recorder trace for the friction of closely packed surfactant monolayers (L-α-dimirystoyl-phosphatidyl-ethanolamine, DMPE) on mica (dry boundary friction) showing qualitatively similar behavior to that obtained with a liquid hexadecane film (Fig. 18.30b). In this case, $L = 0$, $v_c \approx 0.1\ \mu m\ s^{-1}$, atmosphere: dry N_2 gas, $T = 25\ °C$. **(b)** Sliding typical of liquid-like monolayers, here shown for calcium alkylbenzene sulfonate in dry N_2 gas at $T = 25\ °C$ and $L = 0$. (After [18.207] with permission. Copyright 1993 American Chemical Society)

that for unlubricated (dry, solid, rough) surfaces, there is a characteristic memory distance that must be exceeded before the system looses all memory of its initial state (original surface topography). The underlying mechanism for a characteristic distance was first used to successfully explain rock mechanics and earthquake faults [18.278] and, more recently, the tribological behavior of unlubricated surfaces of ceramics, paper and elastomeric polymers [18.252, 279]. Recent experiments [18.275, 280, 281] suggest that fluid lubricants composed of complex branched-chained or polymer molecules may also have characteristic distances (in addition to characteristic relaxation times) associated with their tribological behavior – the characteristic distance being the total sliding distance that must be exceeded before the system reaches its steady-state tribological conditions (see Sect. 18.8.3).

Abrupt vs. Continuous Transitions Between Smooth and Stick–Slip Sliding

An understanding of stick–slip is of great practical importance in tribology [18.282], since these spikes are the major cause of damage and wear of moving parts. Stick–slip motion is a very common phenomenon and is also the cause of sound generation (the sound of a violin string, a squeaking door, or the chatter of machinery), sensory perception (taste texture and feel), earthquakes, granular flow, nonuniform fluid flow such as the spurting flow of polymeric liquids, etc. In the previous section, the stick–slip motion arising from freezing–melting transitions in thin interfacial films was described. There are other mechanisms that can give rise to stick–slip friction, which will now be considered. However, before proceeding with this, it is important to clarify exactly what one is measuring during a friction experiment.

Most tribological systems and experiments can be described in terms of an equivalent mechanical circuit with certain characteristics. The friction force F_0, which is generated at the surfaces, is generally measured as F at some other place in the setup. The mechanical coupling between the two may be described in terms of a simple elastic stiffness or compliance, K, or as more complex nonelastic coefficients, depending on the system. The distinction between F and F_0 is important because in almost all practical cases, the applied, measured, or detected force, F, is *not* the same as the "real" or "intrinsic" friction force, F_0, generated at the surfaces. F and F_0 are coupled in a way that depends on the mechanical construction of the system, for example, the axle of a car wheel that connects it to the engine. This coupling can be modeled as an elastic spring of stiffness K and mass m. This is the simplest type of mechanical coupling and is also the same as in SFA and AFM type experiments. More complicated real systems can be reduced to a system of springs and dashpots, as described by *Peachey* et al. [18.283] and *Luengo* et al. [18.47].

We now consider four different models of stick–slip friction, where the mechanical couplings are assumed to be of the simple elastic spring type. The first three mechanisms may be considered traditional or classical mechanisms or models [18.282], the fourth is essentially the same as the freezing–melting phase-transition model described in Sect. 18.8.3.

Rough Surfaces or Surface Topology Model. Rapid slips can occur whenever an asperity on one surface goes over the top of an asperity on the other surface. The extent of the slip will depend on asperity heights and slopes, on the speed of sliding, and on the elastic

compliance of the surfaces and the moving stage. As in all cases of stick–slip motion, the driving velocity v may be constant, but the resulting motion at the surfaces v_0 will display large slips. This type of stick–slip has been described by *Rabinowicz* [18.282]. It will not be of much concern here since it is essentially a noise-type fluctuation, resulting from surface imperfections rather than from the intrinsic interaction between two surfaces. Actually, at the atomic level, the regular atomic-scale corrugations of surfaces can lead to periodic stick–slip motion of the type shown here. This is what is sometimes measured by AFM tips [18.10, 53, 231, 272].

Distance-Dependent or Creep Model. Another theory of stick–slip, observed in solid-on-solid sliding, is one that involves a characteristic *distance*, but also a characteristic time, τ_s, this being the characteristic time required for two asperities to increase their adhesion strength after coming into contact. Originally proposed by *Rabinowicz* [18.282, 284], this model suggests that two rough macroscopic surfaces adhere through their microscopic asperities of characteristic length. During shearing, each surface must first creep this distance (the size of the contacting junctions) after which the surfaces continue to slide, but with a lower (kinetic) friction force than the original (static) value. The reason for the decrease in the friction force is that even though, on average, new asperity junctions should form as rapidly as the old ones break, the time-dependent adhesion and friction of the new ones will be lower than the old ones.

The friction force, therefore, remains high during the creep stage of the slip. But once the surfaces have moved the characteristic distance, the friction rapidly drops to the kinetic value. For any system where the kinetic friction is less than the static force (or one that has a negative slope over some part of its F_0 versus v_0 curve) will exhibit regular stick–slip sliding motion for certain values of K, m, and driving velocity, v.

This type of friction has been observed in a variety of dry (unlubricated) systems such as paper-on-paper [18.285, 286] and steel-on-steel [18.284, 287, 288]. This model is also used extensively in geologic systems to analyze rock-on-rock sliding [18.289, 290]. While originally described for adhering macroscopic asperity junctions, the distance-dependent model may also apply to molecularly smooth surfaces. For example, for polymer lubricant films, the characteristic length would now be the chain–chain entanglement length, which could be much larger in a confined geometry than in the bulk.

Velocity-Dependent Friction Model. In contrast to the two friction models mentioned above, which apply mainly to unlubricated, solid-on-solid contacts, the stick–slip of surfaces with thin liquid films between them is better described by other mechanisms. The velocity-dependent friction model is the most studied mechanism of stick–slip and, until recently, was considered to be the only cause of intrinsic stick–slip. If a friction force decreases with increasing sliding velocity, as occurs with boundary films exhibiting shear thinning, the force (F_s) needed to initiate motion will be higher than the force (F_k) needed to maintain motion.

A decreasing intrinsic friction force F_0 with sliding velocity v_0 results in the sliding surface or stage moving in a periodic fashion, where during each cycle rapid acceleration is followed by rapid deceleration. So long as the drive continues to move at a fixed velocity v, the surfaces will continue to move in a periodic fashion punctuated by abrupt stops and starts whose frequency and amplitude depend not only on the function $F_0(v_0)$, but also on the stiffness K and mass m of the moving stage, and on the starting conditions at $t = 0$.

More precisely, the motion of the sliding surface or stage can be determined by solving the following differential equation:

$$m\ddot{x} = (F_0 - F) = F_0 - (x_0 - x)K$$
$$\text{or} \quad m\ddot{x} + (x_0 - x)K - F_0 = 0\,, \tag{18.47}$$

where $F_0 = F_0(x_0, v_0, t)$ is the intrinsic or "real" friction force at the shearing surfaces, F is the force on the spring (the externally applied or measured force), and $(F_0 - F)$ is the force on the stage. To fully solve (18.47), one must also know the initial (starting) conditions at $t = 0$, and the driving or steady-state conditions at finite t. Commonly, the driving condition is: $x = 0$ for $t < 0$ and $x = vt$ for $t > 0$, where $v =$ constant. In other systems, the appropriate driving condition may be $F =$ constant.

Various, mainly phenomenological, forms for $F_0 = F_0(x_0, v_0, t)$ have been proposed to explain various kinds of stick–slip phenomena. These models generally assume a particular functional form for the friction as a function of velocity only, $F_0 = F_0(v_0)$, and they may also contain a number of mechanically coupled elements comprising the stage [18.291, 292]. One version is a two-state model characterized by two friction forces, F_s and F_k, which is a simplified version of the phase transitions model described below. More complicated versions can have a rich F–v spectrum, as proposed by *Persson* [18.293]. Unless the experimental data is very detailed and extensive, these models cannot generally distinguish between different types of mechanisms.

Neither do they address the basic question of the *origin* of the friction force, since this is assumed to begin with.

Experimental data has been used to calculate the friction force as a function of velocity *within* an individual stick–slip cycle [18.294]. For a macroscopic granular material confined between solid surfaces, the data shows a velocity-weakening friction force during the first half of the slip. However, the data also shows a hysteresis loop in the friction–velocity plot, with a different behavior in the deceleration half of the slip phase. Similar results were observed for a 1–2 nm liquid lubricant film between mica surfaces [18.275]. These results indicate that a purely velocity-dependent friction law is insufficient to describe such systems, and an additional element such as the *state* of the confined material must be considered.

Phase Transitions Model. In recent molecular dynamics computer simulations it has been found that thin interfacial films undergo first-order phase transitions between solid-like and liquid-like states during sliding [18.201, 273]. It has been suggested that this is responsible for the observed stick–slip behavior of simple isotropic liquids between two solid crystalline surfaces. With this interpretation, stick–slip is seen to arise because of the abrupt change in the flow properties of a film at a transition [18.224, 225, 261], rather than the gradual or continuous change, as occurs in the previous example. Other computer simulations indicate that it is the stick–slip that induces a disorder ("shear-melting") in the film, not the other way around [18.295].

An interpretation of the well-known phenomenon of decreasing coefficient of friction with increasing sliding velocity has been proposed by *Thompson* and *Robbins* [18.201] based on their computer simulation. This postulates that it is not the friction that changes with sliding speed v, but rather the time various parts of the system spend in the sticking and sliding modes. In other words, at any instant during sliding, the friction at any local region is always F_s or F_k, corresponding to the "static" or "kinetic" values. The measured frictional force, however, is the sum of all these discrete values averaged over the whole contact area. Since as v increases each local region spends more time in the sliding regime (F_k) and less in the sticking regime (F_s), the overall friction coefficient falls. One may note that this interpretation reverses the traditional way that stick–slip has been explained. Rather than invoking a decreasing friction with velocity to explain stick–slip, it is now the more fundamental stick–slip phenomenon that is producing the apparent decrease in the friction force with increasing sliding velocity. This approach has been studied analytically by *Carlson* and *Batista* [18.296], with a comprehensive rate- and state- dependent friction force law. This model includes an analytic description of the freezing–melting transitions of a film, resulting in a friction force that is a function of sliding velocity in a natural way. This model predicts a full range of stick–slip behavior observed experimentally.

An example of the rate- and state-dependent model is observed when shearing thin films of OMCTS between mica surfaces [18.297, 298]. In this case, the static friction between the surfaces is dependent on the time that the surfaces are at rest with respect to each other, while the intrinsic kinetic friction $F_{k,0}$ is relatively constant over the range of velocities. At slow driving velocities, the system responds with stick–slip sliding with the surfaces reaching maximum static friction before each slip event, and the amplitude of the stick–slip, $F_s - F_k$, is relatively constant. As the driving velocity increases, the static friction decreases as the time at relative rest becomes shorter with respect to the characteristic time of the lubricant film. As the static friction decreases with increasing drive velocity, it eventually equals the intrinsic kinetic friction $F_{k,0}$, which defines the critical velocity v_c, above which the surfaces slide smoothly without the jerky stick–slip motion.

The above classifications of stick–slip are not exclusive, and molecular mechanisms of real systems may exhibit aspects of different models simultaneously. They do, however, provide a convenient classification of existing models and indicate which experimental parameters should be varied to test the different models.

Critical Velocity for Stick–Slip. For any given set of conditions, stick–slip disappears above some critical sliding velocity v_c, above which the motion continues smoothly in the liquid-like or kinetic state. The critical velocity is well described by two simple equations. Both are based on the phase transition model, and both include some parameter associated with the inertia of the measuring instrument. The first equation is based on both experiments and simple theoretical modeling [18.276]:

$$v_c \approx \frac{(F_s - F_k)}{5K\tau_0} , \tag{18.48}$$

where τ_0 is the *characteristic nucleation time* or freezing time of the film. For example, inserting the following typically measured values for a ~1-nm-thick hexadecane film between mica: $(F_s - F_k) \approx 5\,\mathrm{mN}$, spring constant $K \approx 500\,\mathrm{N\,m^{-1}}$, and nucleation time [18.276]

$\tau_0 \approx 5\,\mathrm{s}$, we obtain $v_c \approx 0.4\,\mu\mathrm{m\,s^{-1}}$, which is close to typically measured values (Fig. 18.30b).

The second equation is based on computer simulations [18.273]:

$$v_c \approx 0.1\sqrt{\frac{F_s\sigma}{m}}\,, \tag{18.49}$$

where σ is a molecular dimension and m is the mass of the stage. Again, inserting typical experimental values into this equation, viz., $m \approx 20\,\mathrm{g}$, $\sigma \approx 0.5\,\mathrm{nm}$, and $(F_s - F_k) \approx 5\,\mathrm{mN}$ as before, we obtain $v_c \approx 0.3\,\mu\mathrm{m\,s^{-1}}$, which is also close to measured values.

Stick–slip also disappears above some critical temperature T_c, which is not the same as the melting temperature of the bulk fluid. Certain correlations have been found between v_c and T_c and between various other tribological parameters that appear to be consistent with the principle of time–temperature superposition (see Sect. 18.8.3), similar to that occurring in viscoelastic polymer fluids [18.277, 299, 300].

Recent work on the coupling between the mechanical resonances of the sliding system and molecular-scale relaxations [18.234, 301–303] has resulted in a better understanding of a phenomenon previously noted in various engineering applications: the vibrating of one of the sliding surfaces perpendicularly to the sliding direction can lead to a significant reduction of the friction. At certain oscillation amplitudes and a frequency higher than the molecular-scale relaxation frequency, stick–slip friction can be eliminated and replaced by an ultralow kinetic friction state.

18.9 Role of Molecular Shape and Surface Structure in Friction

The above scenario is already quite complicated, and yet this is the situation for the simplest type of experimental system. The factors that appear to determine the critical velocity v_c depend on the type of liquid between the surfaces (as well as on the surface lattice structure). Small spherical molecules such as cyclohexane and OMCTS have been found to have very high v_c, which indicates that these molecules can rearrange relatively quickly in thin films. Chain molecules and especially branched chain molecules have been found to have much lower v_c, which is to be expected, and such liquids tend to slide smoothly or with erratic stick–slip [18.275], rather than in a stick–slip fashion (see Table 18.5). With highly asymmetric molecules, such as multiply branched isoparaffins and polymer melts, no regular spikes or stick–slip behavior occurs at any speed, since these molecules can never order themselves sufficiently to solidify. Examples of such liquids are some perfluoropolyethers and polydimethylsiloxanes (PDMS).

Recent computer simulations [18.274, 304, 305] of the structure, interaction forces, and tribological behavior of chain molecules between two shearing surfaces indicate that both linear *and* singly or doubly branched chain molecules order between two flat surfaces by aligning into discrete layers parallel to the surfaces. However, in the case of the weakly branched molecules, the expected oscillatory forces do not appear because of a complex cancellation of entropic and enthalpic contributions to the interaction free energy, which results in a monotonically smooth interaction, exhibiting a weak energy minimum rather than the oscillatory force profile that is characteristic of linear molecules. During sliding, however, these molecules can be induced to further align, which can result in a transition from smooth to stick–slip sliding.

Table 18.5 shows the trends observed with some organic and polymeric liquid between smooth mica surfaces. Also listed are the bulk viscosities of the liquids. From the data of Table 18.5 it appears that there is a direct correlation between the shapes of molecules and their coefficient of friction or effectiveness as lubricants (at least at low shear rates). Small spherical or chain molecules have high friction with stick–slip, because they can pack into ordered solid-like layers. In contrast, longer chained and irregularly shaped molecules remain in an entangled, disordered, fluid-like state even in very thin films, and these give low friction and smoother sliding. It is probably for this reason that irregularly shaped branched chain molecules are usually better lubricants. It is interesting to note that the friction coefficient generally decreases as the bulk viscosity of the liquids *increases*. This unexpected trend occurs because the factors that are conducive to low friction are generally conducive to high viscosity. Thus molecules with side groups such as branched alkanes and polymer melts usually have higher bulk viscosities than their linear homologues for obvious reasons. However, in thin films the linear molecules have higher shear stresses, because of their ability to become ordered. The only exception to the

Table 18.5 Effect of molecular shape and short-range forces on tribological properties[a]

Liquid	Short-range force	Type of friction	Friction coefficient	Bulk liquid viscosity (cP)
Organic (water-free)				
Cyclohexane	Oscillatory	Quantized stick–slip	$\gg 1$	0.6
OMCTS[b]	Oscillatory	Quantized stick–slip	$\gg 1$	2.3
Octane	Oscillatory	Quantized stick–slip	1.5	0.5
Tetradecane	Oscillatory ↔ smooth	stick–slip ↔ smooth	1.0	2.3
Octadecane (branched)	Oscillatory ↔ smooth	stick–slip ↔ smooth	0.3	5.5
PDMS[b] ($M = 3{,}700\,\mathrm{g\,mol^{-1}}$, melt)	Oscillatory ↔ smooth	Smooth	0.4	50
PBD[b] ($M = 3{,}500\,\mathrm{g\,mol^{-1}}$, branched)	Smooth	Smooth	0.03	800
Water				
Water (KCl solution)	Smooth	Smooth	0.01–0.03	1.0

[a] For molecularly thin liquid films between two shearing mica surfaces at 20 °C
[b] OMCTS: Octamethylcyclotetrasiloxane, PDMS: Polydimethylsiloxane, PBD: Polybutadiene

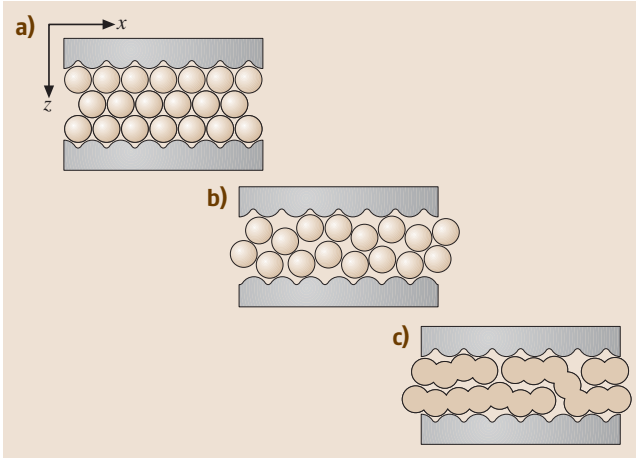


Fig. 18.32a–c Schematic view of interfacial film composed of spherical molecules under a compressive pressure between two solid crystalline surfaces. **(a)** If the two surface lattices are free to move in the x-y-z directions, so as to attain the lowest energy state, they could equilibrate at values of x, y, and z, which induce the trapped molecules to become "epitaxially" ordered into a "solid-like" film. **(b)** Similar view of trapped molecules between two solid surfaces that are not free to adjust their positions, for example, as occurs in capillary pores or in brittle cracks. **(c)** Similar to **(a)**, but with chain molecules replacing the spherical molecules in the gap. These may not be able to order as easily as do spherical molecules even if x, y, and z can adjust, resulting in a situation that is more akin to **(b)**. (After [18.276] with permission. Copyright 1993 American Chemical Society)

above correlations is water, which has been found to exhibit both low viscosity *and* low friction (see Fig. 18.20, and Sect. 18.7.3). In addition, the presence of water can drastically lower the friction and eliminate the stick–slip of hydrocarbon liquids when the sliding surfaces are hydrophilic.

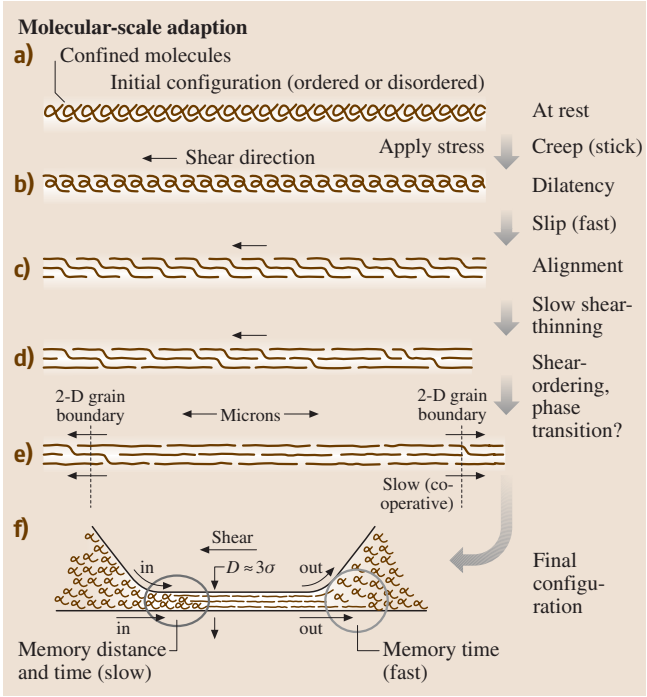


Fig. 18.33a–f Schematic representation of the film under shear. **(a)** The lubricant molecules are just confined, but not oriented in any particular direction. Because of the need to shear, the film dilates **(b)**. The molecules disentangle **(c)** and get oriented in a certain direction related to the shear direction **(d)**. **(e)** Slowly evolving domains grow inside the contact region. These macroscopic domains are responsible for the long relaxation times. **(f)** At the steady-state, a continuous gradient of confinement time and molecular order is established in the contact region, which is different for molecules adsorbed on the upper and lower surfaces. Molecules entering into the contact are not oriented or ordered. The required sliding distance to modify their state defines a characteristic distance. Molecules leaving the contact region need some (short) characteristic time to regain their bulk, unconfined configuration. (After [18.281] with permission. Copyright 2000 American Chemical Society)

If an "effective" viscosity, η_{eff}, were to be calculated for the liquids of Table 18.5, the values would be many orders of magnitude higher than those of the bulk liquids. This can be demonstrated by the following simple calculation based on the usual equation for Couette flow (see (18.44)):

$$\eta_{\text{eff}} = F_{\text{k}} D / A v \,, \tag{18.50}$$

where F_{k} is the kinetic friction force, D is the film thickness, A the contact area, and v the sliding velocity. Using typical values for experiments with hexadecane [18.276]: $F_{\text{k}} = 5\,\text{mN}$, $D = 1\,\text{nm}$, $A = 3 \times 10^{-9}\,\text{m}^2$, and $v = 1\,\mu\text{m}\,\text{s}^{-1}$, one gets $\eta_{\text{eff}} \approx 2{,}000\,\text{Ns}\,\text{m}^{-2}$, or 20,000 Poise, which is $\sim 10^6$ times higher than the bulk viscosity, η_{bulk}, of the liquid. It is instructive to consider that this very high effective viscosity nevertheless still produces a low friction force or friction coefficient μ of about 0.25. It is interesting to speculate that if a 1 nm film were to exhibit bulk viscous behavior, the friction coefficient under the same sliding conditions would be as low as 0.000001. While such a low value has never been reported for any tribological system, one may consider it a theoretical lower limit that could conceivably be attained under certain experimental conditions.

Various studies [18.44, 219, 220, 222] have shown that confinement and load generally increase the effective viscosity and/or relaxation times of molecules, suggestive of an increased glassiness or solid-like behavior (Figs. 18.32 and 18.33). This is in marked contrast to studies of liquids in small confining capillaries where the opposite effects have been observed [18.306, 307]. The reason for this is probably because the two modes of confinement are different. In the former case (confinement of molecules between two structured solid surfaces), there is generally little opposition to any lateral or vertical displacement of the two surface lattices relative to each other. This means that the two lattices can shift in the x-y-z plane (Fig. 18.32a) to accommodate the trapped molecules in the most crystallographically commensurate or epitaxial way, which would favor an ordered, solid-like state. In contrast, the walls of capillaries are rigid and cannot easily move or adjust to accommodate the confined molecules (Fig. 18.32b), which will therefore be forced into a more disordered, liquid-like state (unless the capillary wall geometry and lattice are *exactly* commensurate with the liquid molecules, as occurs in certain zeolites [18.307]).

Experiments have demonstrated the effects of surface lattice mismatch on the friction between surfaces [18.203, 204, 308]. Similar to the effects of lattice mismatch on adhesion (Fig. 18.11), the static friction of a confined liquid film is maximum when the lattices of the confining surfaces are aligned. For OMCTS confined between mica surfaces [18.204] the static friction was found to vary by more than a factor of 4, while

for bare mica surfaces the variation was by a factor of 3.5 [18.308]. In contrast to the sharp variations in adhesion energy over small twist angles, the variations in friction as a function of twist angle were much more broad both in magnitude and angular spread. Similar variations in friction as a function of twist or misfit angles have also been observed in computer simulations [18.309].

Robbins et al. [18.274] computed the friction forces of two clean crystalline surfaces as a function of the angle between their surface lattices. They found that for all non-zero angles (finite "twist" angles) the friction forces fell to zero due to incommensurability effects. They further found that submonolayer amounts of organic or other impurities trapped between two incommensurate surfaces can generate a finite friction force. They, therefore, concluded that any finite friction force measured between incommensurate surfaces is probably due to such "third body" effects.

With rough surfaces, i. e., those that have *random* protrusions rather than being periodically structured, we expect a smearing out of the correlated intermolecular interactions that are involved in film freezing and melting (and in phase transitions in general). This should effectively eliminate the highly regular stick–slip and may also affect the location of the slipping planes [18.123, 252, 279]. The stick–slip friction of "real" surfaces, which are generally rough, may, therefore, be quite different from those of perfectly smooth surfaces composed of the same material. We should note, however, that even between rough surfaces, most of the contacts occur between the tips of microscopic asperities, which may be smooth over their microscopic contact area.

References

18.1 J. N. Israelachvili: Surface Forces and Microrheology of Molecularly Thin Liquid Films. In: *Handbook of Micro/Nanotribology*, ed. by B. Bhushan (CRC Press, Boca Raton 1995) pp. 267–319

18.2 K. B. Lodge: Techniques for the measurement of forces between solids, Adv. Colloid Interface Sci. **19** (1983) 27–73

18.3 J. N. Israelachvili: *Intermolecular and Surface Forces*, 2nd edn. (Academic Press, London 1991)

18.4 P. F. Luckham, B. A. de L. Costello: Recent developments in the measurement of interparticle forces, Adv. Colloid Interface Sci. **44** (1993) 183–240

18.5 P. M. Claesson, T. Ederth, V. Bergeron, M. W. Rutland: Techniques for measuring surface forces, Adv. Colloid Interface Sci. **67** (1996) 119–183

18.6 V. S. J. Craig: An historical review of surface force measurement techniques, Colloids Surf. A **129–130** (1997) 75–94

18.7 J. N. Israelachvili, G. E. Adams: Measurements of forces between two mica surfaces in aqueous electrolyte solutions in the range 0–100 nm, J. Chem. Soc. Faraday Trans. I **74** (1978) 975–1001

18.8 G. Binnig, C. F. Quate, C. Gerber: Atomic force microscope, Phys. Rev. Lett. **56** (1986) 930–933

18.9 W. A. Ducker, T. J. Senden, R. M. Pashley: Direct measurement of colloidal forces using an atomic force microscope, Nature **353** (1991) 239–241

18.10 E. Meyer, R. M. Overney, K. Dransfeld, T. Gyalog: *Nanoscience: Friction and Rheology on the Nanometer Scale* (World Scientific, Singapore 1998)

18.11 J. N. Israelachvili: Measurements of the viscosity of thin fluid films between two surfaces with and without adsorbed polymers, Colloid Polym. Sci. **264** (1986) 1060–1065

18.12 J. P. Montfort, G. Hadziioannou: Equilibrium and dynamic behavior of thin films of a perfluorinated polyether, J. Chem. Phys. **88** (1988) 7187–7196

18.13 A. Dhinojwala, S. Granick: Surface forces in the tapping mode: Solvent permeability and hydrodynamic thickness of adsorbed polymer brushes, Macromolecules **30** (1997) 1079–1085

18.14 B. V. Derjaguin: Untersuchungen über die Reibung und Adhäsion, IV. Theorie des Anhaftens kleiner Teilchen, Kolloid Z. **69** (1934) 155–164

18.15 K. L. Johnson, K. Kendall, A. D. Roberts: Surface energy and the contact of elastic solids, Proc. R. Soc. London A **324** (1971) 301–313

18.16 J. N. Israelachvili, P. M. McGuiggan: Adhesion and short-range forces between surfaces. Part 1: New apparatus for surface force measurements, J. Mater. Res. **5** (1990) 2223–2231

18.17 J. N. Israelachvili: Thin film studies using multiple-beam interferometry, J. Colloid Interface Sci. **44** (1973) 259–272

18.18 Y. L. Chen, T. Kuhl, J. Israelachvili: Mechanism of cavitation damage in thin liquid films: Collapse damage vs. inception damage, Wear **153** (1992) 31–51

18.19 M. Heuberger, G. Luengo, J. Israelachvili: Topographic information from multiple beam interferometry in the surface forces apparatus, Langmuir **13** (1997) 3839–3848

18.20 R. M. Pashley: DLVO and hydration forces between mica surfaces in Li^+, Na^+, K^+, and Cs^+ electrolyte solutions: A correlation of double-layer and hydration forces with surface cation exchange properties, J. Colloid Interface Sci. **83** (1981) 531–546

18.21 R. M. Pashley: Hydration forces between mica surfaces in electrolyte solution, Adv. Colloid Interface Sci. **16** (1982) 57–62

18.22 R. G. Horn, D. T. Smith, W. Haller: Surface forces and viscosity of water measured between silica sheets, Chem. Phys. Lett. **162** (1989) 404–408

18.23 R. G. Horn, D. R. Clarke, M. T. Clarkson: Direct measurements of surface forces between sapphire crystals in aqueous solutions, J. Mater. Res. **3** (1988) 413–416

18.24 W. W. Merrill, A. V. Pocius, B. V. Thakker, M. Tirrell: Direct measurement of molecular level adhesion forces between biaxially oriented solid polymer films, Langmuir **7** (1991) 1975–1980

18.25 J. Klein: Forces between mica surfaces bearing adsorbed macromolecules in liquid media, J. Chem. Soc. Faraday Trans. I **79** (1983) 99–118

18.26 S. S. Patel, M. Tirrell: Measurement of forces between surfaces in polymer fluids, Annu. Rev. Phys. Chem. **40** (1989) 597–635

18.27 H. Watanabe, M. Tirrell: Measurements of forces in symmetric and asymmetric interactions between diblock copolymer layers adsorbed on mica, Macromolecules **26** (1993) 6455–6466

18.28 T. L. Kuhl, D.E. Leckband, D. D. Lasic, J. N. Israelachvili: Modulation and modeling of interaction forces between lipid bilayers exposing terminally grafted polymer chains. In: *Stealth Liposomes*, ed. by D. Lasic, F. Martin (CRC Press, Boca Raton 1995) pp. 73–91

18.29 J. Klein: Shear, friction, and lubrication forces between polymer-bearing surfaces, Annu. Rev. Mater. Sci. **26** (1996) 581–612

18.30 M. Ruths, D. Johannsmann, J. Rühe, W. Knoll: Repulsive forces and relaxation on compression of entangled, polydisperse polystyrene brushes, Macromolecules **33** (2000) 3860–3870

18.31 C. A. Helm, J. N. Israelachvili, P. M. McGuiggan: Molecular mechanisms and forces involved in the adhesion and fusion of amphiphilic bilayers, Science **246** (1989) 919–922

18.32 Y. L. Chen, C. A. Helm, J. N. Israelachvili: Molecular mechanisms associated with adhesion and contact angle hysteresis of monolayer surfaces, J. Phys. Chem. **95** (1991) 10736–10747

18.33 D. E. Leckband, J. N. Israelachvili, F.-J. Schmitt, W. Knoll: Long-range attraction and molecular rearrangements in receptor-ligand interactions, Science **255** (1992) 1419–1421

18.34 J. Peanasky, H. M. Schneider, S. Granick, C. R. Kessel: Self-assembled monolayers on mica for experiments utilizing the surface forces apparatus, Langmuir **11** (1995) 953–962

18.35 C. J. Coakley, D. Tabor: Direct measurement of van der Waals forces between solids in air, J. Phys. D **11** (1978) L77–L82

18.36 J. L. Parker, H. K. Christenson: Measurements of the forces between a metal surface and mica across liquids, J. Chem. Phys. **88** (1988) 8013–8014

18.37 C. P. Smith, M. Maeda, L. Atanasoska, H. S. White, D. J. McClure: Ultrathin platinum films on mica and the measurement of forces at the platinum/water interface, J. Phys. Chem. **92** (1988) 199–205

18.38 S. J. Hirz, A. M. Homola, G. Hadziioannou, C. W. Frank: Effect of substrate on shearing properties of ultrathin polymer films, Langmuir **8** (1992) 328–333

18.39 J. M. Levins, T. K. Vanderlick: Reduction of the roughness of silver films by the controlled application of surface forces, J. Phys. Chem. **96** (1992) 10405–10411

18.40 S. Steinberg, W. Ducker, G. Vigil, C. Hyukjin, C. Frank, M. Z. Tseng, D. R. Clarke, J. N. Israelachvili: Van der Waals epitaxial growth of α-alumina nanocrystals on mica, Science **260** (1993) 656–659

18.41 G. Vigil, Z. Xu, S. Steinberg, J. Israelachvili: Interactions of silica surfaces, J. Colloid Interface Sci. **165** (1994) 367–385

18.42 M. Ruths, M. Heuberger, V. Scheumann, J. Hu, W. Knoll: Confinement-induced film thickness transitions in liquid crystals between two alkanethiol monolayers on gold, Langmuir **17** (2001) 6213–6219

18.43 J. Van Alsten, S. Granick: Molecular tribometry of ultrathin liquid films, Phys. Rev. Lett. **61** (1988) 2570–2573

18.44 J. Van Alsten, S. Granick: Shear rheology in a confined geometry: Polysiloxane melts, Macromolecules **23** (1990) 4856–4862

18.45 A. M. Homola, J. N. Israelachvili, M. L. Gee, P. M. McGuiggan: Measurements of and relation between the adhesion and friction of two surfaces separated by molecularly thin liquid films, J. Tribol. **111** (1989) 675–682

18.46 J. Klein, E. Kumacheva: Simple liquids confined to molecularly thin layers. I. Confinement-induced liquid-to-solid phase transitions, J. Chem. Phys. **108** (1998) 6996–7009

18.47 G. Luengo, F.-J. Schmitt, R. Hill, J. Israelachvili: Thin film rheology and tribology of confined polymer melts: Contrasts with bulk properties, Macromolecules **30** (1997) 2482–2494

18.48 E. Kumacheva: Interfacial friction measurements in surface force apparatus, Prog. Surf. Sci. **58** (1998) 75–120

18.49 A. L. Weisenhorn, P. K. Hansma, T. R. Albrecht, C. F. Quate: Forces in atomic force microscopy in air and water, Appl. Phys. Lett. **54** (1989) 2651–2653

18.50 G. Meyer, N. M. Amer: Simultaneous measurement of lateral and normal forces with an optical-beam-deflection atomic force microscope, Appl. Phys. Lett. **57** (1990) 2089–2091

18.51 E. L. Florin, V. T. Moy, H. E. Gaub: Adhesion forces between individual ligand–receptor pairs, Science **264** (1994) 415–417

18.52 G. U. Lee, D. A. Kidwell, R. J. Colton: Sensing discrete streptavidin–biotin interactions with atomic force microscopy, Langmuir **10** (1994) 354–357

18.53 C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59** (1987) 1942–1945

18.54 R. W. Carpick, M. Salmeron: Scratching the surface: Fundamental investigations of tribology with atomic force microscopy, Chem. Rev. **97** (1997) 1163–1194

18.55 J. P. Cleveland, S. Manne, D. Bocek, P. K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64** (1993) 403–405

18.56 J. E. Sader, J. W. M. Chon, P. Mulvaney: Calibration of rectangular atomic force microscope cantilevers, Rev. Sci. Instrum. **70** (1999) 3967–3969

18.57 Y. Liu, T. Wu, D. F. Evans: Lateral force microscopy study on the shear properties of self-assembled monolayers of dialkylammonium surfactant on mica, Langmuir **10** (1994) 2241–2245

18.58 C. Neto, V. S. J. Craig: Colloid probe characterization: Radius and roughness determination, Langmuir **17** (2001) 2097–2099

18.59 R. G. Horn, J. N. Israelachvili: Molecular organization and viscosity of a thin film of molten polymer between two surfaces as probed by force measurements, Macromolecules **21** (1988) 2836–2841

18.60 R. G. Horn, S. J. Hirz, G. Hadziioannou, C. W. Frank, J. M. Catala: A reevaluation of forces measured across thin polymer films: Nonequilibrium and pinning effects, J. Chem. Phys. **90** (1989) 6767–6774

18.61 O. I. Vinogradova, H.-J. Butt, G. E. Yakubov, F. Feuillebois: Dynamic effects on force measurements. I. Viscous drag on the atomic force microscope cantilever, Rev. Sci. Instrum. **72** (2001) 2330–2339

18.62 D. Leckband, J. Israelachvili: Intermolecular forces in biology, Quart. Rev. Biophys. **34** (2001) 105–267

18.63 E. Evans, D. Needham: Physical properties of surfactant bilayer membranes: Thermal transitions, elasticity, rigidity, cohesion, and colloidal interactions, J. Phys. Chem. **91** (1987) 4219–4228

18.64 E. Evans, D. Needham: Attraction between lipid bilayer membranes in concentrated solutions of nonadsorbing polymers: Comparison of mean-field theory with measurements of adhesion energy, Macromolecules **21** (1988) 1822–1831

18.65 S. E. Chesla, P. Selvaraj, C. Zhu: Measuring two-dimensional receptor-ligand binding kinetics by micropipette, Biophys. J. **75** (1998) 1553–1572

18.66 E. Evans, K. Ritchie, R. Merkel: Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces, Biophys. J. **68** (1995) 2580–2587

18.67 D. M. LeNeveu, R. P. Rand, V. A. Parsegian: Measurements of forces between lecithin bilayers, Nature **259** (1976) 601–603

18.68 A. Homola, A. A. Robertson: A compression method for measuring forces between colloidal particles, J. Colloid Interface Sci. **54** (1976) 286–297

18.69 V. A. Parsegian, N. Fuller, R. P. Rand: Measured work of deformation and repulsion of lecithin bilayers, Proc. Nat. Acad. Sci. USA **76** (1979) 2750–2754

18.70 R. P. Rand, V. A. Parsegian: Hydration forces between phospholipid bilayers, Biochim. Biophys. Acta **988** (1989) 351–376

18.71 S. Leikin, V. A. Parsegian, D. C. Rau, R. P. Rand: Hydration forces, Annu. Rev. Phys. Chem. **44** (1993) 369–395

18.72 S. Chu, J. E. Bjorkholm, A. Ashkin, A. Cable: Experimental observation of optically trapped atoms, Phys. Rev. Lett. **57** (1986) 314–317

18.73 A. Ashkin: Optical trapping and manipulation of neutral particles using lasers, Proc. Nat. Acad. Sci. USA **94** (1997) 4853–4860

18.74 K. Visscher, S. P. Gross, S. M. Block: Construction of multiple-beam optical traps with nanometer-resolution positioning sensing, IEEE J. Sel. Top. Quantum Electron. **2** (1996) 1066–1076

18.75 D. C. Prieve, N. A. Frej: Total internal reflection microscopy: A quantitative tool for the measurement of colloidal forces, Langmuir **6** (1990) 396–403

18.76 J. Rädler, E. Sackmann: On the measurement of weak repulsive and frictional colloidal forces by reflection interference contrast microscopy, Langmuir **8** (1992) 848–853

18.77 P. Bongrand, C. Capo, J.-L. Mege, A.-M. Benoliel: Use of hydrodynamic flows to study cell adhesion. In: *Physical Basis of Cell–Cell Adhesion*, ed. by P. Bongrand (CRC Press, Boca Raton 1988) pp. 125–156

18.78 G. Kaplanski, C. Farnarier, O. Tissot, A. Pierres, A.-M. Benoliel, A.-C. Alessi, S. Kaplanski, P. Bongrand: Granulocyte–endothelium initial adhesion: Analysis of transient binding events mediated by E-selectin in a laminar shear flow, Biophys. J. **64** (1993) 1922–1933

18.79 H. B. G. Casimir, D. Polder: The influence of retardation on the London–van der Waals forces, Phys. Rev. **73** (1948) 360–372

18.80 H. C. Hamaker: The London–van der Waals attraction between spherical particles, Physica **4** (1937) 1058–1072

18.81 E. M. Lifshitz: The theory of molecular attraction forces between solid bodies, Sov. Phys. JETP (English translation) **2** (1956) 73–83

18.82 I. E. Dzyaloshinskii, E. M. Lifshitz, L. P. Pitaevskii: The general theory of van der Waals forces, Adv. Phys. **10** (1961) 165–209

18.83 H. W. Fox, W. A. Zisman: The spreading of liquids on low-energy surfaces. III. Hydrocarbon surfaces, J. Colloid Sci. **7** (1952) 428–442

18.84 B. V. Derjaguin, V. P. Smilga: Electrostatic component of the rolling friction force moment, Wear **7** (1964) 270–281

18.85 R. G. Horn, D. T. Smith: Contact electrification and adhesion between dissimilar materials, Science **256** (1992) 362–364

18.86 R. Budakian, S. J. Putterman: Correlation between charge transfer and stick–slip friction at a metal–insulator interface, Phys. Rev. Lett. **85** (2000) 1000–1003

18.87 M. Grätzel: Photoelectrochemical cells, Nature **414** (2001) 338–344

18.88 R. M. German: *Sintering Theory and Practice* (Wiley, New York 1996)

18.89 R. Budakian, S. J. Putterman: Time scales for cold welding and the origins of stick–slip friction, Phys. Rev. B **65** (2002) 235429/1–5

18.90 U. Landman, W. D. Luedtke, N. A. Burnham, R. J. Colton: Atomistic mechanisms and dynamics of adhesion, nanoindentation, and fracture, Science **248** (1990) 454–461

18.91 U. Landman, W. D. Luedtke, E. M. Ringer: Molecular dynamics simulations of adhesive contact formation and friction, NATO Science Ser. E **220** (1992) 463–510

18.92 W. D. Luedtke, U. Landman: Solid and liquid junctions, Comput. Mater. Sci. **1** (1992) 1–24

18.93 U. Landman, W. D. Luedtke: Interfacial junctions and cavitation, MRS Bull. **18** (1993) 36–44

18.94 B. Bhushan, J. N. Israelachvili, U. Landman: Nanotribology: Friction, wear and lubrication at the atomic scale, Nature **374** (1995) 607–616

18.95 M. R. Sørensen, K. W. Jacobsen, P. Stoltze: Simulations of atomic-scale sliding friction, Phys. Rev. B **53** (1996) 2101–2113

18.96 A. Meurk, P. F. Luckham, L. Bergström: Direct measurement of repulsive and attractive van der Waals forces between inorganic materials, Langmuir **13** (1997) 3896–3899

18.97 E. J. W. Verwey, J. T. G. Overbeek: *Theory of the Stability of Lyophobic Colloids*, 1st edn. (Elsevier, Amsterdam 1948)

18.98 D. Y. C. Chan, R. M. Pashley, L. R. White: A simple algorithm for the calculation of the electrostatic repulsion between identical charged surfaces in electrolyte, J. Colloid Interface Sci. **77** (1980) 283–285

18.99 B. Derjaguin, L. Landau: Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes, Acta Physicochim. URSS **14** (1941) 633–662

18.100 D. Chan, T. W. Healy, L. R. White: Electrical double layer interactions under regulation by surface ionization equilibriums – dissimilar amphoteric surfaces, J. Chem. Soc. Faraday Trans. 1 **72** (1976) 2844–2865

18.101 G. S. Manning: Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties, J. Chem. Phys. **51** (1969) 924–933

18.102 L. Guldbrand, V. Jönsson, H. Wennerström, P. Linse: Electrical double-layer forces: A Monte Carlo study, J. Chem. Phys. **80** (1984) 2221–2228

18.103 H. Wennerström, B. Jönsson, P. Linse: The cell model for polyelectrolyte systems. Exact statistical mechanical relations, Monte Carlo simulations, and the Poisson–Boltzmann approximation, J. Chem. Phys. **76** (1982) 4665–4670

18.104 J. Marra: Effects of counterion specificity on the interactions between quaternary ammonium surfactants in monolayers and bilayers, J. Phys. Chem. **90** (1986) 2145–2150

18.105 J. Marra: Direct measurement of the interaction between phosphatidylglycerol bilayers in aqueous-electrolyte solutions, Biophys. J. **50** (1986) 815–825

18.106 R. G. Horn, J. N. Israelachvili: Direct measurement of structural forces between two surfaces in a nonpolar liquid, J. Chem. Phys. **75** (1981) 1400–1411

18.107 I. K. Snook, W. van Megen: Solvation forces in simple dense fluids I, J. Chem. Phys. **72** (1980) 2907–2913

18.108 W. J. van Megen, I. K. Snook: Solvation forces in simple dense fluids II. Effect of chemical potential, J. Chem. Phys. **74** (1981) 1409–1411

18.109 R. Kjellander, S. Marcelja: Perturbation of hydrogen bonding in water near polar surfaces, Chem. Phys. Lett. **120** (1985) 393–396

18.110 P. Tarazona, L. Vicente: A model for the density oscillations in liquids between solid walls, Mol. Phys. **56** (1985) 557–572

18.111 D. Henderson, M. Lozada-Cassou: A simple theory for the force between spheres immersed in a fluid, J. Colloid Interface Sci. **114** (1986) 180–183

18.112 J. E. Curry, J. H. Cushman: Structure in confined fluids: Phase separation of binary simple liquid mixtures, Tribol. Lett. **4** (1998) 129–136

18.113 M. Schoen, T. Gruhn, D. J. Diestler: Solvation forces in thin films confined between macroscopically curved substrates, J. Chem. Phys. **109** (1998) 301–311

18.114 F. Porcheron, B. Rousseau, M. Schoen, A. H. Fuchs: Structure and solvation forces in confined alkane films, Phys. Chem. Chem. Phys. **3** (2001) 1155–1159

18.115 H. K. Christenson, D. W. R. Gruen, R. G. Horn, J. N. Israelachvili: Structuring in liquid alkanes between solid surfaces: Force measurements and mean-field theory, J. Chem. Phys. **87** (1987) 1834–1841

18.116 M. L. Gee, J. N. Israelachvili: Interactions of surfactant monolayers across hydrocarbon liquids, J. Chem. Soc. Faraday Trans. **86** (1990) 4049–4058

18.117 J. N. Israelachvili, S. J. Kott, M. L. Gee, T. A. Witten: Forces between mica surfaces across hydrocarbon liquids: Effects of branching and polydispersity, Macromolecules **22** (1989) 4247–4253

18.118 H. K. Christenson: Forces between solid surfaces in a binary mixture of non-polar liquids, Chem. Phys. Lett. **118** (1985) 455–458

18.119 H. K. Christenson, R. G. Horn: Solvation forces measured in non-aqueous liquids, Chem. Scr. **25** (1985) 37–41

18.120 J. Israelachvili: Solvation forces and liquid structure, as probed by direct force measurements, Acc. Chem. Res. **20** (1987) 415–421

18.121 H. K. Christenson: Non-DLVO forces between surfaces – solvation, hydration and capillary effects, J. Disp. Sci. Technol. **9** (1988) 171–206

18.122 L. J. D. Frink, F. van Swol: Solvation forces between rough surfaces, J. Chem. Phys. **108** (1998) 5588–5598

18.123 J. Gao, W. D. Luedtke, U. Landman: Structures, solvation forces and shear of molecular films in a rough nano-confinement, Tribol. Lett. **9** (2000) 3–13

18.124 H. E. Stanley, J. Teixeira: Interpretation of the unusual behavior of H_2O and D_2O at low temperatures: Tests of a percolation model, J. Chem. Phys. **73** (1980) 3404–3422

18.125 H. van Olphen: *An Introduction to Clay Colloid Chemistry*, 2nd edn. (Wiley, New York 1977) Chap. 10

18.126 U. Del Pennino, E. Mazzega, S. Valeri, A. Alietti, M. F. Brigatti, L. Poppi: Interlayer water and swelling properties of monoionic montmorillonites, J. Colloid Interface Sci. **84** (1981) 301–309

18.127 B. E. Viani, P. F. Low, C. B. Roth: Direct measurement of the relation between interlayer force and interlayer distance in the swelling of montmorillonite, J. Colloid Interface Sci. **96** (1983) 229–244

18.128 R. G. Horn: Surface forces and their action in ceramic materials, J. Am. Ceram. Soc. **73** (1990) 1117–1135

18.129 B. V. Velamakanni, J. C. Chang, F. F. Lange, D. S. Pearson: New method for efficient colloidal particle packing via modulation of repulsive lubricating hydration forces, Langmuir **6** (1990) 1323–1325

18.130 R. R. Lessard, S. A. Zieminski: Bubble coalescence and gas transfer in aqueous electrolytic solutions, Ind. Eng. Chem. Fundam. **10** (1971) 260–269

18.131 J. Israelachvili, R. Pashley: The hydrophobic interaction is long range, decaying exponentially with distance, Nature **300** (1982) 341–342

18.132 R. M. Pashley, P. M. McGuiggan, B. W. Ninham, D. F. Evans: Attractive forces between uncharged hydrophobic surfaces: Direct measurements in aqueous solutions, Science **229** (1985) 1088–1089

18.133 P. M. Claesson, C. E. Blom, P. C. Herder, B. W. Ninham: Interactions between water-stable hydrophobic Langmuir–Blodgett monolayers on mica, J. Colloid Interface Sci. **114** (1986) 234–242

18.134 Ya. I. Rabinovich, B. V. Derjaguin: Interaction of hydrophobized filaments in aqueous electrolyte solutions, Colloids Surf. **30** (1988) 243–251

18.135 J. L. Parker, D. L. Cho, P. M. Claesson: Plasma modification of mica: Forces between fluorocarbon surfaces in water and a nonpolar liquid, J. Phys. Chem. **93** (1989) 6121–6125

18.136 H. K. Christenson, J. Fang, B. W. Ninham, J. L. Parker: Effect of divalent electrolyte on the hydrophobic attraction, J. Phys. Chem. **94** (1990) 8004–8006

18.137 K. Kurihara, S. Kato, T. Kunitake: Very strong long range attractive forces between stable hydrophobic monolayers of a polymerized ammonium surfactant, Chem. Lett. (1990) 1555–1558

18.138 Y. H. Tsao, D. F. Evans, H. Wennerstrom: Long-range attractive force between hydrophobic surfaces observed by atomic force microscopy, Science **262** (1993) 547–550

18.139 Ya. I. Rabinovich, R.-H. Yoon: Use of atomic force microscope for the measurements of hydrophobic forces between silanated silica plate and glass sphere, Langmuir **10** (1994) 1903–1909

18.140 V. S. J. Craig, B. W. Ninham, R. M. Pashley: Study of the long-range hydrophobic attraction in concentrated salt solutions and its implications for electrostatic models, Langmuir **14** (1998) 3326–3332

18.141 P. Kékicheff, O. Spalla: Long-range electrostatic attraction between similar, charge-neutral walls, Phys. Rev. Lett. **75** (1995) 1851–1854

18.142 H. K. Christenson, P. M. Claesson: Direct measurements of the force between hydrophobic surfaces in water, Adv. Colloid Interface Sci. **91** (2001) 391–436

18.143 P. Attard: Nanobubbles and the hydrophobic attraction, Adv. Colloid Interface Sci. **104** (2003) 75–91

18.144 M. Hato: Attractive forces between surfaces of controlled "hydrophobicity" across water: A possible range of "hydrophobic interactions" between macroscopic hydrophobic surfaces across water, J. Phys. Chem. **100** (1996) 18530–18538

18.145 H. K. Christenson, P. M. Claesson, J. Berg, P. C. Herder: Forces between fluorocarbon surfactant monolayers: Salt effects on the hydrophobic interaction, J. Phys. Chem. **93** (1989) 1472–1478

18.146 N. I. Christou, J. S. Whitehouse, D. Nicholson, N. G. Parsonage: A Monte Carlo study of fluid water in contact with structureless walls, Faraday Symp. Chem. Soc. **16** (1981) 139–149

18.147 B. Jönsson: Monte Carlo simulations of liquid water between two rigid walls, Chem. Phys. Lett. **82** (1981) 520–525

18.148 S. Marcelja, D. J. Mitchell, B. W. Ninham, M. J. Sculley: Role of solvent structure in solution theory, J. Chem. Soc. Faraday Trans. II **73** (1977) 630–648

18.149 D. W. R. Gruen, S. Marcelja: Spatially varying polarization in water: A model for the electric double layer and the hydration force, J. Chem. Soc. Faraday Trans. 2 **79** (1983) 225–242

18.150 B. Jönsson, H. Wennerström: Image-charge forces in phospholipid bilayer systems, J. Chem. Soc. Faraday Trans. 2 **79** (1983) 19–35
18.151 D. Schiby, E. Ruckenstein: The role of the polarization layers in hydration forces, Chem. Phys. Lett. **95** (1983) 435–438
18.152 A. Luzar, D. Bratko, L. Blum: Monte Carlo simulation of hydrophobic interaction, J. Chem. Phys. **86** (1987) 2955–2959
18.153 P. Attard, M. T. Batchelor: A mechanism for the hydration force demonstrated in a model system, Chem. Phys. Lett. **149** (1988) 206–211
18.154 K. Leung, A. Luzar: Dynamics of capillary evaporation. II. Free energy barriers, J. Chem. Phys. **113** (2000) 5845–5852
18.155 D. Bratko, R. A. Curtis, H. W. Blanch, J. M. Prausnitz: Interaction between hydrophobic surfaces with metastable intervening liquid, J. Chem. Phys. **115** (2001) 3873–3877
18.156 P. G. de Gennes: Polymers at an interface. 2. Interaction between two plates carrying adsorbed polymer layers, Macromolecules **15** (1982) 492–500
18.157 J. M. H. M. Scheutjens, G. J. Fleer: Interaction between two adsorbed polymer layers, Macromolecules **18** (1985) 1882–1900
18.158 E. A. Evans: Force between surfaces that confine a polymer solution: Derivation from self-consistent field theories, Macromolecules **22** (1989) 2277–2286
18.159 H. J. Ploehn: Compression of polymer interphases, Macromolecules **27** (1994) 1627–1636
18.160 S. Asakura, F. Oosawa: Interaction between particles suspended in solutions of macromolecules, J. Polym. Sci. **33** (1958) 183–192
18.161 J. F. Joanny, L. Leibler, P. G. de Gennes: Effects of polymer solutions on colloid stability, J. Polym. Sci. Polym. Phys. **17** (1979) 1073–1084
18.162 B. Vincent, P. F. Luckham, F. A. Waite: The effect of free polymer on the stability of sterically stabilized dispersions, J. Colloid Interface Sci. **73** (1980) 508–521
18.163 R. I. Feigin, D. H. Napper: Stabilization of colloids by free polymer, J. Colloid Interface Sci. **74** (1980) 567–571
18.164 P. G. de Gennes: Polymer solutions near an interface. 1. Adsorption and depletion layers, Macromolecules **14** (1981) 1637–1644
18.165 P. G. de Gennes: Polymers at an interface; a simplified view, Adv. Colloid Interface Sci. **27** (1987) 189–209
18.166 S. Alexander: Adsorption of chain molecules with a polar head. A scaling description, J. Phys. (France) **38** (1977) 983–987
18.167 P. G. de Gennes: Conformations of polymers attached to an interface, Macromolecules **13** (1980) 1069–1075
18.168 S. T. Milner, T. A. Witten, M. E. Cates: Theory of the grafted polymer brush, Macromolecules **21** (1988) 2610–2619
18.169 S. T. Milner, T. A. Witten, M. E. Cates: Effects of polydispersity in the end-grafted polymer brush, Macromolecules **22** (1989) 853–861
18.170 E. B. Zhulina, O. V. Borisov, V. A. Priamitsyn: Theory of steric stabilization of colloid dispersions by grafted polymers, J. Colloid Interface Sci. **137** (1990) 495–511
18.171 H. J. Taunton, C. Toprakcioglu, L. J. Fetters, J. Klein: Interactions between surfaces bearing end-adsorbed chains in a good solvent, Macromolecules **23** (1990) 571–580
18.172 J. Klein, P. Luckham: Forces between two adsorbed poly(ethylene oxide) layers immersed in a good aqueous solvent, Nature **300** (1982) 429–431
18.173 J. Klein, P. F. Luckham: Long-range attractive forces between two mica surfaces in an aqueous polymer solution, Nature **308** (1984) 836–837
18.174 P. F. Luckham, J. Klein: Forces between mica surfaces bearing adsorbed homopolymers in good solvents, J. Chem. Soc. Faraday Trans. **86** (1990) 1363–1368
18.175 J. Klein, E. Kumacheva, D. Mahalu, D. Perahia, L. J. Fetters: Reduction of frictional forces between solid surfaces bearing polymer brushes, Nature **370** (1994) 634–636
18.176 J. N. Israelachvili, H. Wennerström: Hydration or steric forces between amphiphilic surfaces?, Langmuir **6** (1990) 873–876
18.177 J. N. Israelachvili, H. Wennerström: Entropic forces between amphiphilic surfaces in liquids, J. Phys. Chem. **96** (1992) 520–531
18.178 M. K. Granfeldt, S. J. Miklavic: A simulation study of flexible zwitterionic monolayers. Interlayer interaction and headgroup conformation, J. Phys. Chem. **95** (1991) 6351–6360
18.179 L. R. Fisher, J. N. Israelachvili: Direct measurements of the effect of meniscus forces on adhesion: A study of the applicability of macroscopic thermodynamics to microscopic liquid interfaces, Colloids Surf. **3** (1981) 303–319
18.180 H. K. Christenson: Adhesion between surfaces in unsaturated vapors – a reexamination of the influence of meniscus curvature and surface forces, J. Colloid Interface Sci. **121** (1988) 170–178
18.181 R. G. Horn, J. N. Israelachvili, F. Pribac: Measurement of the deformation and adhesion of solids in contact, J. Colloid Interface Sci. **115** (1987) 480–492
18.182 H. Hertz: Über die Berührung fester elastischer Körper, J. Reine Angew. Math. **92** (1881) 156–171
18.183 H. M. Pollock, D. Maugis, M. Barquins: The force of adhesion between solid surfaces in contact, Appl. Phys. Lett. **33** (1978) 798–799
18.184 B. V. Derjaguin, V. M. Muller, Yu. P. Toporov: Effect of contact deformations on the adhesion of particles, J. Colloid Interface Sci. **53** (1975) 314–326
18.185 V. M. Muller, V. S. Yushchenko, B. V. Derjaguin: On the influence of molecular forces on the deforma-

tion of an elastic sphere and its sticking to a rigid plane, J. Colloid Interface Sci. **77** (1980) 91–101
18.186 V. M. Muller, B. V. Derjaguin, Y. P. Toporov: On 2 methods of calculation of the force of sticking of an elastic sphere to a rigid plane, Colloids Surf. **7** (1983) 251–259
18.187 D. Maugis: Adhesion of spheres: The JKR–DMT transition using a Dugdale model, J. Colloid Interface Sci. **150** (1992) 243–269
18.188 V. Mangipudi, M. Tirrell, A. V. Pocius: Direct measurement of molecular level adhesion between poly(ethylene terephthalate) and polyethylene films: Determination of surface and interfacial energies, J. Adh. Sci. Technol. **8** (1994) 1251–1270
18.189 H. K. Christenson: Surface deformations in direct force measurements, Langmuir **12** (1996) 1404–1405
18.190 M. Barquins, D. Maugis: Fracture mechanics and the adherence of viscoelastic bodies, J. Phys. D: Appl Phys. **11** (1978) 1989–2023
18.191 J. A. Greenwood, J. B. P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. London A **295** (1966) 300–319
18.192 B. N. J. Persson: Elastoplastic contact between randomly rough surfaces, Phys. Rev. Lett. **87** (2001) 116101/1–4
18.193 K. N. G. Fuller, D. Tabor: The effect of surface roughness on the adhesion of elastic solids, Proc. R. Soc. London A **345** (1975) 327–342
18.194 D. Maugis: On the contact and adhesion of rough surfaces, J. Adh. Sci. Technol. **10** (1996) 161–175
18.195 B. N. J. Persson, E. Tosatti: The effect of surface roughness on the adhesion of elastic solids, J. Chem. Phys. **115** (2001) 5597–5610
18.196 P. M. McGuiggan, J. N. Israelachvili: Adhesion and short-range forces between surfaces. Part II: Effects of surface lattice mismatch, J. Mater. Res. **5** (1990) 2232–2243
18.197 P. M. McGuiggan, J. Israelachvili: Measurements of the effects of angular lattice mismatch on the adhesion energy between two mica surfaces in water, Mat. Res. Soc. Symp. Proc. **138** (1989) 349–360
18.198 C. L. Rhykerd, Jr., M. Schoen, D. J. Diestler, J. H. Cushman: Epitaxy in simple classical fluids in micropores and near-solid surfaces, Nature **330** (1987) 461–463
18.199 M. Schoen, D. J. Diestler, J. H. Cushman: Fluids in micropores. I. Structure of a simple classical fluid in a slit-pore, J. Chem. Phys. **87** (1987) 5464–5476
18.200 M. Schoen, C. L. Rhykerd, Jr., D. J. Diestler, J. H. Cushman: Shear forces in molecularly thin films, Science **245** (1989) 1223–1225
18.201 P. A. Thompson, M. O. Robbins: Origin of stick–slip motion in boundary lubrication, Science **250** (1990) 792–794
18.202 K. K. Han, J. H. Cushman, D. J. Diestler: Grand canonical Monte Carlo simulations of a Stockmayer fluid in a slit micropore, Mol. Phys. **79** (1993) 537–545
18.203 M. Ruths, S. Granick: Influence of alignment of crystalline confining surfaces on static forces and shear in a liquid crystal, 4′-*n*-pentyl-4-cyanobiphenyl (5CB), Langmuir **16** (2000) 8368–8376
18.204 A. D. Berman: Dynamics of molecules at surfaces. Ph.D. Thesis (Univ. of California, Santa Barbara 1996)
18.205 E. Barthel, S. Roux: Velocity dependent adherence: An analytical approach for the JKR and DMT models, Langmuir **16** (2000) 8134–8138
18.206 M. Ruths, S. Granick: Rate-dependent adhesion between polymer and surfactant monolayers on elastic substrates, Langmuir **14** (1998) 1804–1814
18.207 H. Yoshizawa, Y. L. Chen, J. Israelachvili: Fundamental mechanisms of interfacial friction. 1: Relation between adhesion and friction, J. Phys. Chem. **97** (1993) 4128–4140
18.208 N. Maeda, N. Chen, M. Tirrell, J. N. Israelachvili: Adhesion and friction mechanisms of polymer-on-polymer surfaces, Science **297** (2002) 379–382
18.209 F. P. Bowden, D. Tabor: *The Friction and Lubrication of Solids* (Clarendon, London 1971)
18.210 J. A. Greenwood, K. L. Johnson: The mechanics of adhesion of viscoelastic solids, Phil. Mag. A **43** (1981) 697–711
18.211 D. Maugis: Subcritical crack growth, surface energy, fracture toughness, stick–slip and embrittlement, J. Mater. Sci. **20** (1985) 3041–3073
18.212 F. Michel, M. E. R. Shanahan: Kinetics of the JKR experiment, C. R. Acad. Sci. II (Paris) **310** (1990) 17–20
18.213 C. A. Miller, P. Neogi: *Interfacial Phenomena: Equilibrium and Dynamic Effects* (Dekker, New York 1985)
18.214 J. Israelachvili, A. Berman: Irreversibility, energy dissipation, and time effects in intermolecular and surface interactions, Israel J. Chem. **35** (1995) 85–91
18.215 A. N. Gent, A. J. Kinloch: Adhesion of viscoelastic materials to rigid substrates. III. Energy criterion for failure, J. Polym. Sci. A-2 **9** (1971) 659–668
18.216 A. N. Gent: Adhesion and strength of viscoelastic solids. Is there a relationship between adhesion and bulk properties?, Langmuir **12** (1996) 4492–4496
18.217 H. R. Brown: The adhesion between polymers, Annu. Rev. Mater. Sci. **21** (1991) 463–489
18.218 M. Deruelle, M. Tirrell, Y. Marciano, H. Hervet, L. Léger: Adhesion energy between polymer networks and solid surfaces modified by polymer attachment, Faraday Discuss. **98** (1995) 55–65
18.219 S. Granick: Motions and relaxations of confined liquids, Science **253** (1991) 1374–1379
18.220 H. W. Hu, S. Granick: Viscoelastic dynamics of confined polymer melts, Science **258** (1992) 1339–1342
18.221 M. L. Gee, P. M. McGuiggan, J. N. Israelachvili, A. M. Homola: Liquid to solidlike transitions of

molecularly thin films under shear, J. Chem. Phys. **93** (1990) 1895–1906
18.222 J. Van Alsten, S. Granick: The origin of static friction in ultrathin liquid films, Langmuir **6** (1990) 876–880
18.223 H.-W. Hu, G. A. Carson, S. Granick: Relaxation time of confined liquids under shear, Phys. Rev. Lett. **66** (1991) 2758–2761
18.224 J. Israelachvili, M. Gee, P. McGuiggan, P. Thompson, M. Robbins: Melting–freezing transitions in molecularly thin liquid films during shear, Fall Meeting of the Mater. Res. Soc., Boston, MA. 1990, ed. by J. M. Drake, J. Klafter, R. Kopelman (MRS Publications, Pittsburgh, PA1990) 3–6
18.225 J. Israelachvili, P. McGuiggan, M. Gee, A. Homola, M. Robbins, P. Thompson: Liquid dynamics in molecularly thin films, J. Phys.: Condens. Matter **2** (1990) SA89–SA98
18.226 J. Klein, E. Kumacheva: Confinement-induced phase transitions in simple liquids, Science **269** (1995) 816–819
18.227 A. L. Demirel, S. Granick: Glasslike transition of a confined simple fluid, Phys. Rev. Lett. **77** (1996) 2261–2264
18.228 D. Dowson: *History of Tribology*, 2nd edn. (Professional Engineering Publishing, London 1998)
18.229 D. Tabor: The role of surface and intermolecular forces in thin film lubrication, Tribol. Ser. **7** (1982) 651–682
18.230 M. J. Sutcliffe, S. R. Taylor, A. Cameron: Molecular asperity theory of boundary friction, Wear **51** (1978) 181–192
18.231 G. M. McClelland: Friction at weakly interacting interfaces. In: *Adhesion and Friction*, ed. by M. Grunze, H. J. Kreuzer (Springer, Berlin, Heidelberg 1989) pp. 1–16
18.232 J. N. Israelachvili, Y.-L. Chen, H. Yoshizawa: Relationship between adhesion and friction forces, J. Adh. Sci. Technol. **8** (1994) 1231–1249
18.233 D. H. Buckley: The metal-to-metal interface and its effect on adhesion and friction, J. Colloid Interface Sci. **58** (1977) 36–53
18.234 V. Zaloj, M. Urbakh, J. Klafter: Modifying friction by manipulating normal response to lateral motion, Phys. Rev. Lett. **82** (1999) 4823–4826
18.235 A. Dhinojwala, S. C. Bae, S. Granick: Shear-induced dilation of confined liquid films, Tribol. Lett. **9** (2000) 55–62
18.236 L. M. Qian, G. Luengo, E. Perez: Thermally activated lubrication with alkanes: The effect of chain length, Europhys. Lett. **61** (2003) 268–274
18.237 B. J. Briscoe, D. C. B. Evans, D. Tabor: The influence of contact pressure and saponification on the sliding behavior of steric acid monolayers, J. Colloid Interface Sci. **61** (1977) 9–13
18.238 A. Berman, J. Israelachvili: Control and minimization of friction via surface modification, NATO ASI Ser. E: Appl. Sci. **330** (1997) 317–329
18.239 M. Urbakh, L. Daikhin, J. Klafter: Dynamics of confined liquids under shear, Phys. Rev. E **51** (1995) 2137–2141
18.240 M. G. Rozman, M. Urbakh, J. Klafter: Origin of stick–slip motion in a driven two-wave potential, Phys. Rev. E **54** (1996) 6485–6494
18.241 M. G. Rozman, M. Urbakh, J. Klafter: Stick–slip dynamics as a probe of frictional forces, Europhys. Lett. **39** (1997) 183–188
18.242 B. V. Derjaguin: Molekulartheorie der äußeren Reibung, Z. Physik **88** (1934) 661–675
18.243 B. V. Derjaguin: Mechanical properties of the boundary lubrication layer, Wear **128** (1988) 19–27
18.244 B. J. Briscoe, D. C. B. Evans: The shear properties of Langmuir–Blodgett layers, Proc. R. Soc. London A **380** (1982) 389–407
18.245 A. Berman, C. Drummond, J. Israelachvili: Amontons' law at the molecular level, Tribol. Lett. **4** (1998) 95–101
18.246 P. F. Luckham, S. Manimaaran: Investigating adsorbed polymer layer behaviour using dynamic surface force apparatuses – a review, Adv. Colloid Interface Sci. **73** (1997) 1–46
18.247 A. M. Homola, J. N. Israelachvili, P. M. McGuiggan, M. L. Gee: Fundamental experimental studies in tribology: The transition from "interfacial" friction of undamaged molecularly smooth surfaces to "normal" friction with wear, Wear **136** (1990) 65–83
18.248 S. Yamada, J. Israelachvili: Friction and adhesion hysteresis of fluorocarbon surfactant monolayer-coated surfaces measured with the surface forces apparatus, J. Phys. Chem. B **102** (1998) 234–244
18.249 G. Luengo, S. E. Campbell, V. I. Srdanov, F. Wudl, J. N. Israelachvili: Direct measurement of the adhesion and friction of smooth C_{60} surfaces, Chem. Mater. **9** (1997) 1166–1171
18.250 L. Rapoport, Y. Bilik, Y. Feldman, M. Homyonfer, S. R. Cohen, R. Tenne: Hollow nanoparticles of WS_2 as potential solid-state lubricants, Nature **387** (1997) 791–793
18.251 J. Israelachvili, Y.-L. Chen, H. Yoshizawa: Relationship between adhesion and friction forces. In: *Fundamentals of Adhesion and Interfaces*, ed. by D. S. Rimai, L. P. DeMejo, K. L. Mittal (VSP, Utrecht, The Netherlands 1995) pp. 261–279
18.252 P. Berthoud, T. Baumberger, C. G'Sell, J. M. Hiver: Physical analysis of the state- and rate-dependent friction law: Static friction, Phys. Rev. B **59** (1999) 14313–14327
18.253 A. Berman, S. Steinberg, S. Campbell, A. Ulman, J. Israelachvili: Controlled microtribology of a metal oxide surface, Tribol. Lett. **4** (1998) 43–48

18.254 D. Y. C. Chan, R. G. Horn: The drainage of thin liquid films between solid surfaces, J. Chem. Phys. **83** (1985) 5311–5324

18.255 J. N. Israelachvili, S. J. Kott: Shear properties and structure of simple liquids in molecularly thin films: The transition from bulk (continuum) to molecular behavior with decreasing film thickness, J. Colloid Interface Sci. **129** (1989) 461–467

18.256 T. L. Kuhl, A. D. Berman, S. W. Hui, J. N. Israelachvili: Part 1: Direct measurement of depletion attraction and thin film viscosity between lipid bilayers in aqueous polyethylene glycol solutions, Macromolecules **31** (1998) 8250–8257

18.257 S. E. Campbell, G. Luengo, V. I. Srdanov, F. Wudl, J. N. Israelachvili: Very low viscosity at the solid–liquid interface induced by adsorbed C_{60} monolayers, Nature **382** (1996) 520–522

18.258 J. Klein, Y. Kamiyama, H. Yoshizawa, J. N. Israelachvili, G. H. Fredrickson, P. Pincus, L. J. Fetters: Lubrication forces between surfaces bearing polymer brushes, Macromolecules **26** (1993) 5552–5560

18.259 G. Luengo, J. Israelachvili, A. Dhinojwala, S. Granick: Generalized effects in confined fluids: New friction map for boundary lubrication, Wear **200** (1996) 328–335

18.260 E. Kumacheva, J. Klein: Simple liquids confined to molecularly thin layers. II. Shear and frictional behavior of solidified films, J. Chem. Phys. **108** (1998) 7010–7022

18.261 P. A. Thompson, G. S. Grest, M. O. Robbins: Phase transitions and universal dynamics in confined films, Phys. Rev. Lett. **68** (1992) 3448–3451

18.262 Y. Rabin, I. Hersht: Thin liquid layers in shear: Non-Newtonian effects, Physica A **200** (1993) 708–712

18.263 P. A. Thompson, M. O. Robbins, G. S. Grest: Structure and shear response in nanometer-thick films, Israel J. Chem. **35** (1995) 93–106

18.264 M. Urbakh, L. Daikhin, J. Klafter: Sheared liquids in the nanoscale range, J. Chem. Phys. **103** (1995) 10707–10713

18.265 A. Subbotin, A. Semenov, E. Manias, G. Hadziioannou, G. ten Brinke: Rheology of confined polymer melts under shear flow: Strong adsorption limit, Macromolecules **28** (1995) 1511–1515

18.266 A. Subbotin, A. Semenov, E. Manias, G. Hadziioannou, G. ten Brinke: Rheology of confined polymer melts under shear flow: Weak adsorption limit, Macromolecules **28** (1995) 3901–3903

18.267 J. Huh, A. Balazs: Behavior of confined telechelic chains under shear, J. Chem. Phys. **113** (2000) 2025–2031

18.268 C. Drummond, J. Elezgaray, P. Richetti: Behavior of adhesive boundary lubricated surfaces under shear: A new dynamic transition, Europhys. Lett. **58** (2002) 503–509

18.269 J. N. Israelachvili, P. M. McGuiggan, A. M. Homola: Dynamic properties of molecularly thin liquid films, Science **240** (1988) 189–191

18.270 A. M. Homola, H. V. Nguyen, G. Hadziioannou: Influence of monomer architecture on the shear properties of molecularly thin polymer melts, J. Chem. Phys. **94** (1991) 2346–2351

18.271 M. Schoen, S. Hess, D. J. Diestler: Rheological properties of confined thin films, Phys. Rev. E **52** (1995) 2587–2602

18.272 G. M. McClelland, S. R. Cohen: *Chemistry and Physics of Solid Surfaces VIII* (Springer, Berlin, Heidelberg 1990) pp. 419–445

18.273 M. O. Robbins, P. A. Thompson: Critical velocity of stick–slip motion, Science **253** (1991) 916

18.274 G. He, M. Müser, M. Robbins: Adsorbed layers and the origin of static friction, Science **284** (1999) 1650–1652

18.275 C. Drummond, J. Israelachvili: Dynamic phase transitions in confined lubricant fluids under shear, Phys. Rev. E. **63** (2001) 041506/1–11

18.276 H. Yoshizawa, J. Israelachvili: Fundamental mechanisms of interfacial friction. 2: Stick–slip friction of spherical and chain molecules, J. Phys. Chem. **97** (1993) 11300–11313

18.277 J. D. Ferry: *Viscoelastic Properties of Polymers*, 3rd edn. (Wiley, New York 1980)

18.278 A. Ruina: Slip instability and state variable friction laws, J. Geophys. Res. **88** (1983) 10359–10370

18.279 T. Baumberger, P. Berthoud, C. Caroli: Physical analysis of the state- and rate-dependent friction law. II. Dynamic friction, Phys. Rev. B **60** (1999) 3928–3939

18.280 J. Israelachvili, S. Giasson, T. Kuhl, C. Drummond, A. Berman, G. Luengo, J.-M. Pan, M. Heuberger, W. Ducker, N. Alcantar: Some fundamental differences in the adhesion and friction of rough versus smooth surfaces, Tribol. Ser. **38** (2000) 3–12

18.281 C. Drummond, J. Israelachvili: Dynamic behavior of confined branched hydrocarbon lubricant fluids under shear, Macromolecules **33** (2000) 4910–4920

18.282 E. Rabinowicz: *Friction and Wear of Materials*, 2nd edn. (Wiley, New York 1995) Chap. 4

18.283 J. Peachey, J. Van Alsten, S. Granick: Design of an apparatus to measure the shear response of ultrathin liquid films, Rev. Sci. Instrum. **62** (1991) 463–473

18.284 E. Rabinowicz: The intrinsic variables affecting the stick–slip process, Proc. Phys. Soc. **71** (1958) 668–675

18.285 T. Baumberger, F. Heslot, B. Perrin: Crossover from creep to inertial motion in friction dynamics, Nature **367** (1994) 544–546

18.286 F. Heslot, T. Baumberger, B. Perrin, B. Caroli, C. Caroli: Creep, stick–slip, and dry-friction dynamics: Experiments and a heuristic model, Phys. Rev. E **49** (1994) 4973–4988

18.287 J. Sampson, F. Morgan, D. Reed, M. Muskat: Friction behavior during the slip portion of the stick–slip process, J. Appl. Phys. **14** (1943) 689–700
18.288 F. Heymann, E. Rabinowicz, B. Rightmire: Friction apparatus for very low-speed sliding studies, Rev. Sci. Instrum. **26** (1954) 56–58
18.289 J. H. Dieterich: Time-dependent friction and the mechanics of stick–slip, Pure Appl. Geophys. **116** (1978) 790–806
18.290 J. H. Dieterich: Modeling of rock friction. 1. Experimental results and constitutive equations, J. Geophys. Res. **84** (1979) 2162–2168
18.291 G. A. Tomlinson: A molecular theory of friction, Phil. Mag. **7** (1929) 905–939
18.292 J. M. Carlson, J. S. Langer: Mechanical model of an earthquake fault, Phys. Rev. A **40** (1989) 6470–6484
18.293 B. N. J. Persson: Theory of friction: The role of elasticity in boundary lubrication, Phys. Rev. B **50** (1994) 4771–4786
18.294 S. Nasuno, A. Kudrolli, J. P. Gollub: Friction in granular layers: Hysteresis and precursors, Phys. Rev. Lett. **79** (1997) 949–952
18.295 P. Bordarier, M. Schoen, A. Fuchs: Stick–slip phase transitions in confined solidlike films from an equilibrium perspective, Phys. Rev. E **57** (1998) 1621–1635
18.296 J. M. Carlson, A. A. Batista: Constitutive relation for the friction between lubricated surfaces, Phys. Rev. E **53** (1996) 4153–4165
18.297 A. D. Berman, W. A. Ducker, J. N. Israelachvili: Origin and characterization of different stick–slip friction mechanisms, Langmuir **12** (1996) 4559–4563
18.298 A. D. Berman, W. A. Ducker, J. N. Israelachvili: Experimental and theoretical investigations of stick–slip friction mechanisms, NATO ASI Ser. E: Appl. Sci. **311** (1996) 51–67
18.299 K G. McLaren, D. Tabor: Viscoelastic properties and the friction of solids. Friction of polymers and influence of speed and temperature, Nature **197** (1963) 856–858
18.300 K. A. Grosch: Viscoelastic properties and friction of solids. Relation between friction and viscoelastic properties of rubber, Nature **197** (1963) 858–859
18.301 J. Gao, W. D. Luedtke, U. Landman: Friction control in thin-film lubrication, J. Phys. Chem. B **102** (1998) 5033–5037
18.302 M. Heuberger, C. Drummond, J. Israelachvili: Coupling of normal and transverse motions during frictional sliding, J. Phys. Chem. B **102** (1998) 5038–5041
18.303 L. Bureau, T. Baumberger, C. Caroli: Shear response of a frictional interface to a normal load modulation, Phys. Rev. E **62** (2000) 6810–6820
18.304 J. P. Gao, W. D. Luedtke, U. Landman: Structure and solvation forces in confined films: Linear and branched alkanes, J. Chem. Phys. **106** (1997) 4309–4318
18.305 J. Gao, W. D. Luedtke, U. Landman: Layering transitions and dynamics of confined liquid films, Phys. Rev. Lett. **79** (1997) 705–708
18.306 J. Warnock, D. D. Awschalom, M. W. Shafer: Orientational behavior of molecular liquids in restricted geometries, Phys. Rev. B **34** (1986) 475–478
18.307 D. D. Awschalom, J. Warnock: Supercooled liquids and solids in porous glass, Phys. Rev. B **35** (1987) 6779–6785
18.308 M. Hirano, K. Shinjo, R. Kaneko, Y. Murata: Anisotropy of frictional forces in muscovite mica, Phys. Rev. Lett. **67** (1991) 2642–2645
18.309 T. Gyalog, H. Thomas: Friction between atomically flat surfaces, Europhys. Lett. **37** (1997) 195–200

19. Scanning Probe Studies of Nanoscale Adhesion Between Solids in the Presence of Liquids and Monolayer Films

Adhesion between solids is a ubiquitous phenomenon whose importance is magnified at the micrometer and nanometer scales, where the surface-to-volume ratio diverges as we approach the single atom.

Numerous techniques for measuring adhesion at the atomic scale have been developed. Yet significant limitations exist. Instrumental improvements and reliable quantification are still needed.

Recent studies have highlighted the unique and important effect of liquid capillaries, particularly water, at the nanometer scale. The results demonstrate that macroscopic considerations of classic meniscus theory must be, at the very least, corrected to take into account new scaling and geometric relationships unique to the nanometer scale. More generally, a molecular-scale description of wetting and capillary condensation as it applies to nanometer-scale interfaces is clearly desirable, but remains an important challenge.

The measurement of adhesion between self-assembled monolayers has proven to be a reliable means for probing the influence of surface chemistry and local environment on adhesion. To date, however, few systems have been investigated quantitatively in detail. The molecular origins of adhesion down to the single bond level remain to be fully investigated. The most recent studies illustrate that although new information about adhesion in these systems has been revealed, further enhancements of current techniques, as well as the development of new methodologies, coupled with accurate theoretical modeling, are required to adequately tackle these complex measurements.

19.1 **The Importance of Adhesion at the Nanoscale** 605
19.2 **Techniques for Measuring Adhesion** 606
19.3 **Calibration of Forces, Displacements, and Tips** 610
19.3.1 Force Calibration 610
19.3.2 Probe Tip Characterization 611
19.3.3 Displacement Calibration 612
19.4 **The Effect of Liquid Capillaries on Adhesion** 612
19.4.1 Theoretical Background 612
19.4.2 Experimental and Theoretical Studies of Capillary Formation with Scanning Probes 614
19.4.3 Future Directions 618
19.5 **Self-Assembled Monolayers** 618
19.5.1 Adhesion at SAM Interfaces 618
19.5.2 Chemical Force Microscopy: General Methodology 619
19.5.3 Adhesion at SAM-Modified Surfaces in Liquids 620
19.5.4 Impact of Intra- and Inter-Chain Interactions on Adhesion 621
19.5.5 Adhesion at the Single-Bond Level. 622
19.5.6 Future Directions 623
19.6 **Concluding Remarks** 624
References 624

19.1 The Importance of Adhesion at the Nanoscale

The mutual attraction and bonding of two surfaces, which can occur with or without an intervening medium, is a commonly observed phenomenon with far-reaching manifestations and applications in society. The adherence between a raindrop and a window pane, the climbing of a gecko up a vine, the sticking of multiple adhesive note pads to a professor's wall, the force required to separate hook-and-eye (Velcro™) strips, the building

of a sand castle, and the book page turned by a wetted finger are all examples in which adhesion is important. Within the complexity of these examples and others lies a central theme: The mechanical forces between a pair of materials can be fundamentally affected not just by the macroscopic or microscopic structure of the surfaces, but also by the interatomic and intermolecular forces between them.

Adhesion and intermolecular forces have been studied for many years dating back to ancient times [19.1], and active research continues today of topics as broad as insect and reptile locomotion [19.2], interactions between cells in the body [19.3], and the design of self-healing composites [19.4], to name a few. While adhesion is clearly of interest for a wide range of macroscopic applications, its importance becomes dominant at the micrometer and nanometer scales. This is primarily due to the dramatic increase in the surface-to-volume ratio of materials at these scales, an effect that renders friction and interfacial wear critical phenomena [19.5]. For example, the dominating effect of adhesion at this scale has affected the development of microelectromechanical systems (MEMS), in which interfacial forces can prevent devices from functioning properly, since the small flexible parts often emerge from the fabrication line stuck together. Studies of adhesion in MEMS are ongoing, and commercially deployed MEMS devices rely upon sophisticated surface treatments to reduce and control adhesion [19.6–8].

Detailed control of adhesion at the molecular level will be essential for the proposed design of the smaller nanoelectromechanical systems (NEMS). Much has been written about the possibilities for nanoscale machines, sensors, and actuators, etc. It is crucial to understand that a molecule is *all* surface, and, therefore, molecular and nanoscale devices cannot be properly designed, implemented, or characterized unless an understanding of atomic-scale adhesion is thoroughly presented, particularly if these devices are to involve any moving parts that will come into and out of contact with each other. Studying adhesion at the nanoscale is important for other reasons. The protruding asperities in most MEMS materials often have nanoscale dimensions, and, therefore, a complete understanding of adhesion in MEMS requires the investigation of the adhesive properties of the individual nanoscale asperities. In addition, the experimental study of adhesion at the nanoscale is required for the development of detailed atomic-scale models of adhesion. Such progress requires close collaboration between experiment and theory.

There has been significant progress in the experimental study of adhesion at the nanometer scale using scanning probe methods, but numerous challenges exist. A discussion of solid-solid adhesion without an intervening medium is provided in part B and C of this book. This chapter focuses on how adhesion is affected by the ubiquitous presence of water, and how it can be controlled through the application of self-assembled monolayer (SAM) coatings, again in the presence of a liquid medium. In addition, we discuss specific instrumental challenges that are inherent to adhesion measurements. We do not delve into the realm of atomic-scale modeling of adhesion, nor do we discuss the role of more complex coatings such as polymer brushes and blends. Rather, we will focus on critically evaluating the relevant experimental techniques and critically reviewing recent results from studies of water and SAM films, which are perhaps the two most commonly encountered media in nanoscale adhesion applications.

19.2 Techniques for Measuring Adhesion

The experimental study of adhesion at the nanoscale has experienced two renaisances in this century. The first occurred with the development of the surface forces apparatus (SFA) [19.9, 10], and the second with the later development of the AFM [19.11] and other related scanning force techniques.

SFA experiments have contributed profoundly to our understanding of adhesion. The SFA consists of a pair of atomically smooth surfaces, usually mica sheets, mounted on crossed cylinders that can be pressed together to form a controlled circular contact. The applied load, normal displacement, surface separation, contact area, and shear force (if applied) can all be controlled and/or measured [19.9, 12–14]. The SFA can be operated in air, a controlled environment, or under liquid conditions. The surfaces are often treated to attach molecules whose behavior under confinement can be studied. Alternately, the behavior of a confined fluid layer can be observed. The surface separation can be measured and controlled in the Ångstrom (Å) regime. The lateral resolution is limited to the range of several tens of micrometers. However, the true contact

area between the interfaces can be directly measured, which is a key advantage since it allows an adhesive force to be converted to a force (or energy) per unit area (or per molecule), thus separating geometrical contributions to adhesion from chemical contributions. In this chapter, we will refer to relevant results in the context of the scanning probe experiments. As a very brief summary, some of the most important results pertaining to adhesion include: the observation of capillary effects on adhesion [19.15, 16], the presence of hysteresis in adhesion due to pressure-induced restructuring of the interface [19.17], and numerous studies of how interfacial chemistry (hydrophilicity, surface charge, specific chemical groups, etc.) affect adhesive forces [19.1, 18, 19].

AFM instrumentation is described in part B of this book. Key ways in which the AFM differs from the SFA are: (1) the lateral contact size is nm, not μm, due to the fact that the tip is usually < 100 nm in radius, and the contact area at low applied loads will be a fraction of this radius [19.20]; (2) the force resolution is as good as 10^{-10} N or better, as opposed to $\sim 10^{-6}$ N with the SFA; (3) the contact area is not directly observable, although it may be derived through other means; (4) the actual separation between the tip and sample are not directly observable, which is a key disadvantage; (5) the relative separation between the sample and the *cantilever* (not the tip) is controlled in the 0.1 nm range or better; (6) the measurement bandwidth is typically in the kHz regime, but can extend into the MHz regime depending on the data acquisition technique; (7) the operating environment includes ultrahigh vacuum (UHV) and cryogenic to elevated temperatures; (8) there is virtually no limit to the range of materials that can be probed by the AFM; and (9) half of the interface (the tip) is essentially unknown or uncontrolled, whereas with the SFA, both surfaces are well-defined, which is another key challenge for AFM that has yet to be consistently addressed.

The general setup of the AFM is as follows. A small sharp tip (with a radius typically between 10–100 nm) is attached to the end of a compliant cantilever (see Fig. 19.1). The tip is brought into close proximity with a sample surface in a fashion that is identical to STM tip–sample approach mechanisms. Forces acting between the AFM tip and the sample will result in deflections of the cantilever (Fig. 19.2). The cantilever bends vertically (i. e., toward the sample) in response to attractive and/or repulsive forces acting on the tip. The deflection of the cantilever from its equilibrium position is proportional to the normal load applied to the tip by the cantilever. Lateral forces result in a twisting of the cantilever from

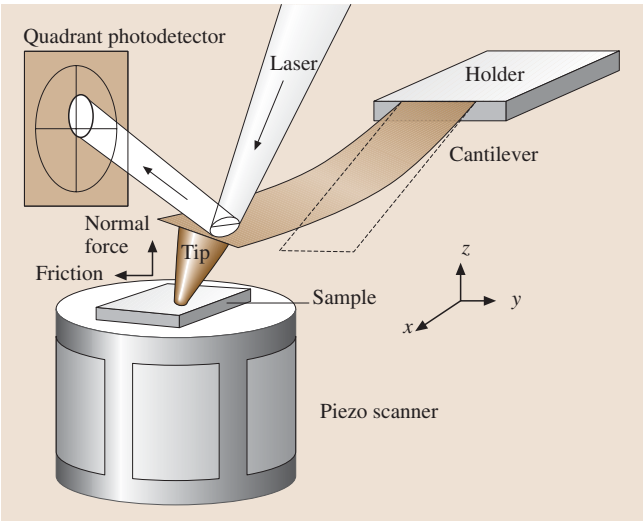


Fig. 19.1 Diagram of the AFM setup for the optical beam deflection method. The tip is in contact with a sample surface. A laser beam is focused on the back of the cantilever and reflects into a four-quadrant photodetector. Normal forces deflect the cantilever up or down, lateral forces twist the cantilever left and right. These deflections are simultaneously and independently measured by monitoring the deflection of the reflected laser beam

its equilibrium position. These measurements can be performed in a variety of environments: ambient air, controlled atmosphere, liquids [19.21], or UHV [19.22–24].

The simplest way (perhaps deceptively so) to measure adhesion with the AFM is through so-called force–displacement plots (also referred to in the literature as force curves and force–distance curves or plots). A force-displacement plot displays the vertical cantilever bending (i. e., the cantilever deflection, which can be calibrated to give the tip–sample interaction force) vs. lever–sample displacement. This displacement is measured between the sample and the rigidly held back end of the cantilever (as opposed to the front end with the tip, which will bend in response to interaction forces). This displacement is altered by varying the vertical position of the piezo tube, which, depending on the type of AFM, displaces the tip or the sample in the direction normal to the sample. Referring to Fig. 19.2a, the stages of acquisition of the force curve are as follows: (a) The lever and sample are initially far apart and no forces act. (b) As the lever is brought close to the sample, the tip senses attractive forces that cause the tip end of the lever to bend downward, thus signifying a negative (attractive) force. These forces may be of electrostatic, van der

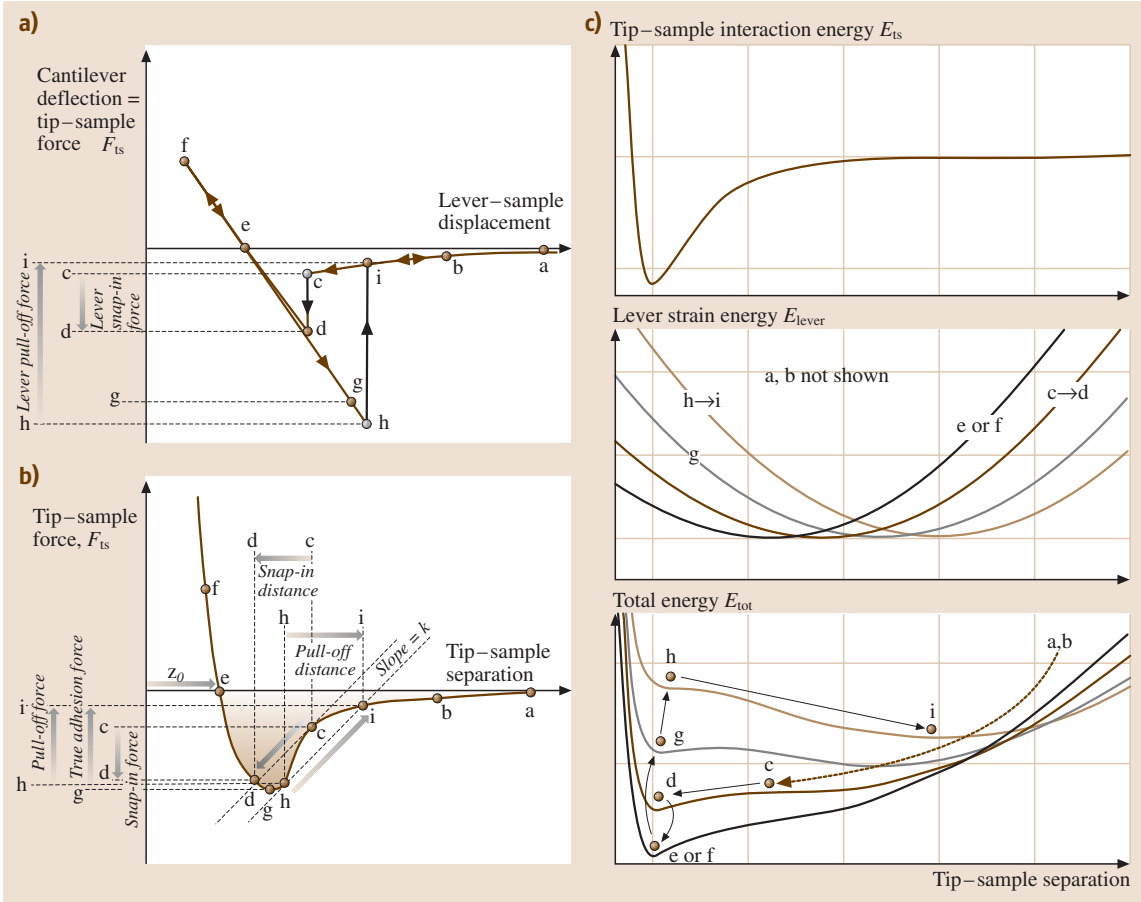


Fig. 19.2 **(a)** A "force-displacement" curve displays the vertical cantilever deflection vs. lever–sample displacement. This displacement is measured between the sample and the rigidly held back end of the cantilever (as opposed to the front end with the tip that will bend in response to interaction forces). **(b)** The true tip–sample interaction as a function of tip–sample separation. **(c)** Tip–sample interaction energy, lever strain energy, and total energy as a function of tip–sample separation

Waals, or other origin. (c) At this point, the attractive force gradient (slope) of the tip-sample interaction force exceeds the spring constant k of the lever, and this instability causes the tip to snap into contact (d) with the sample. The lever-sample displacement can continue to be reduced and eventually crosses the force axis (e), which corresponds to zero externally applied load. As this tip is in repulsive contact with the sample, the front end of the lever is pushed further and further upward, and the force corresponds to the externally applied load. (f) The displacement direction is reversed at a point chosen by the user. As the lever-sample distance is reduced, the force becomes negative. Adhesion between the tip and sample maintains the contact, although there is now a net negative (tensile) load. Eventually the tip passes through the point of maximum adhesion (g), which is not distinguishable on this plot (see Fig. 19.2b). Finally, an unstable point is reached at which the cantilever stiffness exceeds the force gradient of the tip-sample interaction, and the tip snaps out of contact with the sample (h–i). The resulting force relaxation is usually called the pull-off force. Note that the forces and distances are not drawn to scale, and, in particular, the attractive part of the interaction is exaggerated beyond that which often occurs for inert, neutral surfaces.

The force-displacement curve can be measured at a single point, or a series of measurements can be carried out over an area of interest. The so-called adhesion

mapping techniques allow for spatially resolved adhesion measurements to be correlated with other sample properties such as friction, chemical termination, and other types of material heterogeneity.

In order to properly derive adhesion measurements from the AFM, it is critical to understand the nature of the mechanical instability of both the snap-in and pull-off processes. The fundamental point to comprehend is that *the pull-off force is* not *the adhesive bond strength*. In other words, *it is not a direct measure of the actual adhesive forces that were acting between the tip and sample in the absence of applied load*. This important point is often overlooked or misunderstood, and so we discuss it here in some detail.

A cartoon of the interaction force (F_{ts}) between the tip and sample is sketched (not to any particular scale) as a function of the true tip-sample separation in Fig. 19.2b. Superimposed on this is the force-distance relation of the cantilever: a straight line with slope k (N/m), the cantilever stiffness (diagonal dashed lines). Points of instability are shown as gray dots, labels, and arrows, as opposed to stable points, shown as black dots, labels, and arrows. Snap-in occurs when the attractive force gradient $\mathrm{d}F_{ts}/\mathrm{d}z$ just exceeds k, in other words, when the dashed line is tangent to the tip-sample force curve, as shown for point c. This instability is a direct consequence of Newton's 2nd law and is explained further below. Similarly, the pull-off occurs during retraction when k finally just exceeds $\mathrm{d}F_{ts}/\mathrm{d}z$, as shown for point h, where once again, the diagonal line is just tangent to the tip-sample force curve. By definition, since k is finite, this point *cannot* correspond to the point of maximum attractive force (or adhesive force), g. Thus, the force at which pull-off occurs is *not* precisely equal to the true adhesion force. The significance of this deviation is discussed below.

This can also be illustrated from an energetic perspective. In Fig. 19.2c, we show separately the tip-sample interaction energy, E_{ts}, as a function of tip-sample separation (top), followed by the quadratic elastic strain energy, E_{lever}, of the cantilever (middle) and the sum, E_{total}, of the two (bottom). As the lever-sample displacement is varied, the elastic energy curve is shifted to the appropriate position, which, for this example, is represented on the tip-sample separation axis. Far away from the sample, the system resides in a deep minimum, points a and b (not shown). As the tip-sample separation is reduced, this minimum becomes more and more shallow (evolving along the dashed arrow). At the snap-in point c, the minimum is eliminated, since the attractive energy of the tip-sample interaction has overwhelmed the energy minimum of the cantilever. Mathematically this is described as an inflection point, where:

$$\frac{\mathrm{d}^2 E_{tot}}{\mathrm{d}z^2} = 0 \,.$$

Now,

$$E_{tot} = E_{ts} + E_{lever} = E_{ts} + \tfrac{1}{2}k(z - z_0)^2 \qquad (19.1)$$

$\therefore$ snap-in occurs when

$$\frac{\mathrm{d}^2 E_{ts}}{\mathrm{d}z^2} = -\frac{\mathrm{d}F_{ts}}{\mathrm{d}z} = -k \,.$$

This explains why the instability is described by the line of slope k being tangent to the tip-sample force curve, as shown in Fig. 19.2b. The system now rapidly finds the new minimum by following the gray arrow to point d, where a mechanical contact between the tip and sample is now formed.

Note that once the contact is formed, the tip and sample elastically deform, and so, strictly speaking, the tip-sample potential will be distinct from the potential shown before contact occurs. For simplicity, we have left out this change and instead have drawn the potential as a single-valued function. However, such changes in the energy landscape should be considered in a complete description of the problem [19.25].

Nevertheless, it will also hold true that upon retracting the tip from the surface, the stable minimum seen at e, f, and then g becomes more shallow, and eventually disappears as it is overwhelmed by the strain energy of the cantilever. Again, an inflection point is created when

$$\frac{\mathrm{d}^2 E_{attr}}{\mathrm{d}z^2} = -k \,, \qquad (19.2)$$

in other words, when again

$$\frac{\mathrm{d}F_{ts}}{\mathrm{d}z} = k \,. \qquad (19.3)$$

This second instability occurs at point h, the pull-off point. The system then follows the gray arrow to point i, whereby the tip is now out of contact with the sample.

If the spring is sufficiently compliant (low k), or the potential sufficiently stiff (large $\mathrm{d}F/\mathrm{d}z$ shortly past the minimum at g), then the pull-off force *does* nearly correspond to the attractive force minimum F_{min}. In other words, point h would occur very close to point g. But if the cantilever is somewhat stiff, or if the potential is rather compliant (which may be the case for organic, polymeric, biological, or liquid systems), then the pull-off force may differ substantially from F_{min}. For such

cases, this distinction between pull-off force and true adhesion is important, and the limitations imposed by the AFM's intrinsic instabilities become apparent. Of course, if a cantilever is used whose stiffness exceeds the attractive force gradient at all points, one will avoid the instabilities. But since it is the deflection of the cantilever that is used to sense the force, one would have to trade force sensitivity for stability, which is usually an unsuitable compromise.

When a compliant spring or holder is used to manipulate a probe, as is the case for AFM, the technique is generally referred to as being "load-controlled", since the load can be prescribed, but the actual displacement of the probe with respect to the sample cannot (as illustrated by the jump in displacement that occurs during snap-in or pull-off).

In contrast, "displacement-controlled" techniques avoid this instability by effectively eliminating the compliance of the spring or holder. This has been carried out for decades in the mechanical testing community, and over the past 10 years, displacement-controlled scanning probe, have been developed. This is accomplished by displacing the tip by applying force to the tip itself. *Pethica* et al. [19.26,27] use a magnetic coating on the tip and external coils to apply forces to the tip. They refer to the instrument as a "force-controlled microscope". *Houston* et al. [19.28,29] control the force electrostatically and refer to the instrument as an "interfacial force microscope" (IFM). *Lieber* et al. use a variation on Pethica's method, whereby a magnetic coil is used to apply a force to the cantilever [19.30], and adapted it to work in solution.

An example of an adhesion measurement with the IFM is shown in [19.31]. which shows that the instrument is able to measure the entire interaction force curve without instabilities. Thus, this provides a *direct* measurement of the minimum interaction force (as well as forces at all other tip-sample separations) and is a more reliable measure of adhesion. The disadvantage with these techniques is primarily one of inconvenience: The probes require extra manufacturing steps and control electronics. However, given the importance of adhesion in nanoscale science and technology, the extra information gained makes these techniques clearly worth the effort.

Pull-off instabilities can occur even in a displacement-controlled experiment. In this case, the instability is an "intrinsic" instability, whereby the adhesive force gradient competes with the stiffness of the *contact itself* [19.32, 33]. Adhesive materials with low stiffness, such as polymers, may show this behavior.

Finally, the pull-off force may show a time dependence that arises from intrinsic viscoelasticity of the tip or sample materials [19.34], or kinetic effects due to adsorption or reaction of materials at the interface [19.35]. These regimes remain relatively unexplored, but certainly, to compare adhesion measurements between labs, the velocity of approach and retraction, as well as the time in contact ought to be reported for any published experimental results.

A quantitative and reliable examination of adhesion, therefore, requires careful consideration of the mechanics of the contact and the cantilever. For an AFM experiment in which instabilities occur, one can conclude that the pull-off force is a good measure of adhesion only if

1. the materials are fully elastic with little or no viscoelastic character,
2. the interface is chemically stable,
3. the cantilever stiffness is sufficiently low compared with the adhesive force gradient, and
4. the contact stiffness is sufficiently high compared with the adhesive force gradient. Otherwise, a more thorough investigation is required.

19.3 Calibration of Forces, Displacements, and Tips

Whether the forces are to be measured with load-controlled or displacement-controlled techniques, measurements cannot be compared between laboratories unless the forces are properly calibrated. Unfortunately, this can be a rather involved task, and adoption of standards has yet to become widespread and robust. Here we provide a summary of the pertinent issues with references to other work for further reading; these issues are also discussed in part B and C of this book.

19.3.1 Force Calibration

Commercially available AFM cantilevers often come in two forms: V-shaped and rectangular. Silicon and silicon nitride are the typical materials, and reflective coatings are often applied to enhance the reflectivity of the laser beam. The normal force constant of a monolithic rectangular cantilever beam is well-defined [19.36], but it relies on knowledge of all lever dimensions and its mod-

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

ulus. For a rectangular beam of length L, width w, thickness t, Young's modulus E, the stiffness k is given by:

$$k = \frac{Ewt^3}{4L^3} \,. \tag{19.4}$$

The cubic dependence on thickness is particularly problematic since in microfabrication processes t is usually determined by an etching process that is not precisely controlled, and is, therefore, difficult and cumbersome to measure experimentally. Variations in E can also occur depending on the type of cantilever, particularly for silicon nitride cantilevers, although the dependence on E is clearly not as critical. If a bulk value for E is used, the thickness can be determined by measuring the resonance frequency of the cantilever [19.37]. However, this method only works for uncoated monolithic cantilevers, since the metal coatings applied to enhance reflectivity, or for other purposes, will also alter substantially the spring constant [19.38]. In addition, formulae for V-shaped levers are substantially more complicated. Uncoated, single-crystal silicon cantilevers are perhaps the only ones for which the force constant can be reliably determined by using (19.4) and the resonance frequency [19.39]. Otherwise, an experimental, preferably in situ calibration method is strongly encouraged.

Experimental methods to calibrate the force constant of AFM cantilevers have been discussed extensively [19.37, 40–50]. It has been shown repeatedly that the manufacturer's quoted spring constants can be in error by large factors and should simply not be used in any quantitative research effort. We will not delve into the details of these calibration methods, but we do make note of one particularly recent method proposed by *Sader* et al. [19.51] that appears to be reliable and simple to carry out for rectangular levers. It relies on measuring the resonance frequency and the quality factor of the cantilever in air. Use of the hydrodynamic function relates the damping of air to the quality factor, and the dependence of E and t is eliminated from the resulting formula for the force constant.

19.3.2 Probe Tip Characterization

A problem of quite a different nature is that the geometry of the contact formed between the AFM tip and sample surface is not defined if the tip shape and composition is not known. This issue is of crucial importance since one is trying to understand the properties of an interface, and the tip is half of that interface.

The adhesion force between the tip and sample is a meaningless quantity in the absence of any knowledge of the tip shape [19.20]. The only use for such measurements is in cases where the *same* tip is used to compare different samples or different conditions, and verification by cyclic repetition of the experiments is carried out to ensure that the tip itself did not change during the experiment.

There are several in situ methods to characterize the tip shape. A topographic AFM image is actually a convolution of the tip and sample geometry [19.52]. Separation of the tip and sample contributions by contact imaging of known, or at least sharp, sample features allows determination of the tip shape to a significant extent [19.53–64]. Ex-situ tip imaging by transmission electron microscopy has also been performed [19.61, 65]. Some of these measurements have revealed that a *majority* of microfabricated cantilevers possess double tips and other unsuitable tip structures [19.56, 60, 64]. This convincingly demonstrates that tip characterization is absolutely necessary for useful nanotribological measurements with AFM. Thin film coatings applied to microfabricated levers can provide robust, smooth, and even conductive coatings [19.65–67]. Further work in this direction would be useful so as to provide a wider array of dependable tip structures and materials.

In addition to the shape of the tip, its chemical composition is equally important, but is also challenging to determine or control. *Xiao* et al. have shown that the AFM tip is readily chemically modified when scanned in contact with various materials [19.68], even tips that have been coated with self-assembled monolayers in order to control their chemistry. They recommend "running in" the tip with a standard sample to give reproducible results. The stresses that take place in a nano-contact can be very large [19.5], and so modification of both the chemistry and structure of the tip is important to consider.

One class of experiments for which the tip shape and chemistry is not as critical is when a molecule or nanostructure is tethered to the end of the tip and specific interactions are probed [19.69–73]. Another alternative method is to use colloid probes [19.74, 75], whereby colloidal particles are attached to the cantilever on top of (or in place of) the tip. This method requires a unique calibration procedure [19.76], but provides particles whose structure and chemistry can be measured and perhaps controlled prior to attachment to the lever.

19.3.3 Displacement Calibration

Proper signal and spatial calibration also requires knowing the sensitivity of the piezoelectric scanning elements. This can involve complications due to instrumental drift [19.77] and inherent piezoelectric effects, namely nonlinearity, hysteresis, creep, and variations of sensitivity with applied voltage [19.78–82]. Caution must be exercised when determining and relying upon these parameters. Techniques such as laser interferometry [19.83], scanning sloped samples [19.45, 78], scanning known surface step heights [19.84], or the use of pre-calibrated piezoelectrics [19.85] can facilitate piezo calibration.

19.4 The Effect of Liquid Capillaries on Adhesion

The adsorption of water and other liquids onto surfaces, and their subsequent behavior at interfaces, continue to be a vibrant area of research. The importance of liquid-solid interface behavior is massive, encompassing topics as broad as paints, textiles, lubricants, geology, and environmental chemistry, and covering all corners of biology.

19.4.1 Theoretical Background

The ability to measure forces at the nanoscale using scanning probe microscopy has generated much interest in these fields. For any force measurement carried out in ambient laboratory conditions, one must consider the possibility of a capillary neck forming between the tip and sample. The study of such necks may, in turn, provide insight into the behavior of the liquid, which is discussed in other chapters in this book. In order to provide background for these emerging areas, we consider here the fundamental mechanical and chemical aspects of adhesion in the presence of a liquid meniscus. By way of introduction, *Israelachvili* [19.17] and *de Gennes* [19.87] provide rigorous coverage of the terminology, physics, and chemistry of liquid films and their wetting properties.

Water readily adsorbs at many surfaces. At a crack or sharp corner it can condense to form a meniscus if its contact angle is small enough. The small gap between an AFM tip and a surface is, therefore, an ideal occasion for such condensation. Early on, it was realized that liquid condensation plays a significant role in tip-sample interactions [19.88, 89].

The classic theory of capillary condensation starts by considering the thermodynamics of capillary formation. The general geometry of the capillary is shown in Fig. 19.3, adapted from [19.86], with hydrophobic surfaces sketched on the left half ($r_1 > 0$) and hydrophilic surfaces on the right half ($r_1 < 0$). The tip is idealized as a sphere of radius R. The water contact angles with the sample and tip respectively are θ_1 and θ_2. D represents the separation of the tip and sample. The angle ϕ is referred to as the "filling angle". The pressure difference, or Laplace pressure, across a curved interface is given by the Young–Laplace equation [19.17, 86]:

$$\Delta p = \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) , \tag{19.5}$$

where r_1 and r_2 are defined in Fig. 19.3 and γ is the surface tension of the liquid. The resulting force is attractive if $\Delta p < 0$. Note that $r_1 > 0$ and $r_2 < 0$. If the capillary formation is isothermal, then one can derive the Kelvin equation:

$$RT \ln \frac{p}{p_0} = \gamma V \left(\frac{1}{r_1} + \frac{1}{r_2} \right) , \tag{19.6}$$

where p_0 is the saturation pressure of the liquid, V is the molar volume of the liquid, T is the temperature, and R the molar gas constant. The ratio p/p_0 simply corresponds to the relative vapor pressure of the liquid,

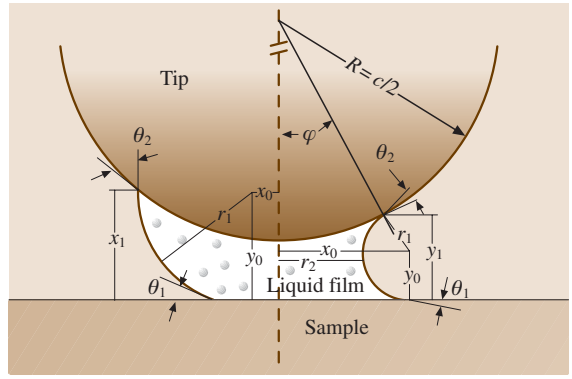


Fig. 19.3 The AFM tip considered as a sphere at a distance D from the sample. The liquid film in between may form a concave (*right*) or convex (*left*) shaped meniscus in the plane shown (adapted after [19.86])

which in the case of water is just the relative humidity (RH). This is often rewritten in terms of the Kelvin radius r_k:

$$r_k^{-1} = \left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{R_M T}{\gamma V} \ln \frac{p}{p_0} \,. \tag{19.7}$$

The Kelvin radius varies logarithmically with the partial pressure of the liquid and is by definition less than zero (negative values correspond to convex curvature). For water at 20 °C, $\gamma V/RT = 0.54$ nm. A graph of r_k vs. p/p_0 is shown in Fig. 19.4. Of particular note is the fact that starting from $p/p_0 = 0.35$ and higher, we find $|r_k| < 2$ nm. The application of this classic theory to capillaries with such small curvature radii is certainly a question to be considered carefully, and we will address it below.

A simple equation for the maximum attractive capillary force F_c between the tip and sample that is commonly used is:

$$F_c = -4\pi\gamma R \cos\theta \,. \tag{19.8}$$

This equation is calculated by considering the Laplace pressure only. The adhesion force is predicted to be independent of RH. Notice that an immediate problem arises, which is that it predicts a finite force even at 0% RH, where there can be no capillary formation. Derivations of (19.8) are presented in several other publications [19.17, 86, 90, 91]. Significantly, this equation and its use with AFM experiments contain several assumptions, *many of which may not be valid for an AFM experiment*:

1. $|r_1| \ll R$, which is equivalent to saying ϕ is small,
2. $|r_1| \ll |r_2|$,
3. $\theta_1 = \theta_2$,
4. $D \ll |r_1|, |r_2|$,
5. the tip is shaped like a perfect sphere,
6. the effect of solid-solid adhesion is negligible with respect to the meniscus force,
7. the meniscus cross sections are perfect, circular arcs,
8. the force from the Laplace pressure dominates the force due to the resolved surface tension of the meniscus,
9. the surface tension γ is independent of the meniscus size,
10. the meniscus volume remains constant as the tip is retracted,
11. the maximum force of attraction is equal to the pull-off force,
12. the tip and sample are perfectly rigid.

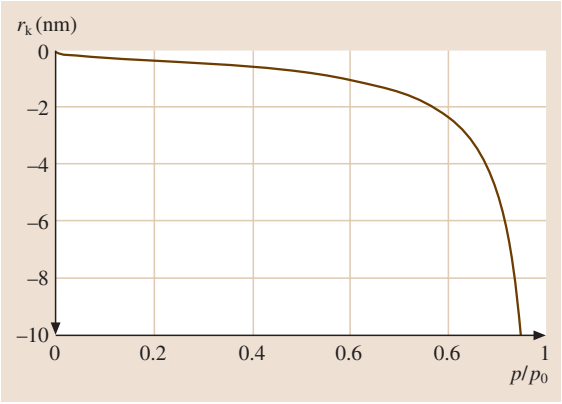


Fig. 19.4 The Kelvin radius of water at 20 °C plotted as a function of the relative humidity

Assumption 1 may not be true, since the tip radius may indeed be small and comparable to the meniscus curvature radii. Assumption 2 may not be true since the small tip geometry may lead the two meniscus curvature radii to be similar. Assumption 3 is by no means true if the tip and sample are made of different materials. Assumption 4 may not be true since both the separation of the tip and sample, as well as the radii r_1 and r_2 may be in the nanometer range. Assumption 5 may be slightly or grossly in error and is a particularly dubious assumption in the absence of tip characterization. In addition, for large menisci, the capillary will grow beyond the end of the tip and start climbing up its shank, which may be pyramidal or conical in shape. Assumption 6 may also be inaccurate if van der Waals or other adhesive forces are significant (this is discussed further in the next section). Assumption 7 is not correct [19.90], and the regimes where it is a reasonable approximation require close scrutiny. Assumption 8 will be inaccurate at high relative vapor pressures [19.17, 91]. The nature of assumptions 1 to 8 and corrections to the theory to account for their violation are presented in the work of *Orr* and *Scriven* [19.90]. This theory remains within the bounds of the classical picture of capillary formation.

Assumption 9 concerns an important scientific question that has not been fully resolved and represents the possible violation of the classical framework of molecular effects at the nanoscale. SFA measurements by *Israelachvili* have indicated that for cyclohexane and other inert organic liquids, γ remains nearly equal to its macroscopic value even for Kelvin radii that, remarkably, correspond to one or two molecular diameters. However, for water, the adhesion force comes to within

10% of the bulk prediction only when $p/p_0 > 0.9$, which corresponds to a Kelvin radius greater than ~ 5 nm in magnitude [19.16]. This deviation from macroscopic thermodynamic predictions alone calls the use of (19.8) into serious question for AFM measurements. Rather, the exploration of this deviation at the molecular scale presents a unique opportunity for scanning probe measurements.

Assumptions 10 and 11 are not assumptions of (19.8) itself, but rather are assumptions that are often used when applying (19.8) to AFM measurements. Equation (19.8) simply gives the maximum force of attraction between the tip and sample. As discussed above, an AFM does *not* measure this quantity. Rather, it measures the force at which an instability occurs. If a capillary has formed between the tip and sample, then the force as a function of distance can be calculated. Calculating this force requires making one of two assumptions: Either the volume of the capillary is conserved (i. e., the rate of displacement is large with respect to the adsorption or desorption kinetics of the liquid), or the Kelvin radius is conserved (i. e., the rate of displacement is slow with respect to the adsorption or desorption kinetics of the liquid, and so the capillary remains in equilibrium). The constant volume assumption 10 has been used in every paper we have reviewed. *Israelachvili*, however, pointed out the difference between these two approaches in his book [19.17] and left the solution of the problem as an exercise to the reader! As with the problem of scale-dependent surface tension mentioned above, the kinetics of capillary formation and dissolution is a relatively unexplored problem and is therefore worthy of further investigation. A recent study of the humidity dependence of friction as a function of sliding speed is an example where this issue is raised [19.92].

Once an assumption about how the meniscus changes with displacement has been made, one still needs to consider the nature of the instability in order to relate the AFM pull-off measurement to the capillary's properties. As stated above and shown in Fig. 19.2, a low lever stiffness k or a strongly varying adhesive force will lead to a pull-off force that is nearly equal to the adhesive force, and so assumption 11 would be valid. However, if k is sufficiently large, or the capillary stiffness sufficiently weak, this assumption will fail. As we shall see below, experimental efforts to investigate this point, and theoretical work to predict it, have yet to be carried out.

Finally, assumption 12, if violated, requires a substantially more complex analysis. The question has been addressed independently by *Maugis* [19.93] and *Fogden* and *White* [19.94]. Both papers provide a nondimensional parameter that allows one to determine the severity of the effect. In the limit of small tips, stiff materials, large (in magnitude) Kelvin radii, and low surface tensions, the effect of elastic deformation is negligible. However, for relatively compliant materials, large tips, and small Kelvin radii, the meniscus can appreciably deform the contact in the immediate vicinity of the meniscus. This can substantially alter the mechanics of adhesion, as well as significantly affect the stresses. The dependence on the Kelvin radius is particularly critical. This effect may be of particular concern with soft materials like polymers or biological specimens. According to *Maugis*, the problem is analogous to the adhesive contact problem for solids, discussed by *Johnson* et al. [19.95] and studied further in many papers since [19.25, 96, 97].

19.4.2 Experimental and Theoretical Studies of Capillary Formation with Scanning Probes

There have been several experimental and theoretical investigations of how pull-off forces are affected by liquid capillaries. These studies have mostly focused on the effect of relative humidity. We do not present an exhaustive review here, but rather summarize a few key results that highlight the important trends observed and the outstanding questions that remain.

Early on, it was realized that water capillary formation occurred readily if at least one of the two surfaces in contact was hydrophilic. Higher adhesion will lead to higher contact forces and therefore larger elastic contact areas, and this can degrade the lateral spatial resolution of the AFM, as observed by *Thundat* et al. [19.98]. Furthermore, several observations have confirmed the expected result that capillary formation was readily prevented for hydrophobic surfaces. For example, *Binggeli* and *Mate* [19.88, 99] showed that tungsten tips in contact with clean silicon wafers with a hydrophilic native oxide exhibited strong adhesion and long pull-off lengths, whereas surfaces treated with a hydrophobic perfluoropolyether (such as Z-dol) showed pull-off forces reduced by a factor of 2 to 3. These results were confirmed by *Bhushan* and *Sundararajan* [19.100] who investigated the pull-off force between silicon nitride tips and Si(100) with and without Z-dol coatings for a range of relative humidities. As another example, MoO_3 films, likely hydrophilic, showed a sixfold increase in pull-off force measured with silicon tips from 0 to 50% RH [19.101].

While these examples illustrate some basic trends, the need for more finely resolved measurements as

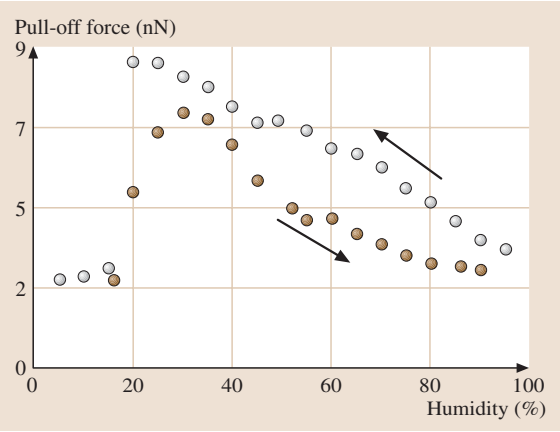


Fig. 19.5 Pull-off force between a silicon nitride tip and the muscovite mica surface as a function of RH. Two sets of data are shown: increasing (*grey circles*) and decreasing (*brown circles*) humidity (after [19.35])

a function of humidity has been addressed only quite recently. For example, *Xu* et al. measured adhesion between a silicon nitride tip and (hydrophilic) muscovite mica in ∼5% RH increments. The result is shown in Fig. 19.5. There are three identifiable regions: constant adhesion at low RH (< 20%), increasing adhesion, and then decreasing adhesion. Some hysteresis is seen between experiments conducted with increasing and decreasing RH, but the overall trend is preserved. *Xu* et al. correlated their measurements with detailed studies of the growth of molecular water films on the mica surface, which they could image directly using scanning polarization force microscopy (Fig. 19.6). They proposed that below 20%, capillary condensation does not occur, and, indeed, they saw no evidence of a water film at these low humidities. Above 20% RH, a strongly bound molecular water layer was formed on the bare mica surface (Fig. 19.6). In the presence of the tip, a capillary meniscus condensed. Above 40% RH, the pull-off force decreased. Recognizing one of the limitations of (19.8), *Xu* et al. attribute this to the violation of assumption 5, listed above. They argue that for a pyramidal AFM tip, r_1 and r_2 become comparable in magnitude (and remain opposite in sign) once the capillary reaches the shank of the tip, leading to a near cancellation of the Laplace pressure given in (19.5). This argument is certainly plausible, although a rigorous proof is not provided, and the other limitations of (19.8) are not discussed in relation to this issue. Nevertheless, the correlation between the onset of adhesion increase and the formation of the molecu-

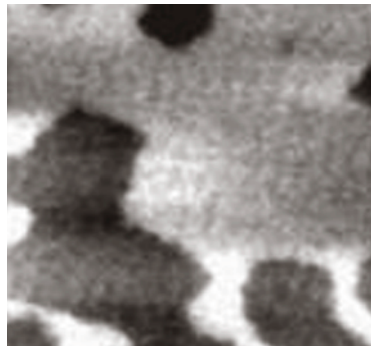


Fig. 19.6 Scanning polarization for microscopy image of water structures on the muscovite mica surface. A degree of polygonal shape to the boundaries is seen. The signal represents two distinct phases of water that are present as a molecular film below 45% RH (after [19.35])

lar water film, as seen directly in their dramatic images, is an extremely convincing case where the classical assumptions of (19.8) must be modified to account for the molecular structure of the water film.

Further considerations of the limitations to the (19.8) were measured, discussed, and modeled in a detailed paper by *Xiao* and *Qian* [19.91]. Adhesion measurements, collected in large numbers for good statistics, were carried out with the same silicon nitride tip on two different surfaces: hydrophilic SiO_2 and a hydrophobic layer of *n*-octadecyltrimethoxysilane (OTE) on SiO_2. Contact angle measurements to confirm the assertion of hydrophilicity were not presented, however, the OTE surface was confirmed to be hydrophobic with a water contact angle of 108°. The results are shown in Fig. 19.7. The hydrophobic surface shows no dependence on RH, whereas the hydrophilic bare SiO_2 surface

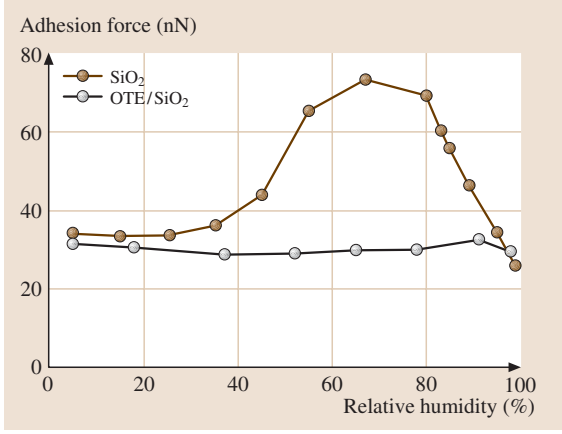


Fig. 19.7 Pull-off force as a function of RH for adhesion between a silicon nitride tip and SiO_2 (*brown circles*) and OTE/SiO_2 (*grey circles*) samples (after [19.91])

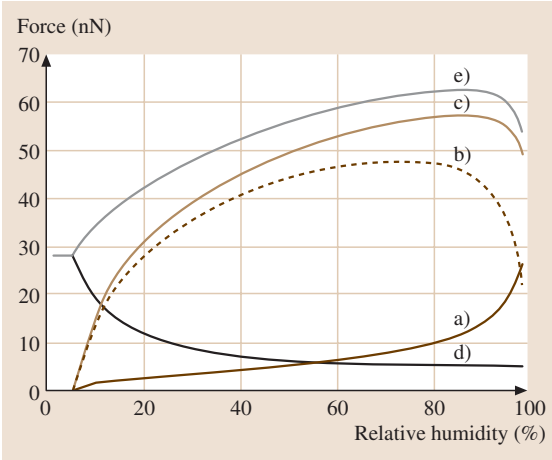


Fig. 19.8 Contributions to the adhesion force as a function of humidity: a) resolved surface tension force; b) Laplace pressure force; c) total capillary force (Laplace + surface tension); d) van der Waals force; e) total adhesion force. (After [19.91])

shows three regimes similar to the result of *Xu* et al.: constant pull-off force (>∼ 30% RH), increasing pull-off force (30%–70% RH), and then decreasing pull-off force. *Xiao* and *Qian* discuss the limitations of (19.8) in substantial detail. In particular, they consider the violation of assumptions 1 through 8 listed above. This includes a treatment of van der Waals adhesion, taking into consideration the effect of electrostatic screening of this force by the water itself. Equations for this more general case are presented. With all these aspects taken into account, they are only able to fit the model to their data qualitatively (Fig. 19.8). At low RH (< 10%), the van der Waals force dominates and the adhesion is initially constant. At intermediate values, the Laplace pressure contribution increases and then begins to saturate. The contribution from the resolved surface tension becomes significant at high RH (above ∼ 80%), and the contribution from the Laplace pressure begins to drop strongly around this same point. Qualitatively, this reproduces their results (and the aforementioned ones) by producing regimes of constant, increasing, and then decreasing adhesion. However, the humidities at which the transitions occur, and the relative changes in adhesion, do not match the data. Somewhat better agreement at high RH (> 70%) is found by considering alternate (blunt) tip shapes. Nevertheless, the most significant discrepancy occurs for low RH, where the extent of the constant adhesion force is underestimated by the classical theory. The authors attribute this to the failure of the classical continuum theory to properly describe the properties of a molecular-scale meniscus. Another interesting point of this study is that at low RH, the adhesion is very similar for both samples, a fact that the authors attribute to the dominance of the van der Waals force for both samples (which is largely determined by the substrate and not affected significantly by the OTE film).

Slightly more recently, *He* et al. have studied capillary forces for a variety of tip-sample pairs [19.102]. Hydrophilic tips (silicon and silicon nitride with no surface treatment) and hydrophobic tips (coated with *n*-octadecyltricholorsilane) were used. The hydrophobic character of the tip was asserted based on a water contact angle measurement of 105.5°, presumably taken on

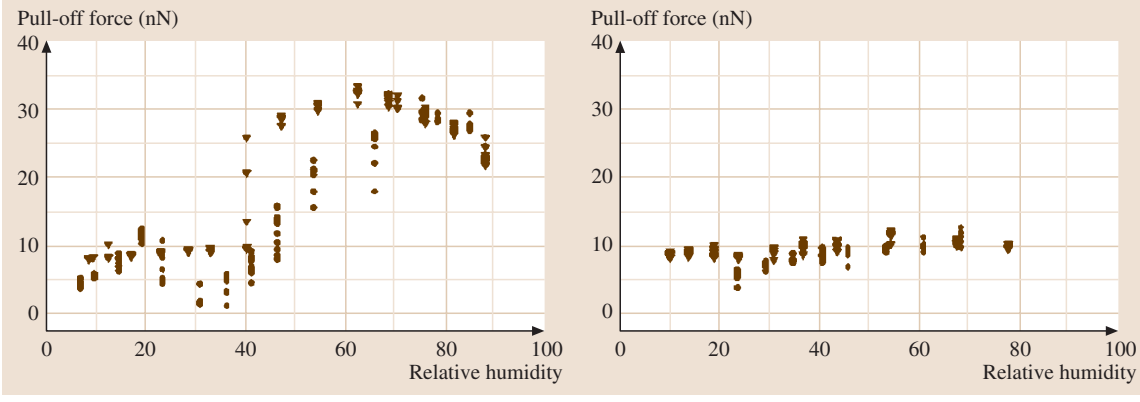


Fig. 19.9 **(a)** Pull-off force vs. RH measured between a hydrophilic tip and a flat silicon sample. *Circles*: measured when increasing RH. *Triangles*: measured when decreasing RH. **(b)** Pull-off force vs. RH measured between a sharp SFM tip coated with OTS and a flat silicon sample. The pull-off force is independent of humidity. (After [19.102])

a different region of the cantilever chip. Solvent-cleaned silicon samples and calcium fluoride films were used as hydrophilic substrates. As with the measurements of Xiao and Qian, contact angle measurements for the hydrophilic samples and tips were not presented. Results for hydrophobic and hydrophilic tips are shown in Fig. 19.9.

These results, which were carried out independently and without knowledge of Xiao and Qian's work, show impressive agreement. This is particularly interesting given that here the tip was varied from hydrophilic to hydrophobic (while the sample stayed hydrophilic), whereas in Xiao and Qian's work, the tip presumably remained hydrophilic and the sample was varied from hydrophilic to hydrophobic. For the hydrophobic tip, the pull-off force remains constant, indicating once again that capillary formation was suppressed. However, for the hydrophilic tip, three regimes of adhesion are found (Fig. 19.9). Similar results are found when a hydrophilic glass microsphere is used as a tip. The authors refer to these three regimes as the van der Waals regime, the mixed van der Waals–capillary regime, and the capillary regime. In agreement with Xiao and Qian's assessment, the authors propose that at low RH ($<\sim 35\%$ in this case), the formation of the water meniscus is suppressed and the adhesion is dominated by solid-solid (presumably van der Waals) interactions. They propose, based on the work of *de Gennes* on the theory of spreading [19.87], that a minimum precursor film thickness is required to form the meniscus. The authors also present a calculation of the adhesion force when assumptions 1 and 2 are relaxed.

Note that the results discussed above consistently indicate that only one hydrophilic surface is required to form the water film. This means that should meniscus formation be undesirable in a nanoscale application (as is the case for MEMS), then hydrophobic treatment of *both* contacting surfaces is likely required.

Practical guidance for discerning the dependence of capillary and van der Waals force is provided by *Stifter* et al. [19.86], who have produced a detailed theoretical model of the meniscus as a function of contact angles, tip size, relative humidity, separation, and surface tension (to represent liquids other than water). An example is shown in Fig. 19.11. The model relaxes assumptions

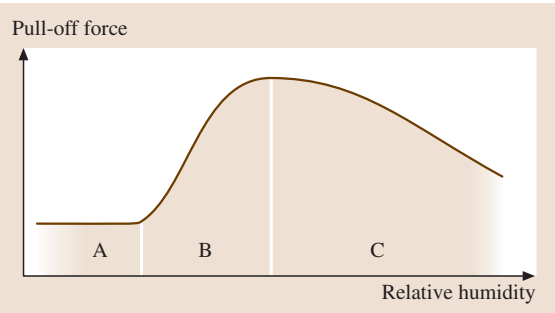


Fig. 19.10 Figure illustrating the distinct regimes of the pull-off force as a function of RH. Regimes A, B, and C are referred to as the van der Waals regime, mixed van der Waals–capillary regime, and capillary regime, respectively. (After [19.102])

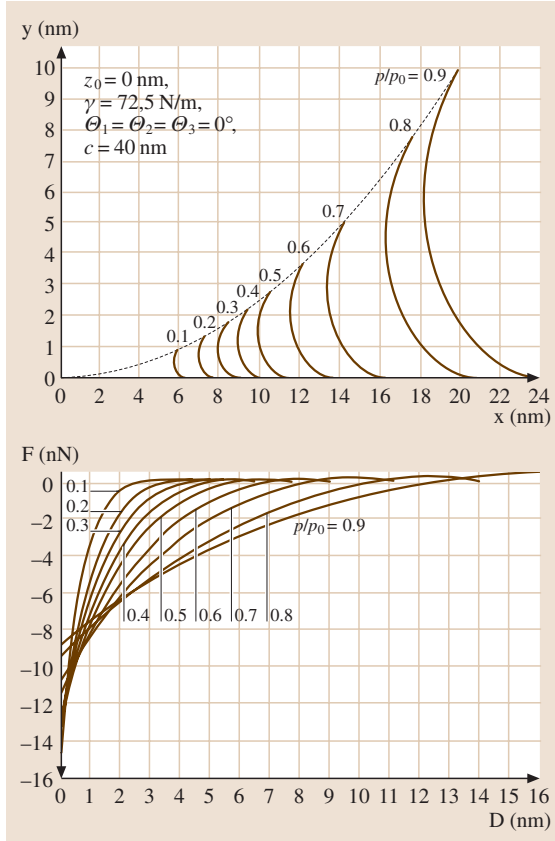


Fig. 19.11 *Top*: geometric shape of the meniscus for a range of RH values. *Bottom*: the meniscus force between the tip and the surface due to the Laplace pressure. The *dashed line* represents a cantilever with force constant 1 N/m. This relatively stiff value is still less stiff than the capillary force curve and suggests that the pull-off force would be very close to the true adhesion force. Of course, summation of the van der Waals and other contributions are necessary to make a full comparison of this sort. (After [19.86])

1 through 4 and 6. These studies show regimes where either van der Waals force or Laplace pressure forces dominate and provide guidance for actual values of the tip geometry and contact angles that minimize the capillary forces. It is not clear how much of an effect the resolved surface tension would have, as that is not included in this particular calculation. Another practical result is that for most of the plots of force vs. displacement, the "stiffness" (gradient) of the curve rarely drops below 1 N/m. This indicates that, within the assumptions of this paper, using cantilevers with force constants of 0.1 N/m will give pull-off forces that are reasonably accurate measurements of the adhesion force. This is consistent with the results of *He* et al., who found no dependence of their measurements when using cantilevers whose nominal force constants ranged from 0.032–0.5 N/m.

19.4.3 Future Directions

The results and modeling so far indicate the possibility of two trends. Hydrophobic surfaces will exhibit little dependence of adhesion on RH, whereas if one surface is hydrophilic, three regimes of behavior occur: constant solid-solid adhesion at low RH, increasing adhesion at intermediate RH, and decreasing adhesion at the highest RH. Some insight into the physical mechanisms behind these regimes has been presented, but it would be desirable to pursue further work in this area.

Perhaps most critically needed is an atomic-scale picture of the menisci and water films present under low partial pressures. This could address the question of why adhesion is initially independent of RH, and when the meniscus itself would start to form. Further theoretical developments that address the assumptions laid out above would help to clarify the picture. It is also important to extend these studies beyond simply the case of water, as the properties of other liquids are also of great interest and could be compared to previous experiments with the SFA and other tools.

Studies that clarify the kinetics of meniscus formation are also needed. There should be a noticeable transition in behavior once the rate of displacement becomes comparable to the appropriate kinetic rates, and this could provide valuable information about these kinetic processes at the nanometer scale.

Finally, as mentioned before, there continues to be a gap in reproducibility and comparability between laboratories that will only be bridged when standard techniques for tip characterization and force and displacement calibration are addressed. Efforts that take these considerations into account are worthy of support.

19.5 Self-Assembled Monolayers

When thinking about adhesion and the related phenomenon of friction, it is important to realize that the interfaces of real surfaces in contact are rarely atomically smooth. Surfaces that appear smooth on the macroscopic scale, upon closer inspection are found to consist of nanometer-scaled asperities (typically on the order of 10 nanometers) whose intentional or accidental interactions ultimately control adhesion, friction, and wear at contacts [19.6, 7, 103–105]. The size of these asperities becomes particularly important when one considers that the true contact area between interfaces for the load distribution is localized through these asperity-asperity interactions in wich extremely high pressures can be produced, sharply increasing local stress fields that can cause materials to yield and shear as they encounter each other during sliding and intermittent contact. In addition to load distribution at nanoscale asperity-asperity contacts, their size will influence surface wetting and adhesion due to capillary forces localized at the contacts [19.7, 104, 105]. The structure of applied lubricant films at such asperities will be highly dependent upon asperity curvature, and defects in lubricant film structure may form more readily here than on atomically flat surfaces.

19.5.1 Adhesion at SAM Interfaces

The minimization of adhesion at such asperity-asperity contacts is a critical issue in MEMS devices [19.6, 7, 104, 105]. In fact, the intentional introduction of surface roughness (on the order of 10 nm RMS) can be employed to lead to reduced stiction during post–processing feature release. These same asperities, however, must later resist wear during controlled or accidental contact during device operation. Thus, the specific details of adhesion and energy dissipation at such asperity-asperity contacts are required for the rational design of such systems.

To function as a protective lubricant layer in such systems, self-assembled monolayers (SAMs) of alkylsilane and fluorosilane compounds with chain lengths

ranging from C_{10} to C_{18} have been useful in the reduction of friction and adhesion in MEMS [19.7, 104–106]. Such direct applications of SAMs as lubricant films, combined with the ease of sample preparation and the ability to generate model surfaces with well-defined film chemistry and structure, have made nanotribological and adhesion studies of SAMs a rich area of research [19.5]. Many of these nanotribological studies have used AFM to examine either SAMs of alkylsilane films on atomically smooth Si wafers, glass, or mica surfaces. Alternatively, many researchers have examined alkanethiol films on atomically smooth Au(111) surfaces. Using this approach, many molecular level details, such as the influence of film chemistry and molecular organization on friction, and the adhesion and wear of SAMs, can be obtained. Developing a clear understanding of the details of adhesion at SAM-modified surfaces allows for the complex links between surface chemistry, adhesion, and friction to be understood at the molecular level. In this section, we overview AFM studies of adhesion on SAM-modified surfaces.

19.5.2 Chemical Force Microscopy: General Methodology

In order to probe adhesion between chemically modified surfaces using AFM, the probe tips and sample surfaces are typically modified via self-assembly of monolayers using organosilanes on such surfaces as mica, glass, or oxidized Si or Si_3N_4 (the latter two being the typical materials of which AFM tips are made), or formed from thiols on Au-coated AFM tips and surfaces (Fig. 19.12) [19.107, 108]. While this has been shown to be a facile method for the modification of AFM tips for chemical force measurements, it should be noted that the details of the packing densities of the monolayers formed on the AFM tips are generally not known. This lack of detail regarding molecular overlayers on AFM probe tips can be a problem that requires careful consideration when using such chemically modified tips for the quantitative determination of adhesion forces and molecular interactions, as the number density of species in the tip sample contact is related to the measured adhesion.

Details of the environment in which the adhesion measurements are carried out are also an important consideration. Under ambient environmental lab conditions, surfaces are contaminated with organic compounds from the air, as well as with a layer of condensed water vapor, which varies with humidity. The condensed water layer can form a contact meniscus between the tip and sample, introducing a capillary force into the measured adhesion [19.109]. The presence of this capillary force can overwhelm the details of the adhesion from the SAM-terminated surfaces to be probed. To eliminate capillary forces, many studies are performed under liquid environments or ultrahigh vacuum. Under liquids, the nature of the solvent will of course impact the measured adhesion for a given pair of interacting surfaces, as solvent exclusion plays an important role. Also, in the case of water, the pH and ionic strength of the water environment can influence the measured adhesion in the presence of any surface-bound charges.

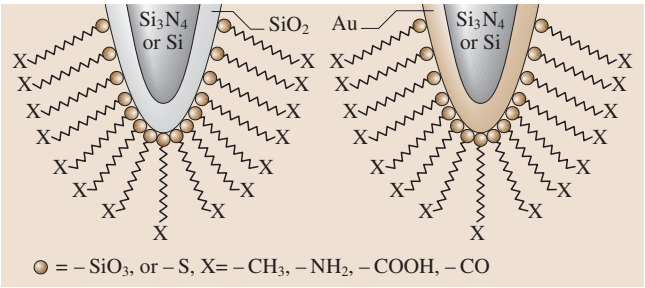


Fig. 19.12 The modification of AFM tips for chemical force microscopy is frequently carried out by chemical functionalization by alkylsilane monolayers on the oxidized surfaces of Si- or Si_3N_4-based tips, or by alkylthiol monolayers assembled on Au-coated (~ 50 nm) AFM tips. The Au tips also typically have a Cr binding layer (~ 5 nm) placed on before Au coating

In quantifying adhesion energies from AFM measurements, the contact mechanics model developed by Johnson, Kendall, and Roberts (JKR model) [19.20, 95] is often employed in the analysis of the adhesion data acquired by force-distance spectroscopy, whereby the number of interacting species (and consequently the average "unit" interaction force or energy) can be derived from the estimated contact area and the average molecular packing density. Using the JKR model, the force of adhesion (AFM pull-off force) is related to the work of adhesion, W_{adh}, and the reduced radius, R, of the tip-surface contact:

$$F_{adh} = -\frac{3}{2}\pi R W_{adh} \,. \tag{19.9}$$

The work of adhesion is a combination of the tip-surface (γ_{ts}), tip-solvent (γ_{tl}) and surface-solvent (γ_{sl}) interfacial energies ($W_{adh} = \gamma_{sl} + \gamma_{tl} - \gamma_{ts}$), and for tip-surface combinations that have the same chemical composition, the surface energy may be estimated directly from the adhesion measurement as W_{adh}. The effective contact radius at separation, r_s, from the JKR

model is given as:

$$r_s = \left(\frac{3\pi W_{adh} R^2}{2K}\right)^{1/3}, \quad (19.10)$$

where K is the reduced elastic moduli of the tip and surface. Using the contact area at separation and the assumed packing density of the molecules at the surfaces in contact, an estimate of the adhesion force or interaction energy on a per molecule basis can be made.

The accuracy of the interfacial energies and per molecule values obtained using this approach, however, must be carefully considered, due to the accumulation of error carried by the imprecise knowledge of the contact, including the tip radius, molecular packing densities of the modified surfaces, as well as the associated elastic properties of the contact at the monolayer level. As the details of the elastic properties of self-assembled monolayers are generally not known, the elastic properties of the contacts are typically assumed to be dominated by those of the underlying substrate, and the bulk values of the surface and/or tip materials (Au, Si, mica, Si_3N_4, SiO_2) are often employed in these calculations. Moreover, as mentioned above, if the molecular packing densities of the monolayers being evaluated are not known (as is the case with a typical AFM tip), then estimations must be used. For contact areas at pull-off approaching 1 nm^2, if the error in packing density is as much as 1 molecule per nm^2, this can lead to an error as high as 25% for a typical alkanethiol, or more in the reported per molecule adhesion force. Such details should be taken into consideration when describing quantitative measurements.

An alternative approach to the measure of adhesive interactions based on Poisson statistics has been promoted by *Beebe* et al. for the statistical evaluation of single bond forces without a priori knowledge of the tip-surface contact details involved [19.119, 121, 122]. A principal limitation of this approach, however, is that a completely homogeneous chemical system is assumed, i. e., that there is only one type of discrete interaction present that gives rise to the observed adhesion. Unfortunately, for many solution phase systems a number of different interactions are typically operative, including energetic exchange with and reorganization of solvent molecules, depending on the solution conditions. These issues have never been thoroughly addressed in any molecular-level measurement of adhesion.

Table 19.1 List of various interactions evaluated by atomic force microscopy adhesion measurements

Chemical Contacts	References
$-CH_3/-CH_3$	[19.110–119]
$-CH_3/-COOH$	[19.110, 114, 116, 118]
$-CH_3/-CONH_2$	[19.118]
$-CH_3/-NH_2$	[19.118]
$-CH_3/-OH$	[19.118]
$-CF_3/-CF_3$	[19.113]
$-CF_3/-CH_3$	[19.113]
$-OH/-OH$	[19.30, 115, 117–121]
$-OH/NH_2$	[19.118]
$-OH/-COOH$	[19.117, 118]
$-OH/-CONH_2$	[19.118]
$-COOH/-COOH$	[19.30, 111, 114, 115, 117, 118, 122]
$-CONH_2/-CONH_2$	[19.118]
$-CONH_2/-COOH$	[19.118]
$-NH_2/-NH_2$	[19.116–118, 123]
$-CH_3/-NH_2$	[19.118, 123]
$-NH_2/-COOH$	[19.118]
$-NH_2/-CONH_2$	[19.118]
$-SO_3H/-SO_3H$	[19.123]
$-CH_3/-SO_3H$	[19.123]
$-NH_2/-SO_3H$	[19.123]
$-PO_3H_2/-PO_3H_2$	[19.116]
$-SH/-SH$	[19.119]

19.5.3 Adhesion at SAM-Modified Surfaces in Liquids

A number of researchers have used AFM to probe interfacial adhesion for a variety of different chemical systems. Most notably, *Lieber* et al. promoted the use of chemically modified substrates and AFM tips to study selective molecular interactions [19.73, 107, 110, 114, 117, 124]. A number of other researchers have adopted a similar methodology, leading to the measurement of adhesion forces and interfacial energies for the interactions of a variety of molecular functional groups in various environments, although to date few specific types of interactions have been thoroughly investigated (Table 19.1) [19.30, 107, 110–115, 120, 122, 125–129].

As described above, measurements in air suffer from difficulties of water vapor, thus, the adhesion of many SAM-terminated surfaces have been evaluated under liquids, and it is these systems that we shall focus on here. The value of the adhesion is of course modified depending on the solvent. For example, the interaction of methyl-terminated interfaces is much stronger in water than in nonpolar solvents, in agreement with the

general concept that upon separation the generation of hydrophobic interfaces in a polar solvent is highly energetically unfavorable. The impact of solvent on adhesion can be addressed basically as a variation on the Hamaker constant.

When ionizable end groups are studied in water, the details of the adhesion measurements and results become more complicated, as the chemical nature of the surfaces are now dependent on the pH and ionic strength of the solution. In these circumstances, multiple interactions, including ionic, van der Waals and double-layer forces, come into play simultaneously. Under these conditions, the general form for the JKR adhesion force may be modified to include the following additional forces:

$$F_{\mathrm{adh}} = -\frac{3}{2}\pi R W_{\mathrm{adh}} - \frac{AR}{6D} + 6\pi R W_{\mathrm{dl}} \,, \qquad (19.11)$$

where the second term now includes the attractive van der Waals component and the third term is the repulsive double layer.

Several groups have utilized the ability of AFM to function as a local probe of ionicity and to carry out local force titration measurements on several functional groups, including −COOH [19.30, 114–118, 122], $-NH_2$ [19.116–118, 123], $-SO_3H$ [19.123] and $-PO_3H_2$ [19.116, 130] (see Table 19.1). The force-distance curves can give a general sense of the local chemical state based on whether the approach curve is attractive or repulsive. In force titrations, the adhesion is measured as a function of pH, and the change in force is dependent upon the equilibrium mixture of charged species within the tip-samples contact at the time of measurement. Peaks in the adhesion force vs. pH allow for the determination of the local pK values for the ionized species, which are shown to be dependent on ionic strength. A force titration of a diprotic acid (11-thioundecyl-1-phosphonic acid) is shown in Fig. 19.13. Shifts in the pK values from those measured in free solution have been ascribed to a variety of factors, including buildup of excess surface charge and solvation effects of the surface-bound ionic species. The localization of charge and change migration are effects that have yet to be fully explored in these systems, however.

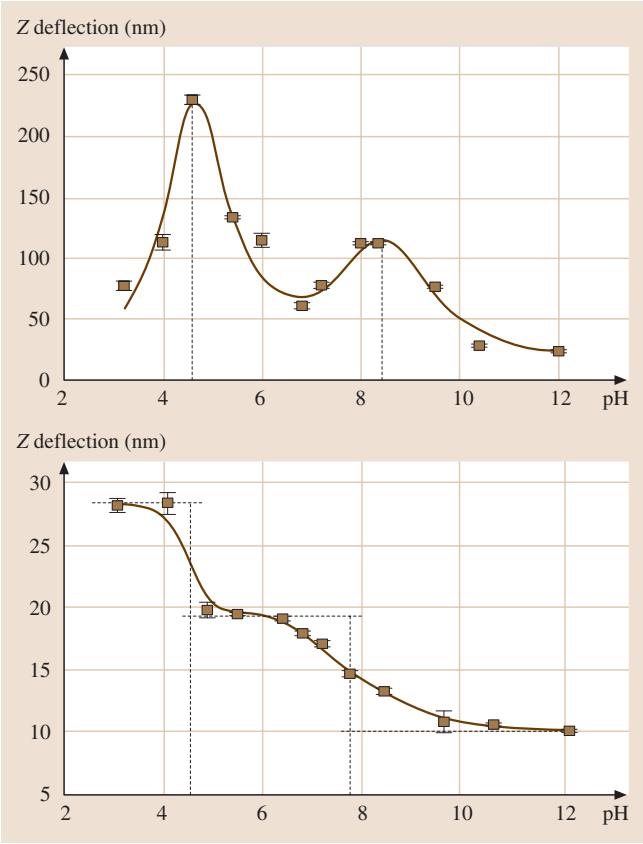


Fig. 19.13a,b Adhesion vs. pH for 11-thioundecyl-1-phosphonic acid in buffer illustrating the impact of pH and ionic strength, **(a)** 10^{-4} M and **(b)** 10^{-1} M on the measured adhesion. The peaks in adhesion provide local measures of the surface pK (adapted after [19.130])

19.5.4 Impact of Intra- and Inter-Chain Interactions on Adhesion

Several studies of the effects of chain length on friction and adhesion have found that, in general, adhesion decreases with increasing chain length, due to increased stability from lateral chain–chain van der Waals interactions within the film, which increases the overall film stiffness. The increased stiffness of the films acts to reduce the effective contact area that develops under compression, consequently reducing adhesion and friction [19.126, 127, 131]. In alkyl chain-based monolayers, the end-group orientation, however, is also dependent on chain length. This results in an "odd-even" effect on the measured adhesion/friction of the monolayer. With $-CH_3$-terminated films, the methyl group orientation differs for odd and even length molecules. This impacts the orientation of the methyl group net dipole and, hence, the local surface free energy [19.73]. In circumstances where interchain hydrogen bonding can occur, additional film stability can be introduced, also yielding an odd–even effect, as observed by *Houston* et al. [19.111, 125].

Studies by AFM of alkanethiol films on Au surfaces have shown that, depending upon the tip size, under varying loads, the tip can readily penetrate the SAM film, displacing the thiol layer from the tip-sample contact [19.132–134]. Upon reduction of the force, the tip again moves out of the SAM film and the surface structure is returned to its original condition. Of key importance in such studies is the mechanism by which the film is displaced. Recently, this has been modeled by *Harrison* using molecular dynamics simulations [19.135]. These studies have clearly demonstrated that in the initial stages of film compression and penetration by an asperity, gauche defects within the typically all-trans configurations of molecules in the SAM layers appear and propagate. The introduction of such defects is the catalyst for the weakening of the chain-chain lateral interactions that help stabilize and maintain film integrity. As this is lost, the asperity can rapidly penetrate the film, and alkanethiol displacement can occur either via chain collapse or bond scission from the surface. *Salmeron* et al. demonstrated that the prevalence of molecular displacement versus film compression depends heavily on the AFM tip size, with sharp tips readily penetrating and displacing surface-bound thiols, while large tips spread the load over more molecules and induce compression over displacement (Fig. 19.14).

The inherent stability of the film structure has also been confirmed by sum-frequency generation spectroscopic studies that have indicated that without the presence of lateral chain-chain interactions, gauche defects appear within the chain structure that reduce overall order and lateral interactions within films [19.136]. The ability to form these defects through poor film order will consequently increase adhesion, friction, and wear of the film. This same spectroscopic study further demonstrated that the appearance of gauche defects can be induced by controlling the local environment, i. e., with the presence of water, causing chains to collapse back upon themselves.

As molecular structure in SAMs moves away from ideally organized layers, chain-chain interactions between the contacting surfaces can also result in entanglements. A recent study of the adhesive interactions of Au(111) surfaces modified with dialkylsulfides [$CH_3(CH_2)_n-S-(CH_2)_9CH_3$; $n = 9, 11, 13, 15, 17$] with varying chain arm lengths probed the combined effect of chain length, solvent, and inter-surface chain entanglement on friction and adhesion using simultaneously modified surfaces and AFM tips [19.137]. This study found that chain-chain interpenetration produced the reverse dependence on chain length on the measured adhesion mentioned above. This work points to the need to examine the nature of inter-surface chain entanglements in nanoscaled systems. Such entanglements should be more prevalent in asperity-asperity contacts where ideal film structure will not be feasible.

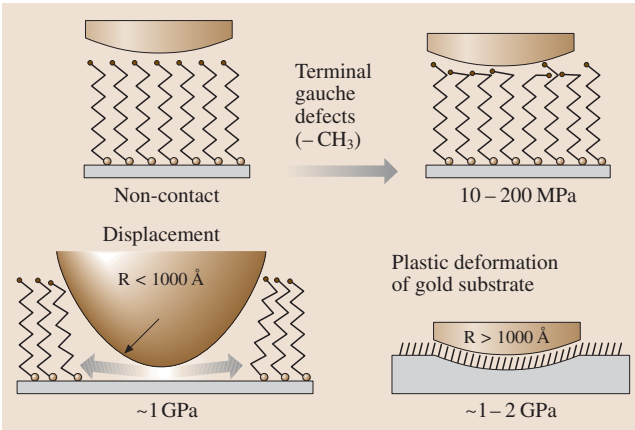


Fig. 19.14 Schematic illustrating the compression of a model lubricant layer under an AMF tip. At low pressures, the gauche defects can form at the tip–sample junction, but the molecules remain in place. Depending on tip size and load, displacement (for sharp tips) or trapping of monolayer molecules (for blunt tips) can occur (adapted after [19.5])

19.5.5 Adhesion at the Single-Bond Level

The ability to resolve the discrete components of interactions is highly desirable. There have been few reports of the direct observation of discrete force components observed with the separation of an AFM tip from a surface. *Beebe* et al., have utilized a statistical method (as described above) for the direct determination of single-bond forces for a variety of interactions [19.119, 121, 122], including biological systems such as biotin-avidin [19.138, 139]. The first report of quantized force measurements was described by *Hoh* et al., which lead to the estimation of single hydrogen bonding forces from studies of glass surfaces in water to be on the order of 10 pN [19.140]. The use of AFM for the study of the energetics of true single chemical bond cleavage has also received little attention. One previous report described discrete covalent bond scission using AFM. In that case, it was proposed that the jumps in the observed pull-off curves were due to sequential scission of chemical bonds contained in a large,

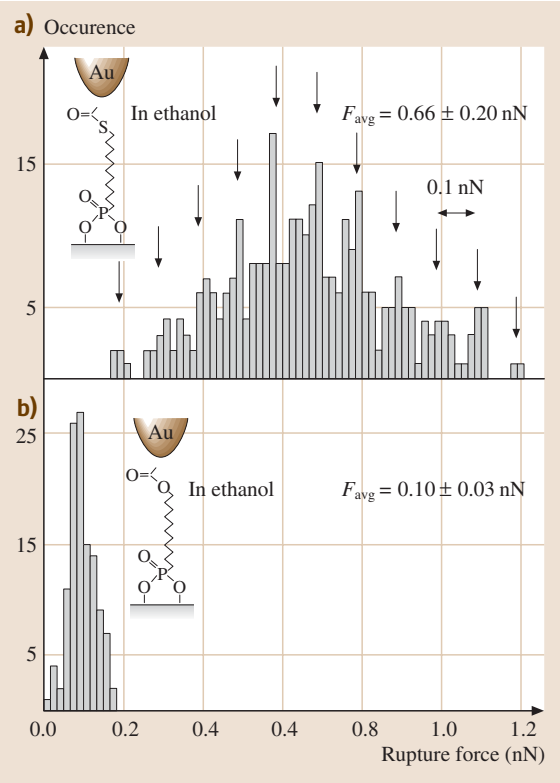


Fig. 19.15 Pull-off force distribution for Au atom abstraction in ethanol. The 100 pN peaks in the distribution represent quantized numbers of Au atoms being abstracted from the sample surface (adapted after [19.129])

multifunctional, polymeric species as it progressively detached from the substrate [19.141]. However, the identification of the relevant chemical bonds involved at each stage was largely based on the known (gas phase) bond strengths of the potentially active functional groups such that solvation effects on the bond energies were ignored – a simplification that profoundly affects the estimated energetics, as energetic exchange with the solvent must be included. More recently, the measurement of discrete bond scission was reported by *Frisbie* et al. for Au-thiol complexes (Fig. 19.15). Here the details of Au atom abstraction were reported with a quantized value of 100 pN (estimated at ∼ 10 kJ/mol, based on an assumed bond rupture length of 1 Å) [19.129]. These studies have demonstrated the feasibility of probing local single bond energetics and have suggested some general requirements for the measurement of adhesion quantization in SAM layers, including the need for a significant negative tip-surface interfacial energy coupled with minimal solvent surface tension [19.129, 142].

19.5.6 Future Directions

Extending AFM adhesion measurements to reactive systems in which chemical bonds can form between the tip and surface affords an expansion of chemical reaction dynamics to solution-based chemistries, whereby the energetic details of single reaction events, previously only accessible for gas phase scattering experiments, may be obtained. Studies of such complex heterogeneous systems will open the door to the evaluation of the energetic pathways of solution-based chemistries for any system in which the appropriate functionalization of surfaces can be exploited. Adhesion has already been demonstrated as a reasonable local probe of surface reaction kinetics, whereby the local changes in the chemical forces may be followed as a function of time during surface chemical reactions [19.143, 144]. In addition to advances in measurements of reactive systems by AFM, complete insight into the operative molecular mechanisms can only be gained when combined with detailed theory that takes into account not only the specific types of interactions present between the surfaces, but also the requisite energetic exchange with local solvent molecules. Advances in computer technologies and in computational theory have made this realistic [19.128].

To advance the area of adhesion measurements at the molecular level, energetic barriers for specific interactions also need to be evaluated, with attention given to the nature of the molecular interactions being probed, *including* the details of energetic exchange with the solvent surrounding the interacting molecules. Studies by AFM of molecular interactions within sharply confined geometries (i. e., ∼ 1 nm^2 contact area) provide an opportunity to evaluate such contributions with molecular detail. Here again, *Lieber* et al. have been advancing the approach of chemically functionalizing carbon nanotubes to reduce both the type and number of specific interacting species [19.73]. This approach may hold some promise for probing well-defined, specific chemical interactions and/or reactions, as long as the nanotubes can be sufficiently stabilized against buckling [19.135, 145, 146].

In addition to the modification of probe geometries to improve localization of interactions for adhesion measurements, one of the principle difficulties in performing AFM measurements of adhesion at SAM surfaces is the unavoidable snap to contact. This makes the details of

the long-range interaction potential almost completely inaccessible. To address this issue, *Lieber* et al. have developed a modified AFM system in which the cantilever is magnetized, providing an additional feedback mechanism to aid in avoiding snap-in [19.30]. When used, this approach provides a smooth approach and retract curve. Similarly, the capacitive coupling feedback mechanism of the interfacial force microscope (IFM) also affords measurements of this transition without snap-in [19.125].

19.6 Concluding Remarks

Scanning probes are powerful tools for determining the fundamental molecular basis of adhesion. Continuum models of adhesion and capillary condensations are useful, but careful attention must be paid to their limits and assumptions. Further progress in these areas requires detailed analysis of the structure and chemistry of both the tip and sample, as well as their environment (solvent, humidity, etc.).

The ability to determine the effects of molecular-scale water menisci or single bond energetics are truly spectacular accomplishments that continue to inspire researchers worldwide to pursue these measurements. With attention paid to previous literature and consideration of issues raised in this chapter, many more discoveries are sure to be made.

References

19.1 I. L. Singer, H. M. Pollock (Eds.): *Fundamentals of Friction* (Kluwer, Dordrecht 1992) p. 351

19.2 K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch et al.: Adhesive force of a single gecko foot-hair, Nature **405** (2000) 681–685

19.3 A. L. Baldwin, G. Thurston: Mechanics of endothelial cell architecture and vascular permeability, Critical Rev. Biomed. Eng. **29** (2001) 247–278

19.4 S. R. White, N. R. Sottos, P. H. Guebelle, J. S. Moore et al.: Autonomic healing of polymer composites, Nature **409** (2001) 794–797

19.5 R. W. Carpick, M. Salmeron: Scratching the surface: Fundamental investigations of tribology with atomic force microscopy, Chem. Rev. **97** (1997) 1163–1194

19.6 R. Maboudian: Adhesion and friction issues associated with reliable operation of MEMS, MRS Bull. **23** (1998) 47–51

19.7 R. Maboudian, W. R. Ashurst, C. Caffaro: Tribological challenges in micromechanical systems, Tribol. Lett. **12** (2002) 95–100

19.8 R. Maboudian, R. T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. **15** (1997) 1

19.9 J. N. Israelachvili: Thin film studies using multiple-beam interferometry, J. Colloid Interface Sci. **44** (1973) 259–272

19.10 J. N. Israelachvili, D. Tabor: The measurement of van der Waals dispersion forces in the range 1.5 to 130 nm, Proc. R. Soc. London A **331** (1972) 19–38

19.11 G. Binnig, C. F. Quate, C. Gerber: Invention of AFM, Phys. Rev. Lett. **56** (1986) 930

19.12 P. Frantz, N. Agraït, M. Salmeron: SFA w/ capacitance, Langmuir **12** (1996) 3289–3294

19.13 J. N. Israelachvili, P. M. McGuiggan, A. M. Homola: Dynamic properties of molecularly thin liquid films, Science **240** (1988) 189–191

19.14 J. Peachey, J. Van Alsten, S. Granick: Design of an apparatus to measure the shear response of ultra-thin liquid, Rev. Sci. Instrum. **62** (1991) 463–473

19.15 L. R. Fisher, J. N. Israelachvili: Direct measurement of the effect of meniscus forces on adhesion: A study of the applicability of macroscopic thermodynamics to microscopic liquid interfaces, Colloids Surf. **3** (1981) 303–319

19.16 L. R. Fisher, J. N. Israelachvili: Experimental studies on the applicability of the Kelvin equation to highly curved concave menisci, J. Colloid Interface Sci. **80** (1981) 528–541

19.17 J. N. Israelachvili: *Intermolecular and Surface Forces* (Academic Press, London 1992)

19.18 S. Granick: Molecular tribology of fluids. In: *Fundamentals of Friction*, ed. by I. L. Singer, H. M. Pollock (Kluwer, Dordrecht 1992) p. 387

19.19 G. Reiter, A. L. Demirel, J. Peanasky, L. L. Cai et al.: Stick to slip transition and adhesion of lubricated surfaces in moving contact, J. Chem. Phys. **101** (1994) 2606

19.20 K. L. Johnson: *Contact Mechanics* (Cambridge Univ. Press, Cambridge 1987)

19.21 O. Marti, B. Drake, P. K. Hansma: Atomic force microscopy of liquid-covered surfaces: Atomic resolution images, Appl. Phys. Lett. **51** (1987) 484–486

19.22 G. J. Germann, S. R. Cohen, G. Neubauer, G. M. McClelland et al.: Atomic scale friction of a diamond tip on diamond (100) and (111) surfaces, J. Appl. Phys. **73** (1993) 163

19.23 L. Howald, E. Meyer, R. Lüthi, H. Haefke et al.: UHV AFM design, Appl. Phys. Lett. **63** (1993) 117

19.24 M. Kageshima, H. Yamada, K. Nakayama, H. Sakama et al.: Development of an ultrahigh vacuum atomic force microscope for investigations of semiconductor surfaces, J. Vac. Sci. Technol. B **11** (1993) 1987–1991

19.25 J. A. Greenwood: Adhesion of elastic spheres, Proc. R. Soc. London A **453** (1997) 1277–1297

19.26 S. P. Jarvis, A. Oral, T. P. Weihs, J. B. Pethica: A novel force microscope and point contact probe, Rev. Sci. Instrum. **64** (1993) 3515

19.27 S. P. Jarvis, H. Yamada, S.-I. Yamamoto, H. Tokumoto: A new force controlled atomic force microscope for use in ultrahigh vacuum, Rev. Sci. Instrum. **67** (1996) 2281

19.28 S. A. Joyce, J. E. Houston: A new force sensor incorporating force-feedback control for interfacial force microscopy, Rev. Sci. Instrum. **62** (1991) 710–715

19.29 S. A. Joyce, J. E. Houston, T. A. Michalske: Differentiation of topographical and chemical structures using an interfacial force microscope, Appl. Phys. Lett. **60** (1992) 1175

19.30 P. D. Ashby, L. W. Chen, C. M. Lieber: Probing intermolecular forces and potentials with magnetic feedback chemical force microscopy, J. Am. Chem. Soc. **122** (2000) 9467–9472

19.31 H. I. Kim, V. Boiadjiev, J. E. Houston, X.-Y. Zhu, J. D. Kiely: Tribological properties of self-assembled monolayers on Au, SiO_x and Si surfaces, Tribol. Lett. **10**(1-2) (2001) 97–101

19.32 J. S. Nelson, B. W. Dodson, P. A. Taylor: Adhesive avalanche in covalently bonded materials, Phys. Rev. B **45** (1992) 4439–4444

19.33 P. A. Taylor, J. S. Nelson, B. W. Dodson: Adhesion between atomically flat metallic surfaces, Phys. Rev. B **44** (1991) 5834–5841

19.34 K. R. Shull: Contact mechanics and the adhesion of soft solids, Mater. Sci. Eng. R: Rep. **R 36** (2002) 1–45

19.35 L. Xu, A. Lio, J. Hu, D. F. Ogletree et al.: Wetting and capillary phenomena of water on mica, J. Phys. Chem. B **102** (1998) 540–548

19.36 S. P. Timoshenko, J. N. Goodier: *Theory of Elasticity* (McGraw Hill, New York 1987)

19.37 J. P. Cleveland, S. Manne, D. Bocek, P. K. Hansma: A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy, Rev. Sci. Instrum. **64** (1993) 403

19.38 J. E. Sader: Parallel beam approximation for V-shaped atomic force microscope cantilevers, Rev. Sci. Instrum. **66** (1995) 4583

19.39 M. Tortonese, M. Kirk: Characterization of application specific probes for SPMs, Proc. SPIE **3009** (1997) 53–60

19.40 T. R. Albrecht, S. Akamine, T. E. Carver, C. F. Quate: Microfabrication of cantilever styli for the atomic force microscope, J. Vac. Sci. Technol. A **8** (1990) 3386–3396

19.41 H.-J. Butt, P. Siedle, K. Seifert, K. Fendler et al.: Scan speed limit in atomic force microscopy, J. Microsc. **169** (1993) 75–84

19.42 Y. Q. Li, N. J. Tao, J. Pan, A. A. Garcia et al.: Direct measurement of interaction forces between colloidal particles using the scanning force microscope, Langmuir **9** (1993) 637–641

19.43 R. Lüthi, E. Meyer, H. Haefke, L. Howald et al.: Nanotribology: An UHV-SFM study on thin films of C_{60} and AgBr, Surf. Sci. **338** (1995) 247–260

19.44 J. M. Neumeister, W. A. Ducker: Lateral, normal, and longitudinal spring constants of atomic force microscopy cantilevers, Rev. Sci. Instrum. **65** (1994) 2527–2531

19.45 D. F. Ogletree, R. W. Carpick, M. Salmeron: Calibration of frictional forces in atomic force microscopy, Rev. Sci. Instrum. **67** (1996) 3298–3306

19.46 J. A. Ruan, B. Bhushan: Atomic-scale friction measurements using friction force microscopy: Part I – general principles and new measurement techniques, Trans. ASME J. Tribol. **116** (1994) 378

19.47 J. E. Sader, I. Larson, P. Mulvaney, L. R. White: Method for the calibration of atomic force microscope cantilevers, Rev. Sci. Instrum. **66** (1995) 3789

19.48 U. D. Schwarz, P. Koster, R. Wiesendanger: Quantitative analysis of lateral force microscopy experiments, Rev. Sci. Instrum. **67** (1996) 2560

19.49 T. J. Senden, W. A. Ducker: Experimental determination of spring constants in atomic force microscopy, Langmuir **10** (1994) 1003

19.50 A. Torii, M. Sasaki, K. Hane, S. Okuma: A method for determining the spring constant of cantilevers for atomic force microscopy, Meas. Sci. Technol. **7** (1996) 179–184

19.51 J. E. Sader, J. W. M. Chon, P. Mulvaney: Calibration of rectangular atomic force microscope cantilevers, Rev. Sci. Instrum. **70** (1999) 3967–3969

19.52 J. S. Villarrubia: Morphological estimation of tip geometry for scanned probe microscopy, Surf. Sci. **321** (1994) 287–300

19.53 F. Atamny, A. Baiker: Direct imaging of the tip shape by AFM, Surf. Sci. **323** (1995) L314

19.54 R. W. Carpick, N. Agraït, D. F. Ogletree, M. Salmeron: Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscope, J. Vac. Sci. Technol. B **14** (1996) 1289–1295

19.55 R. Dixson, J. Schneir, T. McWaid, N. Sullivan et al.: Toward accurate linewidth metrology using atomic force microscopy and tip characterization, Proc. SPIE **2725** (1996) 589–607

19.56 L. S. Dongmo, J. S. Villarrubia, S. N. Jones, T. B. Renegar et al.: Experimental test of blind tip reconstruction for scanning probe microscopy, Ultramicroscopy **85** (2000) 141–153

19.57 K. F. Jarausch, T. J. Stark, P. E. Russell: Silicon structures for in situ characterization of atomic force microscope probe geometry, J. Vac. Sci. Technol. B **14** (1996) 3425–3430

19.58 P. Markiewicz, M. C. Goh: Atomic force microscope tip deconvolution using calibration arrays, Rev. Sci. Instrum. **66** (1995) 3186–3190

19.59 C. Odin, J. P. Aimé, Z. El Kaakour, T. Bouhacina: Tip's finite size effects on atomic force microscopy in the contact mode: Simple geometrical considerations for rapid estimation of apex radius and tip angle based on the study of polystyrene latex balls, Surf. Sci. **317** (1994) 321–340

19.60 S. S. Sheiko, M. Moller, E. M. C. M. Reuvekamp, H. W. Zandbergen: Evaluation of the probing profile of scanning force microscopy tips, Ultramicroscopy **53** (1994) 371–380

19.61 P. Siedle, H.-J. Butt, E. Bamberg, D. N. Wang et al.: Determining the form of atomic force microscope tips. In: *X-Ray Optics and Microanalysis*, ed. by P. B. Kenway, P. J. Duke, G. W. Lorimer et al. (Iop, Manchester 1992) pp. 361–364

19.62 J. S. Villarrubia: Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation, J. Res. Nat. Inst. Stand. Technol. **102** (1997) 425–454

19.63 K. L. Westra, D. J. Thomson: Atomic force microscope tip radius needed for accurate imaging of thin film surfaces, J. Vac. Sci. Technol. B **12** (1994) 3176–3181

19.64 S. Xu, M. F. Arnsdorf: Calibration of the scanning (atomic) force microscope with gold particles, J. Microsc. **3** (1994) 199–210

19.65 U. D. Schwarz, O. Zwörner, P. Köster, R. Wiesendanger: Friction force spectroscopy in the low-load regime with well-defined tips. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997)

19.66 P. Niedermann, W. Hanni, N. Blanc, R. Christoph et al.: Chemical vapor deposition diamond for tips in nanoprobe experiments, J. Vac. Sci. Technol. A **14** (1995) 1233–1236

19.67 S. J. O'Shea, R. N. Atta, M. E. Welland: Characterization of tips for conducting atomic force microscopy, Rev. Sci. Instrum. **66** (1995) 2508–2512

19.68 L. M. Qian, X. D. Xiao, S. Z. Wen: Tip in situ chemical modification and its effects on tribological measurements, Langmuir **16** (2000) 662–670

19.69 E. L. Florin, V. T. Moy, H. E. Gaub: Adhesion forces between individual ligand-receptor pairs, Science **264** (1994) 415–417

19.70 G. U. Lee, L. A. Chrisey, R. J. Colton: Direct measurement of the forces between complementary strands of DNA, Science **266** (1994) 771–773

19.71 V. T. Moy, E. L. Florin, H. E. Gaub: Intermolecular forces and energies between ligands and receptors, Science **266** (1994) 257–259

19.72 O. H. Willemsen, M. M. E. Snel, K. O. van der Werf, B. G. de Grooth et al.: Simultaneous height and adhesion imaging of antibody-antigen interactions by atomic force microscopy, Biophys. J. **75** (1998) 2220–2228

19.73 S.-S. Wong, H. Takano, M. D. Porter: Mapping orientation differences of terminal functional groups by friction force microscopy, Anal. Chem. **70** (1998) 5209–5212

19.74 H. J. Butt: Measuring electrostatic, van der Waals, and hydration forces in electrolyte solutions with an atomic force microscope, Biophys. J. **60** (1991) 1438–1444

19.75 W. A. Ducker, T. J. Senden, R. M. Pashley: Direct measurement of colloidal forces using an atomic force microscope, Nature **353** (1991) 239–241

19.76 V. S. J. Craig, C. Neto: In situ calibration of colloid probe cantilevers in force microscopy: Hydrodynamic drag on a sphere approaching a wall, Langmuir **17** (2001) 6018–6022

19.77 R. Staub, D. Alliata, C. Nicolini: Drift elimination in the calibration of scanning probe microscopes, Rev. Sci. Instrum. **66** (1995) 2513–2516

19.78 J. Fu: In situ testing and calibrating of z-piezo of an atomic force microscope, Rev. Sci. Instrum. **66** (1995) 3785–3788

19.79 J. Garnaes, L. Nielsen, K. Dirscherl, J. F. Jorgensen et al.: Two-dimensional nanometer-scale calibration based on one-dimensional gratings, Appl. Phys. A **66** (1998) 831–835

19.80 S. M. Hues, C. F. Draper, K. P. Lee, R. J. Colton: Effect of PZT and PMN actuator hysteresis and creep on nanoindentation measurements using force microscopy, Rev. Sci. Instrum. **65** (1994) 1561

19.81 J. F. Jorgensen, K. Carneiro, L. L. Madsen, K. Conradsen: Hysteresis correction of scanning tunneling microscope images, J. Vac. Sci. Technol. B **12** (1994) 1702–1704

19.82 J. F. Jorgensen, L. L. Madsen, J. Garnaes, K. Carneiro et al.: Calibration, drift elimination, and molecular structure analysis, J. Vac. Sci. Technol. B **12** (1994) 1698–1701

19.83 M. Jaschke, H. J. Butt: Height calibration of optical lever atomic force microscopes by simple laser interferometry, Rev. Sci. Instrum. **66** (1995) 1258

19.84 L. A. Nagahara, K. Hashimoto, A. Fujishima, D. Snowden-Ifft et al.: Mica etch pits as a height calibration source for atomic force microscopy, J. Vac. Sci. Technol. B **12** (1993) 1694–1697

19.85 H. M. Brodowsky, U.-C. Boehnke, F. Kremer: Wide range standard for scanning probe microscopy height calibration, Rev. Sci. Instrum. **67** (1996) 4198–4200

19.86 T. Stifter, O. Marti, B. Bhushan: Theoretical investigation of the distance dependence of capillary and

van der Waals forces in scanning force microscopy, Phys. Rev. B **62** (2000) 13667–13673
19.87 P. G. de Gennes: Wetting: Statistics and dynamics, Rev. Mod. Phys. **57** (1985) 827–863
19.88 M. Binggeli, C. M. Mate: Influence of capillary condensation of water on nanotribology studied by force microscopy, Appl. Phys. Lett. **65** (1994) 415
19.89 Y. Sugawara, M. Ohta, T. Konishi, S. Morita et al.: Effects of humidity and tip radius on the adhesive force measured with atomic force microscopy, Wear **168** (1993) 13–16
19.90 F. M. Orr, L. E. Scriven, A. P. Rivas: Pendular rings between solids: Meniscus properties and capillary force, J. Fluid Mech. **67** (1975) 723–742
19.91 X. Xiao, Q. Linmao: Investigation of humidity-dependent capillary force, Langmuir **16** (2000) 8153–8158
19.92 E. Riedo, F. Levy, H. Brune: Kinetics of capillary condensation in nanoscopic sliding friction, Phys. Rev. Lett. **88** (2002) 185505/1–4
19.93 D. Maugis, B. Gauthiermanuel: JKR-DMT transition in the presence of a liquid meniscus, J. Adhes. Sci. Technol. **8** (1994) 1311–1322
19.94 A. Fogden, L. R. White: Contact elasticity in the presence of capillary condensation. 1. The nonadhesive Hertz problem, J. Colloid Interface Sci. **138** (1990) 414
19.95 K. L. Johnson, K. Kendall, A. D. Roberts: Surface energy and the contact of elastic solids, Proc. R. Soc. London A **324** (1971) 301–313
19.96 K. Johnson, J. Greenwood: An adhesion map for the contact of elastic spheres, J. Colloid Interface Sci. **192** (1997) 326–333
19.97 K. L. Johnson: Adhesion and friction between a smooth elastic asperity and a plane surface, Proc. R. Soc. London A **453** (1997) 163
19.98 T. Thundat, X. Y. Zheng, G. Y. Chen, R. J. Warmack: Role of relative humidity in atomic force microscopy imaging, Surf. Sci. **294** (1993) 939–943
19.99 M. Binggeli, R. Christoph, H.-E. Hintermann: Observation of controlled, electrochemically induced friction force modulations in the nano-Newton range, Tribol. Lett. **1** (1995) 13
19.100 B. Bhushan, S. Sundararajan: Micro/nanoscale friction and wear mechanisms of thin films using atomic force and friction force microscopy, Acta Mater. **46** (1998) 3793–3804
19.101 W. Gulbinski, D. Pailharey, T. Suszko, Y. Mathey: Study of the influence of adsorbed water on AFM friction measurements on molybdenum trioxide thin films, Surf. Sci. **475** (2001) 149–158
19.102 M. He, A. S. Blum, D. E. Aston, C. Buenviaje et al.: Critical phenomena of water bridges in nanoasperity contacts, J. Chem. Phys. **114** (2001) 1355–1360
19.103 J. A. Greenwood, J. B. P. Williamson: Contact of nominally flat surfaces, Proc. R. Soc. London A **295** (1966) 300
19.104 K. Komvopoulous: Surface engineering and microtribology for microelectromechanical systems, Wear **200** (1996) 305–327
19.105 R. Maboudian, W. R. Ashurst, C. Carraro: Self-assembled monolayers as anti-stiction coatings for MEMS: Characteristics and recent developments, Sens. Actuators A **82** (2000) 219
19.106 R. Maboudian: Surface processes in MEMS technology, Surf. Sci. Rep. **30** (1998) 207–270
19.107 A. Noy, D. V. Vezenov, C. M. Lieber: Chemical force microscopy, Annu. Rev. Mater. Sci. **27** (1997) 381–421
19.108 H. Takano, J. R. Kenseth, S.-S. Wong, J. C. O'Brien et al.: Chemical and biochemical analysis using scanning force microscopy, Chem. Rev. **99** (1999) 2845
19.109 D. L. Sedin, K. L. Rowlen: Adhesion forces measured by atomic force microscopy in humid air, Anal. Chem. **72** (2000) 2183–2189
19.110 C. D. Frisbie, L. F. Rozsnyai, A. Noy, M. S. Wrighton et al.: Functional group imaging by chemical force microscopy, Science **265** (1994) 2071–2074
19.111 H. I. Kim, J. E. Houston: Separating mechanical and chemical contributions to molecular-level friction, J. Am. Chem. Soc. **122** (2000) 12045–12046
19.112 T. Nakagawa, K. Ogawa, T. Kurumizawa: Discriminating molecular length of chemically adsorbed molecules using an atomic force microscope having a tip covered with sensor molecules (an atomic force microscope having chemical sensing function), Jpn. J. Appl. Phys. **32** (1993) 294–296
19.113 T. Nakagawa, K. Ogawa, T. Kurumizawa: Atomic force microscope for chemical sensing, J. Vac. Sci. Techol. **B 12** (1994) 2215–2218
19.114 A. Noy, C. D. Frisbie, L. F. Rozsnyai, M. S. Wrighton et al.: Chemical force microscopy: Exploiting chemically-modified tips to quantify adhesion, friction, and functional group distributions in molecular assemblies, J. Am. Chem. Soc. **117** (1995) 7943–7951
19.115 S. K. Sinniah, A. B. Steel, C. J. Miller, J. E. Reutt-Robey: Solvent exclusion and chemical contrast in scanning force microscopy, J. Am. Chem. Soc. **118** (1996) 8925–8931
19.116 E. W. v.d. Vegte, G. Hadziioannou: Acid-base properties and the chemical imaging of surface-bound functional groups with scanning force microscopy, J. Phys. Chem. **B 101** (1997) 9563–9569
19.117 D. V. Vezenov, A. Noy, L. F. Rozsnyai, C. M. Lieber: Force titrations and ionization state sensitive imaging of functional group distributions in molecular assemblies, J. Am. Chem. Soc. **119** (1997) 2006–2015
19.118 E. W. v.d. Vegte, G. Hadziioannou: Scanning force microscopy with chemical specificity: An extensive study of chemically specific tip-surface interactions and the chemical imaging of surface functional groups, Langmuir **13** (1997) 4357–4368

19.119 L. A. Wenzler, G. L. Moyes, L. G. Olson, J. M. Harris et al.: Single-molecule bond rupture force analysis of interactions between AFM tips and substrates modified with organosilanes, Anal. Chem. **69** (1997) 2855–2861

19.120 T. Ito, M. Namba, P. Buhlmann, Y. Umezawa: Modification of silicon nitride tips with trichlorosilane self-assembled monolayers (SAMs) for chemical force microscopy, Langmuir **13** (1997) 4323–4332

19.121 L. A. Wenzler, G. L. Moyes, G. N. Raikar, R. L. Hansen et al.: Measurements of single-molecule bond rupture forces between self-assembled monolayers of organosilanes with the atomic force microscope, Langmuir **13** (1997) 3761–3768

19.122 T. Han, J. M. Williams, T. P. Beebe: Chemical bonds studied with functionalized atomic force microscopy tips, Anal. Chim. Acta **307** (1995) 365–376

19.123 V. Tsukruk, V. N. Bliznyuk: Adhesive and friction forces between chemically modified silicon and silicon nitride surfaces, Langmuir **14** (1998) 446–455

19.124 D. V. Vezenov, A. V. Zhuk, G. M. Whitesides, C. M. Likeber: Chemical force spectroscopy in heterogeneous systems: Intermolecular interactions involving epoxy polymer, mixed monolayers, and polar solvents, J. Am. Chem. Soc. **124** (2002) 10578–10588

19.125 J. E. Houston, H. I. Kim: Adhesion, friction, and mechanical properties of functionalized alkanethiol self-assembled monolayers, Accounts Chem. Res. **35** (2002) 547–553

19.126 H. I. Kim, M. Graupe, O. Oloba, T. Koini et al.: Molecularly specific studies of the frictional properties of monolayer films: A systematic comparison of CF_3-, $(CH3)_2(CH)$-, and CH_3- terminated films, Langmuir **15** (1999) 3179–3185

19.127 S. Lee, Y. S. Shon, R. Colorado, R. L. Guenard et al.: The influence of packing densities and surface order on the frictional properties of alkanethiol self-assembled monolayers (SAMs) on gold: A comparison of SAMs derived from normal and spiroalkanedithiols, Langmuir **16** (2000) 2220–2224

19.128 Y. Leng, S. Jiang: Dynamic simulations of adhesion and friction in chemical force microscopy, J. Am. Chem. Soc. **124** (2002) 11764–11770

19.129 H. Skulason, C. D. Frisbie: Detection of discrete interactiuons upon rupture of Au microcontacts to self-assembled monolayers terminated with $-S(CO)CH_3$ or $-SH$, J. Am. Chem. Soc. **122** (2000) 9750–9760

19.130 J. Zhang, J. Kirkham, C. Robinson, M. L. Wallwork et al.: Determination of the ionization state of 11-thioundecyl-1-phosphonic acid in self-assembled monolayers by chemical force microscopy, Anal. Chem. **72** (2000) 1973–1978

19.131 A. Lio, D. H. Charych, M. Salmeron: Comparative atomic force microscopy study of the chain length dependence of frictional properties of alkanethiols on gold and alkylsilanes on mica, J. Phys. Chem. **B 101** (1997) 3800

19.132 A. Lio, C. Morant, D. F. Ogletree, M. Salmeron: Atomic force microscopy study of the pressure-dependent structural and frictional properties of n-alkanethiols on gold, J. Phys. Chem. B **101** (1997) 4767–4773

19.133 G.-Y. Liu, M. Salmeron: Reversible displacement of chemisorbed n-alkane thiol molecules on Au(111) surface: An atomic force microscopy study, Langmuir **10** (1994) 367–370

19.134 X. Xiao, J. Hu, D. H. Charych, M. Salmeron: Chain length dependence of the frictional properties of alkylsilane molecules self-assembled on mica studied by atomic force microscopy, Langmuir **12** (1996) 235

19.135 A. B. Tutein, S. J. Stuart, J. A. Harrison: Indentation analysis of linear-chain hydrocarbon monolayers anchored to diamond, J. Phys. Chem. **B 103** (1999) 11357

19.136 R. L. Pizzolatto, Y. J. Yang, L. K. Wolf, M. C. Messmer: Conformational aspects of model chromatographic surfaces studied by sum-frequency generation, Anal. Chim. Acta **397** (1999) 81

19.137 E. W. v.d. Vegte, A. Subbotin, G. Hadziioannou: Nanotribological properties of unsymmetrical n-dialkyl sulfide monolayers on gold: Effect of chain length on adhesion, friction and imaging, Langmuir **16** (2000) 3249–3256

19.138 Y.-S. Lo, N. D. Huefner, W. S. Chan, F. Stevebs et al.: Specific interactions between biotin and avidin studies by atomic force microscopy using the poisson statistical analysis method, Langmuir **15** (1999) 1373–1382

19.139 Y.-S. Lo, J. Simons, T. P. Beebe, Jr.: Temperature dependence of the biotin–avidin bond rupture force studied by atomic force microscopy, J. Phys. Chem. **B 106** (2002) 9847–9857

19.140 J. H. Hoh, J. P. Cleavland, C. B. Prater, J.-P. Revel et al.: Quantitized adhesion detected with the atomic force microscope, J. Am. Chem. Soc. **114** (1992) 4917–4918

19.141 M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann et al.: How strong is a covalent bond?, Science **283** (1999) 1727–1730

19.142 H. Skulason, C. D. Frisbie: Contact mechanics modeling of pull-off measurements: Effect of solvent, probe radius, and chemical binding probability on the detection of single-bond rupture forces by atomic force microscopy, Anal. Chem. **74** (2002) 3096–3104

19.143 H. Schonherr, V. Chechik, C. J. M. Stirling, G. J. Vancso: Monitoring surface reactions at an AFM tip: An approach to following reaction kinetics in self-assembled monolayers on the nanometer scale, J. Am. Chem. Soc. **122** (2000) 3679–3687

19.144 M. P. L. Werts, E. W. v.d. Vegte, G. Hadziioannou: Surface chemical reactions probed with scanning force microscopy, Langmuir **13** (1997) 4939–4942

19.145 S. S. Wong, E. Joselevich, A. T. Woolley, C. Chin Li et al.: Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology, Nature **394** (1998) 52–55

19.146 S. S. Wong, A. T. Woolley, E. Joselevich, C. L. Cheung et al.: Covalently-functionalized single-walled carbon nanotube tips for chemical force microscopy, J. Am. Chem. Soc. **120** (1998) 8557–8558

20. Friction and Wear on the Atomic Scale

Friction is an old subject of research: the empirical da Vinci–Amontons laws are common knowledge. Macroscopic experiments systematically performed by the school of *Bowden* and *Tabor* have revealed that macroscopic friction can be related to the collective action of small asperities. During the last 15 years, experiments performed with the atomic force microscope gave new insight into the physics of single asperities sliding over surfaces. This development, together with complementary experiments by means of surface force apparatus and quartz microbalance, established the new field of nanotribology. At the same time, increasing computing power allowed for the simulation of the processes in sliding contacts consisting of several hundred atoms. It became clear that atomic processes cannot be neglected in the interpretation of nanotribology experiments. Experiments on even well-defined surfaces directly revealed atomic structures in friction forces. This chapter will describe friction force microscopy experiments that reveal, more or less directly, atomic processes in the sliding contact.

We will begin by introducing friction force microscopy, including the calibration of cantilever force sensors and special aspects of the ultra-high vacuum environment. The empirical Tomlinson model largely describes atomic stick-slip results and is therefore presented in detail. We review experimental results regarding atomic friction. These include thermal activation, velocity dependence, as well as temperature dependence. The geometry of the contact plays a crucial role in the interpretation of experimental results, as we will demonstrate, for example, for the calculation of the lateral contact stiffness. The onset of wear on atomic scale has recently come into the scope of experimental studies and is described here. In order to compare the respective results, we present molecular dynamics simulations that are directly related to atomic friction experiments. We close the chapter with a discussion of dissipation measurements performed in noncontact force microscopy, which may become an important complementary tool for the study of mechanical dissipation in nanoscopic devices.

20.1 **Friction Force Microscopy in Ultra-High Vacuum** 632
20.1.1 Friction Force Microscopy 632
20.1.2 Force Calibration 632
20.1.3 The Ultra-high Vacuum Environment 635
20.1.4 A Typical Microscope in UHV 635

20.2 **The Tomlinson Model** 636
20.2.1 One-dimensional Tomlinson Model 636
20.2.2 Two-dimensional Tomlinson Model 637
20.2.3 Friction Between Atomically Flat Surfaces ... 637

20.3 **Friction Experiments on Atomic Scale** 638
20.3.1 Anisotropy of Friction 642

20.4 **Thermal Effects on Atomic Friction** 642
20.4.1 The Tomlinson Model at Finite Temperature 642
20.4.2 Velocity Dependence of Friction 644
20.4.3 Temperature Dependence of Friction 645

20.5 **Geometry Effects in Nanocontacts** 646
20.5.1 Continuum Mechanics of Single Asperities 646
20.5.2 Load Dependence of Friction 647
20.5.3 Estimation of the Contact Area 647

20.6 **Wear on the Atomic Scale** 649
20.6.1 Abrasive Wear on the Atomic Scale . 649
20.6.2 Wear Contribution to Friction 650

20.7 **Molecular Dynamics Simulations of Atomic Friction and Wear** 651
20.7.1 Molecular Dynamics Simulation of Friction Processes 651
20.7.2 Molecular Dynamics Simulations of Abrasive Wear 652

20.8 **Energy Dissipation in Noncontact Atomic Force Microscopy** .. 654

20.9 **Conclusion** 656

References 657

Springer Handbook *of* Nanotechnology
B. Bhushan • © Springer 2004

20.1 Friction Force Microscopy in Ultra-High Vacuum

The *friction force microscope* (FFM, also called lateral force microscope, LFM) exploits the interaction of a sharp tip sliding on a surface to quantify dissipative processes down to the atomic scale (Fig. 20.1).

20.1.1 Friction Force Microscopy

The relative motion of tip and surface is realized by a *scanner* formed by piezoelectric elements, which moves the surface perpendicularly to the tip with a certain periodicity. The scanner can be also extended or retracted in order to vary the normal force, F_N, which is applied on the surface. This force is responsible for the deflection of the *cantilever* that supports the tip. If the normal force F_N increases while scanning due to the local slope of the surface, the scanner is retracted by a feedback loop. On the other hand, if F_N decreases, the surface is brought closer to the tip by extending the scanner. In such way, the surface topography can be determined line by line from the vertical displacement of the scanner. An accurate control of such vertical movement is made possible by a light beam, which is reflected from the rear of the lever into a photodetector. When the bending of the cantilever changes, the light spot on the detector moves up or down and causes a variation of the photocurrent that corresponds to the normal force F_N to be controlled.

Usually the relative sliding of tip and surface is also accompanied by *friction*. A lateral force, F_L, with the opposite direction of the scan velocity, v, hinders the motion of the tip. This force provokes the torsion of the cantilever, and it can be observed with the topography if the photodetector can reveal not only the normal deflection but also the lateral movement of the lever while scanning. In practice this is realized by four-quadrants photodetectors, as shown in Fig. 20.1. We should notice that friction forces also cause the lateral bending of the cantilever, but this effect is negligible if the thickness of the lever is much less than the width.

The FFM was first used by *Mate* et al. in 1987 to reveal friction with atomic features [20.1], i.e. just one year after *Binnig*, *Quate*, and *Gerber* introduced the atomic force microscope [20.2]. In their experiment, Mate used a tungsten wire and a slightly different technique to detect lateral forces (non-fiber interferometry). The optical beam deflection was introduced later by *Marti* et al., and *Meyer* et al. [20.3, 4]. Other methods to measure the forces between tip and surface are given by capacitance detection [20.5], dual fiber interferometry [20.6], and piezolevers [20.7]. In the first method, two plates close to the cantilever reveal capacitance while scanning. The second technique uses two optical fibers to detect the cantilever deflection along two orthogonal directions angled 45° with respect to the surface normal. Finally, in the third method, cantilevers with two Wheatstone bridges at their base reveal normal and lateral forces, which are respectively proportional to the sum and the difference of both bridge signals.

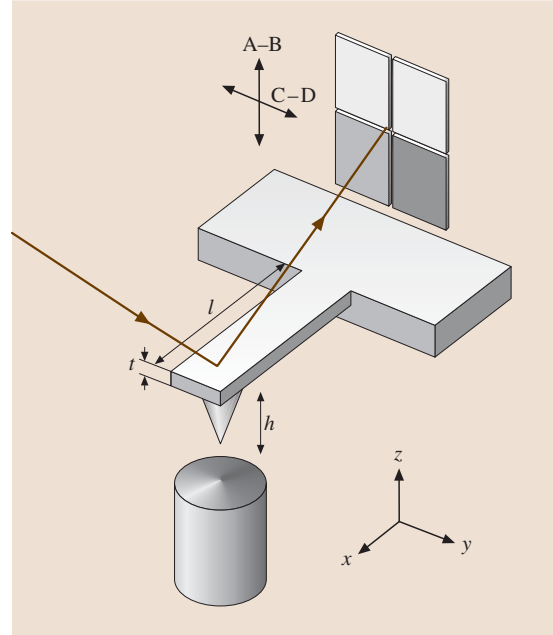


Fig. 20.1 Schematic diagram of a beam-deflection friction force microscope

20.1.2 Force Calibration

The force calibration is relatively simple if rectangular cantilevers are used. Due to possible discrepancies with the geometric values provided by manufacturers, one should use optical and electron microscopes to determine the width, thickness, and length of cantilever, w, t, l, the tip height, h, and the position of the tip with respect to the cantilever. The thickness of the cantilever can also be determined from the resonance frequency of the lever, f_0, using the

relation [20.8]:

$$t = \frac{2\sqrt{12}\pi}{1.875^2}\sqrt{\frac{\rho}{E}}f_0 l^2 , \tag{20.1}$$

In (20.1) ρ is the density of the cantilever and E is its Young's modulus. The normal spring constant, c_N, and the lateral spring constant, c_L, of the lever are given by

$$c_N = \frac{Ewt^3}{4l^3} , \quad c_L = \frac{Gwt^3}{3h^2 l} , \tag{20.2}$$

where G is the shear modulus. Figure 20.2 shows some SEM images of rectangular silicon cantilevers used for FFM. In the case of silicon, $\rho = 2.33\times10^3\,\mathrm{kg/m^3}$, $E = 1.69\times10^{11}\,\mathrm{N/m^2}$ and $G = 0.5\times10^{11}\,\mathrm{N/m^2}$. Thus, for the cantilever in Fig. 20.2, $c_N = 1.9\,\mathrm{N/m}$ and $c_L = 675\,\mathrm{N/m}$.

The next step in force calibration consists of measuring the sensitivity of the photodetector, S_z (nm/V). For beam-deflection FFMs, the sensitivity S_z can be determined by force vs. distance curves measured on hard surfaces (e.g., Al_2O_3), where elastic deformations are negligible and the vertical movement of the scanner equals the deflection of the cantilever. A typical relation between the difference of the vertical signals on the four-quadrant detector, V_N, and the distance from the surface, z, is sketched in Fig. 20.3. When the tip is approached, no signal is revealed until the tip jumps into contact at $z = z_1$. Further extension or retraction of the scanner results in an elastic behavior until the tip jumps again out of contact at a distance $z_2 > z_1$. The slope of the elastic part of the curve gives the required sensitivity, S_z.

The normal and lateral forces are related to the voltage V_N, and the difference between the horizontal signals, V_L, as follows:

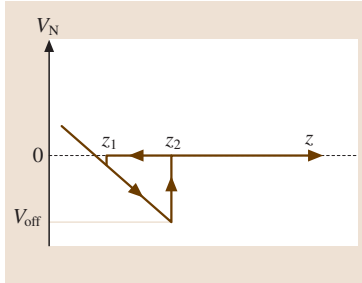


Fig. 20.3 Sketch of a typical force vs. distance curve

$$F_N = c_N S_z V_N , \quad F_L = \frac{3}{2}c_L\frac{h}{l}S_z V_L , \tag{20.3}$$

in which it is assumed that the light beam is positioned above the probing tip.

The normal spring constant c_N can also be calibrated with alternative methods. *Cleveland* et al. [20.10] attached tungsten spheres to the tip, which changes the resonance frequency f_0 according to the formula

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{c_N}{M+m^*}} . \tag{20.4}$$

M is the mass of the added object, and m^* is an effective mass of the cantilever, which depends on its geometry [20.10]. The spring constant can be extrapolated from the frequency shifts corresponding to the different masses attached.

As an alternative, *Hutter* et al. observed that the spring constant c_N can be related to the area of the power spectrum of the thermal fluctuations of the cantilever, P [20.11]. The correct relation is $c_N = 4k_B T/3P$, where $k_B = 1.38\times10^{-3}\,\mathrm{J/K}$ is the Boltzmann's constant and T is the temperature [20.12].

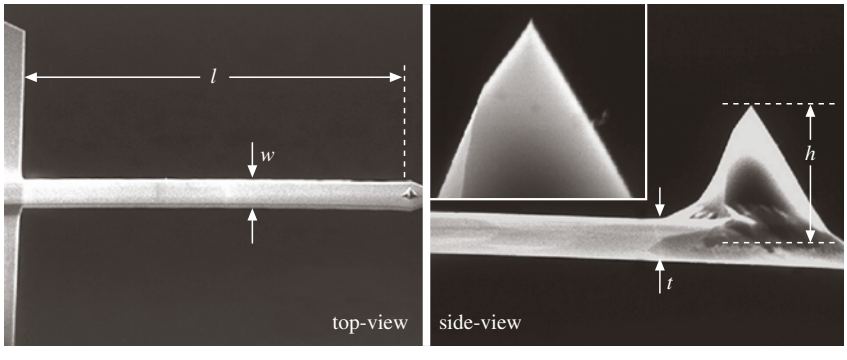


Fig. 20.2 SEM images of a rectangular cantilever. The relevant dimensions are $l = 445\,\mu\mathrm{m}$, $w = 43\,\mu\mathrm{m}$, $t = 4.5\,\mu\mathrm{m}$, $h = 14.75\,\mu\mathrm{m}$. Note that h is given by the sum of the tip height and half of the cantilever thickness. (After [20.9])

Cantilevers with different shape require finite element analysis, although in few cases analytical formulas can be derived. For V-shaped cantilevers, *Neumeister* et al. found the following approximation for the lateral spring constant c_L [20.13]:

$$c_L = \frac{Et^3}{3(1+\nu)h^2} \times \left(\frac{1}{\tan\alpha} \ln \frac{w}{d\sin\alpha} + \frac{L\cos\alpha}{w} - \frac{3\sin 2\alpha}{8} \right)^{-1}, \tag{20.5}$$

where the geometrical quantities L, w, α, d, t, and h are defined in Fig. 20.4. The expression for the normal constant is more complex and can be found in the cited reference.

Surfaces with well-defined profiles provide an alternative in situ calibration of lateral forces [20.14]. We present a slightly modified version of the method [20.15]. Figure 20.5 shows a commercial grating formed by alternated faces with opposite inclinations with respect to the scan direction. When the tip slides on the inclined planes, the normal force, F_N, and the lateral force, F_L, with respect to the surface, are different from the two components $F_\perp$ and $F_\parallel$, which are separated by the photodiode (see Fig. 20.6a).

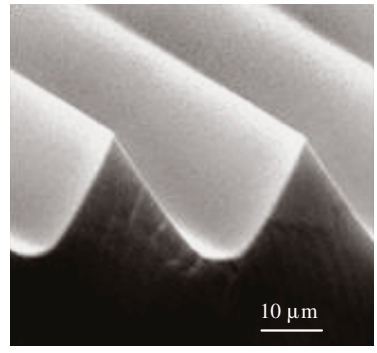


Fig. 20.5 Silicon grating formed by alternated faces angled ±55° (courtesy Silicon-MDT Ltd., POB 50, 103305, Moscow, Russia)

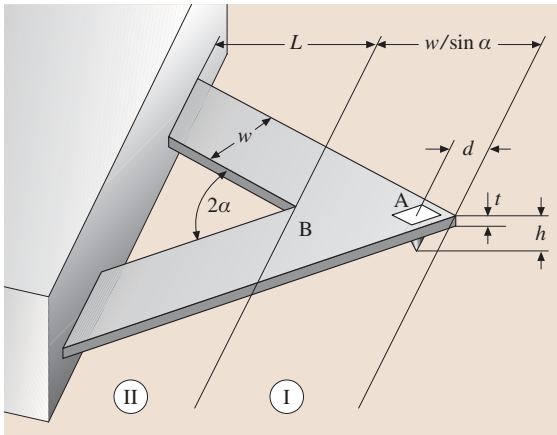


Fig. 20.4 Geometry of a V-shaped cantilever. (After [20.13])

If the linear relation $F_L = \mu F_N$ holds (see Sect. 20.5), the component $F_\parallel$ can be expressed in terms of $F_\perp$:

$$F_\parallel = \frac{\mu + \tan\theta}{1 - \mu\tan\theta} F_\perp \,. \tag{20.6}$$

The component $F_\perp$ is kept constant by the feedback loop. The sum and the difference of $F_\parallel$ on the two planes (1) and (2) are given by

$$F_+ \equiv F_\parallel^{(1)} + F_\parallel^{(2)} = \frac{2\mu\left(1+\tan^2\theta\right)}{1-\mu^2\tan^2\theta} F_\perp$$
$$F_- \equiv F_\parallel^{(1)} - F_\parallel^{(2)} = \frac{2\left(1+\mu^2\right)\tan\theta}{1-\mu^2\tan^2\theta} F_\perp \,. \tag{20.7}$$

The values of F_+ and F_- (in volts) can be measured by scanning the profile back and forth (Fig. 20.6b). If F_+ and F_- are recorded with different values of $F_\perp$, one can determine the conversion ratio between volts and nanonewtons as well as the coefficient of friction, μ.

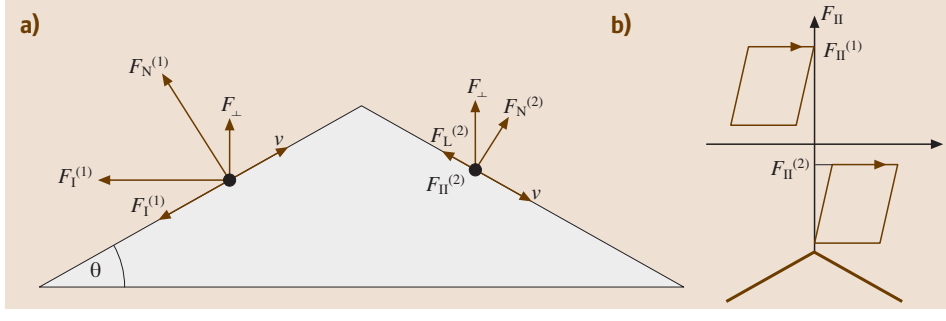


Fig. 20.6 **(a)** Forces acting on a FFM tip sliding on the grating in Fig. 20.5; **(b)** Friction loops acquired on the two faces

those between like atoms, there is a mixing instability leading to a cold welded junction.

Lancon [23.69] reported that there is the possibility for elastic instabilities under carefully designed conditions between two incommensurate gold surfaces. The importance of the interfacial interactions with respect to the intrabulk interactions was increased through the application of a pressure normal to the interface. At 4 GPa, an elastic Aubry transition was observed concurrently with finite static friction. The normal pressures employed in those simulations were well above the yield strength of gold. However, plastic flow at the contact's circumference could be avoided by the use of flat surfaces that filled out the whole simulation cell parallel to the interface. Additional examples of elastic instabilities may be found for very compliant solids, but the majority of simple solids will not pin elastically on an atomic scale. Note, however, that cold welding may still occur on long time scales due to diffusion. On such time scales, the interface geometry may rearrange to produce locally commensurate regions that lower the interfacial free energy and allow geometrical pinning.

23.2.5 Hydrodynamic Lubrication and Its Confinement-Induced Breakdown

So far, we have been concerned with dry friction. However, most surfaces are covered by dirt and/or lubricants. Hydrodynamics and elasto-hydrodynamics have been very successful in describing lubrication by micron-thick films [23.70]. These continuum theories begin to break down as atomic structure becomes important. Experiments and simulations reveal a sequence of dramatic changes in the static and dynamic properties of fluid films as their thickness decreases from microns down to molecular scales. These changes have important implications for the function of boundary lubricants and the origins of static friction.

Let us consider two flat, lubricated walls separated by a distance d and sliding with a relative parallel velocity v. Let us further assume that the lubricant's viscosity μ in the center of the contact equals that of the bulk fluid phase. The shear stress must be uniform in steady state, because any imbalance would lead to accelerations. Since the velocity gradient times the viscosity μ gives the stress in the fluid, the kinetic friction per unit area is

$$F_k = \frac{v\mu}{d+\mathcal{S}} , \qquad (23.13)$$

where we introduce the so-called slip length $\mathcal{S}$. It represents the distance into the wall at which the velocity gradient would extrapolate to zero; see Fig. 23.7f for a graphical illustration of $\mathcal{S}$. The calculation of slip length from velocity profiles has some ambiguity, and we refer to work by *Bocquet* et al. [23.71] for the least ambiguous resolution to the definition of the slip length known to us.

Macroscopic experiments are generally well-described by a "no-slip" boundary condition (BC), where $\mathcal{S} = 0$ and the tangential component of the fluid velocity equals that of the solid at the surface. The one prominent exception is contact-line motion, where an interface between two fluids moves along a solid surface. This motion would require an infinite force in hydrodynamic theory, unless slip occurred near the contact line.

As d decreases, any deviation from $\mathcal{S} = 0$ becomes more evident in (23.13). Studies of confined films show

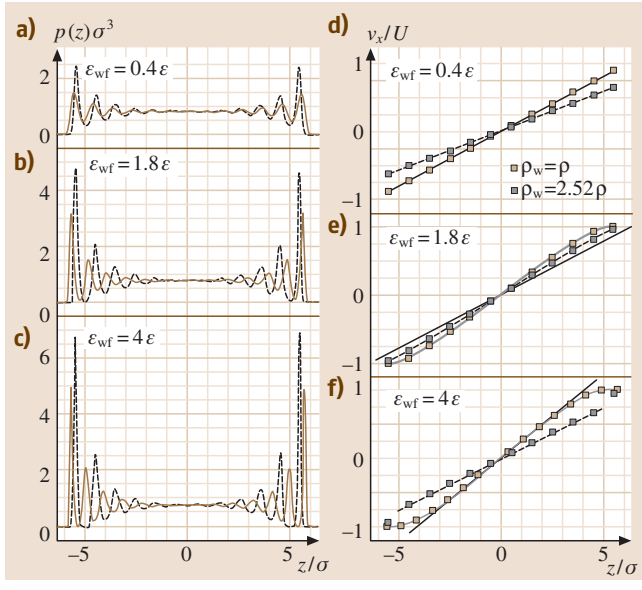


Fig. 23.7a–f Panels **(a)–(c)** show density as a function of position z relative to two walls whose atoms are centered at $z/\sigma = \pm 6.4$. Values of ϵ_{wf} are indicated. *Solid lines* are for equal wall and fluid densities ($\rho = 0.81\sigma^{-3}$), and *dotted lines* are for $\rho_w = 2.52\rho$. *Squares* in panels **(d)–(f)** show the average velocity in each layer as a function of z, and *solid* and *dotted lines* are fits through these values for low and high density walls, respectively. For this data, the walls were moved in opposite directions at speed $U = 1\sigma/t_{LJ}$. The *dashed line* in **(e)** represents the flow expected from hydrodynamics with a no-slip BC ($\mathcal{S} = 0$). As shown in **(f)**, the slope (*dashed line*) far from the walls is used to define $\mathcal{S}$. (From [23.8] after [23.68])

that slip may occur ($\mathcal{S} > 0$) or layers may stick onto the solid ($\mathcal{S} < 0$) [23.68, 71–73]. These deviations from the no-slip BC have been correlated with structural changes in the fluid near the walls. The first observation was that layering occurs in planes parallel to the walls. It is induced by the sharp cutoff in fluid density at the wall and the pair correlation function $g(r)$ between fluid atoms. An initial fluid layer forms at the preferred wall-fluid spacing. Additional fluid molecules tend to lie in a second layer, at the preferred fluid-fluid spacing. This layer induces a third, and so on. Layering can alter the BC and the viscosity of the confined fluid. As suggested by the contrasting behavior of commensurate and incommensurate solids, the BC depends sensitively on the degree to which the walls imprint their in-plane density modulations into the lubricant.

Some of the trends that are observed in the degree and extent of layering are illustrated in Fig. 23.7 [23.68, 72]. The fluid density is plotted as a function of the distance between walls for a model that is also considered in almost all of the other studies of flow BCs described below. The fluid consists of spherical molecules interacting through a Lennard-Jones potential (23.1) with characteristic energy and length scales ϵ and σ. They are confined by crystalline walls containing discrete atoms. In this case, the walls were planar (001) surfaces of an fcc crystal. Wall and fluid atoms also interact with a Lennard-Jones potential, but with a different binding energy ϵ_{wf}.

The net adsorption potential from the walls can be increased by raising ϵ_{wf}, or by increasing the density of the walls ρ_w, so that more wall atoms attract the fluid. Figure 23.7 shows that both increases lead to increases in the height of the first density peak. The height also increases with the pressure in the fluid [23.73], since that forces atoms into steeper regions of the adsorption potential. The height of subsequent density peaks decreases smoothly with distance from the wall, and only four or five well-defined layers are seen near each wall in Fig. 23.7. The rate at which the density oscillations decay is determined by the decay length of structure in the bulk pair-correlation function of the fluid. Since all panels of Fig. 23.7 have the same conditions in the bulk, the decay rate is the same. The adsorption potential only determines the initial height of the peaks.

As mentioned above, solid surfaces also induce density modulations within the plane of the layers [23.7]. These modulations can be quantified by the 2-D static structure factor at the shortest reciprocal lattice vector Q of the substrate [23.35, 68]. When normalized by the number of atoms in the layer, N_l, this becomes an intensive variable that would correspond to the Debye–Waller factor in a crystal. The maximum possible value, $S(Q)/N_l = 1$, corresponds to fixing all atoms exactly at crystalline lattice sites. In a strongly ordered case such as $\rho_w = \rho$ in Fig. 23.7c, the small oscillations about lattice sites in the first layer only decrease $S(Q)/N_l$ to 0.71. This is well above the value of 0.6 that is typical of bulk 3-D solids at their melting point [23.74] and indicates that the first layer has crystallized onto the wall. The values of $S(Q)/N_l$ in the second and third layers are 0.31 and 0.07, respectively, and atoms in these layers exhibit typical fluid diffusion.

There is some correlation between the factors that produce strong layering and those that produce strong in-plane modulations. For example, chain molecules have several conflicting length scales that tend to frustrate both layering and in-plane order [23.36, 75, 76]. However, the dependence of in-plane order on the density of substrate atoms is more complicated than for layering. When $\rho_w = \rho$, the fluid atoms can naturally sit on the sites of a commensurate lattice, and $S(Q)$ is large. When the substrate density ρ_w is increased by a factor of 2.52, the fluid atoms no longer fit easily into the corrugation potential. The degree of induced in-plane order drops sharply, although the layering becomes stronger (Fig. 23.7). Sufficiently strong adsorption potentials may eventually lead to crystalline order in the first layers and stronger layering. However, this may actually increase slip.

Thompson et al. [23.68] found that all of their results for $\mathcal{S}$ collapsed onto a universal curve when plotted against the structure factor $S(Q)/N_l$. When one or more layers crystallized onto the wall, the same collapse could be applied as long as the effective wall position was shifted by a layer and the Q for the outer wall layer was used. The success of this collapse at small $S(Q)/N_l$ can be understood from a perturbation theory [23.6, 77] developed to quantitatively explain the kinetic friction on adsorbed monolayers [23.3].

One of the surprising features of Fig. 23.7 is that the viscosity remains the same even in regions near the wall where there is pronounced layering. Any change in viscosity would produce a change in the velocity gradient, since the stress is constant. However, the flow profiles in panel (d) remain linear throughout the cell. The profile for the dense walls in panel (f) is linear up to the last layer, which has crystallized onto the wall. From Fig. 23.7 it is apparent that local density variations by at least a factor of seven can be accommodated without a viscosity change.

Bitsanis et al. [23.78] examined the change in viscosity with film thickness. They found that results for film thicknesses $h > 4\sigma$ could be fit using the bulk viscosity for the average density. However, as h decreased below 4σ, the viscosity diverged much more rapidly than any model based on bulk viscosity could explain. These observations were consistent with surface force apparatus (SFA) experiments on nanometer thick films of a wide variety of small molecules [23.4, 5]. Layering in molecularly thin films gave rise to oscillations in the energy, normal force, and effective viscosity as the film thickness decreased. Most films of one to three molecular layers exhibited a yield stress characteristic of solid-like behavior, even though the molecules form a simple Newtonian fluid in the bulk. Deviations from the no-slip condition might cause μ_{eff} to differ from the bulk viscosity by an order of magnitude. However, they could not explain the observed changes of μ_{eff} by more than five orders of magnitude, or the even more dramatic changes by 10 to 12 orders of magnitude in the characteristic viscoelastic relaxation time determined from the shear rate dependence of μ_{eff} [23.5].

More recent experimental and theoretical studies confirm that most fluids exhibit a glass transition when confined to channels less than 3 to 10 molecular diameters [23.33, 79]. *Thompson* et al. [23.34, 36] considered changes in flow with film thickness in a wide range of model systems. As in experiment, they found dramatic divergences in the viscosity and relaxation time that suggested confinement induced a glass transition. Later work by *Baljon* et al. [23.33, 80] indicates that the same glass transition is produced by changes in thickness, pressure, or temperature. In particular, the shear-rate-dependent viscoelastic response of systems of all sizes, pressures, and temperatures can be collapsed onto the same universal curve above the glass transition [23.33].

23.2.6 Submonolayer Films

Physisorbed molecules such as short hydrocarbon chains can be expected to sit on any surface exposed to atmospheric conditions. Even in ultrahigh vacuum, special surface treatments are needed to remove strongly physisorbed species from surfaces. Given the generic observation of glassy behavior in thin films, *Robbins* et al. [23.81] suggested that these submonolayer contaminants could provide a general explanation for the universal observation of static friction. Subsequent simulations have confirmed that physisorbed molecules qualitatively alter tribological behavior. For example, they lead to static friction between two incommensurate walls [23.67, 82] or two disordered walls [23.47] that would slide without friction if they were clean. The theoretical explanation of this effect can be understood from a simple analytical model considered by *Müser* et al. [23.47]. In order to make two surfaces pin, a nonvanishing fraction of the density modulations must share the same period and lock into phase. This happens automatically for commensurate surfaces, but not for incommensurate surfaces unless they are unrealistically compliant. However, a submonolayer of atoms or molecules that does not form strong covalent bonds with the confining solids can simultaneously lock to the density modulations of two incommensurate walls and hence give rise to a finite friction coefficient.

Simulation results for the friction due to physisorbed layers are consistent with a large number of experimental observations. While Amontons's laws state that the friction is independent of the macroscopic area, it is well established that the friction is proportional to the real area of molecular contact A_{real} [23.30, 83, 84]. Experiments where both the friction and molecular area can be measured [23.4, 85–89] show that the local shear stress τ_s needed to initiate sliding rises linearly with the normal pressure P:

$$\tau_s = \tau_0 + \alpha P \,. \tag{23.14}$$

This gives a macroscopic friction force $F_s = \tau_0 A_{\text{real}} + \mu L$, where the total load is just $L = PA_{\text{real}}$. This friction law is actually more general than Amontons's laws, reducing to it at large loads and small areas, but explaining why the friction between adhesive surfaces (like tape) does not vanish at zero load.

For a wide range of parameters [23.82, 90], the friction calculated for physisorbed monolayers follows (23.14) up to the gigapascal pressures that are present in real contacts. Moreover, the results are remarkably insensitive to parameters that are not controlled in experiments such as the crystallographic orientation of the solids, the thickness of the adsorbed layer, and the direction of sliding. The only exceptions occur for the unlikely case of commensurate walls [23.90, 91], as discussed further below. Subsequent studies of kinetic friction [23.39, 92] explain both the close relation between experimental values of static friction and the slow variation with velocity. In particular, the logarithmic rate dependence seen in τ for many systems [23.83] can be explained through thermally activated hopping of physisorbed molecules [23.92].

Müser et al. [23.47] have constructed a simple analytical theory that captures many of the results and conclusions from the above simulations. At the pres-

sures of interest, the atoms act like hard spheres. In order to slide, the bounding surfaces must lift up over these spheres against the normal pressure. The situation is analogous to the simple cartoon in Fig. 23.2, but with atomic-scale ramps, and the required stress increases linearly with pressure for the same reason. The offset, τ_0, is related to adhesive interactions between the two surfaces that increase the effective load [23.90]. Unlike the original model of geometrical interlocking, the atomic-scale ramps rearrange during sliding. The dissipation during these rearrangements is responsible for kinetic friction [23.47, 92], which, as we now discuss, is very different for commensurate and incommensurate walls.

The characteristic motion of individual boundary lubricant atoms depends on the relative displacement for commensurate walls, because the distribution of metastable states depends on the relative alignment of the two walls [23.90, 92]. For example, no "pops" of the lubricant atoms are observed when the two commensurate solids are aligned directly above each other. This behavior is in contrast with that for incommensurate surfaces for which the distribution of metastable states does not change as the slider is translated with respect to the substrate. Stressing the analogy to the Prandtl–Tomlinson model, static and kinetic friction are related for incommensurate, boundary-lubricated solids, while they are not necessarily related for commensurate, boundary-lubricated solids. Indeed, commensurate, boundary-lubricated surfaces exhibit a dramatic decrease in friction at the transition from stick-slip motion to smooth sliding, while this dramatic decrease is absent for incommensurate, boundary-lubricated walls [23.39].

It has been noted recently, that dimensionality not only plays an important role for elastic objects, but also for boundary lubricants [23.39]. It was predicted that kinetic friction would be Stokes-like (i. e., vanish linearly with velocity) between incommensurate surfaces, if it were possible to confine lubricant particles into trenches of (sub)nanometer depth and width, so that the motion of the lubricant embedded between the surfaces becomes 1-D. At the same time, however, static friction would remain finite.

The above results suggest that adsorbed molecules and other "third-bodies" may prove key to understanding macroscopic friction measurements. It will be interesting to extend these studies to more realistic molecular potentials and to rough surfaces. To date, realistic potentials have only been used for "between-sorbed" particles between commensurate surfaces. This work is reviewed elsewhere [23.8, 10, 11].

23.3 Stick–Slip Dynamics

The dynamics of sliding systems can be very complex and depend on many factors, including the types of metastable states in the system, the times needed to transform between states, and the mechanical properties of the device that imposes the stress. At high rates or stresses, systems usually slide smoothly. At low rates, the motion often becomes intermittent, with the system alternately sticking and slipping forward [23.30, 32]. Everyday examples of such stick-slip motion include the squeak of hinges and the music of violins.

The alternation between stuck and sliding states of the system reflects changes in the way energy is stored. While the system is stuck, elastic energy is pumped into the system by the driving device. When the system slips, this elastic energy is released into kinetic energy and eventually dissipated as heat. The system then sticks once more, begins to store elastic energy, and the process continues. Both elastic and kinetic energy can be stored in all the mechanical elements that drive the system. The whole coupled assembly must be included in any analysis of the dynamics.

The simplest type of intermittent motion is the atomic-scale stick-slip that occurs in the multistable regime ($\lambda > 1$) of the Prandtl–Tomlinson model. Energy is stored in the springs while atoms are trapped in a metastable state, and converted to kinetic energy as they pop to the next metastable state. This phenomenon is quite general and has been observed in many simulations of wearless friction [23.8, 10], as well as in the motion of atomic force microscope tips [23.1, 2]. In these cases, motion involves a simple ratcheting over the surface potential through a regular series of hops between neighboring metastable states. The slip distance is determined entirely by the periodicity of the surface potential.

Many examples of stick-slip involve a rather different type of motion that can lead to intermittency and chaos [23.93–97]. Instead of jumping between neighboring metastable states, the system slips for very long distances before sticking. For example, *Gee* et al. [23.4] and *Yoshizawa* et al. [23.98] observed slip distances of many microns in their studies of confined films with the

SFA. This distance is much larger than any characteristic periodicity in the potential and varied with velocity, load, and the mass and stiffness of the SFA. The fact that the SFA does not stick after moving by a lattice constant indicates that sliding has changed the state of the system in some manner, so that it can continue sliding even at forces less than the yield stress.

One simple property that depends on past history is the amount of stored kinetic energy. This can carry a system over potential energy barriers even when the stress is below the yield stress. The simplest example is the underdamped Prandtl–Tomlinson model, which has been thoroughly studied in the mathematically equivalent case of an underdamped Josephson junction [23.99]. One finds a hysteretic response function, where static and moving steady-states coexist over a range of forces between $F_{\min}$ and the static friction F_s. There is a minimum stable steady-state velocity $v_{\min}$ corresponding to $F_{\min}$. At lower velocities, the only steady state is linearly unstable because $\partial v/\partial F < 0$ – pulling harder slows the system [23.30, 32]. If the top wall of the Tomlinson model is pulled at an average velocity less than $v_{\min}$ by a sufficiently compliant system, it will exhibit large-scale stick-slip motion.

Experimental systems typically have many other internal degrees of freedom that may depend on past history. Macroscopic models normally assume that a single unspecified state variable is relevant and add dynamical rules for this variable to create a phenomenological rate-state model [23.94, 100]. This approach has captured experimental behavior from a wide range of systems, including surface force apparatus experiments [23.38, 101], paper sliding on paper [23.93], and earthquakes [23.102]. However, in most cases, there is no microscopic model for the state variable and its time dependence.

One set of systems where microscopic pictures for the state variable have been considered are confined films in SFA experiments. Confined films have structural degrees of freedom that can change during sliding, and this provides an alternative mechanism for stick-slip motion [23.35]. Some of these structural changes are illustrated in Fig. 23.8, which shows stick-slip motion of a two-layer film of simple spherical molecules. The bounding walls were held together by a constant normal load. A lateral force was applied to the top wall through a spring k attached to a stage that moved with fixed velocity v in the x direction. The equilibrium configuration of the film at $v = 0$ is a commensurate crystal that resists shear. Thus at small times, the center of mass of the top wall remains pinned at $X = 0$. The force grows linearly

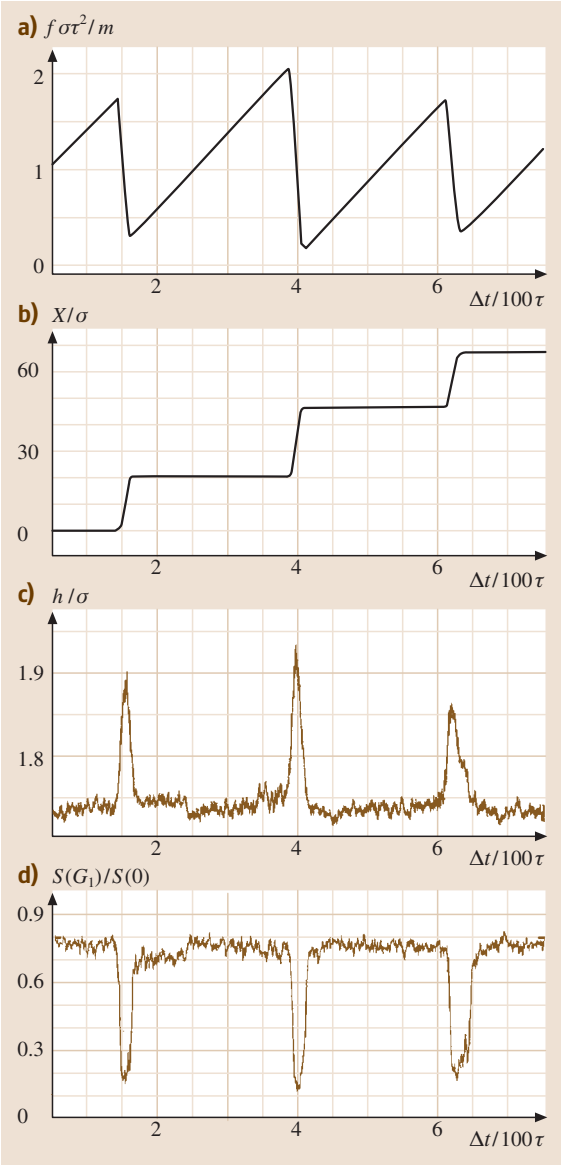


Fig. 23.8a–d Time dependence of **(a)** force per unit area f, **(b)** lateral position of the top wall X, **(c)** wall spacing h, and **(d)** degree of crystallinity $S(G_1)/S(0)$ as a function of time during stick-slip motion of walls confining two layers of spherical molecules. (After [23.35])

with time, $F = kv$, as the stage advances ahead of the wall. When F exceeds F_s, the wall slips forward. The force drops rapidly because the slip velocity $\dot{X}$ is much greater than v. When the force drops sufficiently, the

film recrystallizes, the wall stops, and the force begins to rise once more.

One structural change that occurs during each slip event is dilation by about 10% (Fig. 23.8c). *Dhinojwala* et al. [23.103] have recently confirmed that dilation occurs during slip in SFA experiments. The increased volume makes it easier for atoms to slide past each other and is part of the reason the sliding friction is lower than F_s. The system may be able to keep sliding in this dilated state as long as it takes more time for the volume to contract than for the wall to advance by a lattice constant. Dilation of this type plays a crucial role in the yield, flow, and stick-slip dynamics of granular media [23.104–106]. In some simulations, the dilation is localized at a single plane – either a wall/film interface or a plane within the film [23.33, 107]. In some cases, there is also in-plane ordering of the film to enable it to slide more easily over the wall. This ordering remains after sliding stops and provides a mechanism for the long-term memory seen in some experiments [23.4, 98, 108].

The degree of crystallinity throughout the film may also change during sliding. Deviations from an ideal crystalline structure can be quantified by the Debye–Waller factor $S(G_1)/S(0)$ (Fig. 23.8d), where G_1 is one of the shortest reciprocal lattice vectors and $S(0) = N$ is the total number of atoms in the film. When the system is stuck, $S(G_1)/S(0)$ has a large value that is characteristic of a 3-D crystal. During each slip event, $S(G_1)/S(0)$ drops dramatically. The minimum values are characteristic of simple fluids that would show a no-slip boundary condition (Sect. 23.2.5). The atoms also exhibit rapid diffusion that is characteristic of a fluid. The periodic melting and freezing transitions that occur during stick-slip are induced by shear, not by the negligible changes in temperature. Shear-melting transitions at constant temperature have been observed in both theoretical and experimental studies of bulk colloidal systems [23.29, 109]. While the above simulations of confined films used a fixed number of particles, *Lupowski* et al. [23.110] found equivalent results at fixed chemical potential.

Very similar behavior has been observed in simulations of sand [23.104], chain molecules [23.33], incommensurate or amorphous walls [23.35], and minimal models with single molecules between surfaces [23.95–97]. These systems transform between glassy and fluid states during stick-slip motion. As in equilibrium, the structural differences between glass and fluid states are small. However, there are strong changes in the self-diffusion and other dynamic properties when the film goes from the static glassy to sliding fluid state. *Buldum* and *Ciraci* [23.111] found stick-slip motion due to periodic structural transformations in the bottom layers of a pyramidal Ni(111) tip sliding on an incommensurate Cu(110) surface.

23.4 Conclusions

The discussion of computer simulations in this chapter aimed to build a general atomistic understanding of tribological processes rather than to elucidate a specific sliding system. The advantage of studying simplified models, as opposed to chemically realistic systems, is that system properties can be varied at will. This allows the role of adhesion, surface curvature, contaminant layers and other factors to be isolated and explored. For example, to explore the effects of adhesion one can compare results from a simulation that includes the long-range attractive interactions between atoms with those from a simulation where all attractive interactions between surfaces are eliminated, leaving all other parameters unchanged. Care should be taken to avoid simplifications that are known to yield atypical results. In particular, the use of commensurate, perfectly aligned, crystalline surfaces has been shown to produce qualitatively different results, even in the presence of a lubricating film! Surface curvature must also be included in a way that properly models stress distributions from the bounding solids.

One central question is how contacting surfaces lock together to produce a static friction force. This is particularly perplexing in the context of many nanotribological systems where friction occurs without significant plastic deformation or wear. Many theories for wearless solid friction assume that surfaces deform elastically to conform to each other, and that atoms advance in a series of elastic instabilities during sliding. However, the simulations discussed above show that simple solids should hardly ever show elastic instabilities on an atomic scale without concurrent plastic flow or material transfer. Despite extensive efforts to design an atomistic system whose static friction mechanism is entirely elastic, no such model could be found as long as the ingredients were: three-dimensional incommensurate solids,

flat or smoothly curved contacts, and two-body interactions between all atoms. Under these conditions the relative lateral motion is only opposed by a viscous-like force that goes to zero for small sliding velocities. When interfacial interactions are increased in an attempt to produce elastic instabilities, plastic flow is also observed.

Wearless friction between clean solid surfaces is only found in simulations of very small contacts. The geometric interlocking in such contacts is particularly relevant to AFM experiments. When only a few atoms on the tip make contact, the potential from the substrate does not average to zero. Static friction can then occur without elastic instabilities within the tip. The Prandtl–Tomlinson model can be applied to these systems with the cantilever stiffness replacing the internal stiffness of the material in the tip. As the diameter of the contact increases, due to blunting of the tip or increased load, the friction coefficient produced by geometric interlocking decreases. In the specific case of non-adhesive contacts, the friction between an amorphous tip and a clean crystalline substrate scales as $F \propto L^{2/3}$ and μ vanishes as $L^{-1/3}$. Thus geometric interlocking is not expected to yield significant friction when the contact area contains more than $\sim 10^3$ atoms.

Most contacting solids are separated by contaminant layers, or third bodies, that introduce many additional structural degrees of freedom. The intervening atoms can rearrange into a local free energy minimum for any geometry of the bounding solids without paying the large elastic energy cost that limits rearrangements within these solids. Simulations that include these layers reproduce many details of SFA experiments. They also provide an explanation for Amontons's laws, and exceptions to them, in larger contacts.

Many important problems remain to be addressed by simulations. Of particular interest are the roles of realistic, random surface roughness and subsurface plastic deformation. Studies of these phenomena will require new multiscale approaches that treat long-range elastic and plastic deformations in the solids. Other important problems will require extensions in the time available to simulations. Examples include cold-welding through inter-surface diffusion, wear, and fatigue. Addressing these problems will be a continuing challenge to the computer modeling community.

References

23.1 C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang: Atomic-scale friction of a tungsten tip on a graphite surface, Phys. Rev. Lett. **59** (1987) 1942–1945

23.2 R. W. Carpick, M. Salmeron: Scratching the surface: Fundamental investigations of tribology with atomic force microscopy, Chem. Rev. **97** (1997) 1163–1194

23.3 J. Krim: Friction at the atomic scale, Sci. Am. **275**(4) (1996) 74–80

23.4 M. L. Gee, P. M. McGuiggan, J. N. Israelachvili, A. M. Homola: Liquid to solid transitions of molecularly thin films under shear, J. Chem. Phys. **93** (1990) 1895–1906

23.5 S. Granick: Motions and relaxations of confined liquids, Science **253** (1992) 1374–1379

23.6 M. Cieplak, E. D. Smith, M. O. Robbins: Molecular origins of friction: The force on adsorbed layers, Science **265** (1994) 1209–1212

23.7 M. Schoen, C. L. Rhykerd, D. J. Diestler, J. H. Cushman: Fluids in micropores: I. Structure of a simple classical fluid in a slit-pore, J. Chem. Phys. **87** (1987) 5464–5476

23.8 M. O. Robbins, M. H. Müser: Computer simulations of friction, lubrication and wear. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 717–765

23.9 M. H. Müser, M. Urbakh, M. O. Robbins: Statistical mechanics of static and low-velocity kinetic friction, Adv. Chem. Phys. **126** (2003) 187–272

23.10 J. A. Harrison, S. J. Stuart, D. W. Brenner: Atomic-scale simulation of tribological and related phenomena. In: *Handbook of Micro/Nanotribology*, ed. by B. Bhushan (CRC, Boca Raton 1999) pp. 525–594

23.11 J. A. Harrison, G. T. Gao, R. J. Harrison, G. M. Chateauneuf, P. T. Mikulski: The friction of model self-assembled monolayers. In: *Encyclopedia of Nanoscience and Nanotechnology*, ed. by H. S. Nalwa (American Scientific, Los Angeles 2004)

23.12 M. P. Allen, D. J. Tildesley: *Computer Simulation of Liquids* (Clarendon, Oxford 1987)

23.13 D. Frenkel, B. Smit: *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego 1996)

23.14 J. P. Ryckaert, A. Bellemans: Molecular dynamics of liquid alkanes, Faraday Discuss. **66** (1978) 95–106

23.15 W. Paul, D. Y. Yoon, G. D. Smith: An optimized united atom model for simulations of polymethylene melts, J. Chem. Phys. **103** (1995) 1702–1709

23.16 W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn: Simulation of polymer melts. I. Coarse graining procedure for polycarbonates, Acta Polym. **49** (1998) 61–74

23.17 W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn: Simulation of polymer melts. II. From coarse grained models back to atomistic description, Acta Polym. **49** (1998) 75–79

23.18 P. Flory: *Statistical Mechanics of Chain Molecules* (Hanser, München 1988)

23.19 K. Binder: *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford, New York 1995)

23.20 D.W. Brenner, O.A. Shenderova, J.A. Harrison, S.J. Stuart, B. Ni, S.B. Sinnot: Second generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons, J. Phys. C **14** (2002) 783–802

23.21 S.J. Stuart, A.B. Tutein, J.A. Harrison: A reactive potential for hydrocarbons with intermolecular interactions, J. Chem. Phys. **112** (2000) 6472–6486

23.22 M.S. Daw, M.I. Baskes: Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals, Phys. Rev. B **29** (1984) 6443–6453

23.23 S. Nosé: Constant temperature molecular dynamics methods, Prog. Theor. Phys. Supp. **103** (1991) 1–46

23.24 T. Schneider, E. Stoll: Molecular dynamics study of a three-dimensional one-component model for distortive phase transitions, Phys. Rev. B **17** (1978) 1302–1322

23.25 G.S. Grest, K. Kremer: Molecular dynamics simulations for polymers in the presence of a heat bath, Phys. Rev. A **33** (1986) 3628–3631

23.26 S.S. Sarman, D.J. Evans, P.T. Cummings: Recent developments in non-Newtonian molecular dynamics, Phys. Rep. **305** (1998) 1–92

23.27 D.J. Evans, G.P. Morriss: Shear thickening and turbulence in simple fluids, Phys. Rev. Lett. **56** (1986) 2172–2175

23.28 W. Loose, G. Ciccotti: Temperature and temperature control in nonequilibrium-molecular-dynamics simulations of the shear flow of dense liquids, Phys. Rev. A **45** (1992) 3859–3866

23.29 M.J. Stevens, M.O. Robbins: Simulations of shear-induced melting and ordering, Phys. Rev. E **48** (1993) 3778–3792

23.30 F.P. Bowden, D. Tabor: *The Friction and Lubrication of Solids* (Clarendon, Oxford 1986)

23.31 R. Khare, J.J. de Pablo, A. Yethiraj: Rheology of confined polymer melts, Macromolecules **29** (1996) 7910–7918

23.32 E. Rabinowicz: *Friction and Wear of Materials* (Wiley, New York 1965)

23.33 M.O. Robbins, A.R.C. Baljon: Response of thin oligomer films to steady and transient shear. In: *Microstructure and Microtribology of Polymer Surfaces*, ed. by V.V. Tsukruk, K.J. Wahl (American Chemical Society, Washington 2000) pp. 91–117

23.34 P.A. Thompson, G.S. Grest, M.O. Robbins: Phase transitions and universal dynamics in confined films, Phys. Rev. Lett. **68** (1992) 3448–3451

23.35 P.A. Thompson, M.O. Robbins: Origin of stick-slip motion in boundary lubrication, Science **250** (1990) 792–794

23.36 P.A. Thompson, M.O. Robbins, G.S. Grest: Structure and shear response in nanometer-thick films, Israel J. Chem. **35** (1995) 93–106

23.37 M. Porto, V. Zaloj, M. Urbakh, J. Klafter: Macroscopic versus microscopic description of friction: From Tomlinson model to shearons, Tribol. Lett. **9** (2000) 45–54

23.38 B.N.J. Persson: *Sliding Friction: Physical Principles and Applications* (Springer, Berlin 1998)

23.39 M.H. Müser: Nature of mechanical instabilities and their effect on kinetic friction, Phys. Rev. Lett. **89** (2002) 224301: 1–4

23.40 M.H. Müser: Dry friction between flat surfaces: Wearless multistable elasticity vs. material transfer and plastic deformation, Tribol. Lett. **10** (2001) 15–22

23.41 J.B. Sokoloff: Static friction between elastic solids due to random asperities, Phys. Rev. Lett. **86** (2001) 3312–3315

23.42 B.N.J. Persson, E. Tosatti: Theory of friction: Elastic coherence length and earthquake dynamics, Solid State Commun. **109** (1999) 739–744

23.43 F.F. Abraham, N. Bernstein, J.Q. Broughton, D. Hess: Dynamic fracture of silicon: Concurrent simulation of quantum electrons, classical atoms, and the continuum solid, MRS Bull. **25** (2000) 27–32

23.44 L. Wenning, M.H. Müser: Friction laws for elastic nano-scale contacts, Europhys. Lett. **54** (2001) 693–699

23.45 U. Landman, W.D. Luedtke, N.A. Burnham, R.J. Colton: Atomistic mechanisms and dynamics of adhesion, nanoindentation, and fracture, Science **248** (1990) 454–461

23.46 D. Dowson: *History of Tribology* (Longman, New York 1979)

23.47 M.H. Müser, L. Wenning, M.O. Robbins: Simple microscopic theory of Amontons's laws for static friction, Phys. Rev. Lett. **86** (2001) 1295–1298

23.48 M. Enachescu, R.J.A. van den Oetelaar, R.W. Carpick, D.F. Ogletree, C.J.F. Flipse, M. Salmeron: An AFM study of an ideally hard contact: The diamond (111)/tungsten carbide interface, Phys. Rev. Lett. **81** (1998) 1877–1880

23.49 C.A.J. Putman, M. Igarashi, R. Kaneko: Single-asperity friction in friction force microscope: The composite model, Appl. Phys. Lett. **66** (1995) 3221–3223

23.50 P. Schwarz, O. Zwörner, P. Köster, R. Wiesendanger: Quantitative analysis of the frictional properties of solid materials at low loads. II. Mica and germanium sulfide, Phys. Rev. B **56** (1997) 6987–7000

23.51 L. Prandtl: Ein Gedankenmodell zur kinetischen Theorie der festen Körper, Z. Angew. Math. Mech. **8** (1928) 85–106

23.52 G. A. Tomlinson: A molecular theory of friction, Philos. Mag. Series **7** (1929) 905–939
23.53 G. M. McClelland, S. R. Cohen: Tribology at the atomic scale. In: *Chemistry and Physics of Solid Surfaces*, Vol. VII, ed. by R. Vanselow, R. Rowe (Springer, Berlin, Heidelberg 1990) pp. 419–445
23.54 Y. I. Frenkel, T. Kontorova: On the theory of plastic deformation and twinning, Zh. Eksp. Teor. Fiz. **8** (1938) 1340
23.55 O. M. Braun, Y. S. Kivshar: Nonlinear dynamics of the Frenkel–Kontorova model, Phys. Rep. **306** (1998) 1–108
23.56 M. O. Robbins: Jamming, friction, and unsteady rheology. In: *Jamming and Rheology: Constrained Dynamics on Microscopic and Macroscopic Scales*, ed. by A. J. Liu, S. R. Nagel (Taylor and Francis, London 2000)
23.57 L. W. Bruch, M. W. Cole, E. Zaremba: *Physical Adsorption: Forces and Phenomena* (Oxford, New York 1997)
23.58 F. C. Frank, J. H. van der Merwe: One-dimensional dislocations. I. Static theory, Proc. R. Soc. A **198** (1949) 205–225
23.59 S. Aubry: The new concept of transitions by breaking of analyticity in a crystallographic model. In: *Solitons and Condensed Matter Physics*, ed. by A. R. Bishop, T. Schneider (Springer, Berlin, Heidelberg 1979) pp. 264–290
23.60 M. Hirano, K. Shinjo: Atomistic locking and friction, Phys. Rev. B **41** (1990) 11837–11851
23.61 J. Belak, I. F. Stowers: The indentation and scraping of a metal surface: A molecular dynamics study. In: *Fundamentals of Friction: Macroscopic and Microscopic Processes*, ed. by I. L. Singer, H. M. Pollock (Kluwer, Dordrecht 1992) pp. 511–520
23.62 M. R. Sørensen, K. W. Jacobsen, P. Stoltze: Simulations of atomic-scale sliding friction, Phys Rev. B **53** (1996) 2101–2113
23.63 A. N. Kolmogorov, V. H. Crespi: Smoothest bearings: Interlayer sliding in multiwalled carbon nanotubes, Phys. Rev. Lett. **85** (2000) 4727–4730
23.64 A. Volmer, T. Natterman: Towards a statistical theory of solid dry friction, Z. Phys. B **104** (1997) 363–371
23.65 C. Mak, J. Krim: Quartz-crystal microbalance studies of disorder induced lubrication, Faraday Discuss. **107** (1997) 389–397
23.66 K. W. Jacobsen, J. K. Norskov, M. J. Puska: Interatomic interactions in the effective-medium theory, Phys. Rev. B **35** (1987) 7423–7442
23.67 M. H. Müser, M. O. Robbins: Conditions for static friction between flat crystalline surfaces, Phys. Rev. B **61** (2000) 2335–2342
23.68 P. A. Thompson, M. O. Robbins: Shear flow near solids: Epitaxial order and flow boundary conditions, Phys. Rev. A **41** (1990) 6830–6837
23.69 F. Lancon: Aubry transition in a real material: Prediction for its existence in an incommensurate gold/gold interface, Eur. Phys. Lett. **57** (2002) 74–79
23.70 D. Dowson, G. R. Higginson: *Elastohydrodynamic Lubrication* (Pergamon, Oxford 1968)
23.71 L. Bocquet, J.-L. Barrat: Hydrodynamic boundary conditions and correlation functions of confined fluids, Phys. Rev. Lett. **70** (1993) 2726–2729
23.72 P. A. Thompson, M. O. Robbins: Simulations of contact-line motion: Slip and the dynamic contact angle, Phys. Rev. Lett. **63** (1989) 766–769
23.73 J.-L. Barrat, L. Bocquet: Large slip effect at a nonwetting fluid-solid interface, Phys. Rev. Lett. **82** (1999) 4671–4674
23.74 M. J. Stevens, M. O. Robbins: Melting of Yukawa systems: A test of phenomenological melting criteria, J. Chem. Phys. **98** (1993) 2319–2324
23.75 J. Gao, W. D. Luedtke, U. Landman: Structure and solvation forces in confined films: Linear and branched alkanes, J. Chem. Phys. **106** (1997) 4309–4318
23.76 A. Koike, M. Yoneya: Chain length effects on frictional behavior of confined ultrathin films of linear alkanes under shear, J. Phys. Chem. B **102** (1998) 3669–3675
23.77 E. D. Smith, M. Cieplak, M. O. Robbins: The friction on adsorbed monolayers, Phys. Rev. B **54** (1996) 8252–8260
23.78 I. Bitsanis, S. A. Somers, H. T. Davis, M. Tirrell: Microscopic dynamics of flow in molecularly narrow pores, J. Chem. Phys. **93** (1990) 3427–3431
23.79 A. L. Demirel, S. Granick: Glasslike transition of a confined simple fluid, Phys. Rev. Lett. **77** (1996) 2261–2264
23.80 A. R. C. Baljon, M. O. Robbins: Energy dissipation during rupture of adhesive bonds, Science **271** (1996) 482–484
23.81 M. O. Robbins, E. D. Smith: Connecting molecular-scale and macroscopic tribology, Langmuir **12** (1996) 4543–4547
23.82 G. He, M. H. Müser, M. O. Robbins: Adsorbed layers and the origin of static friction, Science **284** (1999) 1650–1652
23.83 J. H. Dieterich, B. D. Kilgore: Direct observation of frictional contacts: New insights for state-dependent properties, Pure Appl. Geophys. **143** (1994) 238–302
23.84 P. Berthoud, T. Baumberger: Shear stiffness of a solid-solid multicontact interface, Proc. R. Soc. Lond. A **454** (1998) 1615–1634
23.85 A. L. Demirel, S. Granick: Transition from static to kinetic friction in a model lubricated system, J. Chem. Phys. **109** (1998) 6889–6897
23.86 I. L. Singer: Solid lubrication processes. In: *Fundamentals of Friction: Macroscopic and Microscopic Processes* (Elsevier, Amsterdam 1992) pp. 237–261
23.87 B. J. Briscoe: Friction and wear of organic solids and the adhesion model of friction, Philos. Mag. A **43** (1981) 511–527

23.88 B. J. Briscoe, A. C. Smith: The interfacial shear strength of molybdenum disulphide and graphite films, Trans. ASLE **25** (1982) 349–354

23.89 A. Berman, C. Drummond, J. N. Israelachvili: Amontons's law at the molecular level, Tribol. Lett. **4** (1998) 95–101

23.90 G. He, M. O. Robbins: Simulations of the static friction due to adsorbed molecules, Phys. Rev. B **64** (2001) 035413

23.91 J. Gao, X. C. Zeng, D. J. Diestler: Nonlinear effects of physisorption on static friction, J. Chem. Phys. **113** (2000) 11293–11296

23.92 G. He, M. O. Robbins: Simulations of the kinetic friction due to adsorbed molecules, Tribol. Lett. **10** (2001) 7–14

23.93 F. Heslot, T. Baumberger, B. Perrin, B. Caroli, C. Caroli: Creep, stick-slip, and dry-friction dynamics: Experiments and a heuristic model, Phys. Rev. E **49** (1994) 4973–4988

23.94 A. Ruina: Slip instability and state variable friction laws, J. Geophys. Res. **88** (1983) 10359–10370

23.95 M. G. Rozman, M. Urbakh, J. Klafter: Stick-slip motion and force fluctuations in a driven two-wave potential, Phys. Rev. Lett. **77** (1996) 683–686

23.96 M. G. Rozman, M. Urbakh, J. Klafter: Stick-slip dynamics as a probe of frictional forces, Europhys. Lett. **39** (1997) 183–188

23.97 M. G. Rozman, M. Urbakh, J. Klafter, F. J. Elmer: Atomic scale friction and different phases of motion of embedded systems, J. Phys. Chem. B **102** (1998) 7924–7930

23.98 H. Yoshizawa, J. N. Israelachvili: Fundamental mechanisms of interfacial friction. 1. Stick-slip friction of spherical and chain molecules, J. Phys. Chem. **97** (1993) 11300–11313

23.99 D. E. McCumber: Effect of ac impedance on the dc voltage-current characteristics of superconductor weak-link junctions, J. App. Phys. **39**(7) (1968) 3113–3118

23.100 J. H. Dieterich: Modeling of rock friction. 2. Simulation of pre-seismic slip, J. Geophys. Res. **84** (1979) 2169–2175

23.101 A. A. Batista, J. M. Carlson: Bifurcations from steady sliding to stick slip in boundary lubrication, Phys. Rev. E **57** (1998) 4986–4996

23.102 J. H. Dieterich, B. D. Kilgore: Implications of fault constitutive properties for earthquake prediction, Proc. Natl. Acad. Sci. USA **93** (1996) 3787–3794

23.103 A. Dhinojwala, S. C. Bae, S. Granick: Shear-induced dilation of confined liquid films, Tribol. Lett. **9** (2000) 55–62

23.104 P. A. Thompson, G. S. Grest: Granular flow: Friction and the dilatancy transition, Phys. Rev. Lett. **67** (1991) 1751–1754

23.105 H. M. Jaeger, S. R. Nagel, R. P. Behringer: Granular solids, liquids, and gases, Rev. Mod. Phys. **68** (1996) 1259–1273

23.106 S. Nasuno, A. Kudrolli, J. P. Gollub: Friction in granular layers: Hysteresis and precursors, Phys. Rev. Lett. **79** (1997) 949–952

23.107 A. R. C. Baljon, M. O. Robbins: Stick-slip motion, transient behavior, and memory in confined films. In: *Micro/Nanotribology and Its Applications*, ed. by B. Bhushan (Kluwer, Dordrecht 1997) pp. 533–553

23.108 A. L. Demirel, S. Granick: Friction fluctuations and friction memory in stick-slip motion, Phys. Rev. Lett. **77** (1996) 4330–4333

23.109 B. J. Ackerson, J. B. Hayter, N. A. Clark, L. Cotter: Neutron scattering from charge stabilized suspensions undergoing shear, J. Chem. Phys. **84** (1986) 2344–2349

23.110 M. Lupowski, F. van Swol: Ultrathin films under shear, J. Chem. Phys. **95** (1991) 1995–1998

23.111 A. Buldum, S. Ciraci: Interplay between stick-slip motion and structural phase transitions in dry sliding friction, Phys. Rev. B **55** (1997) 12892–12895

24. Mechanics of Biological Nanotechnology

One of the most compelling areas to be touched by nanotechnology is biological science. Indeed, we will argue that there is a fascinating interplay between these two subjects, with biology as a key beneficiary of advances in nanotechnology as a result of a new generation of single molecule experiments that complement traditional assays involving statistical assemblages of molecules. This interplay runs in both directions, with nanotechnology continually receiving inspiration from biology itself. The goal of this chapter is to highlight some representative examples of the exchange between biology and nanotechnology and to illustrate the role of nanomechanics in this field and how mechanical models have arisen in response to the emergence of this new field. Primary attention will be given to the particular example of the processes that attend the life cycle of bacterial viruses. Viruses feature many of the key lessons of biological nanotechnology, including self assembly, as evidenced in the spontaneous formation of the protein shell (capsid) within which the viral genome is packaged, and a motor-mediated biological process, namely, the packaging of DNA in this capsid by a molecular motor that pushes the DNA into the capsid. We argue that these processes in viruses are a compelling real-world example of nature's nanotechnology and reveal the nanomechanical challenges that will continue to be confronted at the nanotechnology-biology interface.

24.1 **Science at the Biology–Nanotechnology Interface** 740
24.1.1 Biological Nanotechnology 740
24.1.2 Self-Assembly as Biological Nanotechnology 740
24.1.3 Molecular Motors as Biological Nanotechnology 740
24.1.4 Molecular Channels and Pumps as Biological Nanotechnology 741
24.1.5 Biologically Inspired Nanotechnology 742
24.1.6 Nanotechnology and Single Molecule Assays in Biology 743
24.1.7 The Challenge of Modeling the Bio-Nano Interface 744

24.2 **Scales at the Bio-Nano Interface** 746
24.2.1 Spatial Scales and Structures 747
24.2.2 Temporal Scales and Processes 749
24.2.3 Force and Energy Scales: The Interplay of Deterministic and Thermal Forces 750

24.3 **Modeling at the Nano-Bio Interface** 752
24.3.1 Tension Between Universality and Specificity 752
24.3.2 Atomic-Level Analysis of Biological Systems 753
24.3.3 Continuum Analysis of Biological Systems 753

24.4 **Nature's Nanotechnology Revealed: Viruses as a Case Study** 755

24.5 **Concluding Remarks** 760

References 761

Nanotechnology is the seat of a broad variety of interdisciplinary activity with applications as diverse as optoelectronics, microfluidics, and medicine. However, one of the richest interdisciplinary areas only now beginning to be harvested is the interface between the biological sciences and nanotechnology. The aim of the present chapter is threefold. First, we wish to illustrate the synergy that exists between biological nanotechnology (e.g., molecular motors, transmembrane pumps, etc.) and biologically inspired nanotechnology (e.g., synthetic proteins, artificial viruses aimed at delivering drugs, biofunctionalized cantilevers, etc.). Secondly, through a series of order-of-magnitude estimates and an associated examination of the units for describing spatial dimensions, temporal processes, forces at the nanoscale, and the energy budget in nanoscale systems, we aim to build an intuitive sense of the workings of nanotechnology in the biological set-

ting. Finally, through several specific case studies, we hope to illustrate some of the challenges faced in modeling the mechanical processes of biological nanotechnology and show how such challenges have been met thus far and how they might be met in the future.

24.1 Science at the Biology–Nanotechnology Interface

24.1.1 Biological Nanotechnology

Though the innovations leading to the adoption of the expression "nanotechnology" are indeed impressive, the perspective of this chapter is that one need only look inward to the way that our muscles move, to how we digest and synthesize molecules of dazzling complexity, to the way in which we think the thoughts that permit us to fill the shelves of libraries with scientific journals to realize that the greatest nanotechnology of all is that which is revealed in the living world. That is, one of the central thrusts is the idea that the microscopic workings of life offer an inspiring vision of nanotechnology. Clearly, the diversity of the examples of "nanotechnology" seen in the living world can (and do) fill the pages of learned texts. A guided tour of the machinery of the cell can be found in *Alberts* et al. [24.1], while a vision of the cell as an assemblage of "protein machines" has been argued for by *Alberts* [24.2]. The ambition of the present section is to provide several cursory examples of the nanotechnological marvels that power the living realm.

24.1.2 Self-Assembly as Biological Nanotechnology

One of the most intriguing nanotechnological tricks of the living world is the central role played by self-assembly in such systems. Whether we consider the spontaneous assembly of viruses, either in test tubes or in the interior of an infected cell, or the fusion of one membrane-bound region to another through vesicle fusion, spontaneous formation of functional "materials" is a key part of the biological repertoire. To be more concrete, we note that self-assembly in biological systems takes place in a number of different guises. First, the self-assembly of linear assemblies is a key part of the cytoskeletal assembly process with G-actin associating to form actin filaments and, similarly, tubulin monomers joining to form microtubules. This is also the process that is used by certain bacteria such as *Listeria* for locomotion. Both of these examples are described more fully in [24.3]. This same type of process is taken to the next level of sophistication in simple viruses such as tobacco mosaic virus that involve not only the assembly of protein monomers, but also the genetic message in the form of long RNA molecules. A second broad class of self-assembly processes is associated with the formation of containers such as liposomes or viral capsids. In the case of liposomes, lipid molecules such as phosphatidylcholine spontaneously organize in a way that sequesters the hydrophobic tails from the surrounding water. Similarly, protein subunits spontaneously assemble to form viral capsids [24.4]. These structures play a variety of roles in the biological realm, from serving as containers for different macromolecules to providing for concentration gradients of small ions that are maintained and utilized by molecular motors such as ATPsynthase [24.1].

Beyond the simplistic description of self-assembly advanced above, a second key feature of biological self-assembly must also be considered. In particular, a variety of self-assembly processes in the biological realm are templated (or coded). As is well-known, proteins are a hugely versatile class of molecules all based upon the same fundamental building blocks. Interestingly, the enormous diversity of protein action is founded upon 20 distinct amino acid building blocks, and the template for assembling a given protein is carried in the form of messenger RNA, which is then read and used as the basis for protein synthesis by the ribosome. One of the most exciting developments in modern nanotechnology is the attempt to exploit templated self-assembly processes for the purposes of creating new materials. An example in the context of protein-based materials can be found in the article of *van Hest* and *Tirrell* [24.5], where the machinery of the cell has been tricked into incorporating artificial amino acids into synthetic proteins.

24.1.3 Molecular Motors as Biological Nanotechnology

One of the hallmarks of life is change and motion. At the cellular level, such motion is effected through a dizzying variety of mechanisms, most of which when viewed at the molecular level are seen to be the result of the action of molecular motors [24.3]. For example, muscle con-

traction reflects the coherent action of huge numbers of myosin motor molecules as they march along actin filaments. Similarly, the motion of certain bacteria can be traced to the rotation of a rotary motor embedded in the cell wall, which is attached to filaments known as flagella [24.6]. In a similar vein, just as our modern society is replete with examples of systems aimed at allowing for communication and transport between widely separated geographic locations, so, too, has the living world had to answer these same challenges. As will be shown in the section of this chapter aimed at making estimates of various scales and processes in nanotechnology, one of the key mechanisms for communication and transport is diffusion. That is, a chemical concentration at one place can make its presence known at a distance x with a characteristic time, $t_{\text{diffusive}} \approx x^2/2D$, where D is the diffusion constant. However, as will be shown later in the chapter, there are many cellular processes that cannot wait as long as $t_{\text{diffusion}}$. As a result, a host of molecular motors and an associated transport system (the elements of the cytoskeleton) permit active transport. For example, kinesin is a motor molecule that transports material along long, relatively rigid polymeric assemblies known as microtubules [24.1].

Even more incredible are rotary machines like those associated with ATPsynthase and bacterial flagella. ATPsynthase is a membrane-embedded machine that rotates and, in so doing, synthesizes new ATP molecules (adenosine triphosphate, energy currency of the cell) [24.1, 7–9]. Similarly, the bacterial flagellar motor rotates with the result that the attached bacterial flagellum rotates and thereby induces motion of the cell [24.3, 6]. The exquisite details of the construction of these devices are themselves breathtaking. The rotary motors in bacteria are constructed from several components that are much like a rotor and stator and perform periodic motions by deriving energy from the flow of protons across a membrane [24.6]. These rotary motors are very powerful, as evidenced by F_1-ATPsynthase, which can generate a torque large enough to rotate a molecule of actin 100 times its own length [24.10]. In addition, bacterial motors have another layer of sophistication in that stimuli from the external environment, which are sensed by the bacteria through the pores on its membrane, can change the direction of rotation of the motor in a process known as chemotaxis [24.3]. Feedback and signal transmission are implemented in engineering devices by means of complex circuitry, whereas in the living world these functions are often accomplished by means of chemistry and conformational changes in large molecules.

All of these molecular machines involve a rich interplay between chemistry, thermodynamics, and mechanics. From a structural perspective, most molecular motors are proteins with different subunits performing different functions [24.11]. Often there is some region within the motor where chemical energy is derived from the hydrolysis of adenosine triphosphate (ATP). This chemical energy is then converted to mechanical energy through a conformational change in the protein. These examples serve to call the reader's attention to the importance and variety of motor molecules found throughout the living world and which almost any sensible definition of nanotechnology would have to include as particulary sophisticated examples.

24.1.4 Molecular Channels and Pumps as Biological Nanotechnology

From a structural perspective, one of the most intriguing features of cellular systems is their division into a number of separate membrane-bound compartments. We have already touched upon this compartmentalization in the context of self-assembly, and note that the role of membranes and the proteins bound within them is described clearly in *Alberts* et al. [24.1]. The presence of such compartments, often marked by large concentration gradients with respect to the surrounding medium, hints at another nanotechnological wonder of the living world, namely, the presence of a wide range of transmembrane channels that mediate the exchange of material between these different compartments. Certain passive versions of these channels are gated by various mechanisms such as the arrival of signaling molecules or tension in the membrane within which they are found [24.12–14]. Once the channel is in the open state, ions pass through passively, by diffusion. Active versions of ion channels that transfer ions such as Na^+ and Ca^{2+} are similarly critical for the functioning of a cell, both in the case of unicellular and multicellular organisms. For example, in the case of Na^+ ions, typical concentrations within the cell can be as much as a factor of 10–20 less than those in the extracellular milieu. Such concentration gradients imply the need for sophisticated active "devices" which can do work against such gradients. One of the most remarkable of machines of this type is the Na^+–K^+ pump [24.1]. This machine is powered by hydrolysis of ATP (i. e., the consumption of ATP fuel), and it can pump ions up a potential gradient. In particular, this pump hydrolyzes ATP and pumps Na^+ ions *out* of the cell against a very steep concentration gradient while pumping K^+ ions *in* again against a steep gradient.

The point of this brief discussion has been to illustrate the first of several perspectives that we will bring to bear on the question of the biology-nanotechnology interface. Thus far, we have noted that nature is replete with examples of macromolecules and macromolecular assemblies that perform nanotechnological tasks and in this capacity serve as examples of biological nanotechnology. Next, we wish to examine the ways in which biological phenomena can inspire nanotechnology itself.

24.1.5 Biologically Inspired Nanotechnology

As noted in the beginning of this chapter, biological systems are nanotechnologically relevant for several reasons. First, as shown above, the living world is full of examples of nanotechnological devices. However, a second key point is that nanotechnology has been driven and inspired by the example of biological systems and the need (for example, in medicine) to influence biological systems at the scale of a single cell. In addition, preliminary steps have been taken to harness the nanotechnology of biological systems and use it to perform useful functions.

One compelling example of a proof-of-principle, biologically inspired device emerged from work aimed at exploring the function of ATPsynthase, the rotary device already described above. Fluorescently labelled filamentous proteins from the cytoskeleton, known as actin, were attached to the putative rotary component of the F_1 subunit. The point of this exercise is that such filaments are observable using light microscopy. It was then observed that the long actin filament rotated like a propeller when the ATPsynthase performed ATP hydrolysis [24.15].

A second example, also involving ATPsynthase, is suggested by a set of beautiful experiments done by *Racker* and others [24.16, 17]. The idea is illustrated schematically in Fig. 24.1. Two different protein machines are "reconstituted" in an artificial membrane-bound region known as a liposome. One of these devices is known as bacteriorhodopsin and has the capacity to pump hydrogen ions when it is exposed to light. Since both the bacteriorhodopsin and ATPsynthase are embedded in the same membrane, the ATPsynthase can then exploit the light-induced proton gradient to perform ATP synthesis. Again, these experiments were undertaken not for their role as possible devices, but rather to probe the nature of various molecular machines. Nevertheless, we view them as a provocative demonstration of both the manipulation and use of such machines in artificial environments. As such, they provide an inspiring vision of the possibilities for biologically inspired nanotechnology.

Another fascinating example of biologically inspired nanotechnology is that of biofunctionalized cantilevers (see Fig. 24.2). A typical example is provided by the experiments of *Wu* et al. [24.18], who have demonstrated the use of a biofunctionalized cantilever as a scheme for detecting small concentrations of biologically interesting molecules such as prostrate-specific antigen and single-stranded DNA. One surface of the cantilever is coated with an antibody, and then it is placed in environments containing different concentrations of the antigen. The key ideas from a mechanical perspective are: i) the difference in surface energy between the top and bottom surfaces of the cantilever induces spontaneous bending and ii) the binding of molecules of interest to target molecules initially present on the surface leads to surface energy differences and bending that can be detected by optical means. In this way the concentration of the molecules of interest can be measured. The specificity and sensitivity of the method makes it viable for use in the laboratory, as well as for commercial purposes. We return to this example in our discussion of the modeling challenges posed by problems at the interface between biology and nanotechnology.

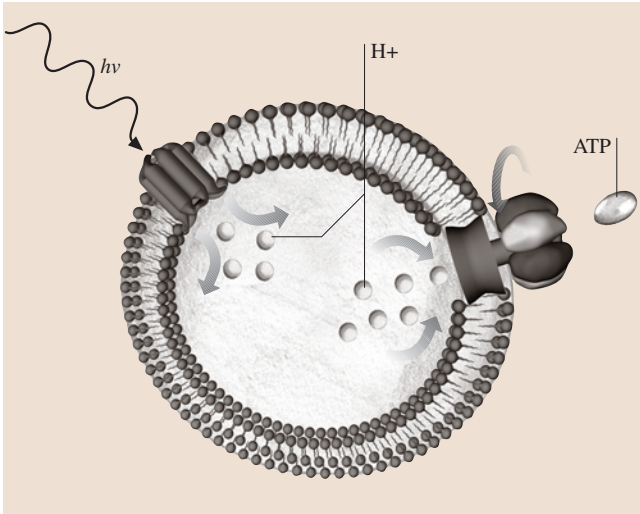


Fig. 24.1 Schematic of the experiment in which bacteriorhodopsin (*top of figure*) and ATPsynthase (*right of figure*) are artificially reconstituted in a liposome, and they act in unison to produce ATP molecules

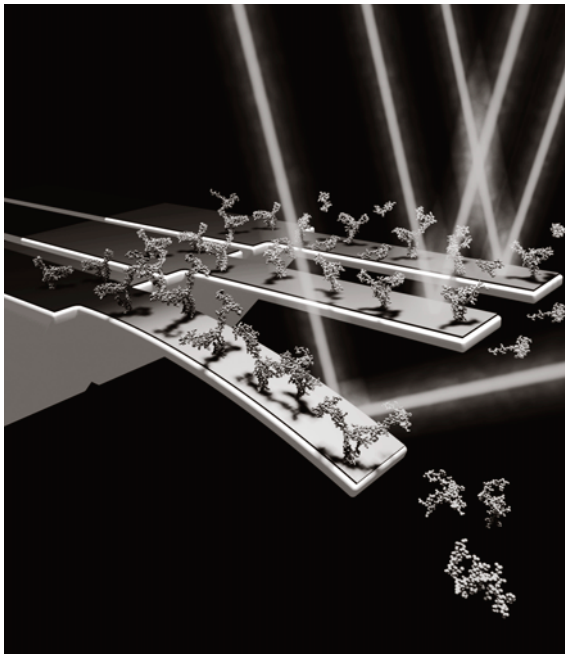


Fig. 24.2 Schematic of the use of biofunctionalized cantilevers as a tool for detecting molecules of biological interest (figure courtesy of Arun Majumdar, Berkeley)

24.1.6 Nanotechnology and Single Molecule Assays in Biology

One of the key refrains of the nanotechnological era is Feynman's quip that "there is plenty of room at the bottom" [24.19]. The benefits of miniaturization are evident at every turn in applications ranging from our cars to our computers. Associated with the development of the inspiring new techniques made possible by nanotechnology has been the emergence of a host of scientific opportunities. One of the arenas to benefit from these new techniques is biology. As scientists and engineers have taken the plunge to the nanotechnological "bottom" foreseen by Feynman, opportunities have constantly arisen to manipulate biological systems in ways that were heretofore unimagined, culminating in a new era of single molecule biology. Single molecule experiments have, in fact, presented us with a view of Feynman's "room at the bottom" as being filled with very complicated machines whose functioning makes life possible.

Single molecule assays complement statistical/collective studies involving a large number of molecular actors by revealing the prominent role of fluctuations at the sub-cellular scale. For example, a photospectrometer measures the optical response of a huge collection of molecules, whereas optical tweezers pull on a single molecule of DNA and enable us to follow the changes in conformation or the breakage of bonds. In fact, experimental methods are so advanced that it is now possible to manipulate a single molecule even as we watch it on a screen as it jiggles around in different conformations. In what follows, we give several examples of how nanotechnology has reached out to help create single molecule biology and in the process has led to the advent of new quantitative opportunities to investigate biological systems.

Atomic-Force Microscopy

One of the tools that has revolutionized nanotechnology, in general, and single molecule biology, in particular, is the atomic-force microscope. The AFM has helped create the field of single molecule force spectroscopy [24.20]. We note that mechanics has a long tradition of using force-extension data (much like the electrical engineer uses current-voltage data) to probe the inner workings of various materials. It is now possible to apply forces of known magnitude on a macromolecule and study how it deforms under the force. This furnishes structural information and provides insights into the energy landscape the molecule needs to navigate as it undergoes force-induced conformational changes. The energy of deformation associated with such molecules is primarily determined by weak forces such as hydrogen bonds, van der Waals contacts, and hydration effects. On a more philosophical note, these experiments force us to think in terms of forces and not energy, complementing the traditional views held in molecular biology, and they can lead to many new insights about the relation between structure and function in proteins, polynucleotides, and other macromolecular entities.

The AFM has been used in a wide variety of single molecule experiments on many of the key classes of molecules found in the living world, including nucleic acids, proteins, and carbohydrates. One fascinating example is the use of atomic force microscopy to examine the mechanical properties of the muscle protein titin [24.21]. The experiment is illustrated schematically in Fig. 24.3, which shows that there is a series of force-extension signatures (increasing force and extension followed by a precipitous load drop) that correspond to the unbinding of the individual domains that make up this protein.

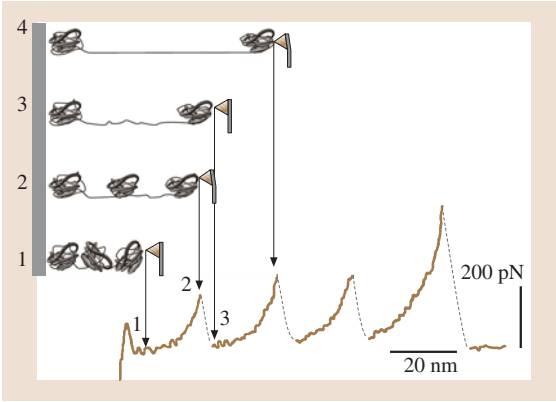


Fig. 24.3 Schematic illustrating the way that the AFM has been harnessed as a nanotechnological analogue of the Instron machine for the measurement of the force-extension properties of single molecules. Four snapshots in the life history of a globular protein subject to loading are shown, as well as the measured force-extension curve. Each sawtooth corresponds to the unfolding of a single protein domain. (Figure adapted from [24.22])

Optical Tweezers

Another instrument that has been used with great success in the realm of biophysics for the purposes of performing nanomechanical measurements on macromolecules and their assemblies is the optical tweezer [24.23,24]. While the AFM is relatively stiff and applies large forces (on the order of 100–1,000 pN), optical tweezers are compliant and can measure smaller forces (on the order of 0.1–10 pN). Some of the most interesting experiments performed with optical tweezers concern the functioning of molecular motors, which by themselves are marvels of nanotechnology. For example, *Svoboda* et al. [24.25] attached kinesin to an optically trapped bead and observed its movement along a microtubule. An example of the type of data to emerge from such experiments is shown in Fig. 24.4. One of the conclusions to emerge from such experiments is that kinesin can exert forces on the order of 5–7 pN before it stalls. Such experiments also permit an examination of the effect of changing the concentration of ATP on the functioning of kinesin [24.26]. Similar experiments have been performed on RNA polymerase as it advances along DNA to deduce not only the stall force, but also its velocity as a function of the constraining force [24.27]. Such measurements provide a mechano-chemical basis of biological function and go a long way in revealing connections between chemical kinetics and mechanical processes at the molecular level.

As noted in the abstract, one of the most fascinating examples of the biology-nanotechnology interface is that of bacterial viruses, known as bacteriophages. The life cycle of a large class of bacteriophages is characterized by self-assembly processes that lead to the formation of the protein shell of the virus followed by active packaging of the viral DNA within this shell by a motor. The structure of this so-called "viral portal motor" has recently been solved using X-ray crystallography [24.28]. In a recent experiment, *Smith* et al. used optical tweezers to study the characteristics of the DNA packaging process of the ϕ-29 bacteriophage [24.29]. One of the conclusions of this experiment is that the motor has to act against an increasing resistive force as more and more of the DNA is packed inside the capsid. From a quantitative perspective, this experiment yields the force and the rate of packing as a function of the fraction of the genome packed. It is also important to note that bacteriophages are not only an obscure subject of quiet enquiry, but are also the basis of a huge range of cloning products (see, for example, the lambda ZAP vectors of Stratagene) used for doing experiments with recombinant DNA, and more generally, viruses are being explored as the basis of gene therapy.

Our discussion thus far has been aimed at providing a rough overview of the vast landscape that sits at the interface between biology and nanotechnology. It is hoped that the few representative case studies set forth above suffice to illustrate our basic thesis, namely, that biological nanotechnology represents nanotechnology at its best.

24.1.7 The Challenge of Modeling the Bio–Nano Interface

As highlighted in the previous two subsections, there have been huge advances at the interface between nanotechnology and biology. We have argued that there are two distinct representations of the interface between biology and nanotechnology, and each has its own associated set of modeling challenges. The argument of the present discussion is that another key part of the infrastructure that must attend these developments is that associated with the modeling of these systems. One of the intriguing ways in which modeling at the biology-nanotechnology interface is assuming greater importance is that with increasing regularity, experimental data on biological systems is of a quantitative character. As a result, the models that are put forth to greet these experiments must similarly be of a quantitative character [24.30].

As an example of the type of modeling challenges that must be faced in contemplating the types of problems described above, we return to the example of biofunctionalized cantilevers as a problem in nanomechanics [24.18, 31]. The basic physics behind the use of biofunctionalized cantilevers as sensors is a competition between elastic bending energy and the surface free energy difference between the upper and lower faces of the cantilever. The face with the lower free energy per unit area tends to increase its area by bending the cantilever. The amount of bending, on the other hand, is limited by the elastic energy cost. The utility of this device derives from the fact that the difference in the surface free energy is affected by specific binding of target molecules to probe molecules that are initially deposited on one side of the cantilever.

To provide a quantitative model of the biofunctionalized cantilever, we construct an energy functional that takes into account the elastic energy of beam bending and the surface energy. Both contributions to the total energy can be written as functionals of $u(x)$, the deflection of the cantilever. Note that in the case in which the two surfaces have the same free energy per unit area, $\gamma_{\text{up}} = \gamma_{\text{down}}$, the equilibrium configuration corresponds to $u(x) = 0$. The case of interest here is that in which the two surfaces have different energies.

We recall that the energy associated with beam bending is of the form [24.32]

$$E_{\text{bend}}\,[u(x)] = \frac{EI}{2}\int_0^L \left[u''(x)\right]^2 \mathrm{d}x\,, \tag{24.1}$$

where E is Young's modulus, L is the length of the beam, and $I = wt^3/12$ is the areal moment of inertia; t and w are the beam thickness and width. The main approximation we have made in writing (24.1) is that the cross section of the beam remains unchanged by the bending process. The surface contribution to the total energy is associated with the changes of the areas of the upper and lower surfaces by virtue of the beam deforming. In particular, we have

$$E_{\text{surf}}[u(x)] = \gamma_{\text{up}} w \left[L - \frac{t}{2}\int_0^L u''(x)\,\mathrm{d}x\right] + \gamma_{\text{down}} w \left[L + \frac{t}{2}\int_0^L u''(x)\,\mathrm{d}x\right]\,, \tag{24.2}$$

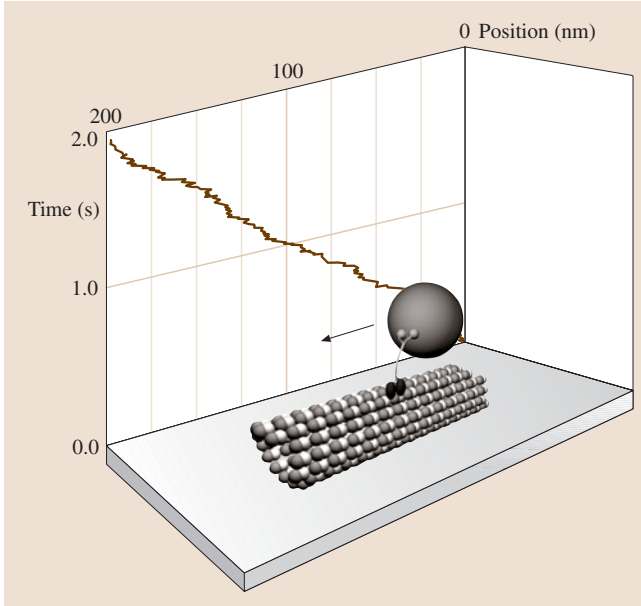


Fig. 24.4 Schematic illustrating the use of optical tweezers to measure the speed of the molecular motor kinesin in its journey along a microtubule

where the terms in parentheses are the arc lengths along the top and bottom surface of the beam, respectively. The physical content of this functional is the idea that if $u'' > 0$ (concave upward) then the upper surface area will shrink while the lower surface area will increase. Neglecting uninteresting constant terms, the total energy functional is

$$E_{\text{tot}}[u(x)] = \frac{EI}{2}\int_0^L \left[u''(x)\right]^2 \mathrm{d}x + \frac{tw}{2}(\gamma_{\text{down}} - \gamma_{\text{up}})\int_0^L u''(x)\,\mathrm{d}x\,. \tag{24.3}$$

Our goal is to find the displacement profile $u(x)$ that yields the minimum value for the total energy. The general mathematical framework for effecting this minimization is the calculus of variations. For the functional in (24.3), a more direct route to the result can be obtained by completing the square. Namely, we note that the E_{tot} involves a term quadratic in $u''(x)$ and a second term that is linear in the same function. As a result, the

total energy can be rewritten in the form

$$E_{\text{tot}}[u(x)] = \frac{EI}{2}\int_0^L \mathrm{d}x\left[u''(x)+\frac{tw\Delta\gamma}{2EI}\right]^2 , \quad (24.4)$$

where we have introduced $\Delta\gamma = \gamma_{\text{down}} - \gamma_{\text{up}}$, and we neglect an uninteresting constant term. Clearly the $u(x)$ that minimizes the energy is one for which the integrand vanishes everywhere, or

$$u''(x)+\frac{tw\Delta\gamma}{2EI} = 0 . \quad (24.5)$$

At this point, we are left with a standard differential equation whose solution, given the boundary conditions $u(0) = 0$ and $u'(0) = 0$, is

$$u(x) = -\frac{tw}{4EI}\Delta\gamma x^2 . \quad (24.6)$$

The physical meaning of this solution is hinted at by observing that if $\gamma_{\text{down}} > \gamma_{\text{up}}$ then the beam curves downward with the result that the area of the lower surface is reduced.

Measurements of *Wu* et al. [24.31] indicate an upward cantilever deflection when single stranded DNA (ssDNA), up to 20 nucleotides in length, hybridizes with complementary strands of ssDNA, which were deposited initially so as to functionalize the beam. This effect might be attributed to the difference in elastic properties of ssDNA and its double stranded counterpart. Namely, under physiological conditions, ssDNA has a persistence length equal to two nucleotides, while dsDNA is much stiffer (due to hydrogen bonding and base stacking interactions between the two strands) and has a persistence length of 150 nucleotides. Therefore, deposition of the flexible ssDNA molecules initially leads to bending of the cantilever downward due to entropic repulsion between the ssDNA chains. After hybridization, rigid dsDNA strands are formed, there is no longer any entropy to be gained by increasing the area of the top surface, and the beam bends back upwards. Remarkably, *Wu* et al. demonstrate that their biofunctionalized cantilever is sensitive to ssDNA differing in length by a single nucleotide!

To gain further insight into the physics of the biofunctionalized cantilever, it is instructive to examine some quantitative aspects of the experiment using (24.6). Namely, *Wu* et al. find a deflection of $u(L) = 12\,\text{nm}$ when a 20-nucleotide strand of ssDNA hybridizes with a 20-nucleotide-long complementary target. Using the quoted numbers for the cantilever, $E = 180\,\text{GPa}$, $L = 200\,\mu\text{m}$, $t = 0.5\,\mu\text{m}$, $w = 20\,\mu\text{m}$, leads to $\Delta\gamma = 4.5\,\text{fJ}/\mu\text{m}^2$. From this result we can estimate the change in surface free energy due to beam bending,

$$\Delta E_{\text{surf}} \approx wL\Delta\gamma . \quad (24.7)$$

We find $\Delta E_{\text{surf}} = -18\,\text{pJ}$, which, given the quoted areal chain density of $6\times10^{12}\,\text{cm}^{-2}$, leads to a decrease in free energy of 75 pN nm, or $18\,k_{\text{B}}T$, per chain. This is comparable to the entropy of a 20 nt ssDNA, which is on the order of $10\,k_{\text{B}}$, lending support to the idea that entropic repulsion between ssDNA strands is implicated in cantilever bending.

While this example gives a feel for the way in which quantitative models have been put forth to respond to biologically inspired nanotechnologies, the remainder of the chapter will emphasize attempts to construct nanomechanical models of biological nanotechnology itself. We turn first to a discussion of the various scales that arise in thinking about the biology-nanotechnology interface and then conclude with several modeling examples.

24.2 Scales at the Bio–Nano Interface

Every scientific discipline has a preferred set of units that lends itself to building intuition concerning the system at hand. For example, an astronomer thinks of distances between stars in light years, not kilometers. Though most of us have an intuitive sense of the meaning of a kilometer, by the time we add more than six zeros, all intuition is lost. At terrestrial scales, we talk of distances between cities in terms of the flying time or the driving time between them and usually not in terms of hundreds of miles. Similar choices must be faced in the biological setting. For example, a biologist might characterize the complexity of an organism by the size (in kilo basepairs) of the genome and not by the organism's physical size. The aim of the present section is to highlight some of the key scales and units that reveal themselves at the interface between biology and nanotechnology. Indeed, we go further and assert that as yet we are still in the process of searching for the most suitable units to

characterize the biology-nanotechnology interface. Although our attempt to determine such units and scales might involve seemingly complicated interconversions, such as measuring distances in terms of time (via diffusion), or measuring concentrations in terms of distances, we hold that the approximate *numerical* characterization of the scales of interest is of crucial importance to the endeavor of considering nanomechanics at the biology-nanotechnology interface. In particular, the right choice of units can assist us in building intuition about these systems. Our goal in this section is to emphasize the scales in length, time, force, energy, and power that are relevant in contemplating the nanomechanics of biological systems.

24.2.1 Spatial Scales and Structures

We begin with a discussion of the length scales that arise in contemplating nanomechanics at the biology-nanotechnology interface. In this case, the prefix *nano* leads justifiably to a consideration of the nanometer as one possible choice as the fundamental unit of length. However, to prepare ourselves for the question of how best to describe the dimensions of the spatial structures of interest here, it is important to consider the *hierarchy* of length scales that arise in the nano-bio arena. After examining this hierarchy of structures, we reformulate these length scales in terms of the volumes of these structures measured in units of the volume of a typical bacterial cell, and conclude the present section with a discussion of the way that chemical concentrations can also be interpreted as determining a length scale.

As noted above, a first step in developing intuition concerning the spatial scales found at the biology-nanotechnology interface is through reference to the hierarchy of scales and structures that arise in this arena. The shortest distance in this hierarchy of scales that will interest us is that associated with the size of individual atoms. We recall that the size of a hydrogen atom is roughly 0.1 nm. The scale characterizing the linkage between atoms is that of typical bond lengths that range from roughly 0.1–0.3 nm. A step further in this hierarchy brings us to the basic building blocks of the biological world such as amino acids, nucleotides, individual sugars, lipid molecules, etc. A typical length scale that characterizes these building blocks is the nanometer itself. For example, we have already made reference to the importance of lipid-bilayer membranes in bounding different regions of the cell. Phosphatidylcholine is one of the molecular building blocks of many such membranes. It has a polar head group and a hydrocarbon tail with an overall dimension of 2–3 nm. Similarly, the dimensions of single amino acids and nucleotides are on the order of 1 nm as well.

As is well-known, individual nucleotides are assembled to form nucleic acids such as DNA, amino acids are assembled to form proteins, sugars combine to form polysaccharides, and lipids self-assemble to form membranes. One way of estimating the size of the resulting molecules is by taking the scale of the individual units and scaling up with the number of such units. The various molecular actors of relevance to the present discussion can also be characterized by a length, scale known as the persistence length, which gives a rough description of the length over which the molecule behaves as a stiff rod. For example, the persistence length of DNA is ≈ 50 nm, while that of the cytoskeletal filaments is in the range of 15 μm for actin and 6 mm for microtubules [24.11]. We note that in our later discussions of viruses as a profound example of biological nanotechnology, the ratio of the persistence length of DNA to the size of the viral capsid (the container within which the DNA is packaged) will serve as a measure of the energetic cost of packing the DNA within the capsid and will signal the need for molecular motors to take active part in the packaging process.

The next scale above that of the various macromolecular building blocks are assemblies of such molecules in the form of various molecular machines which are some of the most compelling examples of nature's nanotechnology. The ability to begin formulating mechanistic models of such machines is founded upon key advances in both X-ray crystallography and cryo-electron microscopy [24.33, 34]. Examples of such machines and their associated dimensions include: the machine responsible for making the message carried by DNA readable by the protein synthesis machinery, namely, RNA polymerase (≈ 15 nm) [24.35], the machine that produces ATP, the energy currency of the cell, namely, ATPsynthase (≈ 10 nm) [24.9], and the machine that carries out protein synthesis, namely, the ribosome (≈ 25 nm) [24.36]. A second class of assemblies of particular relevance to the present article are viruses, representative examples of which include lambda phage (≈ 27 nm), tobacco mosaic virus (≈ 250 nm in length), and the HIV virus (≈ 110–125 nm) [24.37].

From the standpoint of cellular function, the next level of structural organization is associated with the various organelles within the cell, structures such as the cell nucleus (3–10 μm), the mitochondria that serve

as the power plant of the cell ($\approx 1{,}000$ nm), the Golgi apparatus wherein modifications are made to newly synthesized macromolecular components ($\approx 1{,}000$ nm), and so on [24.1]. These organelles should be thought of as factories in which many molecular motors of the type described earlier do their job simultaneously. At larger scales yet, and constituting a higher level of overall organization, life as nanotechnology is revealed as cells themselves, the fundamental unit of life that is self-replicating and self-sustaining. We will make special reference to one particular bacterial cell, namely, *E. coli* with typical dimensions of 1 μm. This should be contrasted with a typical eucaryotic cell such as yeast, which has linear dimensions roughly a factor of 10 larger than the *E. coli* cell.

We note that the various biological structures described above should be seen with reference to the sorts of man-made structures to which they are interfaced. For example, earlier in the chapter we mentioned the importance of optical tweezers as a means of communicating forces to macromolecules and their assemblies. Typical dimensions for the optical beads used in optical tweezers are on the scale of 500–1,000 nm. We note that though such dimensions are characteristic of organelles, they are much larger than the individual molecules they are used to study. A second way in which individual macromolecules are communicated with is through small tips such as those found on an AFM. In this case, the size of the tip can be understood through reference to its radius of curvature, which is typically of the order of 50 nm.

Our discussion thus far has centered on the use of a single characteristic length to describe the spatial extent of biological structures. In our quest to develop intuition about the typical spatial scales found at the biology-nanotechnology interface we note that a second way to evaluate such scales is through reference to the typical *volumes* of the various structures of interest. As noted in the introduction to this section, we claim that a key way to develop intuition is by making sure to use appropriate and revealing units. For the present purposes, we argue that one useful unit of volume is that of an *E. coli* bacterium, which if idealized as a cylinder with diameter 1 μm and height 2 μm, has a volume of $\approx 1.5\times10^9$ nm^3. In particular, we will measure the sizes of the various entities described above in terms of how many of them can fit into a single *E. coli* bacterium. Our estimates of the comparative sizes of various structures of interest are given in Table 24.1. Note that we do not claim that this is literally how many of each of these entities are found in an *E. coli* bacterium, but rather we seek to give an impression of the relative volumes of some of the different structures that arise when contemplating biological nanotechnology.

Table 24.1 Sizes of entities in comparison to the size of *E. coli*

Entity	Size (nm)	# fitting in an *E. coli*
Amino acid	1	5.0×10^9
Nucleotide	1	5.0×10^9
Monosaccharide	1	5.0×10^9
Phosphatidyl choline	2–3	2.0×10^8
Proteins	5–6	3.0×10^7
Ribosome	20	1.0×10^6
Mitochondria	500–1,000	3–4
T4 phage	100	5,000

The idea of counting the number of molecules that fit into a given volume is actually quite standard. This is exactly what chemists do when they invoke the notion of concentration: measuring the strength of an acid or base using pH amounts to expressing the concentration of H^+ ions in solution. Similarly, when we refer to the molarity of a given solution, it is a statement of how many copies of a given molecule will be found per liter of solution. These ideas are pertinent for our examination of the scales that arise in contemplating biological nanotechnology. As an example, we consider the action of molecular pumps such as that which maintains the concentration gradient in Ca^{2+} ions between the cellular interior and the extracellular medium. Note that this is described in detail in chapter 12 of *Alberts* et al. [24.1]. The Ca^{2+} pump is responsible for insuring that the intercellular concentration of Ca^{2+} ions is 10^{-7} smaller than that in the cellular exterior. One explanation for this low a concentration of Ca^{2+} in the cell is the role played by these ions in signaling other activities. The low concentration might serve as a scheme for increasing the signal-to-noise ratio. The reported intracellular concentration of such ions is roughly 10^{-7} mM. It is interesting to ask how many Ca^{2+} ions this corresponds to in a eucaryotic cell. The volume of a eucaryotic cell measuring 20 μm in diameter is about 4×10^{-12} liters, which translates into ≈ 250 Ca^{2+} ions in the cytoplasm. For the case of *E. coli*, which has a volume approximately 1,000 smaller, this concentration would correspond to one ion for every four cells. This number gives us far greater insight than the standard molar (M = mol/liter) method of expressing concentrations. In a similar vein, we argue that distances between in-

dividual ions/molecules in solution are perhaps a better way of thinking about pH and molarity when using them in the context of biological nanosystems. To this end, Fig. 24.5 shows a plot of pH and molarity versus average distance d in nm between individual ions/molecules. Note that, in addition, we have also translated these concentrations into number of copies per *E. coli* cell and the time for an ion to diffuse over the average separation distance. For the time estimate we make use of the typical diffusion constant for small ions in water at 25 °C, $D = 2{,}000\,\mu\text{m}^2/\text{s}$.

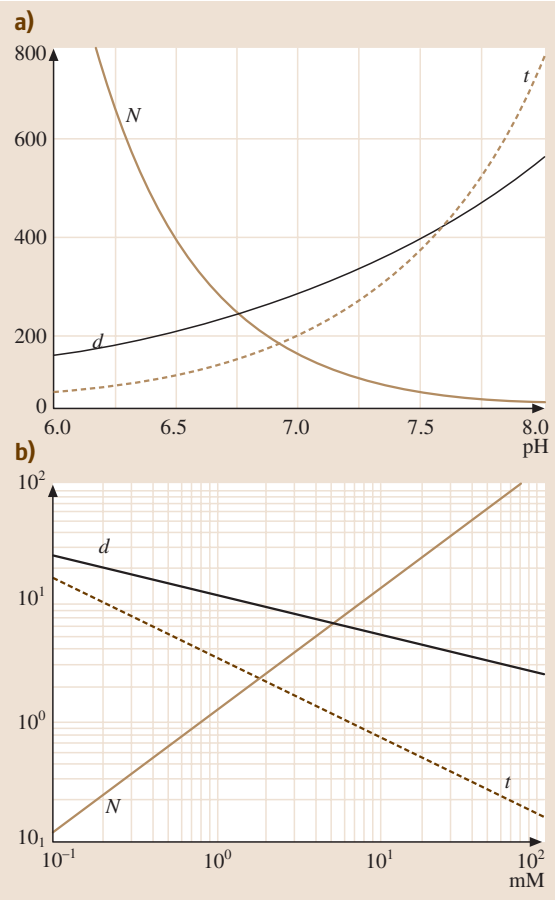


Fig. 24.5 **(a)** Representations of pH in terms of average distance between ions (d in nm), the number of ions per one *E. coli* cell volume (N), and the typical time for an ion to diffuse over a distance equal to their average separation (t in 100 ns). **(b)** Concentration unit millimolar (mM) represented in terms of d (nm), N (millions of molecules), and t (10 ns)

We have seen that the world of biological nanotechnology utilizes length scales that span a rather wide range, from fractions of a nanometer to tens of microns. This provides a challenge to modeling, whereby methods with atomic-scale resolution need to be combined in a consistent and seamless way with coarse-grained continuum descriptions to provide a complete picture. An example of such modeling will be provided in later sections when we take up the mechanical aspect of the viral life cycle.

24.2.2 Temporal Scales and Processes

We have seen that a study of spatial scales is a first step in our quest to understand the units that are suited to characterizing the biology-nanotechnology interface. We note that our understanding of spatial structures both in the living world, as well as those used in intriguing man-made technologies such as microelectronics have been built around important advances such as electron microscopy, X-ray crystallography, nuclear magnetic resonance, etc. that have revolutionized structural biology and materials science alike. On the other hand, one of the key challenges that remain in really appreciating the structure-properties paradigm, which is as central to biology as it is to materials science, is the need to acknowledge the dynamical evolution of these structures.

We note that just as there is a hierarchy of spatial scales that are important to consider in describing biological nanotechnology, so too is there a hierarchy of *temporal processes* that also demand a careful consideration of relevant units. An impressive representation of the temporal hierarchy that must be faced in contemplating biological systems is given by *Chan* and *Dill* [24.38] in their Fig. 5. They organize their temporal hierarchy according to factors of 1,000, starting at the femtosecond time scale and ending with time scales representative of the cell cycle itself. One of the elementary dynamical processes undergone by all of the molecular actors of the living world is thermal vibration. Vibrations of atoms occur at the time scales of 10^{-15} to 10^{-12} s. As will be noted later, from a modeling perspective, this depressing fact manifests itself in the necessity of using time steps of the same order when integrating the equations of motion describing molecular dynamics. The reason that this is unfortunate is that almost everything of dynamical interest occurs at time scales much longer than the femtosecond time scale characteristic of molecular jiggling, thus making molecular dynamics investigations computationally very expensive.

We follow *Chan* and *Dill* [24.38] in examining how successive thousandfold increases in our temporal resolution smears out features such as vibrational motion and brings into focus more interesting processes characteristic of macromolecular function. Indeed, the side chains of amino acids rotate with a characteristic period on the order of 10^{-9} s. Another thousandfold increase in time scale begins to bring key biological processes such as polymerization into view. Just as a length scale on the order of 10–100 nm is perhaps most characteristic of the length scales of biological nanotechnology, the time scale that is most relevant for considering the processes associated with biological nanotechnology is something between a microsecond and a millisecond. To drive home this point, we reconsider the machines described in the previous section from the length scale perspective, but now with an eye toward how fast they perform their function. Several of the molecular machines considered in the previous section mediate polymerization reactions. For example, RNA polymerase produces messenger RNA molecules by the repeated addition of nucleotides to a chain of ever increasing length. Such messenger RNA molecules serve in turn as the template for synthesis of new proteins by the ribosome, which reads the message contained in the RNA and adds individual amino acids onto the protein. From the temporal perspective of the present section, it is of interest to examine the rate at which new monomers are added to these polymers – roughly 10 per s in the case of RNA polymerase [24.39], and 2 per s in the case of the ribosome in eucaryotic cells [24.1] (bacterial ribosomes are 10 times faster). Once proteins are synthesized, in order to assume their full cellular responsibilities, they must fold into their native state, a process that occurs on time scales on the order of a millisecond.

In addition to translational machines like RNA polymerase and the ribosome, there are a host of intriguing rotary machines such as ATP synthase and the flagellar rotary motor. The flagellar motor, which is the means of locomotion for several bacteria, rotates at about 100 rpm [24.6]. ATP synthase makes use of a proton gradient across a lipid bilayer to provide rotation at a rate of roughly 6,000 rpm [24.41], approaching that of turbojet engines.

Thus far, our discussion has emphasized the rate of a variety of active processes of importance to biological nanotechnology. It is of interest to contrast these scales with those pertaining to diffusion. We note that in many instances in the biological setting the time scale of interest is that determined by the time it takes for diffusive communication of two spatially separated regions. In particular, as noted earlier in the chapter, the operative time scale is given by $t_{\text{diffusive}} = x^2/2D$, where D is the diffusion constant and x is the distance over which the diffusion has taken place.

To get a sense for all of the time scales described in this section, Table 24.2 presents a variety of results considered above in tabular form, but now with all times measured in units of the time it takes for ATPsynthase to make one rotation (≈ 10 ms). Our reason for adopting ATPsynthase rotation as defining a unit of time is that it exhibits motions associated with one of the most important of life's processes, namely, the synthesis of new ATP molecules. In addition, our later discussion of units associated with both energy and power will once again appeal to the central role played by ATP in biological nanotechnology.

24.2.3 Force and Energy Scales: The Interplay of Deterministic and Thermal Forces

Another set of scales of great relevance and importance to the nano-bio interface concerns the nature of the forces that act in this setting. Thus far, we have examined the spatial extent and the cycle time of a variety of examples of biological nanotechnology. As a next step in our examination of scales, we consider the forces

Table 24.2 Time scales of various events in biological nanotechnology measured in units of the average time taken for ATPsynthase to make a single rotation

Process	Time scale (units of ATPase rotation)	Ref.
1/(frequency of amino acid addition) – eukaryotic ribosome	50	[24.1]
1/(frequency of monomer addition) – RNA polymerase	10	[24.39]
Time between motion reset of *E. coli*	10	[24.6]
Kinesin step	1	[24.26]
Diffusion time for protein (ion) to cross an *E. coli*	0.5 (0.02)	[24.11]
Vibrational period of nanocantilever	10^{-3} – 10^{-5}	[24.40]

and energies associated with these structures. Two outstanding sources for developing a feel for the relevant numbers are the books of *Wrigglesworth* [24.42] and *Smil* [24.43]. Perhaps the most compelling feature when thinking about forces at these scales is the interplay between deterministic and thermal forces. To substantiate this claim, we note that at room temperature the fundamental energetic quantity is $k_B T \approx 4.1$ pN nm. The reason this number is of interest to the current endeavor is the realization that many of the molecular motors that have thus far been investigated act with forces on the piconewton scale over distances of nanometers.

The fact that the piconewton is the relevant unit of force can be gleaned from a simple estimate. Namely, consider a typical skeletal muscle in the human arm. The cross-sectional size of the muscle is of order 3 cm. The muscle consists of cylindrical rods of protein called myofibrils, which are roughly 2 μm in diameter, while the myofibrils themselves are made of strands of actin and myosin filaments, which total some 60 nm in diameter [24.1]. This gives 10^{12} myosin filaments per cross-sectional area of the muscle. As each myosin filament over the length of a single sarcomere (the contractile unit of myofibrils, some 2.5 μm long) contains some 300 myosin heads, lifting a 30 kg load corresponds to a force of 1 pN per myosin head. This is certainly an underestimate, since not all myosin heads are attached to the actin filament at the same time. Since our estimate leaves out many details, we might wonder how it compares to the measured forces exerted by molecular motors. Sophisticated experiments with optical tweezers have revealed that actin-myosin motors stall at a force of around 5 pN [24.44]; RNA polymerase stalls at a force of about 20 pN [24.39]; the portal motor of the ϕ-29 bacteriophage exerts forces up to 50 pN [24.29].

As noted earlier, RNA polymerase is a molecular motor that moves along DNA while transcribing genes into messenger RNA. The DNA itself is an elastic object that is deformed by forces exerted on it by various proteins. At forces less than 0.1 pN its response is that of an entropic spring with a stretch modulus of 0.1 pN, while at forces exceeding 10 pN its stretch modulus is determined by hydrogen bonding of the base pairs and is roughly 1,000 pN [24.45]. All the above mentioned data reinforces the argument that piconewton is the relevant unit of force in the nanomechanical world of the cell.

The concept of "stress" is closely related to that of force. Stress is a continuum mechanical concept of force per unit area, and it has been used with great success in solid and fluid mechanics. We extend the idea of stress to the nanomechanical level to see what numbers we arrive at. From the data above we can deduce that a single myosin fiber sustains a stress of about 10^{-2} pN/nm^2, which is the same as 10^{-2} MPa. Migratory animal cells such as fibroblasts, which are responsible for scavenging and destroying undesirable products in tissues, can generate a maximum stress on their substrate on the order of 3.0×10^{-2} MPa [24.46]. DNA can sustain stresses in the excess of 20 pN/nm^2, or 20 MPa, at which point an interesting structural transition accompanied by an overall increase of its contour length is observed [24.45]. Engineering materials such as steel and aluminium, on the other hand, can sustain stresses of about 100 MPa. This goes to show that nanotechnology in the context of biological systems is built from rather soft materials.

It is of interest to translate our intuition concerning piconewton forces into corresponding energetic terms. The kinesin motor advances 8 nm in each step against forces as high as 5 pN [24.25]. This translates to a work done on the order of 40 pN nm. A myosin motor suffers displacements of about 15 nm with forces in the piconewton range. ATP hydrolysis (to ADP) releases energy (at pH 7 and room temperature) of about 50 pN nm [24.1]. When a titin molecule is pulled, it unfolds under forces of 30–300 pN, causing discrete expansions of 10–30 nm [24.22], implying energies in the range of 300–9,000 pN nm. Experimental data provides a compelling argument in favor of thinking of the pN nm as a unit of energy for nanomechanics. However, the observation that $k_B T = 4.1$ pN nm at room temperature gives important insight, since it reveals that thermal forces and entropic effects play a competing role in biological nanomechanics. This provides nature with unique design challenges, whereby molecular motors that can perform useful work must do so in the presence of strong thermal fluctuations for the normal functioning of the cell. Operation of motors in such a noisy environment is governed by laws that are probabilistic in nature. This is manifest in single molecule experiments that observe motors stalling and sometimes reversing direction.

Energy conversion is crucial to any developmental or evolutionary process. The steam engine powered the industrial revolution. However, long before thinking beings with man-made machines founded the industrial revolution, power generation had already become a central part of life's nanotechnology. Indeed, the development of ATP-synthase is one of the cornerstones for the evolution of higher life-forms. An inevitable concomitant of evolution is the necessity for faster and more efficient operations. The industrial revolution led to the emergence of bigger and faster modes of transport; biological evolution led to the emergence of complex and

intelligent organisms. Invariably, a machine or an organism is limited in its abilities by the speed at which it can convert one form of energy (usually chemical) to other forms of energy (usually mechanical). This is why a study of their "power plants" becomes important.

Power plants (or engines) are usually characterized by their force-velocity curves and compared using their power-to-weight (P/W) ratio. For example, the myosin motor has a P/W ratio of 2×10^4 W/kg, the bacterial flagellar motor stands at 100 W/kg, an internal combustion engine is at about 300 W/kg, and a turbojet engine stands at 3,000 W/kg [24.47]. These figures tell us that linear motors like myosin and kinesin are extremely powerful machines.

24.3 Modeling at the Nano–Bio Interface

We have already provided a number of different views of the biology-nanotechnology interface, all of which reveal the insights that can emerge from model building. Indeed, one of the key thrusts of this entire chapter is the view that as the type of data that emerges concerning biological systems becomes increasingly quantitative, it must be responded to with models that are also quantitative [24.30]. The plan of this section is to show how atomistic and continuum analyses each offer insights into problems of nanotechnological significance, but under some circumstances, both are found wanting and it is only through a synthesis of both types of models that certain problems will surrender. The plan of this section is to examine the advantages and difficulties associated with adopting both atomistic and continuum perspectives and then to hint at the benefits of seeking mixed representations.

24.3.1 Tension Between Universality and Specificity

One of the key insights concerning model building in nanomechanics, whether we are talking about the nanoscale tribological questions pertinent to magnetic recording or the operations of molecular motors, is that such questions live in the no-man's-land between traditional continuum analysis at one extreme and all-atom approaches such as molecular dynamics on the other. Indeed, there is much discussion about the breakdown of continuum mechanics in modeling the mechanics of systems at the nanoscale. This dichotomy between continuum theories, which treat matter as continuously distributed, and atomic-level models, which explicitly acknowledge the graininess of matter, can be restated in a different (and perhaps more enlightening) way. In particular, it is possible to see atomistic and continuum theories as offering complementary views of the same underlying physics. Continuum models are suitable for characterizing those features of a system that can be thought of as averages over the underlying microscopic fluctuations. By way of contrast, atomistic models reveal the details that a continuum model will never capture, and in particular, they shed light on the specificity of the problem at hand.

The perspective adopted here is that continuum models and atomistic models each reveal important features of a given problem. For example, in contemplating the competition between fracture and plasticity at crack tips, a continuum analysis provides critical insights into the nature of the elastic fields surrounding a defect such as a crack. These fields adopt a fundamental and universal form at large scales with all detailed material features buried in simple material parameters. By way of contrast, the precise details of the dissipative processes occurring at a crack tip (in particular, the competition between bond breaking with the creation of new free surface and dislocation nucleation) require detailed atomic-level descriptions of the energetics of bond stretching and breaking. In the biological setting, similar remarks can be made. In certain instances, the description of biological polymers as random coils suffices and yields insights into features such as the mean size of the polymer chain as a function of its length. On the other hand, if our objective is mechanistic understanding of processes such as how phosphorylation of a particular protein induces conformational change, this is an intrinsically atomic-level question. The language we invoke to describe this dichotomy is the use of the terms *universality* and *specificity*, where, as described above, insights of a universal character refer to those features of systems that are generic, while specificity refers to the features of systems that depend upon precise details such as whether or not a particular molecule is bound at a particular site. This fundamental tension between atomistic and continuum perspectives is elaborated in *Phillips* et al. [24.48].

24.3.2 Atomic-Level Analysis of Biological Systems

As already described in the introduction, one of the intriguing roles of nanotechnology in the biological setting is that it has brought the Instron technology of traditional solid mechanics to the nanoscale and has permitted the investigation of the force-extension characteristics of nanoscale systems (macromolecules and their assemblies, in particular). As noted earlier, mechanical force spectroscopy [24.20] is emerging as a profound tool for exploring the connection between structure, force, and chemistry, in much the same way that conventional stress-strain tests provided insights into the connection between structure and properties of conventional materials. Figure 24.3 gives one such example in the case of the muscle protein titin. Similar insights have been obtained through systematic examination of the force-extension properties of DNA [24.49].

The objective of this section is to call the reader's attention to the types of modeling that can be done from an all-atom perspective. What exactly is meant by a model in this setting? We begin by noting that for the purposes of the general discussion given here, the same basic ideas are present whether one is modeling tribological processes such as the sliding of adjacent surfaces, or attempting to examine the operation of a protein machine such as ATPsynthase. The set of degrees of freedom considered by the atomic-scale modeler is the full set of atomic positions $(\boldsymbol{R}_1, \boldsymbol{R}_2, \cdots \boldsymbol{R}_N)$, which we also refer to as $\{\boldsymbol{R}_i\}$. For most purposes, one proceeds through reference to a *classical* potential energy function, $E_{\text{tot}}(\{\boldsymbol{R}_i\})$, which is a rule that assigns an energy for every configuration $\{\boldsymbol{R}_i\}$. While it would be most appealing to be able to perform a full quantum mechanical analysis, such calculations are computationally prohibitive. Given the potential energy, the forces on each and every atom can be computed where, for example, the force in the αth Cartesian direction on the ith atom is given by

$$F_{i,\alpha} = -\frac{\partial E_{\text{tot}}}{\partial R_{i,\alpha}} \,. \tag{24.8}$$

For those interested in finding the energy minimizers, such forces can be used in conjunction with methods such as the conjugate gradient method or the Newton–Raphson method. Alternatively, many questions are of a dynamic character, and in these cases, Newton's equations of motion are integrated, thus permitting an investigation of the temporal evolution of the system of interest. We note that we have neglected to discuss subtleties of how one maintains the system at constant temperature, and we leave such subtleties to the curious reader, who can learn more about them in *Frenkel* and *Smit* [24.50].

To give a flavor of where such calculations can lead we note that the same force-extension characteristics already shown in Fig. 24.3 have been computed in a molecular dynamics simulation [24.51]. One of the insights to emerge from these calculations was the particular dynamical pathway, namely, the breaking of a particular collection of hydrogen bonds during the rupture process of each of the immunoglobulin domains. We further note that in the case of titin there has been an especially pleasing synergy between the atomic-scale calculations and the corresponding force-extension measurements. In particular, in response to the suggestion that it was a particular set of hydrogen bonds that were impugned in the rupture process, mutated versions of the titin protein were created in which the number of such hydrogen bonds was changed with the result that the rupture force was changed according to expectation [24.52]. This can be seen as a primitive example of the ultimate goal of tailoring new materials (both biological and otherwise) through appropriate computer modeling.

24.3.3 Continuum Analysis of Biological Systems

We have already noted that there are many appealing features to strictly continuum analyses of material systems. In particular, models based upon continuum mechanics result in a mathematical formulation that permits us to uncork the traditional tools associated with partial differential equations and functional analysis. In keeping with our argument that it is the role of models to serve our intuition, the ability to write down continuum models raises the possibility of obtaining analytic solutions to problems of interest, an eventuality that is nearly impossible once the all-atom framework has been adopted.

Macromolecules as Elastic Rods

Whether we contemplate the information carrying nucleic acids, the workhorse proteins, or energy storing sugars, ultimately, the molecular business of the living world is dominated by long chain molecules. For the model builder, such molecules suggest two complementary perspectives, each of which contains a part of the whole truth. On the one hand, polymer physicists have gained huge insights [24.53] by thinking of long chain molecules as random walks. Stated simply, the key virtue

of the random-walk description of long chain molecules is that it reflects the overwhelming importance of entropy in governing the geometric conformations that can be adopted by macromolecules. The other side of the same coin considers long chain molecules as elastic rods with a stiffness that governs their propensity for bending. We will take up this perspective in great detail in the final section of this chapter when we consider the energetics of DNA packaging in viruses.

For our present purposes, we examine one case study in treating macromolecules as elastic rods. The example of interest here is chosen in part because it reflects another fascinating aspect of biological nanotechnology, namely, regulation and control. It is well-known that genes are switched on and off as they are needed. These topics are described beautifully in the work of *Ptashne* [24.54]. The basic idea can be elucidated through the example of a particular set of genes in *E. coli*: the *lac* operon. There are a set of enzymes exploited by this bacterial cell when it needs to digest the sugar lactose. The gene that codes for these enzymes is only turned on when lactose is present and certain other sugars are absent. The nanotechnological solution to the problem of regulating this gene has been solved by nature through the presence of a molecule known as the *lac* repressor, which binds onto the DNA in the region of this *lac* gene and prevents the gene from being expressed. More specifically, the *lac* repressor binds onto two different regions of the DNA molecule simultaneously, forming a loop in the intervening region, and thus rendering that region inoperative for transcription.

We have belabored the difficulties that attend the all-atom simulation of most problems of nanotechnological interest – system sizes are too small and simulated times are too short. On the other hand, clearly it is of great interest to probe the atomic-level dynamics of the way that molecules such as *lac* repressor interact with DNA and thereby serve as gene regulators. A compromise position adopted by *Balaeff* et al. [24.55] was to use elasticity theory to model the structure and energetics of DNA, which then served effectively as a boundary condition for their all-atom calculation of the properties of the *lac* repressor/DNA complex. The simulation cell is shown in Fig. 24.6, which also shows the looped region of DNA that serves as a boundary condition for the atomic-level model of the complex.

Membranes as Elastic Media

One way to classify biological structures is along the lines of their dimensionality. In the previous discussion, we examined the sense in which many of the key macromolecules of the living world can be thought of as one-dimensional rods. The next level in this dimensional hierarchy is to consider the various membranes that compartmentalize the cell and which can be examined from the perspective of two-dimensional elasticity. Just as there are a huge number of modeling questions to be posed concerning the structure and function of long chain molecules such as nucleic acids and proteins, there is a similar list of questions that attend the presence of a host of different membranes throughout the cell. To name a few, we first remind the reader that such membranes are full of various proteins that serve in a variety of different capacities, some of which depend upon the mechanical state of the membrane itself. In a different vein, there has been great interest in examining the factors that give rise to the equilibrium shapes of cells [24.56, 57]. In both instances, the logic associated

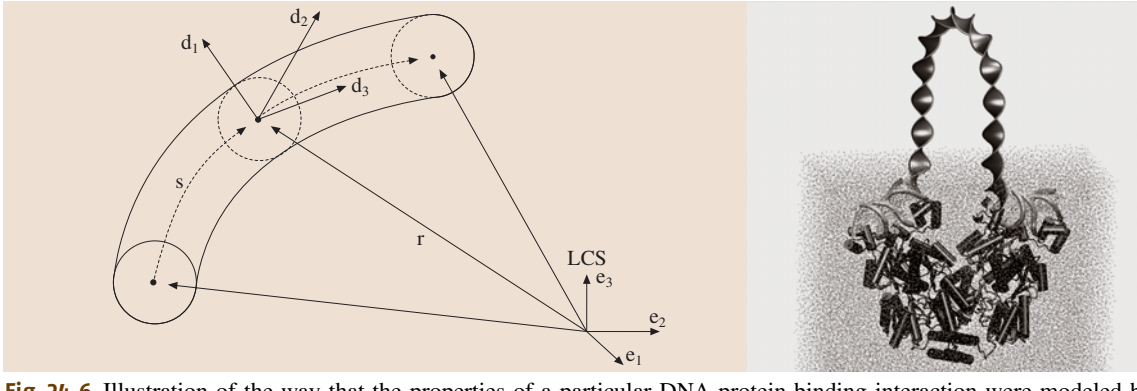


Fig. 24.6 Illustration of the way that the properties of a particular DNA-protein binding interaction were modeled by virtue of an elastic treatment of the looped region of the DNA [24.48]

with building models of these phenomena centers on the construction of an elastic Hamiltonian that captures the energetics of deformations expressed in terms of surface area, curvature, and variations in thickness. Also, just as there has been great progress in measuring the properties of single proteins and nucleic acids, there has been considerable progress in examining the force response of membranes as well [24.58].

The prospect of bringing the tools of traditional continuum theory to bear on problems of biological significance is indeed a daunting one. As noted earlier, to our way of thinking one of the biggest challenges posed by biological problems is the dependence of these systems on detailed molecular structures and dynamics – what we have earlier characterized as "specificity", one of the hallmarks of biological action.

24.4 Nature's Nanotechnology Revealed: Viruses as a Case Study

Over the course of this chapter, we have presented different facets of the relationship between biology and nanotechnology and the modeling challenges they encompass. For example, we have noted that two key aspects of nature's nanotechnologies are the exploitation of self-assembly processes for the construction of molecular machines and the role of active processes mediated by molecular motors. We have also argued for the synergistic role of nanotechnology in producing new methods for the experimental analysis of biological systems with examples ranging from new molecular dyes to the use of optical tweezers. Finally, we have described some of the modeling challenges posed by contemplating the biology-nanotechnology interface and the sorts of coarse-grained models that have arisen to meet these challenges. In this final section, we present a discussion of viruses as a case study at the confluence of these different themes.

Though the importance of viruses from a health perspective are well-known even to the casual observer, they are similarly important both from the technological perspective (as will be explained below) and as a compelling and profound example of nature's nanotechnology par excellence. To appreciate the sense in which viruses serve as a compelling example of biological nanotechnology, we begin by reviewing the nature of the viral life cycle with special reference to the case of bacterial viruses (i. e., viruses that infect bacteria), known as bacteriophages. For concreteness, we consider the life cycle of a bacteriophage such as the famed lambda phage that infects *E. coli*. The life cycle of such viruses is shown schematically in Fig. 24.7. Upon an encounter with the *E. coli* host, the virus attaches to a receptor (protein) embedded in the bacterial membrane and ejects its DNA into the host cell, leaving an empty capsid as refuse from the process. As an aside, it is worth noting that experiments such as the famed Hershey–Chase experiment used tagged DNA and tagged proteins on viruses to settle the question of whether proteins or nucleic acids are the carriers of genetic information. The outcome of these experiments was the conclusion that DNA is the genetic material. For the purposes of the present discussion, the other interesting outcome of the Hershey–Chase experiment is that it provides insights into the mechanistic process associated with delivery of the viral genome.

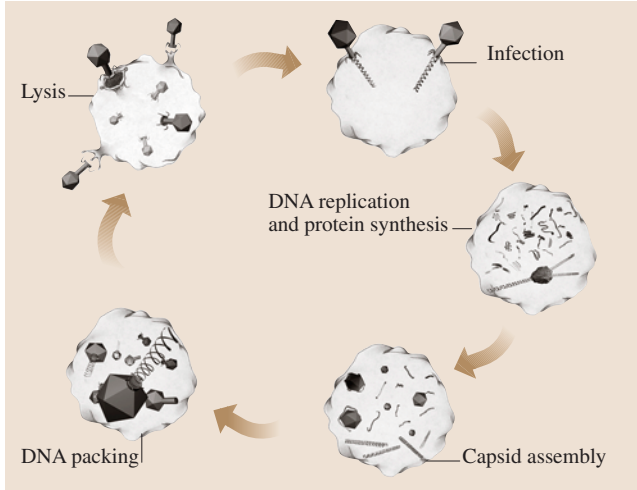


Fig. 24.7 Schematic representation of the viral life cycle illustrating the various nanotechnological actions in this cycle. During the infection stage, the viral genome is delivered to the victimized bacterial cell. During the process labeled DNA replication and protein synthesis, the machinery of the host cell is enlisted to produce renegade DNA and proteins to be used in the assembly of new viral particles. The capsid assembly process involves the attachment of the various protein building blocks that make up the capsid. DNA packing refers to the active packaging of DNA into the newly formed capsids by a molecular motor. Finally, once the assembly process is complete, the infectious phage particles are released from the victimized cell to repeat their infectious act elsewhere

Once the viral genome has been delivered to the host cell (we now oversimplify), the replication machinery of *E. coli* is hijacked to do the virus' bidding. In particular, the genes coded for in the viral DNA are expressed, and the proteins needed to make copies of the phage particle (i. e., the ingredients for an eventual self-assembly process) are created. Interestingly, part of the gene products associated with this process are the components of the molecular motor, which is responsible for packaging the replicated viral DNA into the new protein capsids that will eventually become the next generation of viruses. Indeed, once assembled, this motor takes the replicated viral DNA and packs it into the viral capsid. Recall our insistence that another of the important themes presented by biological examples of nanotechnology is the huge role played by active processes that reflect mechanochemical coupling. Once the packaging process is completed, and the remaining parts of the self-assembly process have been effected (such as the attachment of the viral sheath and legs whereby the phage attaches to host cells), enzymes are released that breakdown the cell wall of the infected cell with the ultimate result that what started as a single bacteriophage has in less than an hour become on the order of 100 new phage particles ready to infect new cells.

As noted above, a second sense in which viruses are deserving of case study status in the current chapter is the role viruses such as lambda phage play in biotechnology itself. For example, from the standpoint of both cloning and the construction of genomic libraries, the use of viruses is commonplace. From a more speculative perspective, viruses are also drawing increasing attention from the standpoint of gene therapy (as a way of delivering DNA to specific locations) and, more generally, as small-scale containers. To substantiate these assertions, we briefly consider the construction of genomic libraries and the role played by viruses in these manipulations. We pose the following question: Given that the length of the human genome is on the order of 10^9 base pairs, how can one organize and store all of this genetic information for the purposes of experiments such as the sequencing of the human genome? One answer to this challenge is the use of bacteriophage to deliver phage DNA to *E. coli* cells, but with the subtlety that the delivered DNA fragments have ligated within them fragments of the human genome whose lengths are on the order of 10 kb (kb = kilo base-pairs). The particulars of this procedure involve first cutting the DNA into fragments of roughly 10 kb in length using a class of enzymes known as restriction enzymes. The result of this operation is that the original genome is now separated into a random collection of fragments. These fragments are then mixed with the original lambda phage DNA that has also been cut at a single site such that the genomic fragments and the lambda fragments are complementary. Using a second enzyme known as ligase, the genomic DNA is joined to the original lambda phage fragments so that the resulting DNA resembles the original lambda DNA, but now with a 10 kb fragment inserted in the middle. These cloned lambda DNA molecules are then packaged into the lambda phage using a packaging reaction (see the website www.stratagene.com for an example of these products). The resulting lambda phage, now fully packed with cloned DNA, are used to infect *E. coli* cells, and the cloned DNA, once in the bacterial cell, circularizes into a DNA fragment known as a plasmid. The plasmid is then passed from one generation of *E. coli* to the next. This is the so-called prophage pathway in which, unlike the lytic pathway shown in Fig. 24.7, the bacteriophage is latent and does not destroy the cell [24.1]. Note that different *E. coli* cells are infected with viruses containing different cloned fragments. As a result, the collection of all such *E. coli* cells constitute a library of the original genomic DNA.

We round out our introductory discussion on viruses with a discussion of the compelling recent experiments that have been performed to investigate the problem of viral packing and, similarly, how ideas like those described in this chapter can be used to model these processes. *Smith* et al. [24.29] used optical tweezers to measure the force applied by the packaging motor of the ϕ-29 bacteriophage during the DNA packaging process. In particular, they measured the force and rate of packaging as a function of the amount of DNA packed into the viral head with the result that as more DNA is packed, the resistive force due to the packed DNA increases, and the packing rate is reduced.

As noted above, the viral problem is interesting not only because it exemplifies many of the features of biological nanotechnology introduced throughout this chapter, but also illustrates the way in which model building has arisen in response to experimental insights. The various competing energies that are implicated in the DNA packing process have been described by *Riemer* and *Bloomfield* [24.59]. The energetics of viral packing is characterized by a number of different factors, including: i) the entropic-spring effect that causes the DNA in solution to adopt a more spread out configuration than that in the viral capsid, ii) the energetics of elastic bending, which results from inducing curvature in the DNA on a scale that is considerably smaller than the persistence length of $\xi_p \approx 50$ nm, and iii) those

factors related to the presence of charge both on the DNA itself and in the surrounding solution. As shown by *Riemer* and *Bloomfield* [24.59], the entropic contribution is smaller by a factor of 10 or more relative to the bending energies and those mediated by the charges on the DNA and the surrounding solution, and hence we make no further reference to it. As a result, just like in earlier work [24.60, 61], we examine the interplay of elastic and interaction forces, though we neglect surface terms originating from DNA-capsid and DNA-solvent interactions.

We note that the viral packing process involves DNA segments with lengths on the order of 10 μm and takes place on the time scale of minutes. As a result, from a modeling perspective, it is clear that such problems are clearly out of reach of conventional molecular dynamics. As a result, we exploit a continuum description of the DNA packing process with the proviso that such models will ultimately need to be refined to account for the sequence dependence of the elasticity of DNA.

A mathematical description of the energetics of viral packing must account for two competing factors, namely, the energy cost to bend the DNA and place it in the capsid and the repulsive interaction between adjacent DNA segments that are too close together. The structural picture of the packaged DNA is inspired by experiments that indicate that the DNA is packed in concentric rings from large radii inwards [24.62–64]. The bending energy cost of accumulating hoops within the capsid is given by

$$E_{\mathrm{el}} = \pi \xi_{\mathrm{p}} k_{\mathrm{B}} T \sum_i \frac{N(R_i)}{R_i} \,, \tag{24.9}$$

where $N(R_i)$ is the number of hoops that are packed at the radius R_i [24.59]. The basic idea of this expression is that we are adding up the elastic energy on a hoop by hoop basis, with each hoop penalized by the usual energy cost to bend an elastic beam into a circular arc. The presence of $N(R_i)$ reflects the fact that because of the shape of the capsid, as the radius gets smaller the DNA can pack higher up into the capsid, thus increasing the number of allowed hoops. To make analytic progress with the expression for the stored elastic energy given above, we convert it into an integral of the form

$$E_{\mathrm{el}} = \frac{\pi \xi_{\mathrm{p}} k_{\mathrm{B}} T}{\sqrt{3} d_{\mathrm{s}}/2} \int_R^{R_{\mathrm{out}}} \frac{N(R')}{R'} \, \mathrm{d}R' \,. \tag{24.10}$$

The summation $\sum_i$ has been replaced by an integral $\int_R^{R_{\mathrm{out}}} \mathrm{d}R'/(\sqrt{3} d_{\mathrm{s}}/2)$, where the integration bounds are the inner and outer radii of the inner spool and $\sqrt{3} d_{\mathrm{s}}/2$ is the horizontal spacing between adjacent strands of the DNA. The geometrical factor $\sqrt{3}/2$ owes its presence to the hexagonal packing of the DNA strands. R is the radius of the innermost stack of hoops, R_{out} is the radius of the capsid, $N(R')$ is the number of hoops at radius R', d_{s} is the spacing between adjacent hoops, and ξ_{p} is the persistence length of DNA. We note that this expression is a manifestation of the first term in the energy functional in (24.3), used earlier in the context of the biofunctionalized cantilever and now specialized to the geometry of a partially filled capsid. Just as the energy functional in the cantilever context reflected a competition between different contributions to the total energy (in that case, elastic bending energy and surface terms), the DNA packing problem reflects a similar competition, this time between the elastic bending and the interactions between nearby DNA segments by virtue of the charge along the DNA. In particular, the interaction energy between the DNA hoops packed in the viral capsid scales with the length and is given by

$$E_{\mathrm{int}} = F_0 L (c^2 + c d_{\mathrm{s}}) \exp(-d_{\mathrm{s}}/c) \,, \tag{24.11}$$

where c and F_0 are constants for a given solvent and L is the length of the packed DNA. This form of the interaction energy was shown to be very robust for a variety of solvents containing monovalent and divalent cations in a series of experiments by *Rau* et al. [24.65, 66] and *Parsegian* et al. [24.67]. The forces arising from this kind of an energy are purely repulsive in character. Interestingly, in this setting, rather than using temperature as a control parameter, it is much more common to change the circumstances in such experiments by tuning concentrations of various chemical constituents. In particular, when trivalent or tetravalent cations are present in solution there is an effective attractive interaction between DNA segments. In that case, the forces are repulsive only when d_{s} is smaller than a certain critical separation d_0 (and attractive otherwise) and the form of the interaction energy changes to

$$E_{\mathrm{int}} = F_0 L (c^2 + c d_{\mathrm{s}}) \exp\left(\frac{d_0 - d_{\mathrm{s}}}{c}\right) \,. \tag{24.12}$$

Once both the elastic and interaction contributions to the energy have been reckoned, we are in a position to compute the total energy. If we recognize that the length L of the packed DNA is given by

$$L = \frac{2}{\sqrt{3} d_{\mathrm{s}}} \int_R^{R_{\mathrm{out}}} 2\pi R' N(R') \, \mathrm{d}R' \,, \tag{24.13}$$

then the total energy for the packed DNA, in terms of L and d_s, is given by

$$E(L, d_s) = \frac{\pi \xi_p k_B T}{\sqrt{3} d_s/2} \int_R^{R_{out}} \frac{N(R')}{R'} \, dR' \tag{24.14}$$

$$+F_0(c^2 + c d_s) \exp(-d_s/c) \frac{2}{\sqrt{3} d_s} \int_R^{R_{out}} 2\pi R' N(R') \, dR' \, .$$

The spacing at a given length L is determined by requiring that the packed DNA be in a minimum energy configuration, which is equivalent to asking that $\partial E/\partial d_s = 0$. Hence, for a given geometry and length packed, one obtains d_s and thereby calculates the energy. Thus the energy is known as a function of the length packed, and to compute the force as a function of length of DNA packed one need only differentiate this energy with respect to the length packed. The result of this procedure is an expression for the force of the form

$$F(L) = F_0(c^2 + c d_s) \exp\left(-\frac{d_s}{c}\right) + \frac{\xi_P k_B T}{2R^2} \, . \tag{24.15}$$

This is a generic expression valid for a capsid of any shape. The effect of the geometry is captured in the variation of d_s as a function of the length packed, as well as through the inner packing radius R. The resulting force-packing curves are shown in Fig. 24.8. In particular, this figure shows the packing force as a function of a fraction of the genome packed for a number of different solutions, with the different curves revealing the large role played by positive counterions in dictating the overall energetics of these processes.

On the basis of calculations like those given above, *Kindt* et al. [24.61] estimate that the pressure inside the capsid of the phage is on the order of 35 atmospheres. A crude estimate by *Smith* et al. [24.29] gives a figure close to 60 atmospheres. These numbers are intriguing in their own right, but more importantly they demonstrate the promise that proteins and other biological materials hold as candidates for engineering materials. To further explore the structural integrity of viruses in their role as pressurized protein shells, we now turn to an examination of the rupture stress of viral capsids.

Throughout this chapter, one of our main arguments has been the idea that whether it is our ambition to model conventional materials or their biological counterparts, modeling at the nanoscale implies challenges that can not be met either by purely continuum ideas or by traditional all-atom thinking. Indeed, one set of powerful methods is built around the attempt to borrow those features of continuum and atomistic thinking that are most robust, while rejecting those that are inapplicable at the atomic scale. As an example of this type of thinking, we examine another question drawn from the viruses-as-nanotechnology setting. In particular, since viruses have been considered a means of transporting material other than the genetic material of the virus itself, it is of interest to consider the maximum internal pressure viral capsids can sustain without rupture, as this will determine how much material can be safely packaged. To that end, we use continuum mechanics to estimate the stresses within the capsid walls. These stresses are then mapped onto atomic-level forces by appealing to the details of the protein structure of the monomers making up the capsid and a knowledge of the forces that link them. By relating the continuum and atomistic calculations, we then determine the maximum sustainable internal pressure.

We imagine the capsid to be a hollow sphere loaded by a pressure p_i from inside and a pressure p_o from outside. The inner and outer radii are R_i and R_o, respectively. As a representative example, bacteriophage

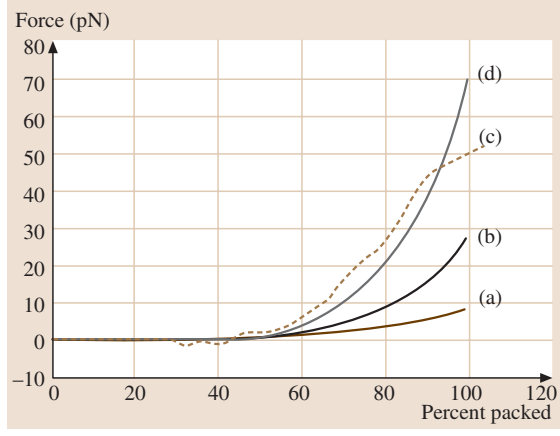


Fig. 24.8 Force as a function of the percent packed for a cylindrical capsid under purely repulsive solvent conditions. The dimensions of the capsid and the length of genome packed were chosen to represent the $\phi 29$ phage: $R_{out} = 210$ Å, $z = 540$ Å, and $L = 6.6$ µm. Curve (c) shows the experimental results of *Smith* et al. [24.29], while theoretical curves (a), (b), and (d) are given by (24.15), with $F_0 = 10 \times 55{,}000$, $40 \times 55{,}000$, and $100 \times 55{,}000$ pN/nm^2, respectively

GA is characterized geometrically by $R_i = 123$ Å and $R_o = 145$ Å. Evaluation of a number of different capsid types suggests that treating capsids as though they have a mean thickness of roughly 15 Å suffices for the level of modeling being considered here. For the purposes of computing the internal stresses within the capsid, we begin with a statement of equilibrium from continuum mechanics that requires that at every point in the capsid

$$\nabla \cdot \sigma = \mathbf{0}\,, \tag{24.16}$$

where σ is the stress tensor comprising three normal stresses and three shear stresses. For a problem with spherical symmetry, like that considered here, the stresses reduce to a radial stress σ_R and a circumferential stress σ_T. Solution of the equilibrium equations results in stresses of the form [24.32]

$$\sigma_R = \frac{C}{r^3} + D, \quad \sigma_T = -\frac{C}{2r^3} + D\,. \tag{24.17}$$

By using the boundary conditions $\sigma_R|_{r=R_i} = -p_i$ and $\sigma_R|_{r=R_o} = -p_o$, the constants C and D can be determined with the result

$$\sigma_R = \frac{p_o R_o^3 \left(r^3 - R_i^3\right)}{r^3 \left(R_i^3 - R_o^3\right)} + \frac{p_i R_i^3 \left(R_o^3 - r^3\right)}{r^3 \left(R_i^3 - R_o^3\right)}\,, \tag{24.18}$$

$$\sigma_T = \frac{p_o R_o^3 \left(2r^3 + R_i^3\right)}{2r^3 \left(R_i^3 - R_o^3\right)} - \frac{p_i R_i^3 \left(2r^3 + R_o^3\right)}{2r^3 \left(R_i^3 - R_o^3\right)}\,. \tag{24.19}$$

The stress σ_T is our primary concern, since it acts so as to tear the sphere apart. By looking at the expressions above, we can see that this stress is maximum at $r = R_i$ and the maximum value is given by

$$\sigma_T^{\max} = \frac{3p_o R_o^3 - p_i \left(2R_i^3 + R_o^3\right)}{2\left(R_i^3 - R_o^3\right)}\,. \tag{24.20}$$

We note that elasticity theory in and of itself is unable to comment on $\sigma_T^{\max}$, since this is effectively a material parameter that characterizes the contacts between the various protein monomers that make up the capsid. As a result, we first examine how the rupture strength depends upon capsid dimensions in abstract terms and then turn to a concrete estimate of $\sigma_T^{\max}$ itself from several complementary perspectives. If (24.20) is rewritten using $p_i = p_o + \Delta p$ and $R_o = R_i + \Delta R$, and, further, it is realized that for typical capsid dimensions $\Delta R/R_i \ll 1$, it can be shown by rearranging (24.20) and by considering the case in which $\sigma_T^{\max}$ is much larger than p_o, the maximum sustainable internal pressure is given by

$$p_i^{\max} = \frac{2\sigma_T^{\max}\Delta R}{R_i}\,, \tag{24.21}$$

where $p_i^{\max}$ is really the quantity of interest, namely, the maximum sustainable internal pressure.

To make concrete progress to the point where we can actually estimate the rupture stress in atm, we need to consider capsids for which the structure is known and for which, at least approximately, the bonds between the monomers making up the capsid are understood. We note that in the language of fracture mechanics, what we seek is a cohesive surface model that provides a measure of the energy of interaction between two surfaces as a function of their separation [24.68]. There are a number of different ways to go about estimating the effective interaction between the monomers making up the capsid, one of which is by appealing to atomic-level calculations like those made by *Reddy* et al. [24.69]. The Viper website has systematized such information for a number of capsids, and one of the avenues we take to estimate $\sigma_T^{\max}$ is to appeal directly to their calculations. To that end, we assume that the energy of interaction per unit area as a function of separation x between adjacent monomers making up the capsid can be written in the form

$$E(x) = V_0\left[\frac{1}{4}\left(\frac{x^*}{x}\right)^{12} - \frac{1}{2}\left(\frac{x^*}{x}\right)^{6}\right]\,. \tag{24.22}$$

The motivation for this functional form is the idea that the energy of interaction between adjacent monomeric units making up the capsid is the result of van der Waals contacts, and hence our cohesive surface law has inherited the properties of the underlying atomic force fields. To proceed to an estimate of $\sigma_T^{\max}$ itself, we must determine the parameters in the cohesive model described above. To that end, we note that *Reddy* et al. [24.69] have computed the association energies of various inequivalent contacts throughout a number of different icosahedral capsids. Their calculations result in a roughly constant value of ≈ -45 cal/mole Å^2 as the association energy, which in the language of our cohesive potential results in $V_0 \approx 12.5$ pN/Å. This may be seen by noting that the association energy is given by $E(x^*) = -V_0/4$.

Once we have fixed x^*, which amounts to choosing a particular value for the equilibrium separation between two monomers, the material parameter $\sigma_T^{\max}$ is obtained by evaluating $\partial E(x)/\partial x$ at a value of x corresponding to the point of inflection ($\partial^2 E(x)/\partial x^2 = 0$) in the cohesive surface function. For the cohesive surface function used

above, this results in a maximum stress of the form

$$\sigma_T^{\max} = \frac{7^{7/6} \times 18}{13^{13/6}} \frac{V_0}{x^*} . \tag{24.23}$$

With $\sigma_T^{\max}$ in hand, the maximum sustainable pressure is obtained from (24.21), and the results are shown in Fig. 24.9. These estimates suggest that the pressures within capsids as a result of packed DNA while large, are still smaller than our estimated rupture stresses. To more completely examine this question we have undertaken finite element elasticity calculations to examine the stresses in capsids exhibiting irregularities in both shape and thickness. In addition, further atomistic analysis of $\sigma_T^{\max}$ is needed with special reference to its dependence on distance between protein units. It would also be of interest to examine mutant versions of the monomeric units making up the capsid to see the implications of such mutations for $\sigma_T^{\max}$.

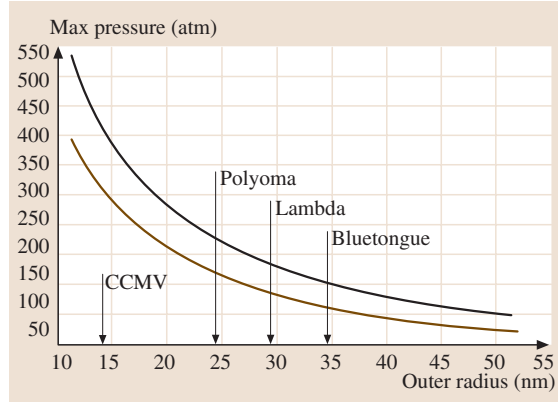


Fig. 24.9 Rupture pressure as a function of capsid radius for $x^* = 3$ Å (*upper curve*) and $x^* = 4$ Å (*lower curve*). The width of the capsid walls was set to $\Delta R = 1.5$ nm, while $V_0 = 180\,\text{cal/molÅ}^2$

24.5 Concluding Remarks

One of the most compelling areas to be touched by nanotechnology is biological science. Indeed, we have argued that there is a fascinating interplay between these two subjects, with biology as a key beneficiary of advances in nanotechnology as a result of a new generation of single molecule experiments that complement traditional assays involving statistical assemblages of molecules. This interplay runs in both directions with nanotechnology continually receiving inspiration from biology itself. The goal of this chapter has been to highlight some representative examples of the interplay between biology and nanotechnology and to illustrate the role of nanomechanics in this field and how mechanical models have arisen in response to the emergence of this new field. Primary attention has been given to the particular example of the processes that attend the life cycle of bacterial viruses. Viruses feature many of the key lessons of biological nanotechnology, including self-assembly, as evidenced in the spontaneous formation of the protein shell (capsid) within which the viral genome is packaged and a motor-mediated biological process, namely, the packaging of DNA in this capsid by a molecular motor that pushes the DNA into the capsid. We argue that these processes in viruses are a compelling real-world example of nature's nanotechnology and reveal the nanomechanical challenges that will continue to be confronted at the nanotechnology-biology interface.

Though this chapter represents something of a departure from the rest of the chapters in this volume, it advances the view that biological nanotechnology serves as an inspiring vision of what nanotechnology can offer and reflects the authors' views that biologically inspired nanotechnology will play an ever increasing role as Feynman's view that "there is plenty of room at the bottom" continues to play out. Moreover, developing simple models of nature's nanotechnology can provide important insights into viable strategies for making machines at the nanoscale.

We have argued that the advance of single molecule technology presents both scientific and technological possibilities. Both classes of questions imply significant modeling demands, just as earlier advances in other settings, such as in materials for microelectronics applications, did. We claim that these modeling challenges are perhaps more acute as a result of the vicious chemical and structural nonhomogeneity of biological systems. As a result, we see an ever increasing role for modeling methods that aim to keep atomic-level specificity where needed, while rejecting such resolution elsewhere.

References

24.1 B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter: *Essential Cell Biology* (Garland, New York 1997) Chap. 12

24.2 B. Alberts: The cell as a collection of protein machines: Preparing the next generation of molecular biologists, Cell **92** (1998) 291

24.3 D. Bray: *Cell Movements From Molecules to Motility* (Garland, New York 2001)

24.4 S. J. Flint, L. W. Enquist, R. M. Krug, V. R. Racaniello, A. M. Skalka: *Principles of Virology* (ASM, Washington, DC 2000)

24.5 J. C. M. van Hest, D. A. Tirrell: Protein-based materials, toward a new level of structural control, Chem. Commun. **19** (2001) 1807–1904

24.6 H. Berg: Motile behavior of bacteria, Phys. Today **53** (2000) 24

24.7 J. E. Walker: ATP synthesis by rotary catalysis, Angew. Chem. Int. Edn. **37** (1998) 2308

24.8 P. D. Boyer: The ATP synthase – A splendid molecular machine, Annu. Rev. Biochem. **66** (1997) 717

24.9 M. Yoshida, E. Muneyuki, T. Hisabori: ATP synthase – A marvelous rotary engine of the cell, Nature Rev. Mol. Cell Bio. **2** (2001) 669

24.10 R. Yasuda, H. Noji, K. Kinosita, M. Yoshida: F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps, Cell **93** (1998) 1117

24.11 J. Howard: *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland 2001)

24.12 B. Hille: *Ion Channels of Excitable Membranes* (Sinauer, Sunderland 2001)

24.13 S. Sukharev, S. R. Durell, H. R. Guy: Structural Models of the MscL gating mechanism, Biophys. J. **81** (2001) 917

24.14 S. Sukharev, M. Betanzos, C.-S. Chiang, H. R. Guy: The gating mechanism of the large mechanosensitive channel MscL, Nature **409** (2001) 720

24.15 K. Kinosita, R. Yasuda, H. Noji: F_1-ATPase: A highly efficient rotary ATP machine, Essays Biochem. **35** (2000) 3

24.16 E. Racker, W. Stoeckenius: Reconstitution of purple membrane vesicles catalyzing light-driven proton uptake and adenosine triphosphate formation, J. Biol. Chem. **249** (1974) 662

24.17 G. Groth, J. E. Walker: ATP synthase from bovine heart mitochondria: reconstitution into unilamellar phospholipid vesicles of the pure enzyme in a functional state, Biochem. J. **318** (1996) 351

24.18 G. Wu, R. H. Datar, K. M. Hansen, T. Thundat, R. J. Cote, A. Majumdar: Bioassay of prostrate-specific antigen (PSA) using microcantilevers, Nature Biotech. **19** (2001) 856

24.19 R. P. Feynmann: There's plenty of room at the bottom, Eng. Sci. **23** (1960) 22–36, and www.zyvex.com/nanotech/feynman.html (1959)

24.20 E. Evans: Probing the relation between force-lifetime and chemistry in single molecular bonds, Annu. Rev. Biophys. Biomol. Struct. **30** (2001) 105

24.21 T. E. Fisher, P. E. Marszalek, J. M. Fernandez: Stretching single molecules into novel conformations using the atomic force microscope, Nature Struct. Bio. **7** (2000) 719

24.22 M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. Gaub: Reversible unfolding of individual titin immunoglobulin domains by AFM, Science **276** (1997) 1109

24.23 K. Svoboda, S. M. Block: Biological applications of optical forces, Annu. Rev. Biophys. Biomol. Struct. **23** (1994) 247

24.24 C. Bustamante, J. C. Macosko, G. J. L. Wuite: Grabbing the cat by the tail: Manipulating molecules one by one, Nature Rev. Mol. Cell Bio. **1** (2000) 130

24.25 K. Svoboda, S. M. Block: Force and velocity measured for single kinesin molecules, Cell **77** (1994) 773

24.26 M. J. Schnitzer, S. M. Block: Kinesin hydrolyses one ATP per 8-nm step, Nature **388** (1997) 386

24.27 B. Maier, T. R. Strick, V. Croquette, D. Bensimon: Study of DNA motors by single molecule micromanipulation, Single Mol. **1** (2000) 145

24.28 A. A. Simpson, Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, M. G. Rossmann: Structure of the bacteriophage ϕ-29 DNA packaging motor, Nature **408** (2000) 745

24.29 D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, C. Bustamante: The bacteriophage ϕ-29 portal motor can package DNA against a large internal force, Nature **413** (2001) 748

24.30 D. Bray: Reasoning for results, Nature **412** (2001) 863

24.31 G. Wu, H. Ji, K. Hansen, T. Thundat, R. Datar, R. Cote, M. F. Hagan, A. K. Charkraborty, A. Majumdar: Origin of nanomechanical cantilever motion generated from biomolecular interactions, Proc. Nat. Acad. Sci. **98** (2001) 1560

24.32 L. D. Landau, E. M. Lifshitz: *Theory of Elasticity* (Pergamon, Oxford 1986)

24.33 D. H. Bamford, R. J. C. Gilbert, J. M. Grimes, D. I. Stuart: Macromolecular assemblies: greater than their parts, Curr. Op. Struc. Biology **11** (2001) 107

24.34 J. Frank: Single-particle imaging of macromolecules by cryo-electron microscopy, Annu. Rev. Biophys. Biomol. Struct. **31** (2002) 303

24.35 S. A. Darst: Bacterial RNA polymerase, Curr. Op. Struc. Biology **11** (2001) 155

24.36 N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz: The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, Science **289** (2000) 905

24.37 T. S. Baker, N. H. Olson, S. D. Fuller: Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs, Microbiol. Mol. Bio. Rev. **63** (1999) 862

24.38 H.-S. Chan, K. A. Dill: The protein folding problem, Phys. Today **46** (February 1993) 24

24.39 M. D. Wang, M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, S. M. Block: Force and velocity measured for single molecules of RNA polymerase, Science **282** (1998) 902

24.40 M. L. Roukes: Nanoelectromechanical Systems, Technical Digest of the 2000 Solid-State Sensor and Actuator Workshop (2000)

24.41 K. Adachi, R. Yasuda, H. Noji, H. Itoh, Y. Harada, M. Yoshida, K. Kinosita: Stepping rotation of F_1-ATPase visualized through angle-resolved single-fluorophore imaging, Proc. Nat. Acad. Sci. **97** (2000) 7243

24.42 J. Wrigglesworth: *Energy and Life* (Taylor and Francis, London 1997)

24.43 V. Smil: *Energies* (MIT, Cambridge 1999)

24.44 J. T. Finer, R. M. Simmons, J. A. Spudich: Single myosin molecule mechanics: Piconewton forces and nanometre steps, Nature **368** (1994) 113

24.45 S. B. Smith, Y. Cui, C. Bustamante: Overstretching B-DNA: The elastic response of individual double-stranded and single-sranded DNA molecules, Science **271** (1996) 795

24.46 S. Munevar, Y. Wang, M. Dembo: Traction force microscopy of migrating normal and H-ras transformed 3T3 fibroblasts, Biophys. J. **80** (2001) 1744

24.47 L. Mahadevan, P. Matsudaira: Motility powered by supramolecular springs and ratchets, Science **288** (2000) 95

24.48 R. Phillips, M. Dittrich, K. Schulten: Quasicontinuum representations of atomic-scale mechanics: From proteins to dislocations, Ann. Rev. Mat. Sci. **32** (2002) 219

24.49 T. Strick, J.-F. Allemand, V. Croquette, D. Bensimon: Twisting and stretching single DNA molecules, Prog. Biophys. Mol. Bio. **74** (2000) 115

24.50 D. Frenkel, B. Smit: *Understanding Molecular Simulation* (Academic, San Diego 1996)

24.51 H. Lu, B. Isralewitz, A. Krammer, V. Vogel, K. Schulten: Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation, Biophys. J. **75** (1998) 662

24.52 M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, J. M. Fernandez: Mechanical design of proteins studies by single-molecule force spectroscopy and protein engineering, Prog. Biophys. Mol. Bio. **74** (2000) 63

24.53 A. Y. Grosberg, A. R. Khokhlov: *Giant Molecules* (Academic, San Diego 1997)

24.54 M. Ptashne, A. Gann: *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 2002)

24.55 A. Balaeff, L. Mahadevan, K. Schulten: Elastic rod model of a DNA loop in the Lac Operon, Phys. Rev. Lett. **83** (1999) 4900

24.56 U. Seifert: Configurations of fluid membranes and vesicles, Adv. Phys. **46** (1997) 13

24.57 D. Boal: *Mechanics of the Cell* (Cambridge Univ. Press, Cambridge 2002)

24.58 D. Leckband, J. Israelachvili: Intermolecular forces in biology, Q. Rev. Biophys. **34** (2001) 105

24.59 S. C. Riemer, V. A. Bloomfield: Packaging of DNA in bacteriophage heads: Some considerations on energetics, Biopolymers **17** (1978) 785

24.60 T. Odijk: Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress, Biophys. J. **75** (1998) 1223

24.61 J. Kindt, S. Tzlil, A. Ben-Shaul, W. Gelbart: DNA packaging and ejection forces in bacteriophage, Proc. Nat. Acad. Sci. **98** (2001) 13671

24.62 K. E. Richards, R. C. Williams, R. Calendar: Mode of DNA packing within bacteriophage heads, J. Mol. Bio. **78** (1973) 255

24.63 M. E. Cerritelli, N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, A. C. Steven: Encapsidated conformation of bacteriophage T7 DNA, Cell **91** (1997) 271

24.64 N. H. Olson, M. Gingery, F. A. Eiserling, T. S. Baker: The structure of isomeric capsids of bacteriophage T4, Virology **279** (2001) 385

24.65 D. C. Rau, B. Lee, V. A. Parsegian: Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: Hydration forces between parallel DNA double helices, Proc. Natl. Acad. Sci. **81** (1984) 2621

24.66 D. C. Rau, V. A. Parsegian: Direct measurement of the intermolecular forces between counterion-condensed DNA double helices, Biophys. J. **61** (1992) 246

24.67 V. A. Parsegian, R. P. Rand, N. L. Fuller, D. C. Rau: Osmotic stress for the direct measurement of intermolecular forces, Meth. Enzymol. **127** (1986) 400

24.68 R. Phillips: *Crystals, Defects and Microstructures* (Cambridge Univ. Press, Cambridge 2001)

24.69 V. S. Reddy, H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks, J. E. Johnson: Energetics of quasiequivalence: Computational analysis of protein-protein interactions in icosahedral viruses, Biophys. J. **74** (1998) 546
The parameters can be found at http://www.scripps.edu/pub/olson-web/gmm/autodock/ad305/Using_AutoDock_305.a.html

25. Mechanical Properties of Nanostructures

Knowledge of the mechanical properties of nanostructures is necessary for designing realistic MEMS/NEMS devices. Microelectromechanical systems (MEMS) refer to microscopic devices that have a characteristic length of less than 1 mm but more than 1 μm and combine electrical and mechanical components. Nanoelectromechanical systems (NEMS) refer to nanoscopic devices that have a characteristic length of less than 1 μm and combine electrical and mechanical components.

Elastic and inelastic properties are needed to predict deformation from an applied load in the elastic and inelastic regimes, respectively. The strength property is needed to predict the allowable operating limit. Some of the properties of interest are hardness, elastic modulus, bending strength, fracture toughness, and fatigue strength.

Structural integrity is of paramount importance in all devices. Load applied during the use of devices can result in component failure. Cracks can develop and propagate under tensile stresses, leading to failure. Atomic force microscopy and nanoindenters can be used satisfactorily to evaluate the mechanical properties of micro/nanoscale structures for use in MEMS/NEMS.

The most commonly used materials are single-crystal silicon and silicon-based materials, e.g., SiO_2 and polysilicon films deposited by low-pressure chemical vapor deposition. An early study showed silicon to be a mechanically resilient material in addition to its favorable electronic properties. Single-crystal SiC deposited on large-area silicon substrates is used for high-temperature micro/nanosensors and actuators. Amorphous alloys can be formed on both metal and silicon substrates by sputtering and plating techniques, providing more flexibility in surface-integration. Electroless deposited Ni-P amorphous thin films have been used to construct microdevices, especially using the so-called LIGA techniques. Micro/nanodevices need conductors to provide power, as well as electrical/magnetic signals to make them functional. Electroplated gold films have found wide applications in electronic devices because of their ability to make thin films and process simply. Use of SiC, Ni-P, and Au films, together with silicon and silicon-based materials, opens up new design opportunities for MEMS/NEMS devices.

This chapter presents a review of mechanical property measurements on the nanoscale of various materials of interest and stress and deformation analyses of nanostructures.

25.1 **Experimental Techniques for Measurement of Mechanical Properties of Nanostructures** 765
- 25.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters 765
- 25.1.2 Bending Tests of Nanostructures Using an AFM 765
- 25.1.3 Bending Tests Using a Nanoindenter 769

25.2 **Experimental Results and Discussion** 770
- 25.2.1 Indentation and Scratch Tests of Various Materials Using Micro/Nanoindenters 770
- 25.2.2 Bending Tests of Nanobeams Using an AFM 773
- 25.2.3 Bending Tests of Microbeams Using a Nanoindenter 777

25.3 **Finite Element Analysis of Nanostructures with Roughness and Scratches** 778
- 25.3.1 Stress Distribution in a Smooth Nanobeam 779
- 25.3.2 Effect of Roughness in the Longitudinal Direction 781
- 25.3.3 Effect of Roughness in the Transverse Direction and Scratches 781
- 25.3.4 Effect on Stresses and Displacements for Materials That Are Elastic, Elastic-Plastic, or Elastic-Perfectly Plastic 784

25.4 **Closure** 785

References 786

Microelectromechanical systems (MEMS) refer to microscopic devices that have a characteristic length of less than 1 mm but more than 1 μm and combine electrical and mechanical components. Nanoelectromechanical systems (NEMS) refer to nanoscopic devices that have a characteristic length of less than 1 μm and combine electrical and mechanical components. To put the dimensions in perspective, individual atoms are typically fraction of a nanometer in diameter, DNA molecules are about 2.5 nm wide, biological cells are in the range of thousands of nm in diameter, and human hair is about 75 μm in diameter. The mass of a micromachined silicon structure can be as low as 1 nN, and NEMS can be built with a mass as low as 10^{-20} N with cross sections of about 10 nm. In comparison, the mass of a drop of water is about 10 μN and the mass of an eyelash is about 100 nN. The acronym MEMS originated in the United States. The term commonly used in Europe and Japan is micro/nanodevices, which is used in a much broader sense. MEMS/NEMS terms are also now used in a broad sense. A micro/nanosystem, a term commonly used in Europe, is referred to as an intelligent miniaturized system comprising sensing, processing, and/or actuating functions.

A wide variety of MEMS, including Si-based devices, chemical and biological sensors and actuators, and miniature non-silicon structures (e.g., devices made from plastics or ceramics) have been fabricated with dimensions in the range of a couple to a few thousand microns (see e.g., [25.1–9]). A variety of NEMS have also been produced [25.10–15]. Two of the largest "killer" industrial applications of MEMS are accelerometers (about 85 million units in 2002) and digital micromirror devices (about $400 million in revenues in 2002). BIOMEMS and BIONEMS are increasingly used in commercial applications. The largest applications of BIOMEMS include silicon-based disposable blood pressure sensor chips for blood pressure monitoring (about 20 million units in 2002) and a variety of biosensors.

Structural integrity is of paramount importance in all devices. Load applied during the use of devices can result in component failure. Cracks can develop and propagate under tensile stresses, leading to failure [25.16, 17]. Friction/stiction and wear limit the lifetimes and compromise the performance and reliability of the devices involving relative motion [25.4, 18]. Most MEMS/NEMS applications demand extreme reliability. Stress and deformation analyses are carried out for an optimal design. MEMS/NEMS designers require mechanical properties on the nanoscale. Mechanical properties include elastic, inelastic (plastic, fracture, or viscoelastic), and strength. Elastic and inelastic properties are needed to predict deformation from an applied load in the elastic and inelastic regimes, respectively. The strength property is needed to predict the allowable operating limit. Some of the properties of interest are hardness, elastic modulus, bending strength (fracture stress), fracture toughness, and fatigue strength. Micro/nanostructures have some surface topography and local scratches dependent upon the manufacturing process. Surface roughness and local scratches may compromise the reliability of the devices and their effect needs to be studied.

Most mechanical properties are size-dependent [25.11, 16, 17]. Several researchers have measured mechanical properties of silicon and silicon-based milli- to microscale structures using tensile tests and bending tests [25.19–28], resonant structure tests for measurement of elastic properties [25.29], fracture toughness tests [25.20, 22, 30–34], and fatigue tests [25.32, 35, 36]. Most recently, a few researchers have measured mechanical properties of nanoscale structures using atomic force microscopy (AFM) [25.37, 38] and nanoindentation [25.39, 40]. For stress and deformation analyses of simple geometries and boundary condition, analytical models can be used. For analysis of complex geometries, numerical models are needed. The conventional finite element method (FEM) can be used down to a few tens of nanometers, although its applicability is questionable at the nanoscale. FEM has been used for simulation and prediction of residual stresses and strains induced in MEMS devices during fabrication [25.41], to perform fault analysis in order to study MEMS faulty behavior [25.42], to compute mechanical strain resulting from the doping of silicon [25.43], and to analyze micromechanical experimental data [25.22, 44, 45] and nanomechanical experimental data [25.38]. FEM analysis of nanostructures has been carried out to analyze the effect of types of surface roughness and scratches on stresses in nanostructures [25.46, 47].

The most commonly used materials for MEMS/NEMS are single-crystal silicon and silicon-based materials (e.g., SiO_2 and polysilicon films deposited by low pressure chemical vapor deposition (LPCVD) process) [25.11]. An early study showed silicon to be a mechanically resilient material in addition to its favorable electronic properties [25.48]. Single-crystal 3C-SiC (cubic or β-SiC) films deposited by atmospheric pressure chemical vapor deposition (APCVD) process on large-area silicon substrates are produced for high-temperature micro/nanosensor and actuator appli-

cations [25.49–51]. Amorphous alloys can be formed on both metal and silicon substrates by sputtering and plating techniques, providing more flexibility in surface-integration. Electroless deposited Ni-P amorphous thin films have been used to construct microdevices, especially using the so-called LIGA techniques [25.11, 40]. Micro/nanodevices need conductors to provide power, as well as electrical/magnetic signals to make them functional. Electroplated gold films have found wide applications in electronic devices, because of their ability to make thin films and process simplicity [25.40]. Use of SiC, Ni-P, and Au films together with silicon and silicon-based materials open up new design opportunities for MEMS/NEMS devices.

This chapter presents a review of mechanical property measurements on the nanoscale of various materials of interest, and stress and deformation analyses of nanostructures.

25.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures

25.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters

A nanoindenter is commonly used to measure hardness, elastic modulus, and fracture toughness and to perform micro/nanoscratch studies to get a measure of scratch/wear resistance of materials [25.11, 52].

Hardness and Elastic Modulus

The nanoindenter monitors and records the dynamic load and displacement of the three-sided pyramidal diamond (Berkovich) indenter during indentation with a force resolution of about 75 nN and displacement resolution of about 0.1 nm. Hardness and elastic modulus are calculated from the load-displacement data [25.11]. The peak indentation load depends on the mechanical properties of the specimen; a harder material requires higher load for a reasonable indentation depth.

Fracture Toughness

The indentation technique for fracture toughness measurement on the microscale is based on the measurement of the lengths of median-radial cracks produced by indentation. A Vickers indenter (a four-sided diamond pyramid) is used in a microhardness test. A load on the order of 0.5 N is typically used in making the Vickers indentations. The indentation impressions are examined using an optical microscope with Nomarski interference contrast to measure the length of median-radial cracks, c. The fracture toughness (K_{IC}) is calculated by the following relation [25.53]:

$$K_{IC} = \alpha \left(\frac{E}{H}\right)^{1/2} \left(\frac{P}{c^{3/2}}\right) , \qquad (25.1)$$

where α is an empirical constant depending on the geometry of the indenter, H and E are hardness and elastic moduli, and P is the peak indentation load. For Vickers indenters, α has been empirically found based on experimental data and is equal to 0.016 [25.11]. Both E and H values are obtained from the nanoindentation data. The crack length is measured from the center of the indent to the end of the crack using an optical microscope. For one indent, all crack lengths are measured. The crack length c is obtained from the average values of several indents.

Scratch Resistance

In micro/nanoscratch studies, a conical diamond indenter with a tip radius of about 1 μm and an included angle of 60° is drawn over the sample surface, and the load is ramped up until substantial damage occurs [25.11]. The coefficient of friction is monitored during scratching. In order to obtain scratch depths during scratching, the surface profile of the sample surface is first obtained by translating the sample at a low load of about 0.2 mN, which is insufficient to damage a hard sample surface. The 500-μm-long scratches are made by translating the sample while ramping the loads on the conical tip over different loads depending on the material hardness. The actual depth during scratching is obtained by subtracting the initial profile from the scratch depth measured during scratching. In order to measure the scratch depth after the scratch, the scratched surface is profiled at a low load of 0.2 mN and is subtracted from the actual surface profile before scratching.

25.1.2 Bending Tests of Nanostructures Using an AFM

Quasi-static bending tests of fixed nanobeam arrays are carried out using an AFM [25.37, 38, 54]. A three-sided pyramidal diamond tip (with a radius of about 200 nm)

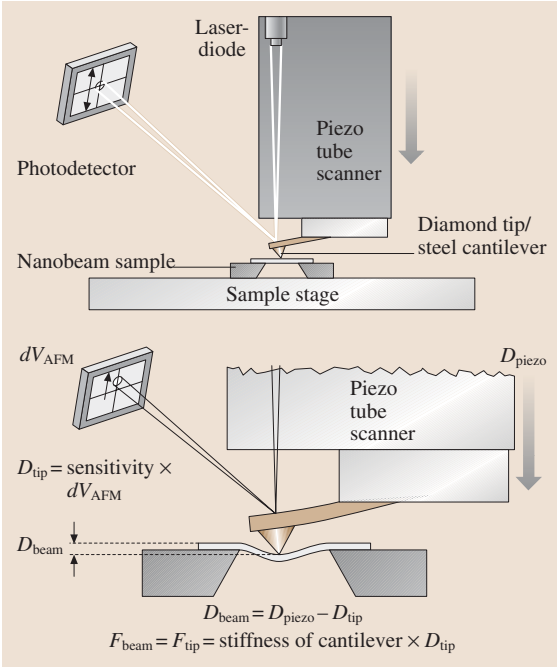


Fig. 25.1 Schematic showing the details of a nanoscale bending test. The AFM tip is brought to the center of the nanobeam and the piezo is extended over a known distance. By measuring the tip displacement, a load-displacement curve of the nanobeam can be obtained [25.38]

mounted on a rectangular stainless steel cantilever is used for the bending tests. The beam stiffness is selected based on the desired load range. The stiffness of the cantilever beams for application of normal load up to 100 μN is about 150–200 N/m.

The wafer with nanobeam array is fixed onto a flat sample chuck using double-stick tape [25.38]. For the bending test, the tip is brought over the nanobeam array with the help of the sample stage of the AFM and a built-in high magnification optical microscope (Fig. 25.1). For the fine positioning of the tip over a chosen beam, the array is scanned in contact mode at a contact load of about 2–4 μN, which results in negligible damage to the sample. After scanning, the tip is located at one end of a chosen beam. To position the tip at the center of the beam span, the tip is moved to the other end of the beam by giving the X-piezo an offset voltage. The value of this offset is determined after several such attempts have been made in order to minimize effects of piezo drift. Half of this offset is then applied to the X-piezo after the tip is positioned at one end of the beam, which usually results in the tip being moved to the center of the span. Once the tip is positioned over the center of the beam span, the tip is held stationary without scanning, and the Z-piezo is extended by a known distance, typically about 2.5 μm, at a rate of 10 nm/s, as shown in Fig. 25.1. During this time, the vertical deflection signal (dV_{AFM}), which is proportional to the deflection of the cantilever (D_{tip}), is monitored. The displacement of the piezo is equal to the sum of the displacements of the cantilever and nanobeam. Hence, the displacement of the nanobeam (D_{beam}) under the point of load can be determined as

$$D_{beam} = D_{piezo} - D_{tip} \,. \tag{25.2}$$

The load (F_{beam}) on the nanobeam is the same as the load on the tip/cantilever (F_{tip}) and is given by

$$F_{beam} = F_{tip} = D_{tip} \times k \,, \tag{25.3}$$

where k is the stiffness of the tip/cantilever. In this manner, a load displacement curve for each nanobeam can be obtained.

The photodetector sensitivity of the cantilever needs to be calibrated to obtain D_{tip} in nm. For this calibration, the tip is pushed against a smooth diamond sample by moving the Z-piezo over a known distance. For the hard diamond material, the actual deflection of the tip can be assumed to be the same as the Z-piezo travel (D_{piezo}), and the photodetector sensitivity (S) for the cantilever setup is determined as

$$S = D_{piezo}/\mathrm{d}V_{AFM}\ \mathrm{nm/V} \,. \tag{25.4}$$

In the measurements, D_{tip} is given as $\mathrm{d}V_{AFM} \times S$.

Since a sharp tip would result in an undesirable large local indentation, *Sundararajan* and *Bhushan* [25.38] used a worn (blunt) diamond tip. Indentation experiments using this tip on a silicon substrate yielded a residual depth of less than 8 nm at a maximum load of 120 μN, which is negligible compared to displacements of the beams (several hundred nm). Hence, we can assume that negligible local indentation or damage is created during the bending process of the beams, and that the displacement calculated from (25.2) is entirely that of the beam structure.

Elastic Modulus and Bending Strength

Elastic modulus and bending strength (fracture stress) of the beams can be estimated by equations based on the assumption that the beams follow linear elastic theory of an isotropic material. This is probably valid since the beams have high length-to-width (ℓ/w) and length-to-thickness (ℓ/t) ratios and also since the length direction

is along the principal stress direction during the test. For a fixed elastic beam loaded at the center of the span, the elastic modulus is expressed as

$$E = \frac{\ell^3}{192I} m \,, \tag{25.5}$$

where ℓ is the beam length, I is the area moment of inertia for the beam cross section, and m is the slope of the load-displacement curve during bending [25.55]. The area moment of inertia for a beam with a trapezoidal cross section is calculated from the following equation:

$$I = \frac{w_1^2 + 4w_1 w_2 + w_2^2}{36(w_1 + w_2)} t^3 \,, \tag{25.6}$$

where w_1 and w_2 are the upper and lower widths, respectively, and t is the thickness of the beam. According to linear elastic theory, for a centrally loaded beam, the moment diagram is shown in Fig. 25.2. The maximum moments are generated at the ends (negative moment) and under the loading point (positive moment), as shown in Fig. 25.2. The bending stresses generated in the beam are proportional to the moments and are compressive or tensile about the neutral axis (line of zero stress). The maximum tensile stress (σ_b, which is the bending strength or fracture stress) is produced on the top surface at both the ends and is given by [25.55]

$$\sigma_b = \frac{F_{max} \ell e_1}{8I} \,, \tag{25.7}$$

where F_{max} is the applied load at failure and e_1 is the distance of the top surface from the neutral plane of the beam cross section and is given by [25.55]

$$e_1 = \frac{t(w_1 + 2w_2)}{3(w_1 + w_2)} \,. \tag{25.8}$$

Although the moment value at the center of the beam is the same as at the ends, the tensile stresses at the center (generated on the bottom surface) are less than those generated at the ends, as per (25.7), because the distance from the neutral axis to the bottom surface is less than e_1. This is because of the trapezoidal cross section of the beam, which results in the neutral axis being closer to the bottom surface than the top (Fig. 25.2).

In the preceding analysis, the beams were assumed to have fixed ends. However, in the nanobeams used by *Sundararajan* and *Bhushan* [25.38], the underside of the beams was pinned over some distance on either side of the span. Hence, a finite element model of the beams was created to see if the difference in the boundary conditions affected the stresses and displacements of the beams. It was found that the difference in the stresses was less than

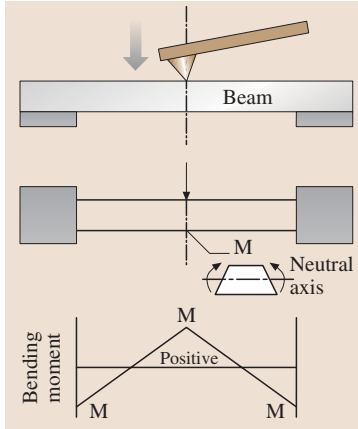


Fig. 25.2 A schematic of the bending moments generated in the beam during a quasi-static bending experiment with the load at the center of the span. The maximum moments occur under the load and at the fixed ends. Due to the trapezoidal cross section, the maximum tensile bending stresses occur at the top surfaces at the fixed ends

1%. This indicates that the boundary conditions near the ends of the actual beams are not that different from that of the fixed ends. Therefore, the bending strength values can be calculated from (25.7).

Fracture Toughness

Fracture toughness is another important parameter for brittle materials such as silicon. In the case of the nanobeam arrays, these are not best suited for fracture toughness measurements, because they do not possess regions of uniform stress during bending. *Sundararajan* and *Bhushan* [25.38] developed a methodology, and its steps are outlined schematically in Fig. 25.3a. First, a crack of known geometry is introduced in the region of maximum tensile bending stress, i. e., on the top surface near the ends of the beam. This is achieved by generating a scratch at high normal load across the width (w_1) of the beam using a sharp diamond tip (radius < 100 nm). A typical scratch thus generated is shown in Fig. 25.3b. By bending the beam as shown, a stress concentration will be formed under the scratch. This will lead to failure of the beam under the scratch once a critical load (fracture load) is attained. The fracture load and relevant dimensions of the scratch are input into the FEM model, which is used to generate the fracture stress plots. Figure 25.3c shows an FEM simulation of one such experiment, which reveals that the maximum stress does occur under the scratch.

If we assume that the scratch tip acts as a crack tip, a bending stress will tend to open the crack in Mode I. In this case, the stress field around the crack tip can be described by the stress intensity parameter K_I (for Mode I) for linear elastic materials [25.56]. In particular, the stresses corresponding to the bending stresses are

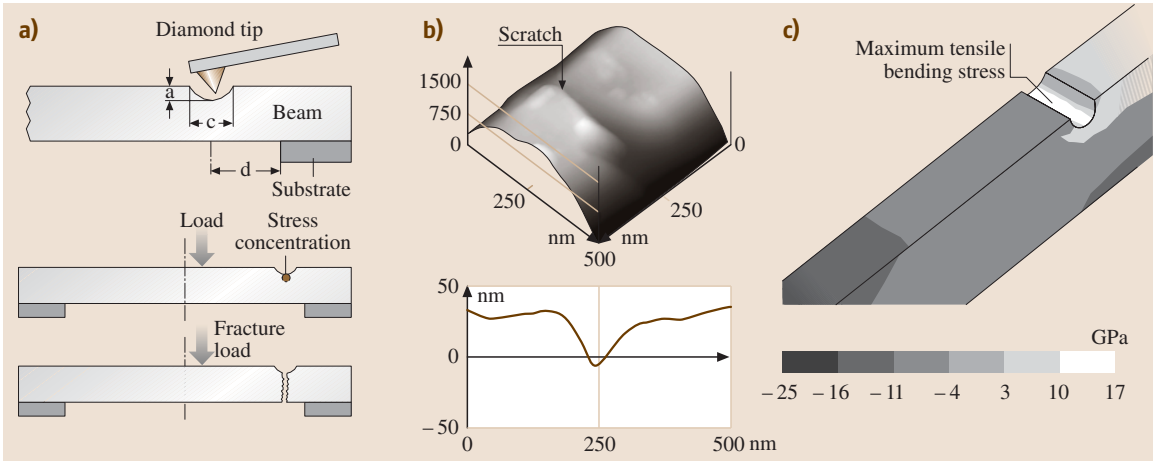


Fig. 25.3 **(a)** Schematic of technique to generate a defect (crack) of known dimensions in order to estimate fracture toughness. A diamond tip is used to generate a scratch across the width of the beam. When the beam is loaded as shown, a stress concentration is formed at the bottom of the scratch. The fracture load is then used to evaluate the stresses using FEM. **(b)** AFM 3-D image and 2-D profile of a typical scratch. **(c)** Finite element model results verifying that the maximum bending stress occurs at the bottom of the scratch [25.38]

described by

$$\sigma = \frac{K_\mathrm{I}}{\sqrt{2\pi r}} \cos\left(\frac{\theta}{2}\right)\left[1+\sin\left(\frac{\theta}{2}\right)\sin\left(\frac{3\theta}{2}\right)\right] \tag{25.9}$$

for every point $p(r,\theta)$ around the crack tip, as shown in Fig. 25.4. If we substitute the fracture stress (σ_f) to the left hand side of (25.9), then the K_I value can be substituted with its critical value, which is the fracture toughness K_IC. Now, the fracture stress can be determined for the point ($r=0$, $\theta=0$), i. e., right under the crack tip as explained above. However, we cannot substitute $r=0$ in (25.9). The alternative is to substitute a value for r, which is as close to zero as possible. For silicon, a reasonable number is the distance between neighboring atoms in the (111) plane, the plane along which silicon exhibits the lowest fracture energy. This value was calculated from silicon unit cell dimensions of 0.5431 nm [25.57] to be 0.4 nm (half of the face diagonal). This assumes that Si displays no plastic zone around the crack tip, which is reasonable since in tension, silicon is not known to display much plastic deformation at room temperature. *Sundararajan* and *Bhushan* [25.38] used values $r=0.4$–1.6 nm (i. e., distances up to four times the distance between the nearest neighboring atoms) to estimate the fracture toughness for both Si and SiO_2 according to the following equation:

$$K_\mathrm{IC} = \sigma_\mathrm{f}\sqrt{2\pi r} \qquad r = 0.4\text{–}1.6\,\mathrm{nm}\,. \tag{25.10}$$

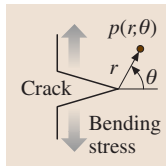


Fig. 25.4 Schematic of crack tip and coordinate systems used in (25.9) to describe a stress field around the crack tip in terms of the stress intensity parameter, K_I [25.38]

Fatigue Strength

In addition to the properties mentioned so far that can be evaluated from quasi-static bending tests, the fatigue properties of nanostructures are also of interest. This is especially true for MEMS/NEMS involving vibrating structures such as oscillators and comb drives [25.58] and hinges in digital micromirror devices [25.59]. To study the fatigue properties of the nanobeams, *Sundararajan* and *Bhushan* [25.38] applied monotonic cyclic stresses using an AFM, Fig. 25.5a. Similar to the bending test, the diamond tip is first positioned at the center of the beam span. In order to ensure that the tip is always in contact with the beam (as opposed to impacting it), the piezo is first extended by a distance D_1, which ensures a minimum stress on the beam. After this extension, a cyclic displacement of amplitude, D_2, is applied continuously until failure of the beam occurs.

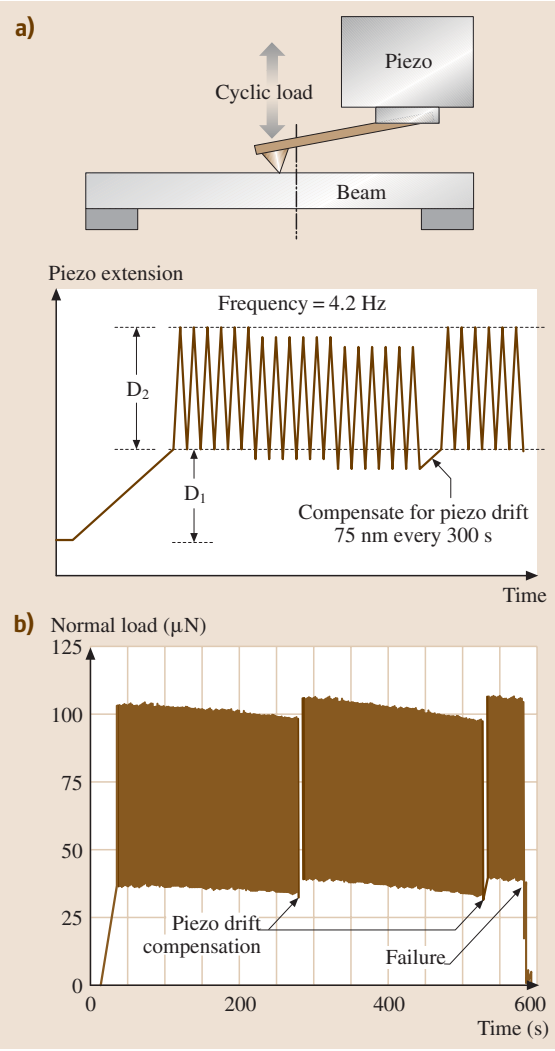


Fig. 25.5 **(a)** Schematic showing the details of the technique to study fatigue behavior of the nanobeams. The diamond tip is located at the middle of the span, and a cyclic load at 4.2 Hz is applied to the beam by forcing the piezo to move in the pattern shown. An extension is made every 300 s to compensate for the piezo drift to ensure that the load on the beam is kept fairly constant. **(b)** Data from a fatigue experiment on a nanobeam until failure. The normal load is computed from the raw vertical deflection signal. The compensations for piezo drift keep the load fairly constant [25.38]

This results in the application of a cyclic load to the beam. The maximum frequency of the cyclic load that could be attained using the AFM by *Sundararajan* and *Bhushan* [25.38] was 4.2 Hz. The vertical deflection signal of the tip is monitored throughout the experiment. The signal follows the pattern of the piezo input up to failure, which is indicated by a sudden drop in the signal. During initial runs, piezo drift was observed that caused the piezo to gradually move away from the beam (i. e., to retract), resulting in a continuous decrease in the applied normal load. In order to compensate for this, the piezo is given a finite extension of 75 nm every 300 s, as shown in Fig. 25.5a. This results in keeping the applied loads fairly constant. The normal load variation (calculated from the vertical deflection signal) from a fatigue test is shown in Fig. 25.5b. The cyclic stress amplitudes (corresponding to D_2) and fatigue lives are recorded for every sample tested. Values for D_1 are set such that minimum stress levels are about 20% of the bending strengths.

25.1.3 Bending Tests Using a Nanoindenter

Quasi-static bending tests of micro/nanostructures are also carried out using a nanoindenter [25.39, 40]. The advantage of the nanoindenter is that loads up to about 500 mN, higher than that in AFM (up to about 100 μN), can be used for structures requiring high loads for experiments. To avoid the indenter tip pushing into the specimen, a blunt tip is used in the bending and fatigue tests. *Li* et al. [25.40] used a diamond conical indenter with a radius of 1 μm and an included angle of 60°. The load position used was at the center of the span for the bridge beams and 10 μm off from the free end of the cantilever beams. An optical microscope with a magnification of 1,500× or an in situ AFM is used to locate the loading position. Then the specimen is moved under the indenter location with a resolution of about 200 nm in the longitudinal direction and less than 100 nm in the lateral direction.

Using the analysis presented earlier, elastic modulus and bending strength of the beams can be obtained from the load-displacement curves [25.40]. For fatigue tests, an oscillating load is applied and contact stiffness is measured during the tests. A significant drop in the contact stiffness during the test is a measure of the number of cycles to failure [25.39].

25.2 Experimental Results and Discussion

25.2.1 Indentation and Scratch Tests of Various Materials Using Micro/Nanoindenters

Studies have been conducted on five different materials: undoped single-crystal Si(100), undoped polysilicon film, SiO_2 film, SiC film, electroless deposited Ni-11.5 wt% P amorphous film, and electroplated Au film [25.40, 50, 51]. A 3-μm thick polysilicon film was deposited by a low pressure chemical vapor deposition (LPCVD) process on a Si(100) substrate. The 1-μm thick SiO_2 film was deposited by a plasma enhanced chemical vapor deposition (PECVD) process on a Si(111) substrate. A 3-μm thick 3C-SiC film was epitaxially grown using an atmospheric pressure chemical vapor deposition (APCVD) process on a Si(100) substrate. A 12-μm thick Ni-P film was electroless plated on a 0.8-mm-thick Al-4.5 wt% Mg alloy substrate. A 3-μm-thick Au film was electroplated on a Si(100) substrate.

Hardness and Elastic Modulus

Hardness and elastic modulus measurements are made using a nanoindenter [25.40]. The hardness and elastic modulus values of various materials at a peak indentation depth of 50 nm are summarized in Fig. 25.6 and Table 25.1. The SiC film exhibits the highest hardness of about 25 GPa and an elastic modulus of about 395 GPa among the samples examined, followed by the undoped Si(100), undoped polysilicon film, SiO_2 film, Ni-P film, and Au film. The hardness and elastic modulus data of the undoped Si(100) and undoped polysilicon film are comparable. For the metal alloy films, the Ni-P film exhibits higher hardness and elastic modulus than the Au film.

Table 25.1 Hardness, elastic modulus, fracture toughness, and critical load results of the bulk single-crystal Si(100) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au

Samples	Hardness (GPa)	Elastic modulus (GPa)	Fracture toughness ($MPa\,m^{1/2}$)	Critical load (mN)
Undoped Si(100)	12	165	0.75	11
Undoped polysilicon film	12	167	1.11	11
SiO_2 film	9.5	144	0.58 (Bulk)	9.5
SiC film	24.5	395	0.78	14
Ni-P film	6.5	130		0.4 (Plowing)
Au film	4	72		0.4 (Plowing)

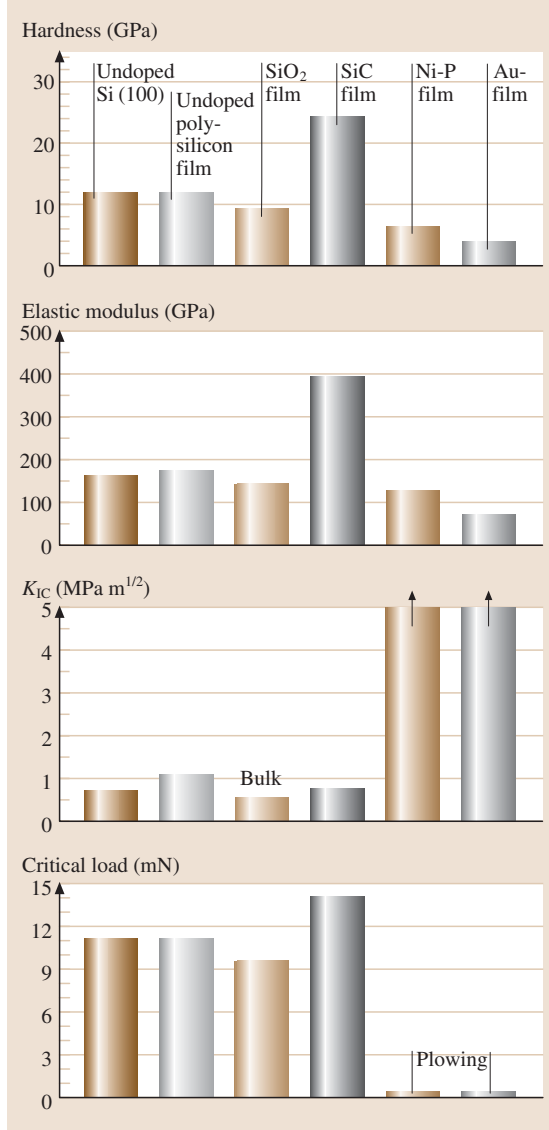


Fig. 25.6 Bar chart summarizing the hardness, elastic modulus, fracture toughness, and critical load (from scratch tests) results of the bulk undoped single-crystal Si(100) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au [25.40]

Fracture Toughness

The optical images of Vickers indentations made using a microindenter at a normal load of 0.5 N held for 15 s on the undoped Si(100), undoped polysilicon film, and SiC film are shown in Fig. 25.7 [25.51]. The SiC film exhibits the smallest indentation mark, followed by the undoped polysilicon film and undoped Si(100). These Vickers indentation depths are smaller than one-third of the film thickness. Thus, the influence of the substrate on the fracture toughness of the films can be ignored. In addition to the indentation marks, radial cracks are observed emanating from the indentation corners. The SiC film shows the longest radial crack length, followed by the undoped Si(100) and undoped polysilicon film. The radial cracks for the undoped Si(100) are straight, whereas those for the SiC and undoped polysilicon film are not straight but go in a zigzag manner. The fracture toughness (K_{IC}) is calculated using (25.1).

The fracture toughness values of all samples are summarized in Fig. 25.6 and Table 25.1. The SiO_2 film used in this study is about 1 μm thick, which is not thick enough for a fracture toughness measurement. The fracture toughness values of bulk silica are listed instead for a reference. The Ni-P and Au films exhibit very high fracture toughness values that cannot be measured by indentation methods. For other samples, the undoped polysilicon film has the highest value, followed by the undoped Si(100), SiC film, and SiO_2 film. For the undoped polysilicon film, the grain boundaries can stop the radial cracks and change the propagation directions of the radial cracks, making the propagation of these cracks more difficult. Values of fracture toughness for the undoped Si(100) and SiC film are comparable. Since the undoped Si(100) and SiC film are single crystal, no grain boundaries are present to stop the radial cracks and change the propagation directions of the radial cracks. This is why the SiC film shows a lower fracture toughness value than the bulk polycrystal SiC materials of 3.6 MPa $m^{1/2}$ [25.60].

Scratch Resistance

The scratch resistance of various materials has been studied using a nanoindenter by *Li* et al. [25.40]. Figure 25.8 compares the coefficient of friction and scratch depth profiles as a function of increasing normal load and optical images of three regions over scratches: at the beginning of the scratch (indicated by A on the friction profile), at the point of initiation of damage at which the coefficient of friction increases to a high value or increases abruptly (indicated by B on the friction profile), and towards the end of the scratch (indicated by C on the friction profile) for all samples. Note that the ramp loads for the Ni-P and Au range are from 0.2 to 5 mN, whereas the ramp loads for other samples are from 0.2 to 20 mN. All samples exhibit a continuous increase in the coefficient of friction with increasing normal load from the beginning of the scratch. The continuous increase in the coefficient of friction during scratching is attributed to the increasing plowing of the sample by the tip with increasing normal load, as shown in the SEM images in Fig. 25.8. The abrupt increase in the coefficient of friction is associated with catastrophic failure, as well as significant plowing of the tip into the sample. Before the critical load, the coefficient of friction of the undoped polysilicon, SiC, and SiO_2 films increased at a slower rate and was smoother than that of the other samples. The undoped Si(100) exhibits some bursts in the friction profiles before the critical load. At the critical load, the SiC and undoped polysilicon films exhibit a small increase in the coefficient of friction, whereas the undoped Si(100) and undoped polysilicon film exhibit a sudden increase in the coefficient of friction. The Ni-P and Au films show a continuous increase in the coefficient of friction, indicating the behavior of a ductile metal. The bursts in the friction profile might result from the plastic deformation and material pile-up in front of the scratch tip. The Au film exhibits a higher coefficient of friction than the Ni-P film, because it has lower hardness and elastic modulus values than the Ni-P film.

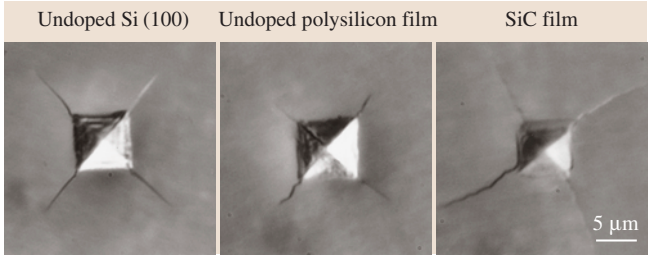


Fig. 25.7 Optical images of Vickers indentations made at a normal load of 0.5 N held for 15 s on the undoped Si(100), undoped polysilicon film, and SiC film [25.51]

The SEM images show that below the critical loads the undoped Si(100) and undoped polysilicon film were damaged by plowing, associated with the plastic flow of the material and formation of debris on the sides of the scratch. For the SiC and SiO_2 films, in region A, a plowing scratch track was found without any debris on the side of the scratch, which is probably responsible for the smoother curve and slower increase in the

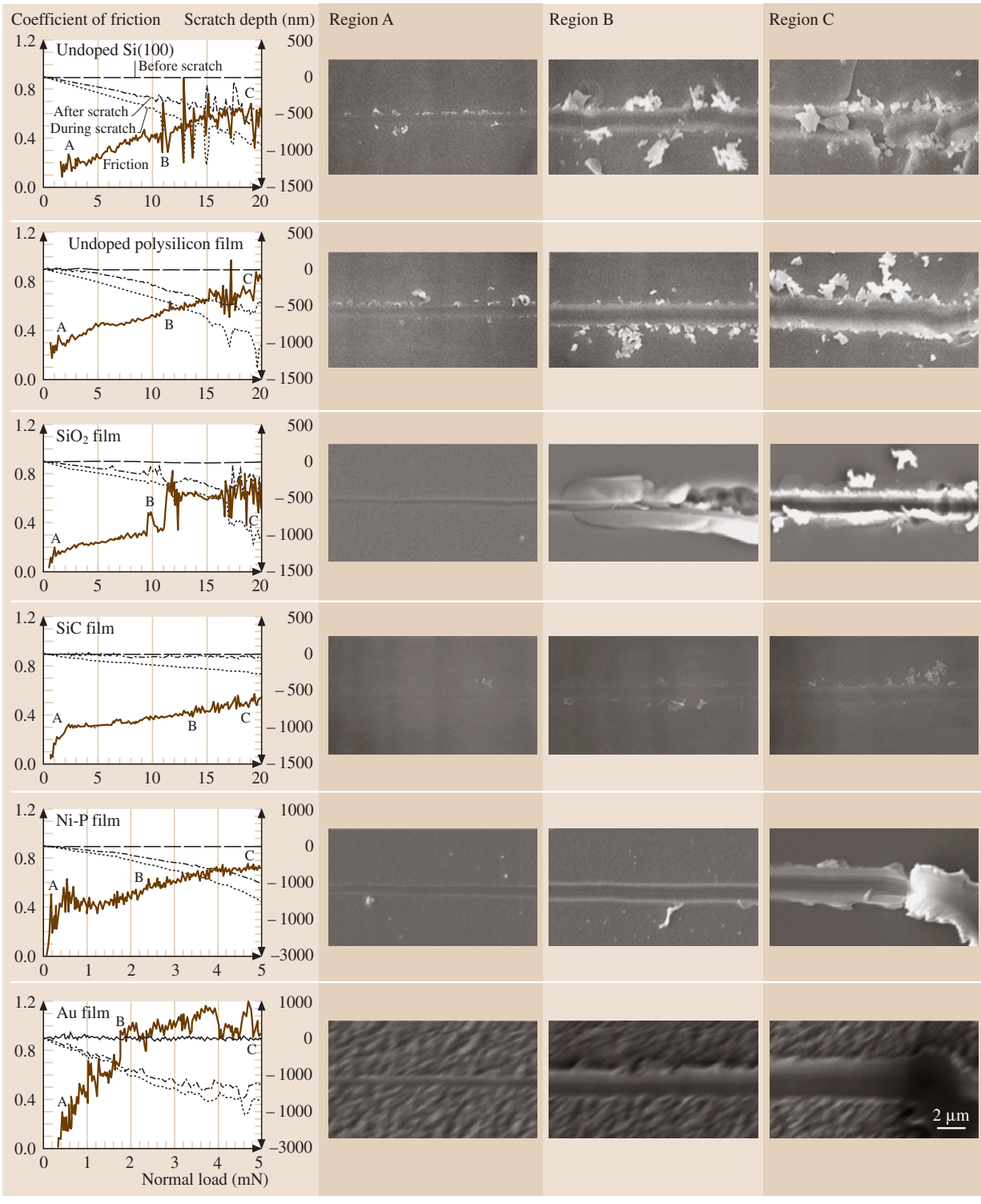

Coefficient of friction
Scratch depth (nm)
Region A
Region B
Region C
Undoped Si(100)
Before scratch
After scratch
During scratch
Friction
Undoped polysilicon film
SiO2 film
SiC film
Ni-P film
Au film
Normal load (mN)
2 µm

Fig. 25.8 Coefficient of friction and scratch depth profiles as a function of increasing normal load and optical images of three regions over scratches: at the beginning of the scratch (indicated by A on the friction profile), at the point of initiation of damage at which the coefficient of friction increases to a high value or increases abruptly (indicated by B on the friction profile), and towards the end of the scratch (indicated by C on the friction profile) for all samples [25.40]

coefficient of friction before the critical load. After the critical load, for the SiO_2 film, delamination of the film from the substrate occurred, followed by cracking along the scratch track. For the SiC film, only several small debris particles were found without any cracks on the side of the scratch, which is responsible for the small increase in the coefficient of friction at the critical load. For the undoped Si(100), cracks were found on the side of the scratch right from the critical load and up, which is probably responsible for the big bursts in the friction profile. For the undoped polysilicon film, no cracks were found on the side of the scratch at the critical load. This might result from grain boundaries, which can stop the propagation of cracks. At the end of the scratch, some of the surface material was torn away and cracks were found on the side of the scratch in the undoped Si(100). A couple of small cracks were found in the undoped polysilicon and SiO_2 films. No crack was found in the SiC film. Even at the end of the scratch, less debris was found in the SiC film. A curly chip was found at the end of the scratch in both Ni-P and Au films. This is a typical characteristic of ductile metal alloys. The Ni-P and Au films were damaged by plowing right from the beginning of the scratch with material pile-up at the side of the scratch.

The scratch depth profiles obtained during and after the scratch on all samples with respect to initial profile, after the cylindrical curvature is removed, are plotted in Fig. 25.8. Reduction in scratch depth is observed after scratching as opposed to during scratching. This reduction in scratch depth is attributed to an elastic recovery after removal of the normal load. The scratch depth after scratching indicates the final depth, which reflects the extent of permanent damage and plowing of the tip into the sample surface and is probably more relevant for visualizing the damage that can occur in real applications. For the undoped Si(100), undoped polysilicon film, and SiO_2 film, there is a large scatter in the scratch depth data after the critical loads, which is associated with the generation of cracks, material removal, and debris. The scratch depth profile is smooth for the SiC film. It is noted that the SiC film exhibits the lowest scratch depth among the samples examined. The scratch depths of the undoped Si(100), undoped polysilicon film, and SiO_2 film are comparable. The Ni-P and Au films exhibit much larger scratch depths than other samples. The scratch depth of the Ni-P film is smaller than that of the Au film.

The critical loads estimated from friction profiles for all samples are compared in Fig. 25.6 and Table 25.1. The SiC film exhibits the highest critical load of about 14 mN, as compared to other samples. The undoped Si(100) and undoped polysilicon film show comparable critical load of about 11 mN, whereas the SiO_2 film shows a low critical load of about 9.5 mN. The Ni-P and Au films were damaged by plowing right from the beginning of the scratch.

25.2.2 Bending Tests of Nanobeams Using an AFM

Bending tests have been performed on Si and SiO_2 nanobeam arrays [25.38, 54]. The single-crystal silicon bridge nanobeams were fabricated by bulk micromachining incorporating enhanced-field anodization using an AFM [25.37]. The Si nanobeams are oriented along the [110] direction in the (001) plane. Subsequent thermal oxidation of the beams results in the formation of SiO_2 beams. The cross section of the nanobeams is trapezoidal owing to the anisotropic wet etching process. SEM micrographs of Si and SiO_2 nanobeam arrays and a schematic of the shape of a typical nanobeam are shown in Fig. 25.9. The actual widths and thicknesses of nanobeams were measured using an AFM in tapping mode prior to tests using a standard Si tapping mode tip (tip radius < 10 nm). Surface roughness measurements of the nanobeam surfaces in tapping mode yielded a σ of 0.7 ± 0.2 nm and peak-to-valley (P–V) distance of 4 ± 1.2 nm for Si and a σ of 0.8 ± 0.3 nm and a P–V distance of 3.1 ± 0.8 nm for SiO_2. Prior to testing, the Si nanobeams were cleaned by immersing them in a "piranha etch" solution (3 : 1 solution by volume of 98% sulphuric acid and 30% hydrogen peroxide) for 600 s to remove any organic contaminants.

Bending Strength

Figure 25.10 shows typical load-displacement curves for Si and SiO_2 beams that were bent to failure [25.38, 54]. The upper width (w_1) of the beams is indicated in the figure. Also indicated are the elastic modulus values obtained from the slope of the load-displacement curve

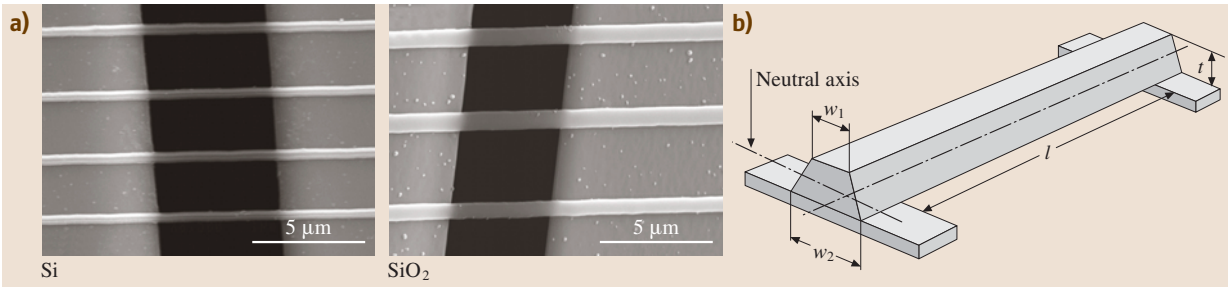


Fig. 25.9 **(a)** SEM micrographs of nanobeam arrays, and **(b)** a schematic of the shape of a typical nanobeam. The trapezoidal cross section is due to the anisotropic wet etching during the fabrication [25.54]

(25.5). All the beams tested showed linear elastic behavior followed by abrupt failure, which is suggestive of brittle fracture. Figure 25.11 shows the scatter in the values of elastic modulus obtained for both Si and SiO_2

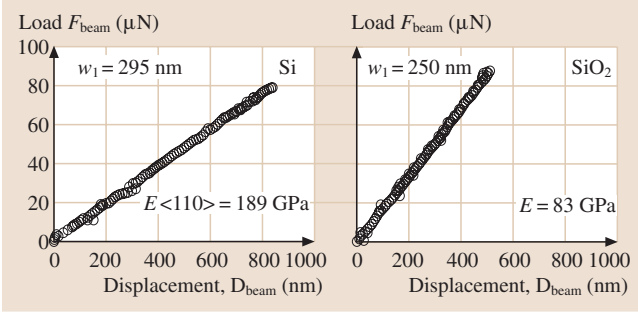


Fig. 25.10 Typical load-displacement curves of silicon and SiO_2 nanobeams. The curves are linear until sudden failure, indicative of brittle fracture of the beams. The elastic modulus (E) values calculated from the curves are shown. The dimensions of the Si beam were $w_1 = 295$ nm, $w_2 = 484$ nm, and $t = 255$ nm, while those of the SiO_2 beam were $w_1 = 250$ nm, $w_2 = 560$ nm, and $t = 425$ nm [25.54]

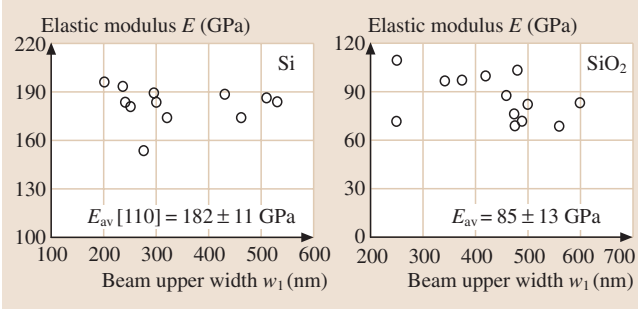


Fig. 25.11 Elastic modulus values measured for Si and SiO_2. The average values are shown. These are comparable to bulk values, which demonstrates that elastic modulus shows no specimen size dependence [25.38]

along with the average values (± standard deviation). The scatter in the values may be due to differences in the orientation of the beams with respect to the trench and the loading point being a little off center with respect to the beam span. The average values are a little higher than the bulk values (169 GPa for Si[110] and 73 GPa for SiO_2 in Table 25.2). However, the values of E obtained from (25.5) have an error of about 20% due to the uncertainties in beam dimensions and spring constant of the tip/cantilever (which affects the measured load). Hence the elastic modulus values on the nanoscale can be considered comparable to bulk values.

Most of the beams when loaded quasi-statically at the center of the span broke at the ends, as shown in Fig. 25.12a, which is consistent with the fact that maximum tensile stresses occur on the top surfaces near the ends. (See FEM stress distribution results in Fig. 25.12b.) Figure 25.13 shows the values of bending strength obtained for different beams. There appears to be no trend in bending strength with the upper width (w_1) of the beams. The large scatter is expected for the strength of brittle materials, since they are dependent on preexisting flaw population in the material and, hence, are statistical in nature. The Weibull distribution, a statistical analysis, can be used to describe the scatter in the bending strength values. The means of the Weibull distributions were found to be 17.9 GPa and 7.6 GPa for Si and SiO_2, respectively. Previously reported numbers of strengths range from 1 to 6 GPa for silicon [25.19, 20, 22–25, 27, 45, 61, 62] and about 1 GPa for SiO_2 [25.34] microscale specimens. This clearly indicates that bending strength shows a specimen size dependence. Strength of brittle materials is dependent on preexisting flaws in the material. Since for nanoscale specimens the volume is smaller than for micro- and macroscale specimens, the flaw population will be smaller as well, resulting in higher values of strength.

Table 25.2 Summary of measured parameters from quasi-static bending tests

Sample	Elastic modulus E (GPa)		Bending strength σ_b (GPa)		Fracture toughness K_{IC} (MPa $\sqrt{m}$)		
	Measured	Bulk value	Measured	Reported (microscale)	Estimated	Reported (microscale)	Bulk value
Si	182±11	169[1]	18±3	< 10[3]	1.67±0.4	0.6–1.65[5]	0.9[6]
SiO_2	85±13	73[2]	7.6±2	< 2[4]	0.60±0.2	0.5–0.9[4]	–

[1] Si(110), [25.63] [2] [25.64] [3] [25.19, 20, 22–25, 45, 61, 62] [4] [25.34] [5] [25.30–33] [6] [25.57]

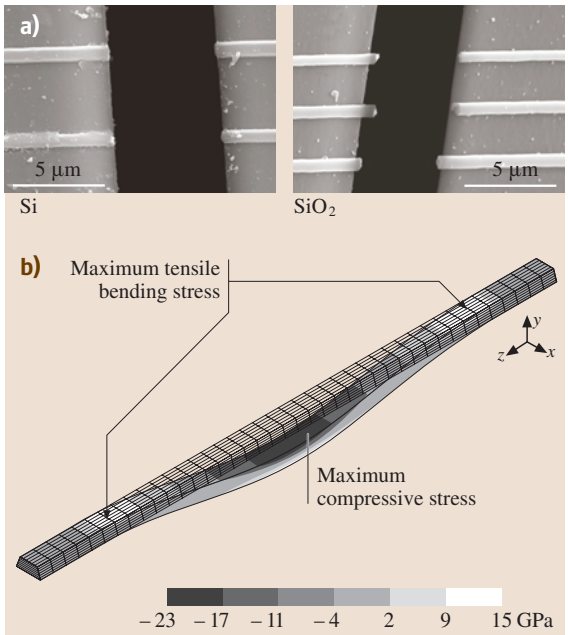


Fig. 25.12 **(a)** SEM micrographs of nanobeams that failed during quasi-static bending experiments. The beams failed at or near the ends, which is the location of maximum tensile bending stress [25.54], and **(b)** bending stress distribution for silicon nanobeam, indicating that the maximum tensile stresses occur on the top surfaces near the fixed ends

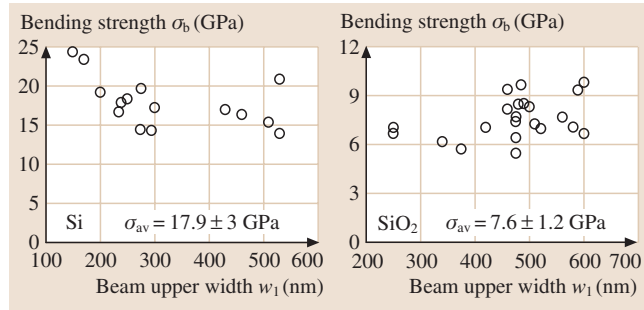


Fig. 25.13 Bending strength values obtained from bending experiments. Average values are indicated. These values are much higher than values reported for microscale specimens, indicating that bending strength shows a specimen size effect [25.38]

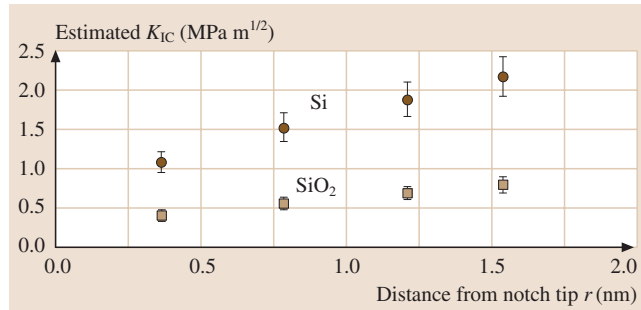


Fig. 25.14 Fracture toughness (K_{IC}) values for increasing values of r corresponding to distance between neighboring atoms in {111} planes of silicon (0.4 nm). Hence r values between 0.4 and 1.6 nm are chosen. The K_{IC} values thus estimated are comparable to values reported by others for both Si and SiO_2 [25.38]

Fracture Toughness

Estimates of fracture toughness calculated using (25.10) for Si and SiO_2 are shown in Fig. 25.14 [25.38]. The results show that the K_{IC} estimate for Si is about 1–2 MPa $\sqrt{m}$, whereas for SiO_2 the estimate is about 0.5–0.9 MPa $\sqrt{m}$. These values are comparable to values reported by others on larger specimens for Si [25.30–33] and SiO_2 [25.34]. The high values obtained for Si could be due to the fact that the scratches, despite being quite sharp, still have a finite radius of about 100 nm. The bulk value for silicon is about 0.9 MPa $\sqrt{m}$ (Table 25.2). Fracture toughness is considered a material property and is believed to be in-

dependent of specimen size. The values obtained in this study, given its limitations, appear to show that fracture toughness is comparable, if not a little higher on the nanoscale.

Fatigue Strength

Fatigue strength measurements of Si nanobeams have been carried out by *Sundararajan* and *Bhushan* [25.38] using an AFM and *Li* and *Bhushan* [25.39] using a nanoindenter. Various stress levels were applied to nanobeams by *Sundararajan* and *Bhushan* [25.38]. The minimum stress was 3.5 GPa for Si beams and 2.2 GPa for SiO_2 beams. The frequency of applied load was 4.2 Hz. In general, the fatigue life decreased with increasing mean stress as well as increasing stress amplitude. When the stress amplitude was less than 15% of the bending strength, the fatigue life was greater than 30,000 cycles for both Si and SiO_2. However, the mean stress had to be less than 30% of the bending strength for a life of greater than 30,000 for Si, whereas even at a mean stress of 43% of the bending strength, SiO_2 beams showed a life greater than 30,000. During fatigue, the beams broke under the loading point or at the ends when loaded at the center of the span. This was different from the quasi-static bending tests, where the beams broke at the ends almost every time. This could be due to the fact that the stress levels under the load and at the ends are not that different and fatigue crack propagation could occur at either location. Figure 25.15 shows a nanoscale *S-N* curve with bending stress (S) as a function of fatigue in cycles (N) with an apparent endurance life at lower stress. This study clearly demonstrates that fatigue properties of nanoscale specimens can be studied.

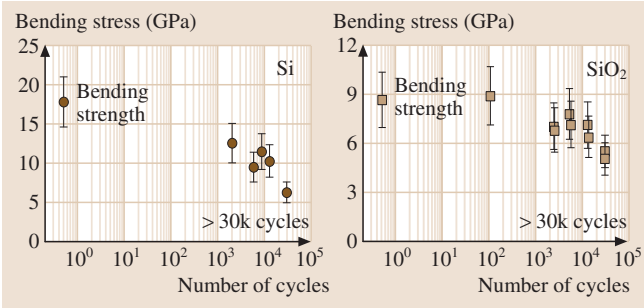


Fig. 25.15 Fatigue test data showing applied bending stress as a function of number of cycles. A single load-unload sequence is considered 1 cycle. The bending strength data points are, therefore, associated with $\frac{1}{2}$ cycle, since failure occurs upon loading [25.38]

SEM Observations of Fracture Surfaces

Figure 25.16 shows SEM images of the fracture surfaces of nanobeams broken during quasi-static bending, as well as fatigue [25.38]. In the quasi-static cases, the maximum tensile stresses occur on the top surface, so it is reasonable to assume that fracture initiated at or

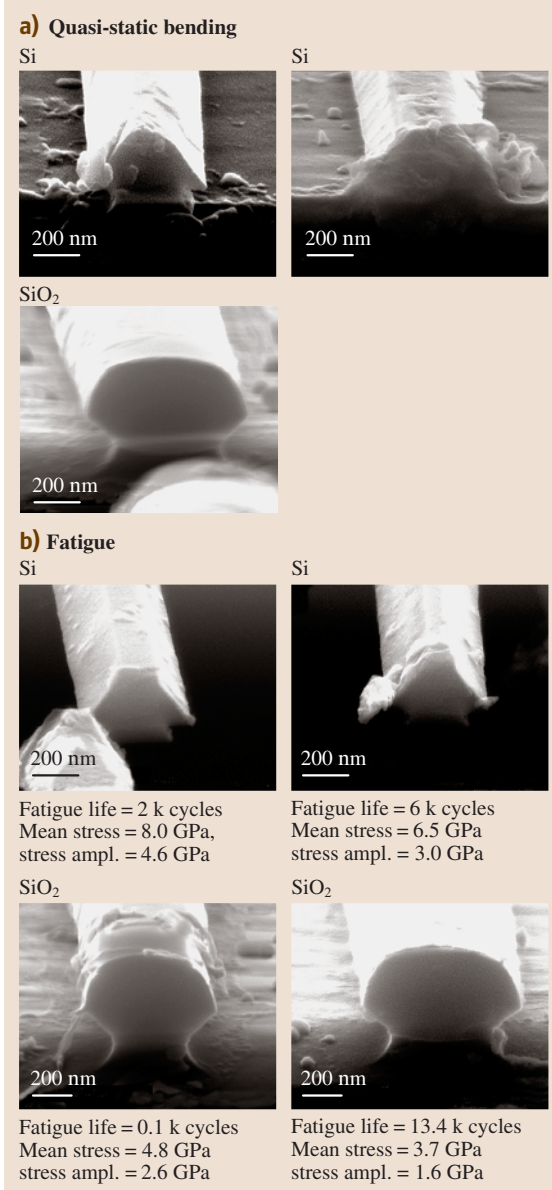


Fig. 25.16a,b SEM micrographs of fracture surfaces of silicon and SiO_2 beams subjected to **(a)** quasi-static bending and **(b)** fatigue [25.38]

near the top surface and propagated downward. The fracture surfaces of the beams suggest a cleavage type of fracture. Silicon beam surfaces show various ledges or facets, which is typical for crystalline brittle materials. Silicon usually fractures along the (111) plane, due to this plane having the lowest surface energy to overcome by a propagating crack. However, failure has also been known to occur along the (110) planes in microscale specimens, despite the higher energy required compared to the (111) planes [25.45]. The plane normal to the beam direction in these samples is the (110) plane, while (111) planes will be oriented at 35° from the (110) plane. The presence of facets and irregularities on the silicon surface in Fig. 25.16a suggest that it is a combination of these two types of fractures that has occurred. Since the stress levels are very high for these specimens, it is reasonable to assume that crack propagating forces will be high enough to result in (110) type failures.

In contrast, the silicon fracture surfaces under fatigue, shown in Fig. 25.16b, appear very smooth without facets or irregularities. This is suggestive of low energy fracture, i. e., of (111) type fracture. We do not see evidence of fatigue crack propagation in the form of steps or striations on the fracture surface. We believe that for the stress levels applied in these fatigue experiments, failure in silicon occurred via cleavage associated with "static fatigue" type of failures.

SiO_2 shows very smooth fracture surfaces for both quasi-static bending and fatigue. This is in contrast to the hackled surface one might expect for the brittle failure of an amorphous material on the macroscale. However, in larger scale fracture surfaces for such materials, the region near the crack initiation usually appears smooth or mirror-like. Since the fracture surface here is so small and very near the crack initiation site, it is not unreasonable to see such a smooth surface for SiO_2 on this scale. There appears to be no difference between the fracture surfaces obtained by quasi-static bending and fatigue for SiO_2.

Summary of Mechanical Properties Measured Using Quasi-Static Bending Tests

Table 25.2 summarizes the various properties measured via quasi-static bending in this study [25.38]. Also shown are bulk values of the parameters, along with values reported on larger scale specimens by other researchers. Elastic modulus and fracture toughness values appear to be comparable to bulk values and show no dependence on specimen size. However, bending strength shows a clear specimen size dependence with nanoscale numbers being twice as large as numbers reported for larger-scale specimens.

25.2.3 Bending Tests of Microbeams Using a Nanoindenter

Bending tests have been performed on Ni-P and Au microbeams [25.40]. The Ni-P cantilever microbeams were fabricated by focused ion beam machining technique. The dimensions were $10 \times 12 \times 50\,\mu m^3$. Notches with a depth of 3 μm and a tip radius of 0.25 μm were introduced in the microbeams to facilitate failure at a lower

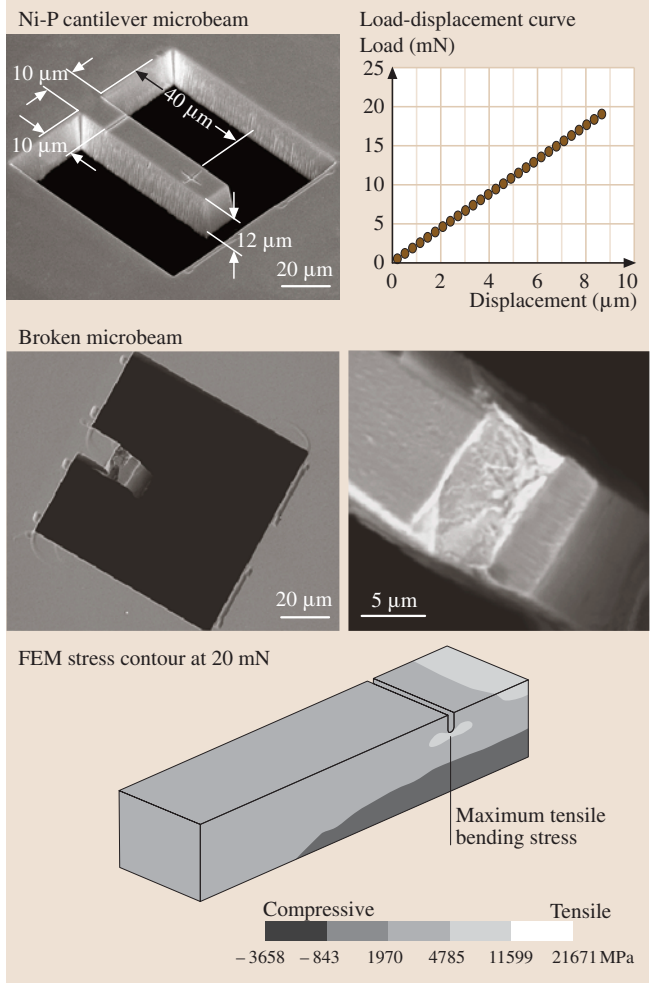


Fig. 25.17 SEM micrographs of the new and broken beams, load-displacement curve, and FEM stress contour for the notched Ni-P cantilever microbeam [25.40]

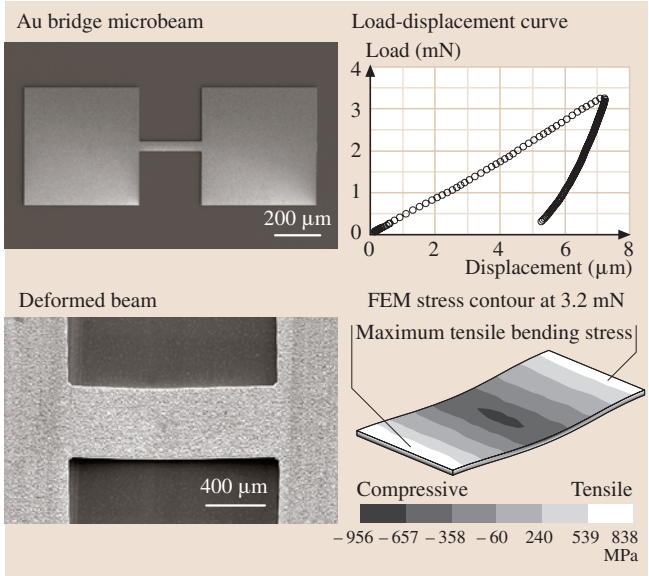


Fig. 25.18 SEM micrographs of the new and deformed beams, load-displacement curve, and FEM stress contour for the Au bridge microbeam [25.40]

load in the bending tests. The Au bridge microbeams were fabricated by electroplating technique.

Figure 25.17 shows the SEM images, load displacement curve, and FEM stress contour for the notched Ni-P cantilever microbeam that was bent to failure [25.40]. The distance between the loading position and the fixed end is 40 µm. The 3-µm-deep notch is 10 µm from the fixed end. The notched beam showed linear behavior followed by abrupt failure. The FEM stress contour shows that there is higher stress concentration at the notch tip. The maximum tensile stress σ_m at the notch tip can be analyzed using Griffith fracture theory as follows [25.53]:

$$\sigma_m \approx 2\sigma_o \left(\frac{c}{\rho}\right)^{1/2} , \qquad (25.11)$$

where σ_o is the average applied tensile stress on the beam, c is the crack length, and ρ is the crack tip radius. Therefore, elastic-plastic deformation will first occur locally at the end of the notch tip, followed by abrupt fracture failure after the σ_m reaches the ultimate tensile strength of Ni-P, even though the rest of the beam is still in the elastic regime. The SEM image of the fracture surface shows that the fracture started right from the notch tip with plastic deformation characteristic. This indicates that although local plastic deformation occurred at the notch tip area, the whole beam failed catastrophically. The present study shows that FEM simulation can predict well the stress concentration and helps in understanding the failure mechanism of the notched beams.

Figure 25.18 shows the SEM images, load-displacement curve and FEM stress contour for the Au bridge microbeam that was deformed by the indenter [25.40]. The recession gap between the beam and substrate is about 7 µm, which is not large enough to break the beam at the load applied. From the load-displacement curve, we note that the beam experienced elastic-plastic deformation. The FEM stress contour shows that the maximum tensile stress is located at the fixed ends, whereas the minimum compressive stress is located around the center of the beam. The SEM image shows that the beam has been permanently deformed. No crack was found on the beam surface. The present study shows a possibility for mechanically forming the Au film into the shape as needed. This may help in designing/fabricating functionally complex smart micro/nanodevices that need conductors for power supply and input/output signals.

25.3 Finite Element Analysis of Nanostructures with Roughness and Scratches

Micro/nanostructures have some surface topography and local scratches dependent upon the manufacturing process. Surface roughness and local scratches may compromise the reliability of the devices and their effect needs to be studied. Finite element modeling is used to perform parametric analysis to study the effect of surface roughness and scratches in different well-defined forms on tensile stresses that are responsible for crack propagation [25.46, 47]. The analysis has been carried out on trapezoidal beams supported at the bottom whose data (on Si and SiO_2 nanobeams) have been presented earlier.

The finite element analysis has been carried out using the static analysis of ANSYS 5.7, which calculates the deflections and stresses produced by the applied loading. The type of element selected for the present study was

the SOLID95, which allows the use of different shapes without much loss of accuracy. This element is 3-D with 20 nodes, each node having three degrees of freedom that imply translation in the x, y, and z directions. Each nanobeam cross section is divided into six elements along the width and the thickness and 40 elements along the length. SOLID95 has plasticity, creep, stress stiffening, large deflection, and large strain capabilities. The large displacement analysis is used for large loads. The mesh is kept finer near the asperities and the scratches in order to take into account variation in the bending stresses. The beam materials studied are made of single-crystal silicon (110) and SiO_2 films whose data have been presented earlier. Based on bending experiments presented earlier, the beam materials can be assumed to be linearly elastic isotropic materials. Young's modulus of elasticity (E) and Poisson's ratio (ν) for Si and SiO_2 are 169 GPa [25.63] and 0.28 [25.57], and 73 GPa [25.64] and 0.17 [25.64], respectively. A sample nanobeam of silicon was chosen for performing most of the analysis, as silicon is the most widely used MEMS/NEMS material. The cross section of the fabricated beams used in the experiment is trapezoidal and supported at the bottom, Fig. 25.9. The following dimensions are used: $w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, and $\ell = 6$ μm. In the boundary conditions, the displacements are constrained in all directions on the bottom surface for 1 μm from each end. A point load applied at the center of the beam is simulated with the load being applied at three closely located central nodes on the beam used. It has been observed from the experimental results that the Si nanobeam breaks at around 80 μN. Therefore, in this analysis, a nominal load of 70 μN is selected. At this load, deformations are large and a large displacement option is used.

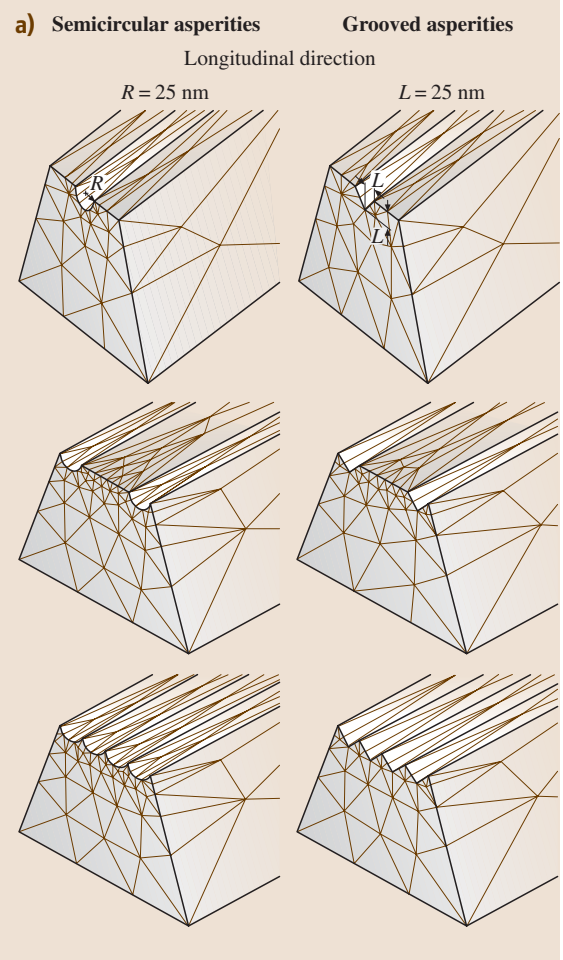


Fig. 25.19 **(a)** Plots showing the geometries of modeled roughness – semicircular and grooved asperities along the nanobeam length with defined geometrical parameters. **(b)** (*See on next page.*) Schematic showing semicircular asperities and scratches in the transverse direction, followed by the illustration of the mesh created on the beam with fine mesh near the asperities and scratches. Also shown are the semicircular asperities and scratches at different pitch values

To study the effect of surface roughness and scratches on the maximum bending stresses the following cases were studied. First, the semicircular and grooved asperities in the longitudinal direction with defined geometrical parameters are analyzed, Fig. 25.19a. Next, semicircular asperities and scratches placed along the transverse direction at a distance c from the end and separated by pitch p from each other are analyzed, Fig. 25.19b. Lastly, the beam material is assumed to be either purely elastic, elastic-plastic, or elastic-perfectly plastic. In the following section, we begin with the stress distribution in smooth nanobeams followed by the effect of surface roughness in the longitudinal and transverse directions and scratches in the transverse direction.

25.3.1 Stress Distribution in a Smooth Nanobeam

Figure 25.20 shows the stress and vertical displacement contours for a nanobeam supported at the bottom and loaded at the center [25.46, 47]. As expected, the maximum tensile stress occurs at the ends, while the maximum compressive stress occurs under the load at

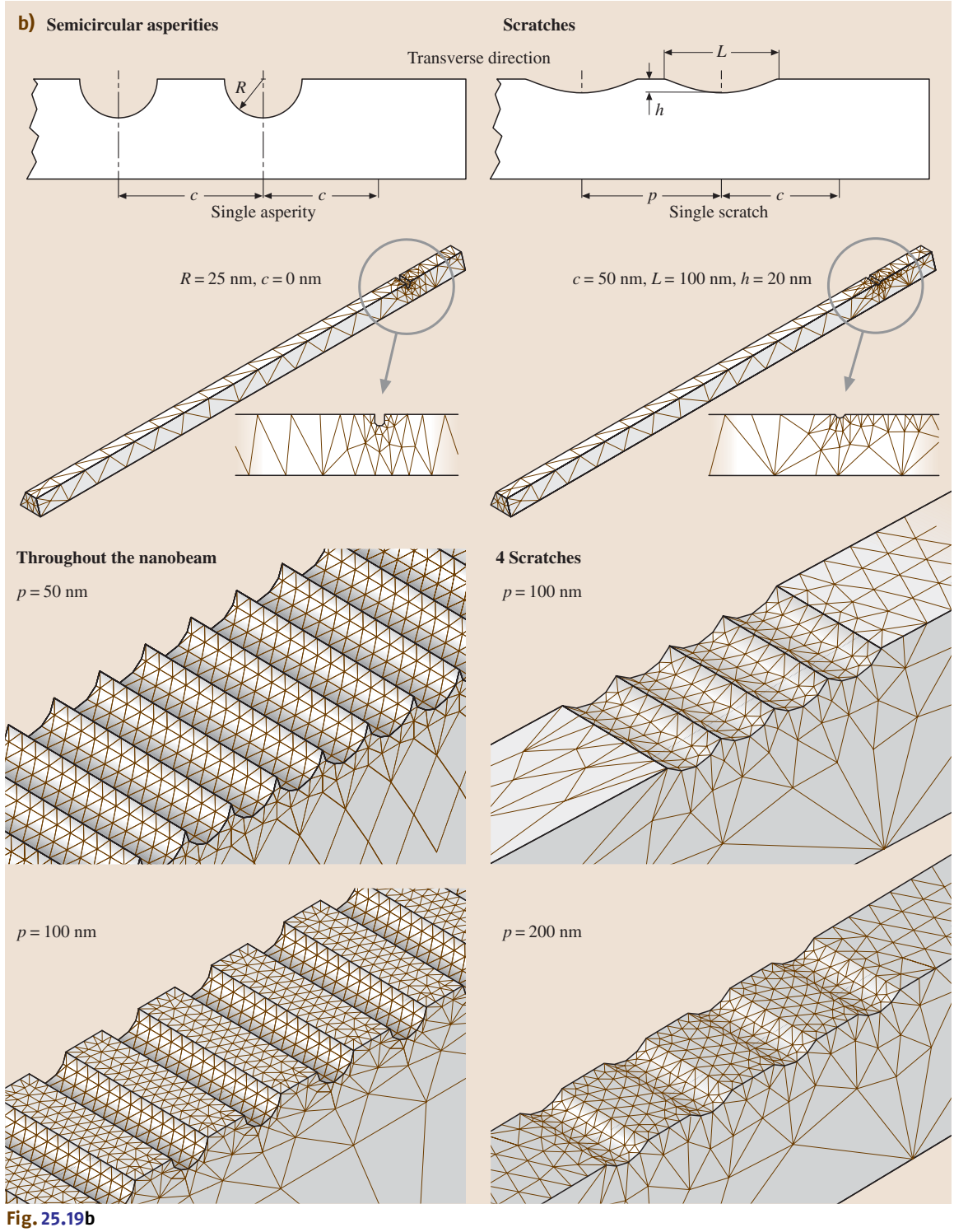

b) Semicircular asperities
Scratches
Transverse direction
R
L
h
c
c
Single asperity
p
c
Single scratch
R = 25 nm, c = 0 nm
c = 50 nm, L = 100 nm, h = 20 nm
Throughout the nanobeam
p = 50 nm
4 Scratches
p = 100 nm
p = 100 nm
p = 200 nm

Fig. 25.19b

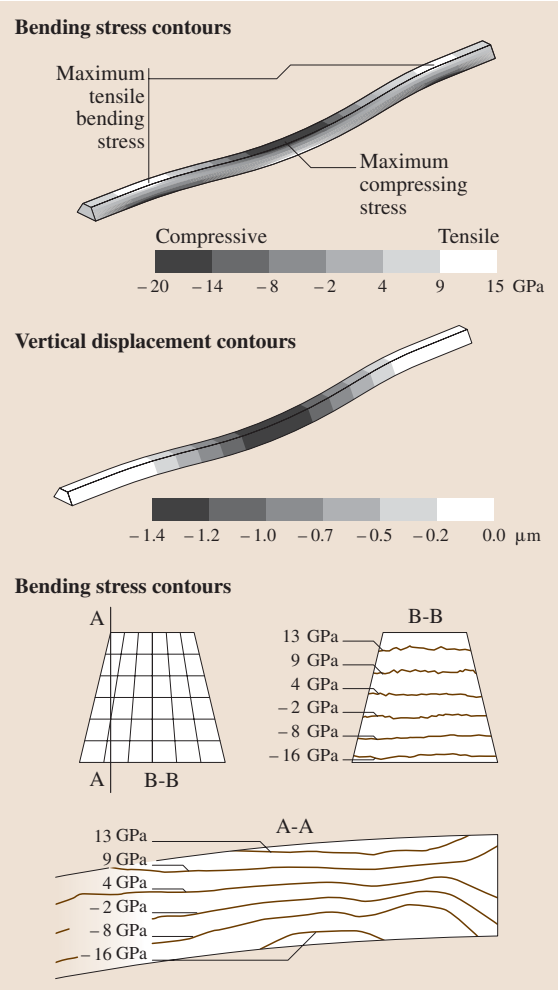


Fig. 25.20 Bending stress contours, vertical displacement contours, and bending stress contours after loading trapezoidal Si nanobeam ($w_1 = 200\,\text{nm}$, $w_2 = 370\,\text{nm}$, $t = 255\,\text{nm}$, $\ell = 6\,\mu\text{m}$, $E = 169\,\text{GPa}$, $\nu = 0.28$) at $70\,\mu\text{N}$ load [25.47]

the center. Stress contours obtained at a section of the beam from the front and side are also shown. In the beam cross section, the stresses remain constant at a given vertical distance from one side to another and change with a change in vertical location. This can be explained by the fact that the bending moment is constant at a particular cross section, so the stress is only dependent on the distance from the neutral axis. However, in cross section A-A the high tensile and compressive stresses are localized near the end of the beam at top and bottom, respectively, whereas the lower values are spread out away from the ends. High value of tensile stresses occurs near the ends because of high bending moment.

25.3.2 Effect of Roughness in the Longitudinal Direction

The roughness in the form of semicircular and grooved asperities in the longitudinal direction on the maximum bending stresses are analyzed [25.47]. The radius R and depth L are kept fixed at 25 nm, while the number of asperities is varied and their effect is observed on the maximum bending stresses. Figure 25.21 shows the variation of maximum bending stresses as a function of asperity shape and the number of asperities. The maximum bending stresses increase as the asperity number increases for both semicircular and grooved asperities. This can be attributed to the fact that as the asperity number increases, the moment of inertia decreases for that cross section. Also, the distance from the neutral axis increases because the neutral axis shifts downwards. Both these factors lead to the increase in the maximum bending stresses, and this effect is more pronounced in the case of semicircular asperity as it exhibits a higher value of maximum bending stress than that in grooved asperity. Figure 25.21 shows the stress contours obtained at a section of the beam from the front side for both cases when we have a single semicircular asperity and when four adjacent semicircular asperities are present. Trends are similar to those observed earlier for a smooth nanobeam (Fig. 25.20).

25.3.3 Effect of Roughness in the Transverse Direction and Scratches

We analyze semicircular asperities when placed along the transverse direction followed by the effect of scratches on the maximum bending stresses in varying numbers and pitch [25.47]. In the analysis of semicircular transverse asperities, three cases were considered, which included a single asperity and asperities throughout the nanobeam surface separated by pitch equal to 50 nm and 100 nm. In all of these cases, c value was kept equal to 0 nm. Figure 25.22 shows that the value of maximum tensile stress is 42 GPa, which is much larger than the maximum tensile stress value with no asperity of 16 GPa, or when the semicircular asperity is present in the longitudinal direction. It is also observed that the maximum tensile stress does not vary with the number of asperities or the pitch, while the maximum compres-

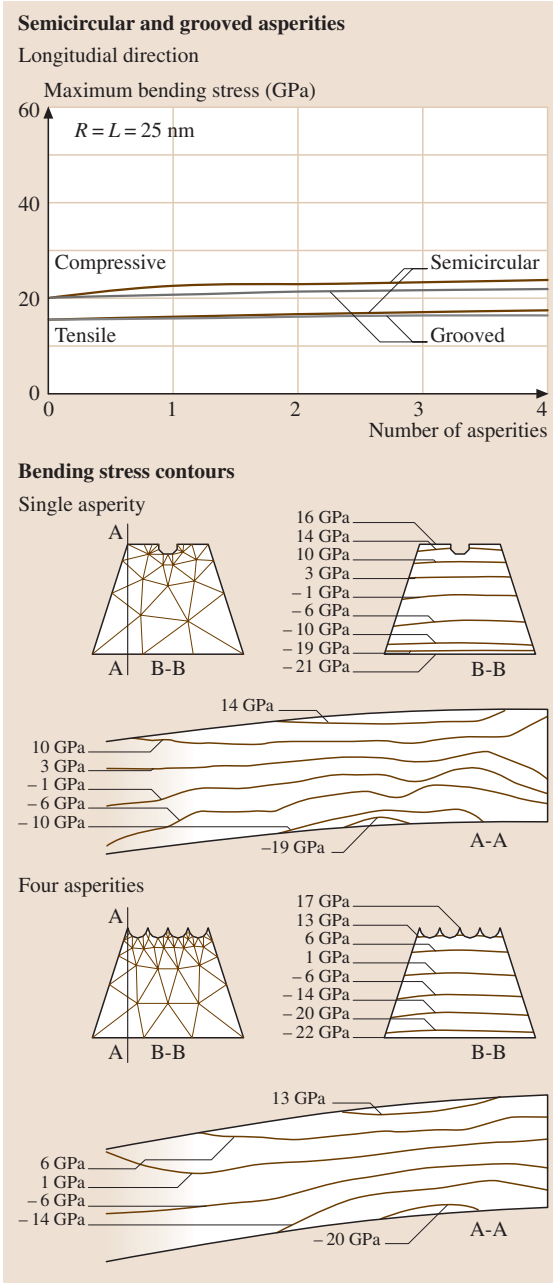


Fig. 25.21 Effect of longitudinal semicircular and grooved asperities in different numbers on maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6\,\mu$m, $E = 169$ GPa, $\nu = 0.28$, load $= 70\,\mu$N). Bending stress contours obtained in the beam with semicircular single asperity and four adjacent asperities of $R = 25$ nm [25.47]

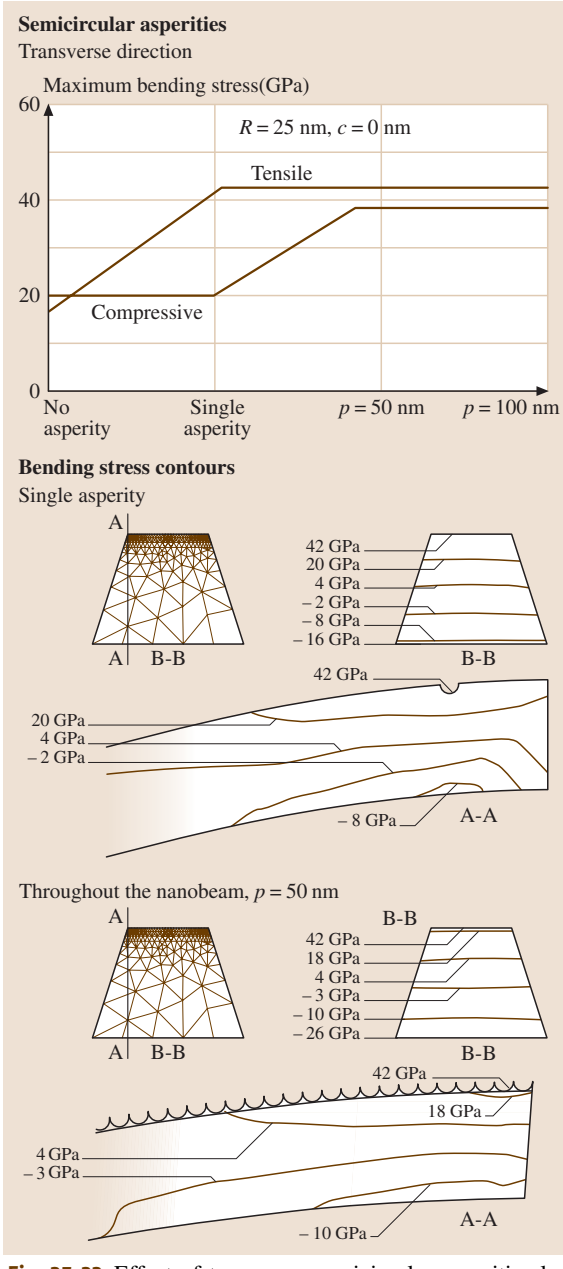


Fig. 25.22 Effect of transverse semicircular asperities located at different pitch values on the maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6\,\mu$m, $E = 169$ GPa, $\nu = 0.28$, load $= 70\,\mu$N). Bending stress contours obtained in the beam with semicircular single asperity and semicircular asperities throughout the nanobeam surface at $p = 50$ nm [25.47]

sive stress does increase dramatically for the asperities present throughout the beam surface from its value when a single asperity is present. Maximum tensile stress occurs at the ends, and an increase in p does not add any asperities at the ends, whereas asperities are added in the central region where compressive stresses are maximum. The semicircular asperities present at the center cause the local perturbation in the stress distribution at the center of the asperity where load is being applied, leading to a high value of maximum compressive stress [25.65]. Figure 25.22 also shows the stress contours obtained at a section of the beam from the front and side for both cases when there is a single semicircular asperity and when asperities are present throughout the beam surface at a pitch equal to 50 nm. Trends are similar to those observed earlier for a smooth nanobeam (Fig. 25.20).

In the study pertaining to scratches, the number of scratches are varied along with the variation in the pitch. Furthermore, the load is applied at the center of the beam and at the center of the scratch near the end, as for all the cases. In all of these, c value was kept equal to 50 nm and L value was equal to 100 nm with h value being 20 nm. Figure 25.23 shows that the value of maximum tensile stress remains almost the same with the number of scratches for both types of loading – that is, when load is applied at the center of the beam and at the center of the scratch near the end. This is because the maximum tensile stress occurs at the beam ends no matter where the load gets applied. But the presence of scratch does increase the maximum tensile stress compared to its value for a smooth nanobeam, although the number of scratches no longer matter as the maximum tensile stress occurring at the nanobeam end is unaffected by the presence of more scratches beyond the first scratch in the direction toward the center. The value of the tensile stress is much lower when the load is applied at the center of scratch, and it can be explained as follows. The negative bending moment at the end near the applied load decreases with load offset after two-thirds of the length of the beam [25.66]. Since this negative bending moment is responsible for tensile stresses, their behavior with the load offset is the same as the negative bending moment. Also, the value of maximum compressive stress when load is applied at the center of the nanobeam remains almost the same as the center geometry is unchanged due to the number of scratches and, hence, the maximum compressive stress occurring below the load at the center is same. On the other hand, when the load is applied at the center of the scratch we observe that the maximum compressive stress in-

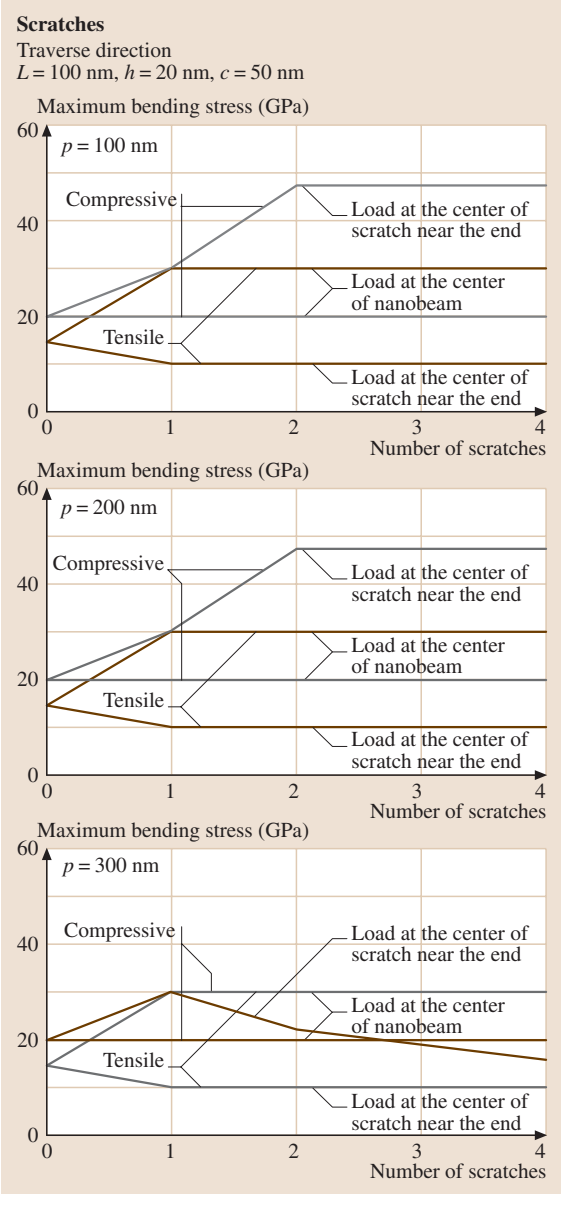


Fig. 25.23 Effect of number of scratches along with the variation in the pitch on the maximum bending stresses after loading trapezoidal Si nanobeams ($w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6\,\mu$m, $E = 169$ GPa, $\nu = 0.28$, load $= 70\,\mu$N). Also shown is the effect of load when applied at the center of the beam and at the center of the scratch near the end [25.47]

creases dramatically because the local perturbation in the stress distribution at the center of the scratch where load is being applied leads to a high value of maximum compressive stress [25.65]. It increases further with the number of scratches and then levels off. This can be attributed to the fact that when there is another scratch present close to the scratch near the end, the stress concentration is more, as the effect of local perturbation in the stress distribution is more significant. However, this effect is insignificant when more than two scratches are present.

Now we address the effect of pitch on the maximum compressive stress when the load is applied at the center of the scratch near the end. When the pitch is up to a value of 200 nm the maximum compressive stress increases with the number of scratches, as discussed earlier. On the other hand, when the pitch value goes beyond 225 nm this effect is reversed. This is because the presence of another scratch no longer affects the local perturbation in the stress distribution at the scratch near the end. Instead, more scratches at a fair distance distribute the maximum compressive stress at the scratch near the end and the stress starts going down. Such observations of maximum bending stresses can help in identifying the number of asperities and scratches allowed separated by an optimum distance from each other.

25.3.4 Effect on Stresses and Displacements for Materials That Are Elastic, Elastic-Plastic, or Elastic-Perfectly Plastic

This section deals with the beam modeled as elastic, elastic-plastic, and elastic-perfectly plastic to observe the variation in the stresses and displacements from an elastic model used so far [25.47]. Figure 25.24 shows the typical stress-strain curves for the three types of deformation regimes and their corresponding load-displacement curves obtained from the model of an Si nanobeam that are found to exhibit the same trends.

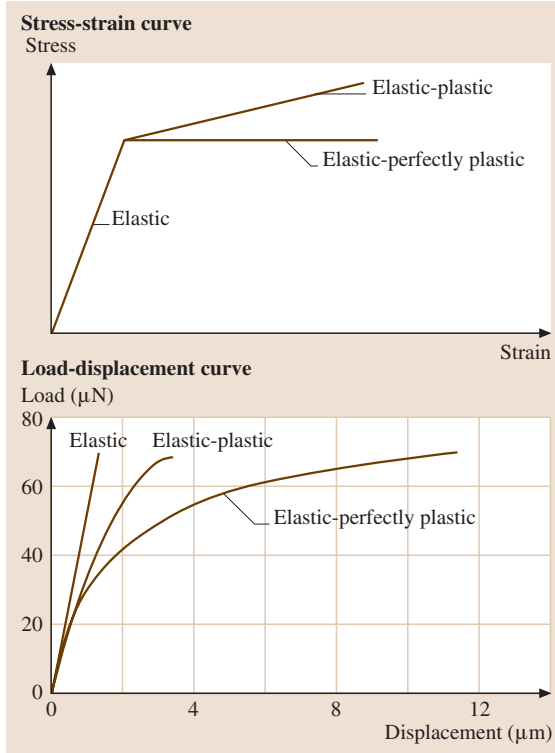


Fig. 25.24 Schematic representation of stress-strain curves and load-displacement curves for material when it is elastic, elastic-plastic, or elastic-perfectly plastic for a Si nanobeam ($w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6\,\mu$m, $E = 169$ GPa, tangent modulus in plastic range $= 0.5\,E$, $\nu = 0.28$) [25.47]

Table 25.3 shows the comparison of maximum von Mises stress and maximum displacements for

Table 25.3 Stresses and displacements for materials that are elastic, elastic-plastic, or elastic-perfectly plastic (load $= 70\,\mu$N, $w_1 = 200$ nm, $w_2 = 370$ nm, $t = 255$ nm, $\ell = 6\,\mu$m, $R = 25$ nm, $E = 169$ GPa, tangent modulus in plastic range $= 0.5E$, $\nu = 0.28$)

	Elastic		Elastic-plastic		Elastic-perfectly plastic	
	Smooth nanobeam	Single semicircular longitudinal asperity	Smooth nanobeam	Single semicircular longitudinal asperity	Smooth nanobeam	Single semicircular longitudinal asperity
Max. von Mises stress (GPa)	18.2	19.3	13.5	15.2	7.8	9.1
Max. displacement (μm)	1.34	1.40	3.35	3.65	11.5	12.3

both smooth nanobeam and nanobeam with a defined roughness, which is single semicircular longitudinal asperity of R value equal to 25 nm for the three different models. It is observed that the maximum value of stress is obtained at a given load for elastic material, whereas the displacement is maximum for elastic-perfectly plastic material. Also, the pattern that the maximum bending stress value increases for a rough nanobeam still holds true in the other models as well.

25.4 Closure

Mechanical properties of nanostructures are necessary in designing realistic MEMS/NEMS devices. Most mechanical properties are scale-dependent. Micro/nanomechanical properties, hardness, elastic modulus, and scratch resistance of bulk materials of undoped single-crystal silicon (Si) and thin films of undoped polysilicon, SiO_2, SiC, Ni-P, and Au are presented. It is found that the SiC film exhibits higher hardness, elastic modulus, and scratch resistance as compared to other materials.

Bending tests have been performed on the Si and SiO_2 nanobeams and Ni-P and Au microbeams using an AFM and a depth-sensing nanoindenter, respectively. The bending tests were used to evaluate elastic modulus, bending strength (fracture stress), fracture toughness (K_{IC}), and fatigue strength of the beam materials. The Si and SiO_2 nanobeams exhibited elastic linear response with sudden brittle fracture. The notched Ni-P beam showed linear deformation behavior followed by abrupt failure. The Au beam showed elastic-plastic deformation behavior. Elastic modulus values of 182±11 GPa for Si(110) and 85±3 GPa for SiO_2 were obtained, which are comparable to bulk values. Bending strength values of 18±3 GPa for Si and 7.6±2 GPa for SiO_2 were obtained, which are twice as large as values reported on larger scale specimens. This indicates that bending strength shows a specimen size dependence. Fracture toughness value estimates obtained were $1.67 \pm 0.4\,\text{MPa}\sqrt{\text{m}}$ for Si and $0.60 \pm 0.2\,\text{MPa}\sqrt{\text{m}}$ for SiO_2, which are also comparable to values obtained on larger specimens. At stress amplitudes less than 15% of their bending strength and at mean stresses of less than 30% of the bending strength, Si and SiO_2 displayed an apparent endurance life of greater than 30,000 cycles. SEM observations of the fracture surfaces revealed a cleavage type of fracture for both materials when subjected to bending, as well as fatigue. The AFM and nanoindenters used in this study can be satisfactorily used to evaluate the mechanical properties of micro/nanoscale structures for use in MEMS/NEMS.

FEM simulations are used to predict the stress and deformation in nanostructures. The FEM has been used to analyze the effect of the type of surface roughness and scratches on stresses and deformation of nanostructures. We find that roughness affects the maximum bending stresses. The maximum bending stresses increase as the asperity number increases for both semicircular and grooved asperities in the longitudinal direction. When the semicircular asperity is present in the transverse direction the maximum tensile stress is much larger than the maximum tensile stress value with no asperity or when the semicircular asperity is present in the longitudinal direction. This observation suggests that the asperity in the transverse direction is more detrimental. The presence of scratches increases the maximum tensile stress. The maximum tensile stress remains almost the same with the number of scratches for two types of loading, that is, when load is applied at the center of the beam or at the center of the scratch near the end, although the value of the tensile stress is much lower when the load is applied at the center of the scratch. This means that the load applied at the ends is less damaging. This analysis shows that FEM simulations can be useful to designers to develop the most suitable geometry for nanostructures.

References

25.1 R. S. Muller, R. T. Howe, S. D. Senturia, R. L. Smith, R. M. White: *Microsensors* (IEEE, New York 1990)

25.2 I. Fujimasa: *Micromachines: A New Era in Mechanical Engineering* (Oxford Univ. Press, Oxford 1996)

25.3 W. S. Trimmer (Ed.): *Micromachines and MEMS, Classic and Seminal Papers to 1990* (IEEE, New York 1997)

25.4 B. Bhushan: *Tribology Issues and Opportunities in MEMS* (Kluwer, Dordrecht 1998)

25.5 G. T. A. Kovacs: *Micromachined Transducers Sourcebook* (WCB McGraw-Hill, Boston 1998)

25.6 S. D. Senturia: *Microsystem Design* (Kluwer, Boston 2001)

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

25.8 T. R. Hsu: *MEMS and Microsystems* (McGraw-Hill, Boston 2002)

25.9 M. Madou: *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd edn. (CRC, Boca Raton 2002)

25.10 K. E. Drexler: *Nanosystems: Molecular Machinery, Manufacturing and Computation* (Wiley, New York 1992)

25.11 B. Bhushan: *Handbook of Micro/Nanotribology*, 2nd edn. (CRC, Boca Raton 1999)

25.12 G. Timp (Ed.): *Nanotechnology* (Springer, New York 1999)

25.13 E. A. Rietman: *Molecular Engineering of Nanosystems* (Springer, New York 2001)

25.14 H. S. Nalwa (Ed.): *Nanostructures Materials and Nanotechnology* (Academic, San Diego 2002)

25.15 W. A. Goddard, D. W. Brenner, S. E. Lyshevski, G. J. Iafrate: *Handbook of Nanoscience, Engineering, and Technology* (CRC, Boca Raton 2003)

25.16 B. Bhushan: *Principles and Applications of Tribology* (Wiley, New York 1999)

25.17 B. Bhushan: *Introduction to Tribology* (Wiley, New York 2002)

25.18 B. Bhushan: Macro- and microtribology of MEMS materials. In: *Modern Tribology Handbook*, ed. by B. Bhushan (CRC, Boca Raton 2001) pp. 1515–1548

25.19 S. Johansson, J. A. Schweitz, L. Tenerz, J. Tiren: Fracture testing of silicon microelements in-situ in a scanning electron microscope, J. Appl. Phys. **63** (1988) 4799–4803

25.20 F. Ericson, J. A. Schweitz: Micromechanical fracture strength of silicon, J. Appl. Phys. **68** (1990) 5840–5844

25.21 E. Obermeier: Mechanical and thermophysical properties of thin film materials for MEMS: Techniques and devices, Micromechan. Struct. Mater. Res. Symp. Proc. **444** (1996) 39–57

25.22 C. J. Wilson, A. Ormeggi, M. Narbutovskih: Fracture testing of silicon microcantilever beams, J. Appl. Phys. **79** (1995) 2386–2393

25.23 W. N. Sharpe, Jr., B. Yuan, R. L. Edwards: A new technique for measuring the mechanical properties of thin films, J. Microelectromech. Syst. **6** (1997) 193–199

25.24 K. Sato, T. Yoshioka, T. Anso, M. Shikida, T. Kawabata: Tensile testing of silicon film having different crystallographic orientations carried out on a silicon chip, Sens. Actuators A **70** (1998) 148–152

25.25 S. Greek, F. Ericson, S. Johansson, M. Furtsch, A. Rump: Mechanical characterization of thick polysilicon films: Young's modulus and fracture strength evaluated with microstructures, J. Micromech. Microeng. **9** (1999) 245–251

25.26 D. A. LaVan, T. E. Buchheit: Strength of polysilicon for MEMS devices, Proc. SPIE **3880** (1999) 40–44

25.27 E. Mazza, J. Dual: Mechanical behavior of a μm-sized single crystal silicon structure with sharp notches, J. Mech. Phys. Solids **47** (1999) 1795–1821

25.28 T. Yi, C. J. Kim: Measurement of mechanical properties for MEMS materials, Meas. Sci. Technol. **10** (1999) 706–716

25.29 H. Kahn, M. A. Huff, A. H. Heuer: Heating effects on the Young's modulus of films sputtered onto micromachined resonators, Microelectromech. Struct. Mater. Res. Symp. Proc. **518** (1998) 33–38

25.30 S. Johansson, F. Ericson, J. A. Schweitz: Influence of surface-coatings on elasticity, residual-stresses, and fracture properties of silicon microelements, J. Appl. Phys. **65** (1989) 122–128

25.31 R. Ballarini, R. L. Mullen, Y. Yin, H. Kahn, S. Stemmer, A. H. Heuer: The fracture toughness of polysilicon microdevices: A first report, J. Mater. Res. **12** (1997) 915–922

25.32 H. Kahn, R. Ballarini, R. L. Mullen, A. H. Heuer: Electrostatically actuated failure of microfabricated polysilicon fracture mechanics specimens, Proc. R. Soc. London A **455** (1999) 3807–3823

25.33 A. M. Fitzgerald, R. H. Dauskardt, T. W. Kenny: Fracture toughness and crack growth phenomena of plasma-etched single crystal silicon, Sens. Actuators A **83** (2000) 194–199

25.34 T. Tsuchiya, A. Inoue, J. Sakata: Tensile testing of insulating thin films: Humidity effect on tensile strength of SiO_2 films, Sens. Actuators A **82** (2000) 286–290

25.35 J. A. Connally, S. B. Brown: Micromechanical fatigue testing, Exp. Mech. **33** (1993) 81–90

25.36 K. Komai, K. Minoshima, S. Inoue: Fracture and fatigue behavior of single-crystal silicon microelements and nanoscopic AFM damage evaluation, Microsyst. Technol. **5** (1998) 30–37

25.37 T. Namazu, Y. Isono, T. Tanaka: Evaluation of size effect on mechanical properties of single-

crystal silicon by nanoscale bending test using AFM, J. Microelectromech. Syst. **9** (2000) 450–459

25.38 S. Sundararajan, B. Bhushan: Development of AFM-based techniques to measure mechanical properties of nanoscale structures, Sens. Actuators A **101** (2002) 338–351

25.39 X. Li, B. Bhushan: Fatigue studies of nanoscale structures for MEMS/NEMS applications using nanoindentation techniques, Surf. Coat. Technol. **163-164** (2003) 521–526

25.40 X. Li, B. Bhushan, K. Takashima, C.W. Baek, Y.K. Kim: Mechanical characterization of micro/nanoscale structures for MEMS/NEMS applications using nanoindentation techniques, Ultramicroscopy **97** (2003) 481–494

25.41 T. Hsu, N. Sun: Residual stresses/strains analysis of MEMS, Proc. of the Int. Conf. on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) 82–87

25.42 A. Kolpekwar, C. Kellen, R.D. (Shawn) Blanton: Fault model generation for MEMS, Proc. of the Int. Conf. on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) 111–116

25.43 H.A. Rueda, M.E. Law: Modeling of strain in boron-doped silicon cantilevers, Proc. of the Int. Conf. on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators, ed. by M. Laudon, B. Romanowicz (Computational Publications, Cambridge 1998) 94–99

25.44 M. Heinzelmann, M. Petzold: FEM analysis of microbeam bending experiments using ultramicro indentation, Comput. Mater. Sci. **3** (1994) 169–176

25.45 C.J. Wilson, P.A. Beck: Fracture testing of bulk silicon microcantilever beams subjected to a side load, J. Microelectromech. Syst. **5** (1996) 142–150

25.46 B. Bhushan, G.B. Agrawal: Stress analysis of nanostructures using a finite element method, Nanotechnology **13** (2002) 515–523

25.47 B. Bhushan, G.B. Agrawal: Finite element analysis of nanostructures with roughness and scratches, Ultramicroscopy **97** (2003) 495–501

25.48 K.E. Petersen: Silicon as a mechanical material, Proc. IEEE **70** (1982) 420–457

25.49 B. Bhushan, S. Sundararajan, X. Li, C.A. Zorman, M. Mehregany: Micro/nanotribological studies of single-crystal silicon and polysilicon and SiC films for use in MEMS devices. In: *Tribology Issues and Opportunities in MEMS*, ed. by B. Bhushan (Kluwer, Dordrecht 1998) pp. 407–430

25.50 S. Sundararajan, B. Bhushan: Micro/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear **217** (1998) 251–261

25.51 X. Li, B. Bhushan: Micro/nanomechanical characterization of ceramic films for microdevices, Thin Solid Films **340** (1999) 210–217

25.52 B. Bhushan, X. Li: Nanomechanical characterization of solid surfaces and thin films, Int. Mater. Rev. **48** (2003) 125–164

25.53 B.R. Lawn, A.G. Evans, D.B. Marshall: Elastic/plastic indentation damage in ceramics: The median/radial system, J. Am. Ceram. Soc. **63** (1980) 574

25.54 S. Sundararajan, B. Bhushan, T. Namazu, Y. Isono: Mechanical property measurements of nanoscale structures using an atomic force microscope, Ultramicroscopy **91** (2002) 111–118

25.55 R.J. Roark: *Formulas for Stress and Strain*, 6th edn. (McGraw-Hill, New York 1989)

25.56 R.W. Hertzberg: *Deformation and Fracture Mechanics of Engineering Materials*, 3rd edn. (Wiley, New York 1989) pp. 277–278

25.57 Anonymous: Properties of silicon, EMIS Datareviews Series No. 4 (1988)

25.58 C.T.C. Nguyen, R.T. Howe: An integrated CMOS micromechanical resonator high-Q oscillator, IEEE J. Solid State Circ. **34** (1999) 440–455

25.59 L.J. Hornbeck: A digital light processing™ update – status and future applications, Proc. SPIE **3634** (1999) 158–170

25.60 M. Tanaka: Fracture toughness and crack morphology in indentation fracture of brittle materials, J. Mater. Sci. **31** (1996) 749

25.61 T. Tsuchiya, O. Tabata, J. Sakata, Y. Taga: Specimen size effect on tensile strength of surface-micromachined polycrystalline silicon thin films, J. Microelectromech. Syst. **7** (1998) 106–113

25.62 T. Yi, L. Li, C.J. Kim: Microscale material testing of single crystalline silicon: Process effects on surface morphology and tensile strength, Sens. Actuators A **83** (2000) 172–178

25.63 B. Bhushan, S. Venkatesan: Mechanical and tribological properties of silicon for micromechanical applications: A review, Adv. Info. Storage Syst. **5** (1993) 211–239

25.64 B. Bhushan, B.K. Gupta: *Handbook of Tribology: Materials, Coatings, and Surface Treatments* (McGraw-Hill, New York 1991)

25.65 S.P. Timoshenko, J.N. Goodier: *Theory of Elasticity*, 3rd edn. (McGraw-Hill, New York 1970)

25.66 J.E. Shigley, L.D. Mitchell: *Mechanical Engineering Design*, 4th edn. (McGraw-Hill, New York 1993)

Part D Molecularl

Part D Molecularly Thick Films for Lubrication

26 Nanotribology of Ultrathin and Hard Amorphous Carbon Films
Bharat Bhushan, Columbus, USA

27 Self-Assembled Monolayers for Controlling Adhesion, Friction and Wear
Bharat Bhushan, Columbus, USA
Huiwen Liu, Columbus, USA

28 Nanoscale Boundary Lubrication Studies
Bharat Bhushan, Columbus, USA
Huiwen Liu, Columbus, USA

29 Kinetics and Energetics in Nanolubrication
Rene M. Overney, Seattle, USA
George W. Tyndall, San Jose, USA
Jane Frommer, San Jose, USA

26. Nanotribology of Ultrathin and Hard Amorphous Carbon Films

Diamond material and its smooth coatings are used for very low wear and relatively low friction. Major limitations of the true diamond coatings are that they need to be deposited at high temperatures, can only be deposited on selected substrates, and require surface finishing. Hard amorphous carbon, commonly known as diamond-like carbon or DLC coatings, exhibit mechanical, thermal, and optical properties close to that of diamond. These can be deposited with a large range of thicknesses by using a variety of deposition processes on a variety of substrates at or near room temperature. The coatings reproduce substrate topography, avoiding the need of post-finishing. Friction and wear properties of some DLC coatings can be very attractive for tribological applications. The largest industrial application of these coatings is in magnetic storage devices.

The prevailing atomic arrangement in the DLC coatings is amorphous or quasi-amorphous with small diamond, graphite, and other unidentifiable micro- or nanocrystallites. Most DLC coatings, except those produced by filtered cathodic arc, contain from a few to about 50 at% hydrogen. Sometimes hydrogen is deliberately incorporated in the sputtered and ion plated coatings to tailor their properties.

EELS and Raman spectroscopies can be successfully used for chemical characterization of amorphous carbon coatings. The prevailing atomic arrangement in the DLC coatings is amorphous or quasi-amorphous with small diamond (sp^3), graphite (sp^2) and other unidentifiable micro- or nanocrystallites. Most DLC coatings except those produced by filtered cathodic arc contain from a few to about 50 at% hydrogen. Sometimes hydrogen is deliberately incorporated in the sputtered and ion plated coatings to tailor their properties.

Amorphous carbon coatings deposited by various techniques exhibit different mechanical and tribological properties. The nanoindenter can be successfully used for measurement of hardness, elastic modulus, fracture toughness, and fatigue life. Microscratch and microwear experiments can be performed using either a nanoindenter or an AFM. Thin coatings deposited by filtered cathodic arc, ion beam, and ECR-CVD hold a promise for tribological applications. Coatings as thin as 5 nm or even thinner provide wear protection. Microscratch, microwear, and accelerated wear testing, if simulated properly, can be successfully used to screen coating candidates for industrial applications. In the examples shown in this chapter, trends observed in the microscratch, microwear, and accelerated macrofriction wear tests are similar to that found in functional tests.

In this chapter, the state-of-the-art of recent developments in the chemical, mechanical, and tribological characterization of ultrathin amorphous carbon coatings is presented.

26.1 **Description of Commonly Used Deposition Techniques** 795
26.1.1 Filtered Cathodic Arc Deposition Technique 798
26.1.2 Ion Beam Deposition Technique 798
26.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique 799
26.1.4 Sputtering Deposition Technique ... 799
26.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique 799

26.2 **Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties** ... 800
26.2.1 EELS and Raman Spectroscopy 800
26.2.2 Hydrogen Concentrations 804
26.2.3 Physical Properties 804
26.2.4 Summary 805

26.3 **Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques** 805
26.3.1 Micromechanical Characterization . 805
26.3.2 Microscratch and Microwear Studies 813
26.3.3 Macroscale Tribological Characterization 822
26.3.4 Coating Continuity Analysis 826

References 827

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

Carbon exists in both crystalline and amorphous forms and exhibits both metallic and nonmetallic characteristics [26.1–3]. Crystalline carbon includes graphite, diamond, and a family of fullerenes, Fig. 26.1. The graphite and diamond are infinite periodic network solids with a planar structure, whereas the fullerenes are a molecular form of pure carbon with a finite network with a nonplanar structure. Graphite has a hexagonal, layered structure with weak interlayer bonding forces and exhibits excellent lubrication properties. The graphite crystal may be visualized as infinite parallel layers of hexagons stacked 0.34 nm apart with a 0.1415-nm interatomic distance between the carbon atoms in the basal plane. The atoms lying in the basal planes are trigonally coordinated and closely packed with strong σ (covalent) bonds to its three carbon neighbors using the hybrid sp^2 orbitals. The fourth electron lies in a p_z orbital lying normal to the σ bonding plane and forms a weak π bond by overlapping side to side with a p_z orbital of an adjacent atom to which carbon is attached by a σ bond. The layers (basal planes) themselves are relatively far apart and the forces that bond them are weak van der Waals forces. These layers can align themselves parallel to the direction of the relative motion and slide over one another with relative ease, thus providing low friction. Strong interatomic bonding and packing in each layer is thought to help reduce wear. The operating environment has a significant influence on lubrication, i. e., low friction and low wear, properties of graphite. It lubricates better in a humid environment than a dry one, resulting from adsorption of water vapor and other gases from the environment, which further weakens the interlayer bonding forces and results in easy shear and transfer of the crystallite platelets to the mating surface. Thus, transfer plays an important role in controlling friction and wear. Graphite oxidizes at high operating temperatures and can be used up to about 430 °C.

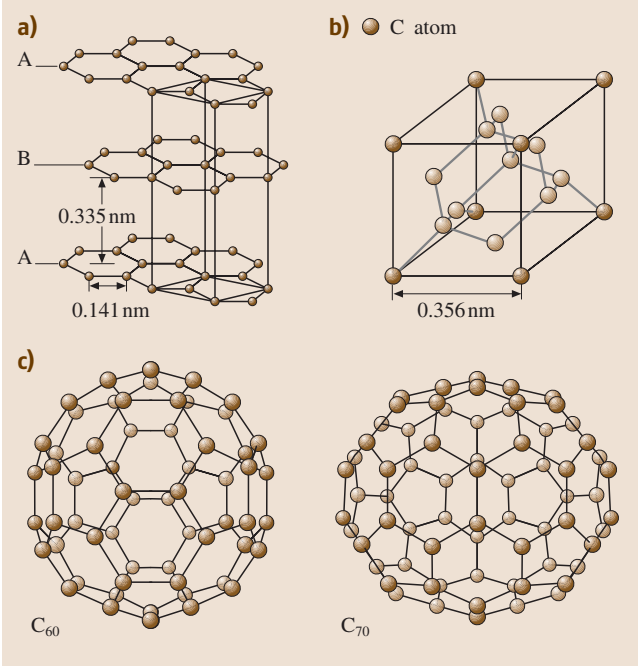


Fig. 26.1a–c The structure of three known forms of crystalline carbon **(a)** hexagonal structure of graphite, **(b)** modified face-centered cubic (fcc) structure, two interpenetrating fcc lattices displaced by one-quarter of the cube diagonal, of diamond (each atom is bonded to four others that form the corners of the pyramidal structure called tetrahedron), and **(c)** the structure of the two most common forms of fullerenes-soccer ball C_{60} and rugby ball C_{70} molecules

One of the fullerene molecules is C_{60}, commonly known as Buckyball. Since the C_{60} molecules are very stable and do not require additional atoms to satisfy chemical bonding requirements, they are expected to have low adhesion to the mating surface and low surface energy. Since C_{60} molecules with a perfect spherical symmetry are weakly bonded to other molecules, C_{60} clusters get detached readily, similar to other layered lattice structures, and either get transferred to the mating surface by mechanical compaction, or are present as loose wear particles that may roll like tiny ball bearings in a sliding contact, resulting in low friction and wear. The wear particles are expected to be harder than as-deposited C_{60} molecules, because of their phase transformation at high asperity contact pressures present in a sliding interface. The low surface energy, spherical shape of C_{60} molecules, weak intermolecular bonding, and high load bearing capacity offer potential for various mechanical and tribological applications. The sublimed C_{60} coatings and fullerene particles as an additive to mineral oils and greases have been reported to be good solid lubricants comparable to graphite and MoS_2 [26.4–6].

Diamond crystallizes in the modified face centered cubic (*fcc*) structure with an interactomic distance of 0.154 nm. The diamond cubic lattice consists of two interpenetrating fcc lattices displaced by one-quarter of the cube diagonal. Each carbon atom is tetrahedrally coordinated, making strong σ (covalent) bonds to its four carbon neighbors using the hybrid sp^3 atomic orbitals, which accounts for its highest hardness (80–104 GPa) and thermal conductivity (900–2,100 W/mK, on the or-

der of five times that of copper) of any known solid, and a high electrical resistivity, optical transmission, and a large optical band gap. It is relatively chemically inert, and it exhibits poor adhesion with other solids with consequent low friction and wear. Its high thermal conductivity allows dissipation of frictional heat during sliding and protects the interface, and the dangling carbon bonds on the surface react with the environment to form hydrocarbons that act as good lubrication films. These are some of the reasons for low friction and wear of the diamond. Diamond and its coatings find many industrial applications: tribological applications (low friction and wear), optical applications (exceptional optical transmission, high abrasion resistance), and thermal management or heat sink applications (high thermal conductivity). The diamond can be used to high temperatures, and it starts to graphitize at about 1,000 °C in ambient air and at about 1,400 °C in vacuum. Diamond is an attractive material for cutting tools, as an abrasive for grinding wheels and lapping compounds, and other extreme wear applications.

The natural diamond, particularly in large sizes, is very expensive and its coatings, a low cost alternative, are attractive. The true diamond coatings are deposited by chemical vapor deposition (CVD) processes at high substrate temperatures (on the order of 800 °C). They adhere best on a silicon substrate and require an interlayer for other substrates. A major roadblock to the widespread use of true diamond films in tribological, optical, and thermal management applications is the surface roughness. Growth of the diamond phase on a non-diamond substrate is initiated by nucleation either at randomly seeded sites or at thermally favored sites, due to statistical thermal fluctuation at the substrate surface. Based on growth temperature and pressure conditions, favored crystal orientations dominate the competitive growth process. As a result, the grown films are polycrystalline in nature with relatively large grain size (> 1 μm) and terminate in very rough surfaces with RMS roughnesses ranging from a few tenths of a micron to tens of microns. Techniques for polishing these films have been developed. It has been reported that the laser polished films exhibit friction and wear properties almost comparable to that of bulk polished diamond [26.7, 8].

Amorphous carbon has no long-range order, and the short-range order of carbon atoms can have one or more of three bonding configurations: – sp^3 (diamond), sp^2 (graphite), or sp^1 (with two electrons forming strong

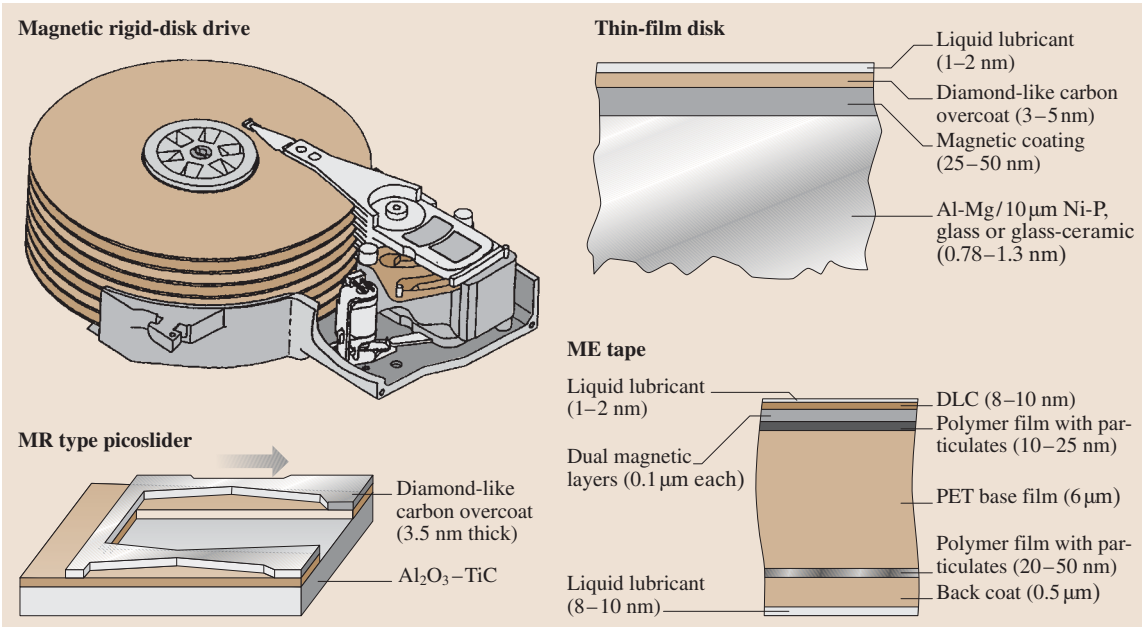


Fig. 26.2 Schematic of a magnetic rigid-disk drive and MR type picoslider and cross-sectional schematics of a magnetic thin-film rigid disk and a metal evaporated (ME) tape

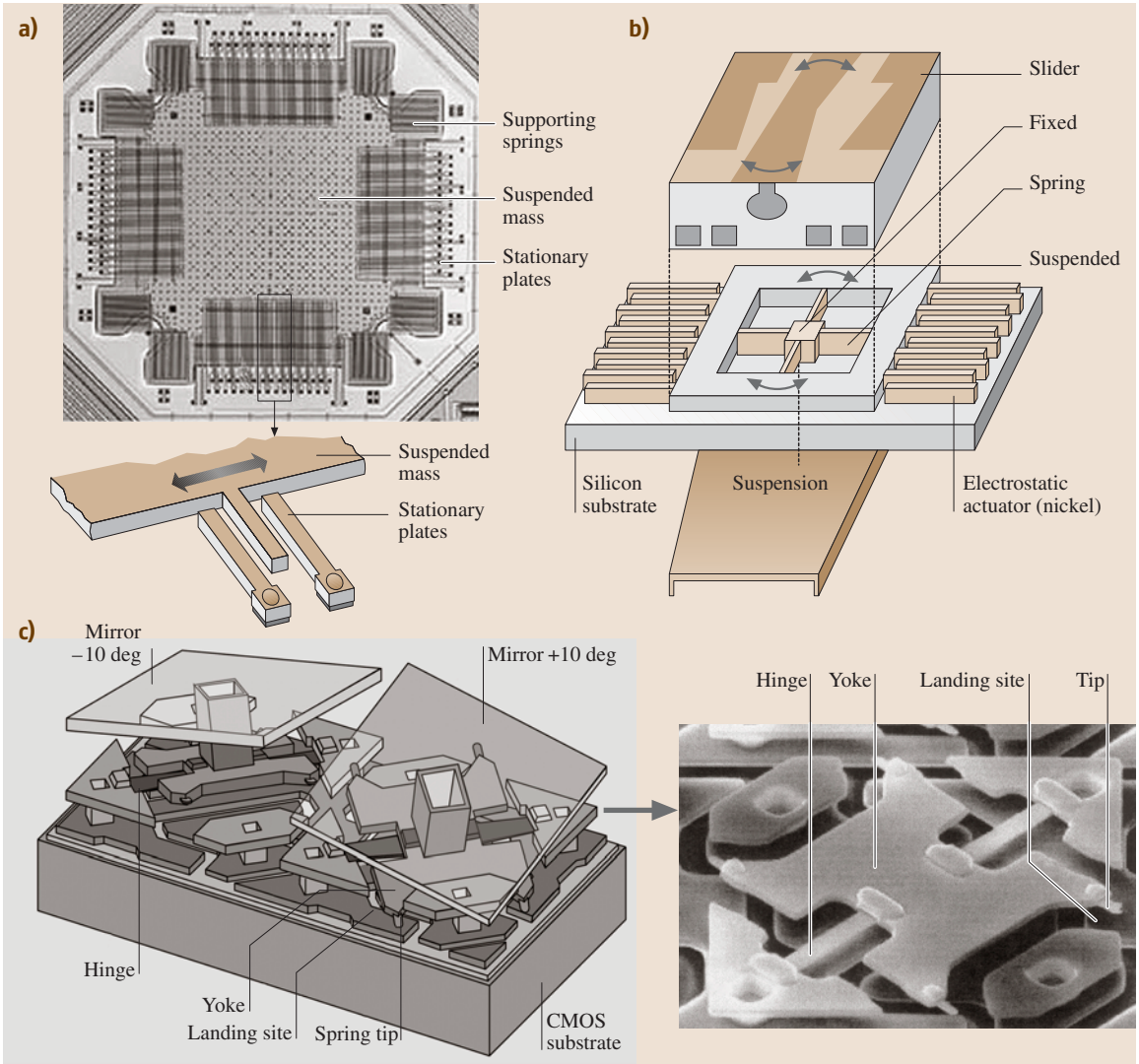


Fig. 26.3a–c Schematics of **(a)** a capacitive-type silicon accelerometer for automotive sensory applications, **(b)** digital micrometer devices for high-projection displays, and **(c)** polysilicon rotary microactuator for magnetic disk drives

σ bonds, and the remaining two electrons left in orthogonal p_y and p_z orbitals to form weak π bonds). Short-range order controls the properties of amorphous materials and coatings. Hard amorphous carbon (a-C) coatings, commonly known as diamond-like carbon or DLC (implying high hardness) coatings, are a class of coatings that are mostly metastable amorphous materials, but include a micro- or nanocrystalline phase. The coatings are random networks of covalently bonded carbon in hybridized tetragonal (sp^3) and trigonal (sp^2) local coordination with some of the bonds terminated by hydrogen. These coatings have been successfully deposited by a variety of vacuum deposition techniques on a variety of substrates at or near room temperature. These coatings generally reproduce substrate topography and do not require any post-finishing. However, these coatings mostly adhere best on silicon substrates. The best adhesion is obtained on substrates that form carbides, e.g., Si, Fe, and Ti. Based on depth profile analyses using Auger and XPS of DLC coatings deposited on silicon

substrates, it has been reported that a substantial amount of silicon carbide (on the order of 5–10 nm in thickness) is present at the carbon-silicon interface for the coatings with good adhesion and high hardness (e.g., [26.9]). For good adhesion of DLC coatings to other substrates, in most cases, an interlayer of silicon is required except for cathodic arc deposited coatings.

There is significant interest in DLC coatings because of their unique combination of desirable properties. These properties include high hardness and wear resistance, chemical inertness to both acids and alkalis, lack of magnetic response, and an optical band gap ranging from zero to a few eV, depending upon the deposition conditions. These are used in a wide range of applications, including tribological, optical, electronic, and biomedical applications [26.1, 10, 11]. The high hardness, good friction and wear properties, versatility in deposition and substrates, and no requirements of post-finishing make them very attractive for tribological applications. Two primary examples include overcoats for magnetic media (thin-film disks and ME tapes) and MR type magnetic heads for magnetic storage devices, Fig. 26.2 [26.12–20] and the emerging field of microelectromechanical systems, Fig. 26.3 [26.21–24]. The largest industrial application of the family of amorphous carbon coatings, typically deposited by DC/RF magnetron sputtering, plasma-enhanced chemical vapor deposition, or ion beam deposition techniques, is in magnetic storage devices. These are employed to protect against wear and corrosion, magnetic coatings on thin-film rigid disks and metal evaporated tapes, and the thin-film head structure of a read/write disk head (Fig. 26.2). To maintain low physical spacing between the magnetic element of a read/write head and the magnetic layer of a media, thicknesses ranging from 3 to 10 nm are employed. Mechanical properties affect friction wear and therefore need to be optimized. In 1998, Gillette introduced Mach 3 razor blades with ultrathin DLC coatings, which has the potential of becoming a very large industrial application. DLC coatings are also used in other commercial applications such as glass windows of supermarket laser barcode scanners and sunglasses. These coatings are actively pursued in microelectromechanical systems (MEMS) components [26.23].

In this chapter, a state-of-the-art review of recent developments in the chemical, mechanical, and tribological characterization of ultrathin amorphous carbon coatings is presented. An overview of the most commonly used deposition techniques is presented, and followed by typical chemical and mechanical characterization data and typical tribological data both from coupon level testing and functional testing.

26.1 Description of Commonly Used Deposition Techniques

The first hard amorphous carbon coatings were deposited by a beam of carbon ions produced in an argon plasma on room-temperature substrates, as reported by *Aisenberg* and *Chabot* [26.25]. Subsequent confirmation by *Spencer* et al. [26.26] led to the explosive growth of this field. Following the first work, several alternative techniques have been developed. The amorphous carbon coatings have been prepared by a variety of deposition techniques and precursors, including evaporation, DC, RF or ion beam sputtering, RF or DC plasma-enhanced chemical vapor deposition (PECVD), electron cyclotron resonance chemical vapor deposition (ECR-CVD), direct ion beam deposition, pulsed laser vaporization and vacuum arc, from a variety of carbon-bearing solids or gaseous source materials [26.1, 27]. Coatings with both graphitic and diamond-like properties have been produced. Evaporation and ion plating techniques have been used to produce coatings with graphitic properties (low hardness, high electrical conductivity, very low friction, etc.), and all techniques have been used to produce coatings with diamond-like properties.

The structure and properties of a coating are dependent upon the deposition technique and parameters. High-energy surface bombardment has been used to produce harder and denser coatings. It is reported that sp^3/sp^2 fractions are in the decreasing order for cathodic arc deposition, pulsed laser vaporization, direct ion beam deposition, plasma-enhanced chemical vapor deposition, ion beam sputtering, and DC/RF sputtering [26.12,28,29]. A common feature to these techniques is that the deposition is energetic, i. e., carbon species arrive with an energy significantly greater than that represented by the substrate temperature. The resultant coatings are amorphous in structure, with hydrogen content up to 50%, and display a high degree of sp^3 character. From the results of previous investigations, it has been proposed that deposition of sp^3-bonded carbon requires that the depositing species have kinetic energies on the order of 100 eV or higher, well above

Table 26.1 Summary of the most commonly used deposition techniques and the kinetic energy of depositing species and deposition rates

Deposition technique	Process	Kinetic energy (eV)	Deposition rate (nm/s)
Filtered cathodic arc (FCA)	Energetic carbon ions produced by a vacuum arc discharge between a graphite cathode and grounded anode	100–2,500	0.1–1
Direct ion beam (IB)	Carbon ions produced from methane gas in an ion source and accelerated toward a substrate	50–500	0.1–1
Plasma-enhanced chemical vapor deposition (PECVD)	Hydrocarbon species, produced by plasma decomposition of hydrocarbon gases (e.g., acetylene), are accelerated toward a DC-biased substrate	1–30	1–10
Electron cyclotron resonance plasma chemical vapor deposition (ECR-CVD)	Hydrocarbon ions, produced by plasma decomposition of ethylene gas in the presence of a plasma in electron cyclotron resonance condition, are accelerated toward a RF-biased substrate	1–50	1–10
DC/RF sputtering	Sputtering of graphite target by argon ion plasma	1–10	1–10

those obtained in thermal processes like evaporation (0–0.1 eV). The species must then be quenched into the metastable configuration via rapid energy removal. Excess energy, such as that provided by substrate heating, is detrimental to the achievement of a high sp^3 fraction. In general, a high fraction of the sp^3-bonded carbon atoms in an amorphous network results in a higher hardness [26.29–36]. The mechanical and tribological properties of a carbon coating depend on the sp^3/sp^2-bonded carbon ratio, the amount of hydrogen in the coating, and adhesion of the coating to the substrate, which are influenced by the precursor material, kinetic energy of the carbon species prior to deposition, deposition rate, substrate temperature, substrate biasing, and the substrate itself [26.29, 33, 35, 37–46]. The kinetic energies and deposition rates involved in selected deposition processes used in the deposition of DLC coatings are compared in Table 26.1 [26.1, 28].

In the studies by *Gupta* and *Bhushan* [26.12, 47], *Li* and *Bhushan* [26.48, 49], and *Sundararajan* and *Bhushan* [26.50], DLC coatings ranging typically in thickness from 3.5 nm to 20 nm were deposited on single-crystal silicon, magnetic Ni-Zn ferrite, and Al_2O_3-TiC substrates (surface roughness $\approx$ 1–3 nm RMS) by filtered cathodic arc (FCA) deposition, (direct) ion beam deposition (IBD), electron cyclotron resonance chemical vapor deposition (ECR-CVD), plasma-enhanced chemical vapor deposition (PECVD), and DC/RF planar, magnetron sputtering (SP) deposition techniques [26.51]. In this chapter, we will limit the presentation of data of coatings deposited by FCA, IBD, ECR-CVD, and SP deposition techniques.

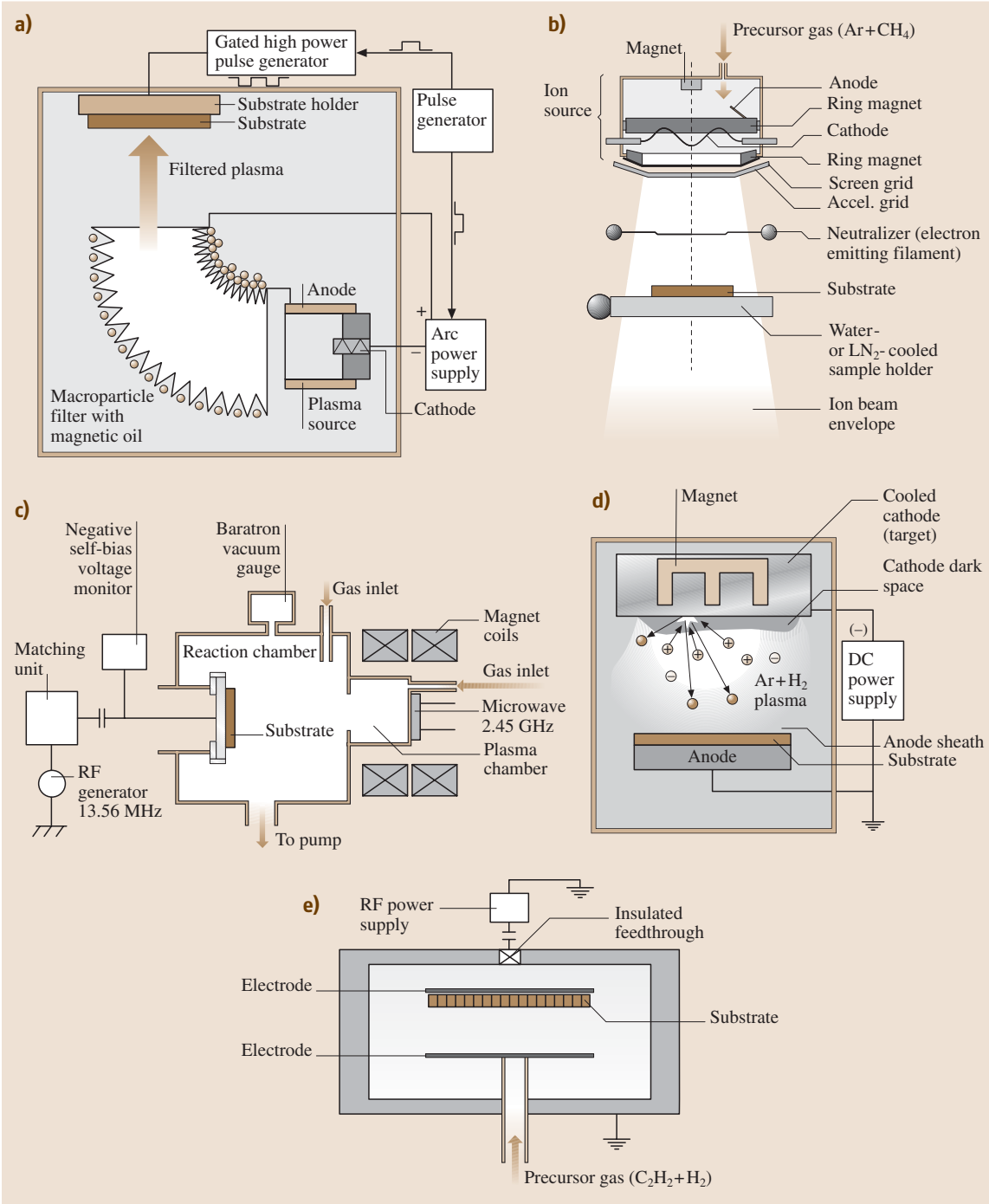


Fig. 26.4a–e Schematic diagrams of deposition by **(a)** filtered cathodic arc deposition, **(b)** ion beam deposition, **(c)** electron cyclotron resonance chemical vapor deposition (ECR-CVD), **(d)** DC planar magnetron sputtering, and **(e)** plasma-enhanced chemical vapor deposition (PECVD)

26.1.1 Filtered Cathodic Arc Deposition Technique

In the filtered cathodic arc deposition of carbon coating [26.29, 52–59], a vacuum arc plasma source is used to form carbon film. In the FCA technique used by *Bhushan* et al. (e.g., [26.12]), energetic carbon ions are produced by a vacuum arc discharge between a planar graphite cathode and grounded anode, Fig. 26.4a. The cathode is a 6-mm-diameter high-density graphite disk mounted on a water-cooled copper block. The arc is driven at an arc current of 200 A, arc duration of 5 ms, and arc repetition rate of 1 Hz. The plasma beam is guided by a magnetic field that transports current between the electrodes to form tiny, rapidly moving spots on the cathode surface. The source is coupled to a 90° bent magnetic filter to remove the macroparticles produced concurrently with the plasma in the cathode spots. The ion current density at the substrate is in the range of 10–50 mA/cm^2. The base pressure is less than 10^{-4} Pa. Compared with electron beam evaporation with auxiliary discharge, much higher plasma density is achieved with the aid of powerful arc discharge. In this process, the cathodic material suffers a complicated transition from the solid phase to an expanding, nonequilibrium plasma via liquid and dense, equilibrium non-ideal plasma phases [26.58]. The carbon ions in the vacuum arc plasma have a direct kinetic energy of 20–30 eV. The high voltage pulses are applied to the substrate mounted on a water-cooled sample holder, and ions are accelerated through the sheath and arrive at the substrate with an additional energy given by the potential difference between the plasma and the substrate. The substrate holder is pulse biased to a negative voltage up to −2 kV with a pulse duration of 1 μs. The negative biasing of −2 kV corresponds to 2 keV kinetic energy of the carbon ions. The use of a pulse bias instead of a DC bias has advantages of applying a much higher voltage and building a surface potential on a nonconducting film. The energy of the ions is varied during the deposition. For the first 10% of the deposition the substrates are pulsed biased to −2 keV with a pulse duty cycle of 25%, i. e., for 25% of the time the energy is 2 keV, for the remaining 75% it is 20 eV, which is the "natural" energy of carbon ions in a vacuum discharge. For the last 90% of the deposition the pulsed bias voltage is reduced to −200 eV with a pulse bias duty cycle of 25%, i. e., the energy is 200 eV for 25% and 20 eV for 75% of the deposition. The high energy at the beginning leads to a good intermixing and adhesion of the films, whereas the lower energy at the later stage leads to hard films. Under the conditions described, the deposition rate at the substrate is about 0.1 nm/s, which is slow. Compared with most gaseous plasma, the cathodic arc plasma is nearly fully ionized, and the ionized carbon atoms have high kinetic energy of carbon ions which help achieving a high fraction of sp^3-bonded carbon ions, which in turn result in a high hardness and higher interfacial adhesion. *Cuomo* et al. [26.42] have reported that based on electron energy loss spectroscopy (EELS) analysis, the sp^3-bonded carbon fraction of cathodic arc coating is 83% compared to 38% for the ion beam sputtered carbon. These coatings are reported to be *nonhydrogenated*.

This technique does not require an adhesion underlayer for non-silicon substrates. However, adhesion of the DLC coatings on the electrically insulating substrate is poor, as negative pulse biasing forms an electrical sheath that accelerates depositing ions to the substrate and enhances the adhesion of the coating to the substrate with associated ion implantation. It is difficult to build potential on an insulating substrate, and lack of biasing results in poor adhesion.

26.1.2 Ion Beam Deposition Technique

In the direct ion beam deposition of carbon coating [26.60–64], as used by *Bhushan* et al. (e.g., [26.12]), the carbon coating is deposited from an accelerated carbon ion beam. The sample is pre-cleaned by ion etching. For the case of non-silicon substrates, a 2–3-nm-thick amorphous silicon adhesion layer is deposited by ion beam sputtering using an ion beam of a mixture of methane and argon at 200 V. For the carbon deposition, the chamber is pumped to about 10^{-4} Pa, and methane gas is fed through the cylindrical ion source and is ionized by energetic electrons produced by a hot-wire filament, Fig. 26.4b. Ionized species then pass through a grid with a bias voltage of about 50 eV, where they gain a high acceleration energy and reach a hot-wire filament, emitting thermionic electrons that neutralize the incoming ions. The discharging of ions is important when insulating ceramics are used as substrates. The species are then deposited on a water-cooled substrate. Operating conditions are adjusted to give an ion beam with an acceleration energy of about 200 eV and a current density of about 1 mA/cm^2. At this operating condition, the deposition rate is about 0.1 nm/s, which is slow. Incidentally, tough and soft coatings are deposited at a high acceleration energy of about 400 eV and at a deposition rate of about 1 nm/s. The ion beam deposited carbon coatings are reported to be hydrogenated (30–40 at% hydrogen).

26.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique

ECR plasma's lack of electrodes and ability to create high densities of charged and excited species at low pressures ($\leq 10^{-4}$ Torr) make it an attractive processing discharge in the coating depositions [26.65]. In the ECR-CVD deposition process of carbon coating described by *Suzuki* and *Okada* [26.66] and used by *Li* and *Bhushan* [26.48, 49] and *Sundararajan* and *Bhushan* [26.50], microwave power is generated by a magnetron operating in continuous mode at a frequency of 2.45 GHz, Fig. 26.4c. The plasma chamber functions as a microwave cavity resonator. The magnetic coils arranged around the plasma chamber generate a magnetic field of 875 G, necessary for electron cyclotron resonance condition. The substrate is placed on a stage that is connected capacitively to a 13.56 MHz RF generator. The process gas is introduced into the plasma chamber and the hydrocarbon ions generated are accelerated by a negative self-bias voltage, which is generated by applying RF power to the substrate. Both the substrate stage and plasma chamber are water-cooled. The process gas used is 100% ethylene and its flow rate is held constant at 100 sccm. The microwave power is 100–900 W. The RF power is 30–120 W. The pressure during deposition is kept close to the optimum value of 5.5×10^{-3} Torr. Before the deposition, the substrates are cleaned using Ar ions generated in the ECR plasma chamber.

26.1.4 Sputtering Deposition Technique

In DC planar magnetron sputtered carbon coating [26.13, 33, 37, 40, 67–71], the carbon coating is deposited by the sputtering of graphite target with Ar ion plasma. In the glow discharge, positive ions from the plasma strike the target with sufficient energy to dislodge the atoms by momentum transfer, which are intercepted by the substrate. First, an about 5-nm-thick amorphous silicon adhesion layer is deposited by sputtering if the deposition is to be carried out on a non-silicon surface. In the process used by *Bhushan* et al. (e.g., [26.12]), the coating is deposited by the sputtering of a graphite target with Ar ion plasma at 300 W power for a 200-mm-diameter target and pressure of about 0.5 Pa (6 mTorr), Fig. 26.4d. Plasma is generated by applying a DC potential between the substrate and a target. *Bhushan* et al. [26.35] reported that sputtered carbon coating contains about 35 at% hydrogen. Hydrogen comes from the hydrocarbon contaminants present in the deposition chamber. To produce hydrogenated carbon coating with a larger concentration of hydrogen, deposition is carried out in Ar and hydrogen plasma.

26.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique

In the RF-PECVD deposition of carbon coating, as used by *Bhushan* et al. (e.g., [26.12]), carbon coating is deposited by adsorption of most free radicals of hydrocarbon to the substrate and chemical bonding to other atoms on the surface. The hydrocarbon species are produced by the RF plasma decomposition of hydrocarbon precusors such as acetylene (C_2H_2), Fig. 26.4e [26.27, 69, 72–75]. Instead of requiring thermal energy in thermal CVD, the energetic electrons in the plasma (at pressures ranging from 1 to 5×10^2 Pa, typically less than 10 Pa) can activate almost any reaction among the gases in the glow discharge at relatively low substrate temperatures ranging from 100 to 600 °C (typically less than 300 °C). To deposit the coating on non-silicon substrates, an about 4-nm-thick amorphous silicon adhesion layer, used to improve adhesion, is first deposited under similar conditions from a gas mixture of 1% silane in argon [26.76]. In the process used by Bhushan and coworkers [26.12], the plasma is sustained in a parallel-plate geometry by a capacitive discharge at 13.56 MHz, at a surface power density on the order of 100 mW/cm^2. The deposition is performed at a flow rate on the order of 6 sccm and a pressure on the order of 4 Pa (30 mTorr) on a cathode-mounted substrate maintained at a substrate temperature of 180 °C. The cathode bias is held fixed at about −120 V with an external DC power supply attached to the substrate (powered electrode). The carbon coatings deposited by PECVD usually contain hydrogen up to 50% [26.35, 77].

26.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties

The chemical structure and properties of amorphous carbon coatings are a function of deposition conditions. It is important to understand the relationship of the chemical structure of amorphous carbon coatings to the properties in order to define useful deposition parameters. Amorphous carbon films are metastable phases formed when carbon particles are condensed on a substrate. The prevailing atomic arrangement in the DLC coatings is amorphous or quasi-amorphous with small diamond (sp^3), graphite (sp^2), and other unidentifiable micro- or nanocrystallites. The coating is dependent upon the deposition process and its conditions contain varying amounts of sp^3/sp^2 ratio and hydrogen. The sp^3/sp^2 ratio of DLC coatings ranges typically from 50% to close to 100% with an increase in hardness with the sp^3/sp^2 ratio. Most DLC coatings, except those produced by filtered cathodic arc, contain from a few to about 50 at% hydrogen. Sometimes hydrogen and nitrogen are deliberately added to produce hydrogenated (a-C:H) and nitrogenated amorphous carbon (a-C:N) coatings, respectively. Hydrogen helps to stabilize sp^3 sites (most of the carbon atoms attached to hydrogen have a tetrahedral structure); therefore, the sp^3/sp^2 ratio for hydrogenated carbon is higher [26.30]. Optimum sp^3/sp^2 in a random covalent network composed of sp^3 and sp^2 carbon sites (N_{sp^2} and N_{sp^3}) and hydrogen is [26.30]:

$$\frac{N_{sp^3}}{N_{sp^2}} = \frac{6X_H - 1}{8 - 13X_H}\,, \tag{26.1}$$

where X_H is the atom fraction of hydrogen. The hydrogenated carbon has a larger optical band gap, higher electrical resistivity (semiconductor), and a lower optical absorption or high optical transmission. The hydrogenated coatings have a lower density, probably because of the reduction of cross-linking due to hydrogen incorporation. However, hardness decreases with an increase of the hydrogen, even though the proportion of sp^3 sites increases (that is, as the local bonding environment becomes more diamond-like) [26.78, 79]. It is speculated that the high hydrogen content introduces frequent terminations in the otherwise strong 3-D network, and hydrogen increases the soft polymeric component of the structure more than it enhances the cross-linking sp^3 fraction.

A number of investigations have been performed to identify the microstructure of amorphous carbon films using a variety of techniques such as Raman spectroscopy, EELS, nuclear magnetic resonance, optical measurements, transmission electron microscopy, and X-ray photoelectron spectroscopy [26.33]. The structure of diamond-like amorphous carbon is amorphous or quasi-amorphous with small graphitic (sp^2) and tetrahedrally coordinated (sp^3) and other unidentifiable nanocrystallites (typically on the order of a couple nm, randomly oriented) [26.33, 80, 81]. These studies indicate that the chemical structure and physical properties of the coatings are quite variable, depending on the deposition techniques and film growth conditions. It is clear that both sp^2 and sp^3-bonded atomic sites are incorporated in diamond-like amorphous carbon coatings and that the physical and chemical properties of the coatings depend strongly on their chemical bonding and microstructure. Systematic studies have been conducted to carry out chemical characterization and investigate how the physical and chemical properties of amorphous carbon coatings vary as a function of deposition parameters (e.g., [26.33, 35, 40]). EELS and Raman spectroscopy are commonly used to characterize the chemical bonding and microstructure. Hydrogen concentration of the coatings is obtained by means of forward recoil spectrometry (FRS). A variety of physical properties of the coatings relevant to tribological performance are measured.

To present the typical data obtained for characterization of typical amorphous carbon coatings and their relationships to physical properties, we present data on several sputtered coatings, RF-PECVD amorphous carbon and microwave-PECVD (MPECVD) diamond coatings [26.33, 35, 40]. The sputtered coatings were DC magnetron sputtered at a chamber pressure of 10 mTorr under sputtering power densities of 0.1 and 2.1 W/cm^2 in a pure Ar plasma, labeled as W1 and W2, respectively. They were prepared at a power density of 2.1 W/cm^2 with various hydrogen fractions of 0.5, 1, 3, 5, 7 and 10% of Ar/H, the gas mixtures labeled as H1, H2, H3, H4, H5, and H6, respectively.

26.2.1 EELS and Raman Spectroscopy

EELS and Raman spectra of four sputtered (W1, W2, H1, and H3) and one PECVD carbon samples were obtained. Figure 26.5 shows the EELS spectra of these carbon coatings. EELS spectra for bulk diamond and polycrystalline graphite in an energy range up to 50 eV are also shown in Fig. 26.5. One prominent peak is seen

at 35 eV in diamond, while two peaks are seen at 27 eV and 6.5 eV in graphite, which are called $(\pi+\sigma)$ and (π) peaks, respectively. These peaks are produced by the energy loss of transmitted electrons to the plasmon oscillation of the valence electrons. The $\pi+\sigma$ peak in each coating is positioned at a lower energy region than that of graphite. The π peaks in the W series and PECVD samples are also seen at a lower energy region than that of the graphite. However, the π peaks in the H-series are comparable to or higher than that of graphite (see Table 26.2). The plasmon oscillation frequency is proportional to the square root of the corresponding electron density to a first approximation. Therefore, the samples in the H-series most likely have a higher density of π electrons than the other samples.

Amorphous carbon coatings contain (mainly) a mixture of sp^2- and sp^3-bonds, even though there is some evidence for the presence of sp-bonds as well [26.82]. The PECVD coatings and the H-series coatings in this study have nearly the same mass density, as seen in Table 26.4, to be presented later, but the former have a lower concentration of hydrogen (18.1%) than the H-series (35–39%), as seen in Table 26.3, to be presented later. The relatively low energy position of π peaks of PECVD coatings, compared to those of the H-series, indicates that the PECVD coatings contain a higher fraction of sp^3-bonds than the sputtered hydrogenated carbon coatings (H-series).

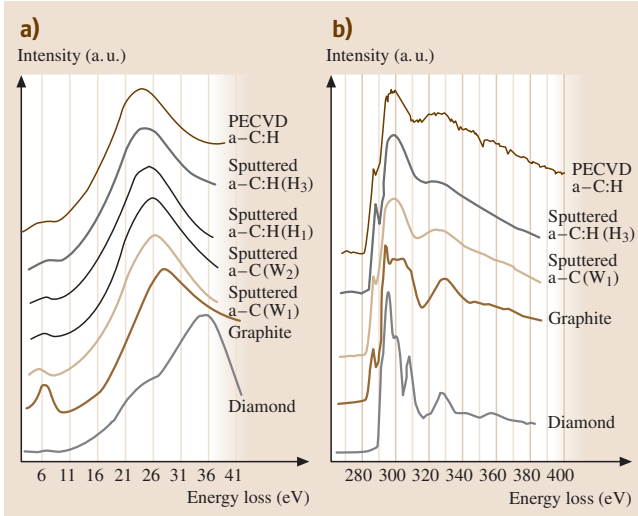


Fig. 26.5 **(a)** Low energy, and **(b)** high energy EELS of DLC coatings produced by DC magnetron sputtering and RF-PECVD techniques. Data for bulk diamond and polycrystalline graphite are included for comparison [26.35]

Figure 26.5b shows the EELS spectra associated with the inner-shell (K-shell) ionization. Again, the spectra for diamond and polycrystalline graphite are included for comparison. Sharp peaks are observed at

Table 26.2 Experimental results from EELS and Raman spectroscopy [26.35]

Sample	EELS peak position		Raman peak position		Raman FWHM[a]		I_D/I_G^d
	π (eV)	$\pi+\sigma$ (eV)	G-band[b] (cm^{-1})	D-band[c] (cm^{-1})	G-band (cm^{-1})	D-band (cm^{-1})	
Sputtered *a*-C coating (W1)	5.0	24.6	1541	1368	105	254	2.0
Sputtered *a*-C coating (W2)	6.1	24.7	1560	1379	147	394	5.3
Sputtered *a*-C:H coating (H1)	6.3	23.3	1542	1334	95	187	1.6
Sputtered *a*-C:H coating (H3)	6.7	22.4	e	e	e	e	e
PECVD *a*-C:H coating	5.8	24.0	1533	1341	157	427	1.5
Diamond coating	...	...	1525[f]	1333[g]	...	8[g]	...
Graphite (for reference)	6.4	27.0	1580	1358	37	47	0.7
Diamond (for reference)	...	37.0	...	1332[g]	...	2[g]	...

[a] Full width at half maximum width
[b] Peak associated with sp^2 "graphite" carbon
[c] Peak associated with sp^2 "disordered" carbon (not sp^3-bonded carbon)
[d] Intensity ratio of the D-band to the G-band
[e] Fluorescence
[f] Includes D and G band, signal too weak to analyze
[g] Peak position and width for diamond phonon

Table 26.3 Experimental results of FRS analysis [26.35]

Sample	Ar/H ratio	C (at% ± 0.5)	H (at% ± 0.5)	Ar (at% ± 0.5)	O (at% ± 0.5)
Sputtered *a*-C coating (W2)	100/0	90.5	9.3	0.2	...
Sputtered *a*-C:H coating (H2)	99/1	63.9	35.5	0.6	...
Sputtered *a*-C:H coating (H3)	97/3	56.1	36.5	...	7.4
Sputtered *a*-C:H coating (H4)	95/5	53.4	39.4	...	7.2
Sputtered *a*-C:H coating (H5)	93/7	58.2	35.4	0.2	6.2
Sputtered *a*-C:H coating (H6)	90/10	57.3	35.5	...	7.2
PECVD *a*-C:H coating	99.5% CH_4	81.9	18.1	...	...
Diamond coating	H_2-1 mole % CH_4	94.0	6.0	...	...

Table 26.4 Experimental results of physical properties [26.35]

Sample	Mass density (g/cm^3)	Nano-hardness (GPa)	Elastic modulus (GPa)	Electrical resistivity (Ohm-cm)	Compressive residual stress (GPa)
Sputtered *a*-C coating (W1)	2.1	15	141	1300	0.55
Sputtered *a*-C:H coating (W2)	1.8	14	136	0.61	0.57
Sputtered *a*-C:H coating (H1)	...	14	96	...	> 2
Sputtered *a*-C:H coating (H3)	1.7	7	35	$> 10^6$	0.3
PECVD *a*-C:H coating	1.6–1.8	33–35	~ 200	$> 10^6$	1.5–3.0
Diamond coating	...	40–75	370–430	...	...
Graphite (for reference)	2.267	Soft	9–15	5×10^{-5a}, 4×10^{-3b}	0
Diamond (for reference)	3.515	70–102	900–1050	10^7–10^{20}	0

[a] Parallel to layer planes
[b] Perpendicular to layer planes

285.5 eV and 292.5 eV in graphite, while no peak is seen at 285.5 eV in diamond. The general features of the K-shell EELS spectra for the sputtered and PECVD carbon samples resemble those of graphite, but with the higher energy features smeared. The observation of the peak at 285.5 eV in the sputtered and PECVD coatings also indicates the presence of sp^2-bonded atomic sites in the coatings. All these spectra peak at 292.5 eV, similar to the spectra of graphite, but the peak in graphite is sharper.

Raman spectra from samples W1, W2, H1, and PECVD are shown in Fig. 26.6. Raman spectra could not be observed in specimens H2 and H3 due to high flourescence signals. The Raman spectra of single crystal diamond and polycrstalline graphite are also shown for comparison in Fig. 26.6. The results of the spectral fits are summarized in Table 26.2. We will focus on the G-band position, which has been shown to be related to the fraction of sp^3-bonded sites. Increasing the power density in the amorphous carbon coatings (W1 and W2) results in a higher G-band frequency, implying a smaller fraction of sp^3-bonding in W2 than in W1. This is consistent with higher density of W1. H1 and PEVCD have still lower G-band positions than W1, implying an even higher fraction of sp^3-bonding, which is presumably caused by the incorporation of H atoms into the lattice. The high hardness of H3 might be attributed to efficient sp^3 cross-linking of small, sp^2-ordered domains.

The Raman spectrum of a MPECVD diamond coating is shown in Fig. 26.6. The diamond Raman peak is at 1,333 cm^{-1} with a line width of 7.9 cm^{-1}. There is a small broad peak around 1,525 cm^{-1}, which is attributed to a small amount of a-C:H. This impurity peak is not intense enough to fit to separate G- and D-bands. The diamond peak frequency is very close to that of natural diamond (1,332.5 cm^{-1}, e.g. Fig. 26.6), indicating that the coating is not under stress [26.83]. The large line width compared to that of natural diamond (2 cm^{-1})

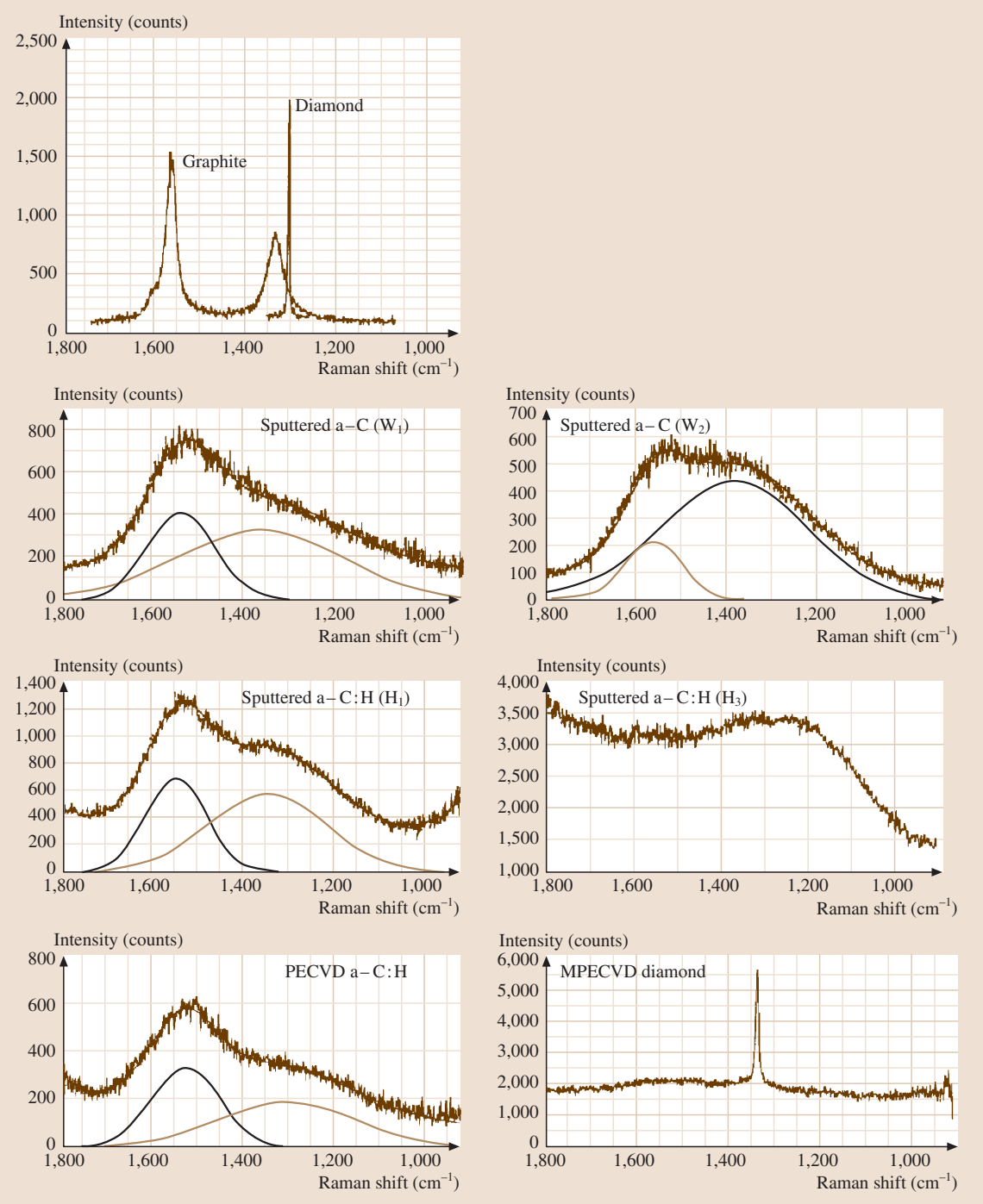


Fig. 26.6 Raman spectra of DLC coatings produced by DC magnetron sputtering and RF-PECVD techniques and a diamond film produced by MPECVD technique. Data for bulk diamond and microcrystalline graphite are included for comparison [26.35]

indicates that the microcrystallites likely have a high concentration of defects [26.84].

26.2.2 Hydrogen Concentrations

FRS analysis of six sputtered (W2, H2, H3, H4, H5, and H6) coatings, one PECVD coating, and one diamond coating was performed. Figure 26.7 shows an overlay of the spectra from the six sputtered samples. Similar spectra were obtained from the PECVD and the diamond films. Table 26.3 shows the H and C fractions, as well as the amount of impurities (Ar and O) in the films in atomic %. Most apparent is the large fraction of H in the sputtered films. Regardless of how much H_2 is in the Ar sputtering gas, the H content of the coatings is about the same, ∼ 35 at%. Interestingly, there is still ∼ 10% H present in the coating sputtered in pure Ar (W2). It is interesting to note that Ar is present only in coatings grown under low (< 1%) H content in the Ar carrier gas. The presence of O in the coatings, combined with the fact that the coatings were prepared approximately nine months before the FRS analysis, caused suspicion that they had absorbed water vapor, and that this may be the cause for the H peak in specimen W2.

All samples were annealed for 24 h at 250 °C in a flowing He furnace and then reanalyzed. Surprisingly, the H content of all coatings measured increased slightly, even though the O content decreased, and W2 still had a substantial amount of H_2. This slight increase in H concentration is not understood. However, since H concentration did not decrease with the oxygen as a result of annealing, it suggests that high H concentration is not due to adsorbed water vapor. The PECVD film has more H (∼ 18%) than the sputtered films initially, but after annealing it has the same fraction as specimen W2, the film sputtered in pure Ar. The diamond film has the smallest amounts of hydrogen, as seen in Table 26.3.

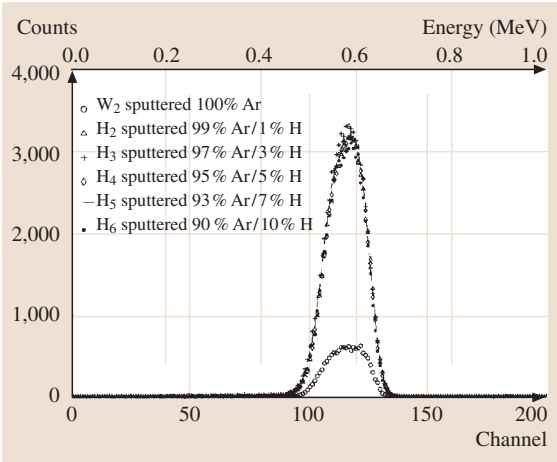


Fig. 26.7 FRS spectra of six DLC coatings produced by DC magnetron sputtering [26.35]

26.2.3 Physical Properties

Physical properties of the four sputtered (W1, W2, H1, and H3) coatings, one PECVD coating, one diamond coating, and bulk diamond and graphite are presented in Table 26.4. The hydrogenated carbon and the diamond coatings have very high resistivity compared to unhydrogenated carbon coatings. It appears that unhydrogenated carbon coatings have a higher density than the hydrogenated carbon coatings, although both groups are less dense than graphite. The density depends on the deposition technique and the deposition parameters. It appears that unhydrogenated sputtered coatings deposited at low power exhibit the highest density. Nanohardness of hydrogenated carbon is somewhat lower than that of the unhydrogenated carbon coatings. PECVD coatings are significantly harder than sputtered coatings. The nanohardness and modulus of elasticity of the diamond coatings are very high compared to that of DLC coatings, even though the hydrogen content is similar. The compressive residual stresses of the PECVD coatings are substantially higher than those of sputtered coatings, which is consistent with the hardness results.

Figure 26.8a shows the effect of hydrogen in the plasma on the residual stresses and the nanohardness for sputtered coatings W2 and H1 to H6. The coatings made with H_2 flow between 0.5 and 1.0% delaminate very quickly, even when only a few tens of nm thick. In pure Ar and at H_2 flows greater than 1%, the coatings appear to be more adhesive. The tendency of some coatings to delaminate can be caused by intrinsic stress in the coating, which is measured by substrate bending. All of the coatings in the figure are in compressive stress. The maximum stress occurs between 0 and 1% H_2 flow, but the stress cannot be quantified in this range because the coatings instantly delaminate upon exposure to air. At higher hydrogen concentrations the stress gradually diminishes. A generally decreasing trend is observed in the hardness of the coatings as hydrogen content increased. The hardness decreases slightly, going from 0% H_2 to 0.5% H_2, and then decreases sharply. These results are probably lower than the true values because of local delamination around the indentation point. This is especially likely for the 0.5% and 1.0% coatings where

delamination is visually apparent, but may also be true to a lesser extent for the other coatings. Such an adjustment would bring the hardness profile into closer correlation with the stress profile. *Weissmantel* et al. [26.68] and *Scharff* et al. [26.85] observed a downturn in hardness for high bias and low pressure of hydrocarbon gas in ion plated carbon coating, and, therefore, presumably low hydrogen content in support of the above contention.

Figure 26.8b shows the effect of sputtering power (with no hydrogen added in the plasma) on the residual stresses and nanohardness for various sputtered coatings. As the power decreases, compressive stress does not seem to change while nanohardness slowly increases. The rate of change becomes more rapid at very low power levels.

The addition of H_2 during sputtering of carbon coatings increases H concentration in the coating. Hydrogen causes the character of the C-C bonds to shift from sp^2 to sp^3, and the rising number of C-H bonds, which ultimately relieves stress and produces a softer "polymer-like" material. Low power deposition, like the presence of hydrogen, appears to stabilize the formation of sp^3 C-C bonds, increasing hardness. These coatings have relieved stress and led to better adhesion. An increase in temperature during deposition at high power density results in graphitization of the coating material, responsible for a decrease in hardness with an increase in power density. Unfortunately, low power also means impractically low deposition rates.

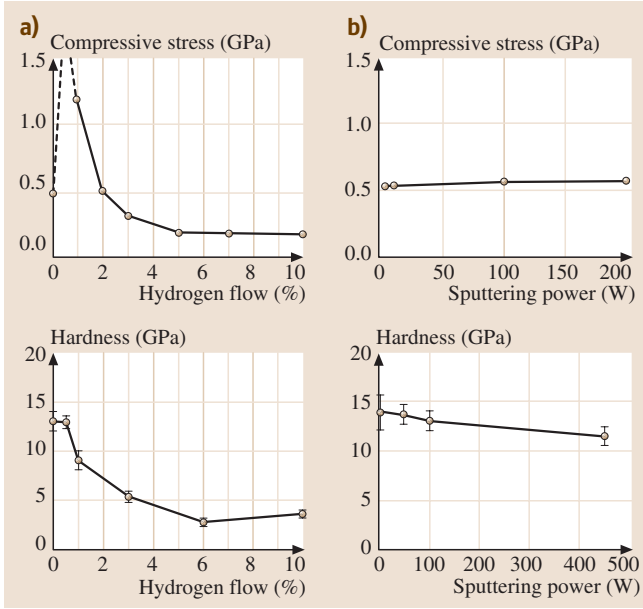


Fig. 26.8a,b Residual compressive stresses and nanohardness **(a)** as a function of hydrogen flow rate; sputtering power is 100 W and target diameter is 75 mm (power density = 2.1 W/cm^2), and **(b)** as a function of sputtering power over a 75-mm-diameter target with no hydrogen added in the plasma [26.40]

26.2.4 Summary

Based on the EELS and Raman data, all DLC coatings have both sp^2 and sp^3 bondings. The sp^2/sp^3 bonding ratio depends on the deposition techniques and parameters. The DLC coatings deposited by sputtering and PECVD contain significant concentrations of hydrogen, while the diamond coating contains only small amounts of hydrogen impurity. Sputtered coatings with no deliberate addition of hydrogen in the plasma contain a significant amount of hydrogen. Regardless of how much hydrogen is in the Ar sputtering gas, the hydrogen content of the coatings increases initially with no further increase.

Hydrogen flow and sputtering power density affect the mechanical properties of these coatings. Maximum compressive residual stress and hardness occur between 0 and 1% hydrogen flow, resulting in rapid delamination. Low sputtering power moderately increases hardness while relieving residual stress.

26.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques

26.3.1 Micromechanical Characterization

Common mechanical characterizations include measurement of hardness and elastic modulus, fracture toughness, fatigue life, and scratch and wear testing. Nanoindentation and atomic force microscopy (AFM) are used for mechanical characterization of ultrathin films.

Hardness and elastic modulus are calculated from the load displacement data obtained by nanoindentation at loads ranging typically from 0.2 to 10 mN using a commercially available nanoindenter [26.23,86].

This instrument monitors and records the dynamic load and displacement of the three-sided pyramidal diamond (Berkovich) indenter during indentation. For the fracture toughness measurement of ultrathin films ranging from 100 nm to a few μm, a nanoindentation-based technique is used in which through-thickness cracking in the coating is detected from a discontinuity observed in the load-displacement curve and energy released during the cracking is obtained from the curve [26.87–89]. Based on the energy released, fracture mechanics analysis is then used to calculate fracture toughness. An indenter with a cube-corner tip geometry is preferred because the through-thickness cracking of hard films can be accomplished at lower loads. In fatigue measurement, a conical diamond indenter having a tip radius of about one micron is used and load cycles of a sinusoidal shape are applied [26.90, 91]. The fatigue behavior of coatings is studied by monitoring the change in contact stiffness, which is sensitive to damage formation.

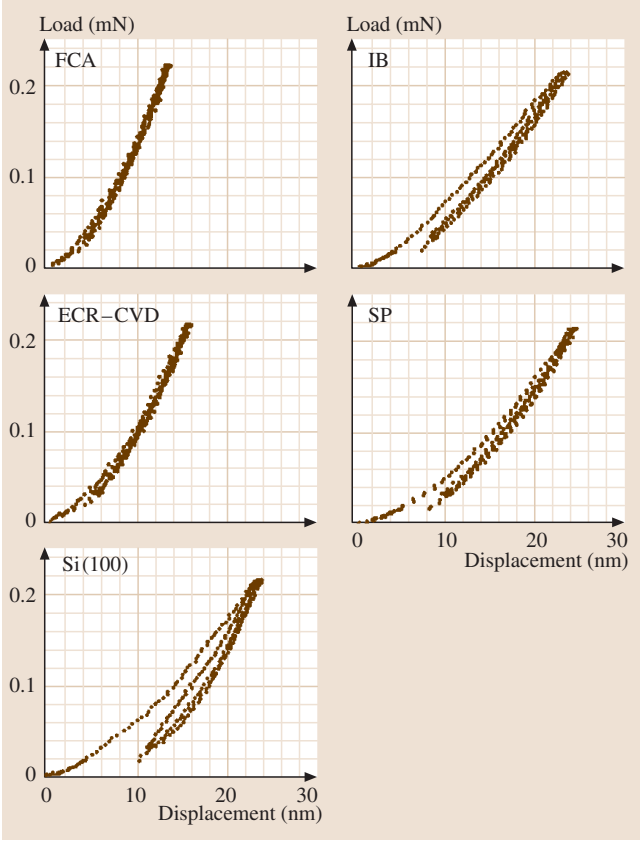


Fig. 26.9 Load versus displacement plots for various 100-nm-thick amorphous carbon coatings on single-crystal silicon substrate and bare substrate

Hardness and Elastic Modulus

For materials that undergo plastic deformation, high hardness and elastic modulus are generally needed for low friction and wear, whereas for brittle materials, high fracture toughness is needed [26.2, 3, 21]. DLC coatings used for many applications are hard and brittle, and values of hardness and fracture toughness need to be optimized.

Representative load-displacement plots of indentations made at 0.2 mN peak indentation load on 100-nm-thick DLC coatings deposited by the four deposition techniques on single-crystal silicon substrate are compared in Fig. 26.9. The indentation depths at the peak load range from about 18 to 26 nm, smaller than that of the coating thickness. Many of the coatings exhibit a discontinuity or pop-in marks in the loading curve, which indicate a sudden penetration of the tip into the sample. A nonuniform penetration of the tip into a thin coating possibly results from formation of cracks in the coating, formation of cracks at the coating-substrate interface, or debonding or delamination of the coating from the substrate.

The hardness and elastic modulus values at a peak load of 0.2 mN on the various coatings and single-crystal silicon substrate are summarized in Table 26.5 and Fig. 26.10 [26.47, 49, 89, 90]. Typical values for the peak and residual indentation depths ranged from 18 to 26 nm and 6 to 12 nm, respectively. The FCA coatings exhibit the highest hardness of 24 GPa and elastic modulus of 280 GPa of various coatings, followed by the ECR-CVD, IB, and SP coatings. Hardness and elastic modulus have been known to vary over a wide range with sp^3-to-sp^2 bonding ratio, which depends on the kinetic energy of the carbon species and amount of hydrogen [26.6, 30, 47, 92, 93]. The high hardness and elastic modulus of the FCA coatings are attributed to the high kinetic energy of carbon species involved in the FCA deposition [26.12, 47]. *Anders* et al. [26.57] also reported a high hardness, measured by nanoindentation, of about 45 GPa for cathodic arc carbon coatings. They observed a change in hardness from 25 to 45 GPa with pulsed bias voltage and bias duty cycle. The high hardness of cathodic arc carbon was attributed to the high percentage (more than 50%) of sp^3 bonding. *Savvides* and *Bell* [26.94] reported an increase in hardness from 12 to 30 GPa and an elastic modulus from 62 to 213 GPa with an increase of sp^3-to-sp^2 bonding ratio, from 3 to 6,

for a C:H coating deposited by low-energy ion-assisted unbalanced magnetron sputtering of a graphite target in an Ar-H_2 mixture.

Bhushan et al. [26.35] reported hardnesses of about 15 and 35 GPa and elastic moduli of about 140 and 200 GPa, measured by nanoindentation, for a-C:H coatings deposited by DC magnetron sputtering and RF-plasma-enhanced chemical vapor deposition techniques, respectively. The high hardness of RF-PECVD a-C:H coatings is attributed to a higher concentration of sp^3 bonding than in a sputtered hydrogenated a-C:H coating. Hydrogen is believed to play a crucial role in the bonding configuration of carbon atoms by helping to stabilize tetrahedra-coordination (sp^3 bonding) of carbon species. *Jansen* et al. [26.78] suggested that the incorporation of hydrogen efficiently passivates the dangling bonds and saturates the graphitic bonding to some extent. However, a large concentration of hydrogen in the plasma in sputter deposition is undesirable. *Cho* et al. [26.33] and *Rubin* et al. [26.40] observed that hardness decreased from 15 to 3 GPa with increased hydrogen content. *Bhushan* and *Doerner* [26.95] reported a hardness of about 10–20 GPa and an elastic modulus of about 170 GPa, measured by nanoindentation, for 100-nm-thick DC-magnetron sputtered a-C:H on the silicon substrate.

Residual stresses measured using a well-known curvature measurement technique are also presented in Table 26.5. The DLC coatings are under significant compressive internal stresses. Very high compressive stresses in FCA coatings are believed to be partly responsible for their high hardness. However, high stresses result in coating delamination and buckling. For this reason, the coatings that are thicker than about 1 μm have a tendency to delaminate from the substrates.

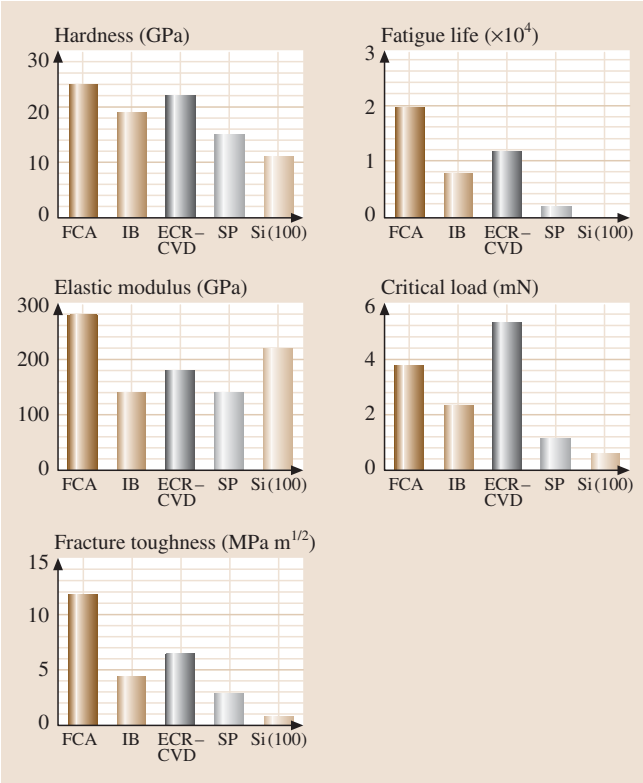


Fig. 26.10 Bar charts summarizing the data of various coatings and single-crystal silicon substrate. Hardness, elastic modulus, and fracture toughness were measured on 100-nm-thick coatings, and fatigue life and critical load during scratch were measured on 20-nm-thick coatings

Fracture Toughness

Representative load-displacement curves of indentations on the 400-nm-thick cathodic arc carbon coating on silicon at various peak loads are shown in Fig. 26.11. Steps are found in all curves, as shown by arrows in Fig. 26.11a. In the 30 mN SEM micrograph, in addition to several radial cracks, ring-like through-thickness cracking is observed with small lips of material overhanging the edge of indentation. The step at about 23 mN in the loading curves of indentations made at 30 and 100 mN peak indentation loads results from the ring-like through-thickness cracking. The step at 175 mN in the loading curve of the indentation made at 200 mN peak indentation load is caused by spalling and second ring-like through-thickness cracking.

Based on *Li* et al. [26.87], the fracture process progresses in three stages: (1) ring-like through-thickness cracks form around the indenter by high stresses in the contact area, (2) delamination and buckling occur around contact area at the coating/substrate interface by high lateral pressure, (3) second ring-like through-thickness cracks and spalling are generated by high bending stesses at the edges of the buckled coating, see Fig. 26.12a. In the first stage, if the coating under the indenter is separated from the bulk coating via the first ring-like through-thickness cracking, a corresponding step will be present in the loading curve. If discontinuous cracks form and the coating under the indenter is not separated from the remaining coating, no step appears in the loading curve, because the coating still supports the indenter and the indenter cannot suddenly advance into the material. In the second stage,

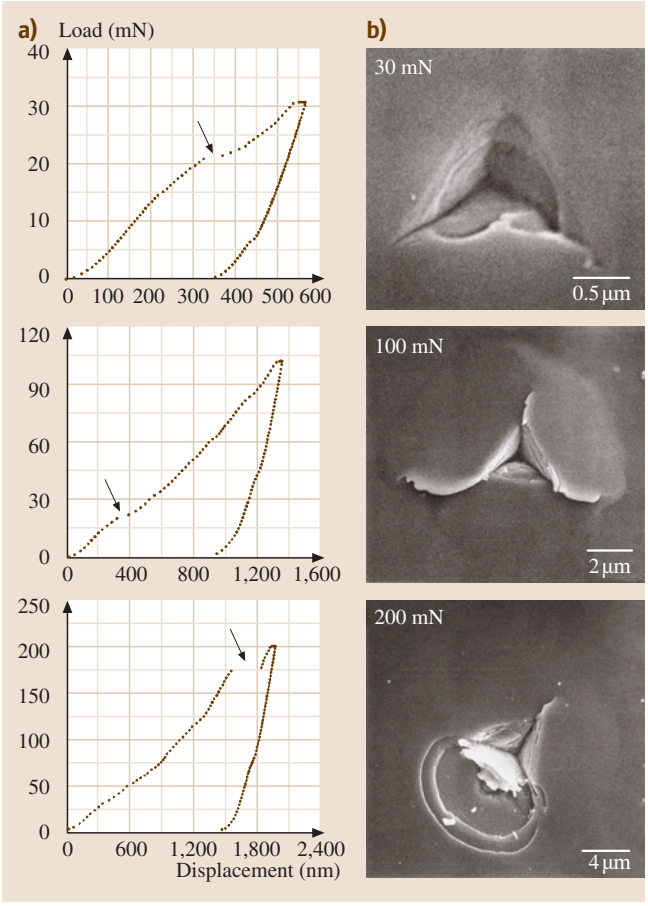


Fig. 26.11 **(a)** Load-displacement curves of indentations made at 30, 100, and 200 mN peak indentation loads using the cube corner indenter, and **(b)** the SEM micrographs of indentations on the 400-nm-thick cathodic arc carbon coating on silicon. *Arrows* indicate steps during loading portion of the load-displacement curve [26.87]

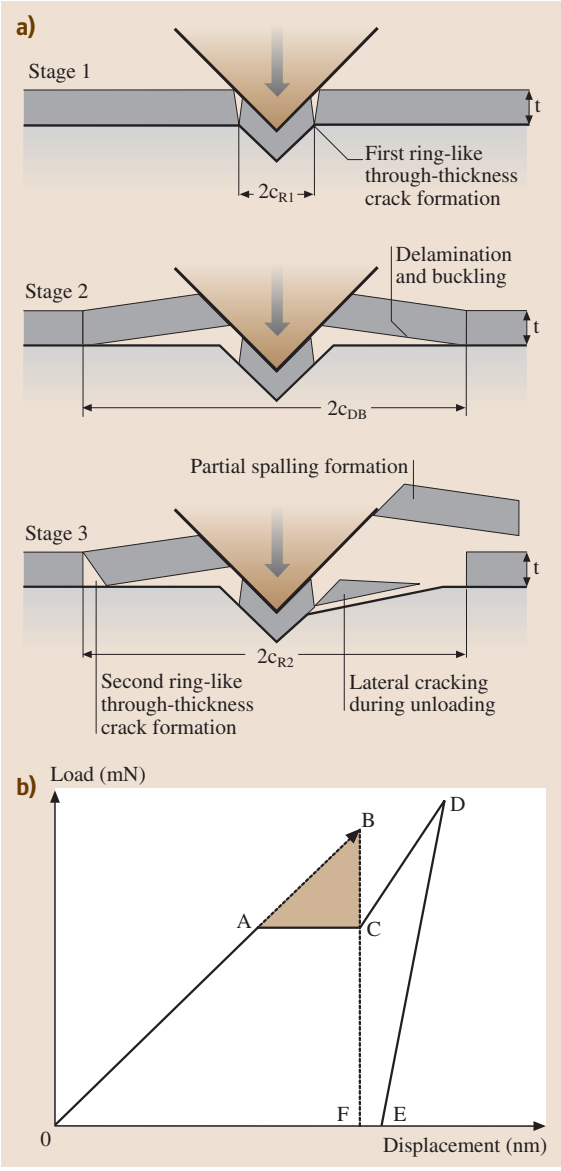


Fig. 26.12 **(a)** Schematic of various stages in nanoindentation fracture for the coatings/substrate system, and **(b)** schematic of a load-displacement curve showing a step during the loading cycle and associated energy release

for the coating used in the present study, the advances of the indenter during the radial cracking, delamination, and buckling are not big enough to form steps in the loading curve, because the coating around the indenter still supports the indenter, but generate discontinuities that change the slope of the loading curve with increasing indentation loads. In the third stage, the stress concentration at the end of the interfacial crack cannot be relaxed by the propagation of the interfacial crack. With an increase in indentation depth, the height of the bulged coating increases. When the height reaches a critical value, the bending stresses caused by the bulged coating around the indenter will result in the second ring-like through-thickness crack formation and spalling at the edge of the buckled coating, as shown in Fig. 26.12a, which leads to a step in the loading curve. This is a single event and results in the separation of the part of the coating around the indenter from the bulk

Table 26.5 Hardness, elastic modulus, fracture toughness, fatigue life, critical load during scratch, coefficient of friction during accelerated wear testing and residual stresses of various DLC coatings on single-crystal silicon substrate

Coating	Hardness[a] [26.48] (GPa)	Elastic modulus[a] [26.48] (GPa)	Fracture toughness[a] [26.89] (MPa $m^{1/2}$)	Fatigue life[b], N_f^d [26.90] $\times 10^4$	Critical load during scratch[b] [26.48] (mN)	Coefficient of friction during accelerated wear testing[b] [26.48]	Compressive residual stress[c] [26.47] (GPa)
Cathodic arc carbon coating (a-C)	24	280	11.8	2.0	3.8	0.18	12.5
Ion beam carbon coating (a-C:H)	19	140	4.3	0.8	2.3	0.18	1.5
ECR-CVD carbon coating (a-C:H)	22	180	6.4	1.2	5.3	0.22	0.6
DC sputtered carbon coating (a-C:H)	15	140	2.8	0.2	1.1	0.32	2.0
Bulk graphite (for comparison)	Very soft	9–15	–	–	–	–	–
Diamond (for comparison)	80–104	900–1050	–	–	–	–	–
Si(100) substrate	11	220	0.75	–	0.6	0.55	0.02

[a] Measured on 100-nm-thick coatings
[b] Measured on 20-nm-thick coatings
[c] Measured on 400-nm-thick coatings
[d] N_f was obtained at a mean load of 10 μN and a load amplitude of 8 μN

coating via cracking through coatings. The step in the loading curve results totally from the coating cracking and not from the interfacial cracking or the substrate cracking.

The area under the load-displacement curve is the work performed by the indenter during elastic-plastic deformation of the coating/substrate system. The strain energy release in the first/second ring-like cracking and spalling can be calculated from the corresponding steps in the loading curve. Fig. 26.12b shows a modeled load-displacement curve. OACD is the loading curve and DE is the unloading curve. The first ring-like through-thickness crack should be considered. It should be emphasized that the edge of the buckled coating is far from the indenter and, therefore, it does not matter if the indentation depth exceeds the coating thickness, or if deformation of the substrate occurs around the indenter when we measure fracture toughness of the coating from the released energy during the second ring-like through-thickness cracking (spalling). Suppose that the second ring-like through-thickness cracking occurs at AC. Now, let us consider the loading curve OAC. If the second ring-like through-thickness crack does not occur, it can be understood that OA will be extended to OB to reach the same displacement as OC. This means that the crack formation changes the loading curve OAB into OAC. For point B, the elastic-plastic energy stored in the coating/substrate system should be OBF. For point C, the elastic-plastic energy stored in the coating/substrate system should be OACF. Therefore, the energy difference before and after the crack generation is the area of ABC, i. e., this energy stored in ABC will be released as strain energy to create the ring-like through-thickness crack. According to the theoretical analysis by *Li* et al. [26.87],

the fracture toughness of thin films can be written as

$$K_{\mathrm{Ic}} = \left[\left(\frac{E}{(1-\nu^2)\,2\pi C_{\mathrm{R}}}\right)\left(\frac{U}{t}\right)\right]^{1/2}, \qquad (26.2)$$

where E is the elastic modulus, ν is the Poisson's ratio, $2\pi C_{\mathrm{R}}$ is the crack length in the coating plane, t is the coating thickness, and U is the strain energy difference before and after cracking.

Using (26.2), the fracture toughness of the coatings is calculated. The loading curve is extrapolated along the tangential direction of the loading curve from the starting point of the step up to reach the same displacement as the step. The area between the extrapolated line and the step is the estimated strain energy difference before and after cracking. C_{R} is measured from SEM micrographs or AFM images of indentations. The second ring-like crack is where the spalling occurs. For example, for the 400-nm-thick cathodic arc carbon coating data presented in Fig. 26.11, U of 7.1 nNm is assessed from the steps in Fig. 26.11a at the peak indentation loads of 200 mN. For C_{R} of 7.0 μm from Fig. 26.11b, $E = 300$ GPa measured using a nanoindenter and an assumed value of 0.25 for ν, fracture toughness values are calculated as $10.9\,\mathrm{MPa}\sqrt{m}$ [26.87, 88]. The fracture toughness and related data for various 100-nm-thick DLC coatings are presented in Fig. 26.10 and Table 26.5.

Nanofatigue

Delayed fracture resulting from extended service is called fatigue [26.96]. Fatigue fracturing progresses through a material via changes within the material at the tip of a crack, where there is a high stress intensity. There are several situations: cyclic fatigue, stress corrosion, and static fatigue. Cyclic fatigue results from cyclic loading of machine components. In a low-flying slider in a magnetic head-disk interface, isolated asperity contacts occur during use and the fatigue failure occurs in the multilayered thin-film structure of the magnetic disk [26.13]. In many MEMS components, impact occurs and the failure mode is cyclic fatigue. Asperity contacts can be simulated using a sharp diamond tip in an oscillating contact with the component.

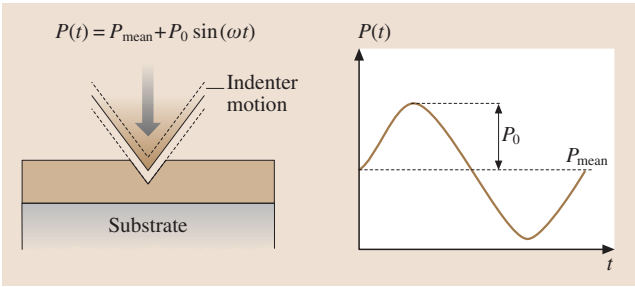


Fig. 26.13 Schematic of a fatigue test on a coating/substrate system using the continuous stiffness measurement technique

Figure 26.13 shows the schematic of a fatigue test on a coating/substrate system using a continuous stiffness measurement (CSM) technique. Load cycles are applied to the coating, resulting in a cyclic stress. P is the cyclic load, P_{mean} is the mean load, P_{o} is the oscillation load amplitude, and ω is the oscillation frequency. The following results can be obtained: (1) endurance limit, i. e., the maximum load below which there is no coating failure for a preset number of cycles; (2) number of cycles at which the coating failure occurs; and (3) changes in contact stiffness measured by using the unloading slope of each cycle, which can be used to monitor the propagation of the interfacial cracks during cyclic fatigue process.

Figure 26.14a shows the contact stiffness as a function of the number of cycles for 20-nm-thick FCA coatings cyclically deformed by various oscillation load amplitudes with a mean load of 10 μN at a frequency of 45 Hz. At 4 μN load amplitude, no change in contact stiffness was found for all coatings. This indicates that 4 μN load amplitude is not high enough to damage the coatings. At 6 μN load amplitude, an abrupt decrease in contact stiffness was found at a certain number of cycles for each coating, indicating that fatigue damage had occurred. With increasing load amplitude, the number of cycles to failure, N_{f}, decreases for all coatings. Load amplitude versus N_{f}, a so-called S-N curve, is plotted in Fig. 26.14b. The critical load amplitude, below which no fatigue damage occurs (an endurance limit), was identified for each coating. This critical load amplitude, together with mean load, are of critical importance to the design of head-disk interfaces or MEMS/NEMS device interfaces.

To compare the fatigue lives of different coatings studied, the contact stiffness as a function of the number of cycles for 20-nm-thick FCA, IB, ECR-CVD and SP coatings cyclically deformed by an oscillation load amplitude of 8 μN with a mean load of 10 μN at a frequency of 45 Hz is shown in Fig. 26.14c. FCA coating has the longest N_{f}, followed by ECR-CVD, IB, and SP coatings. In addition, after the N_{f}, the contact stiffness of the FCA coating shows a slower decrease than the other coatings. This indicates that after the N_{f}, the FCA coating had less damage than the others. The fatigue behavior of FCA and ECR-CVD coatings of differ-

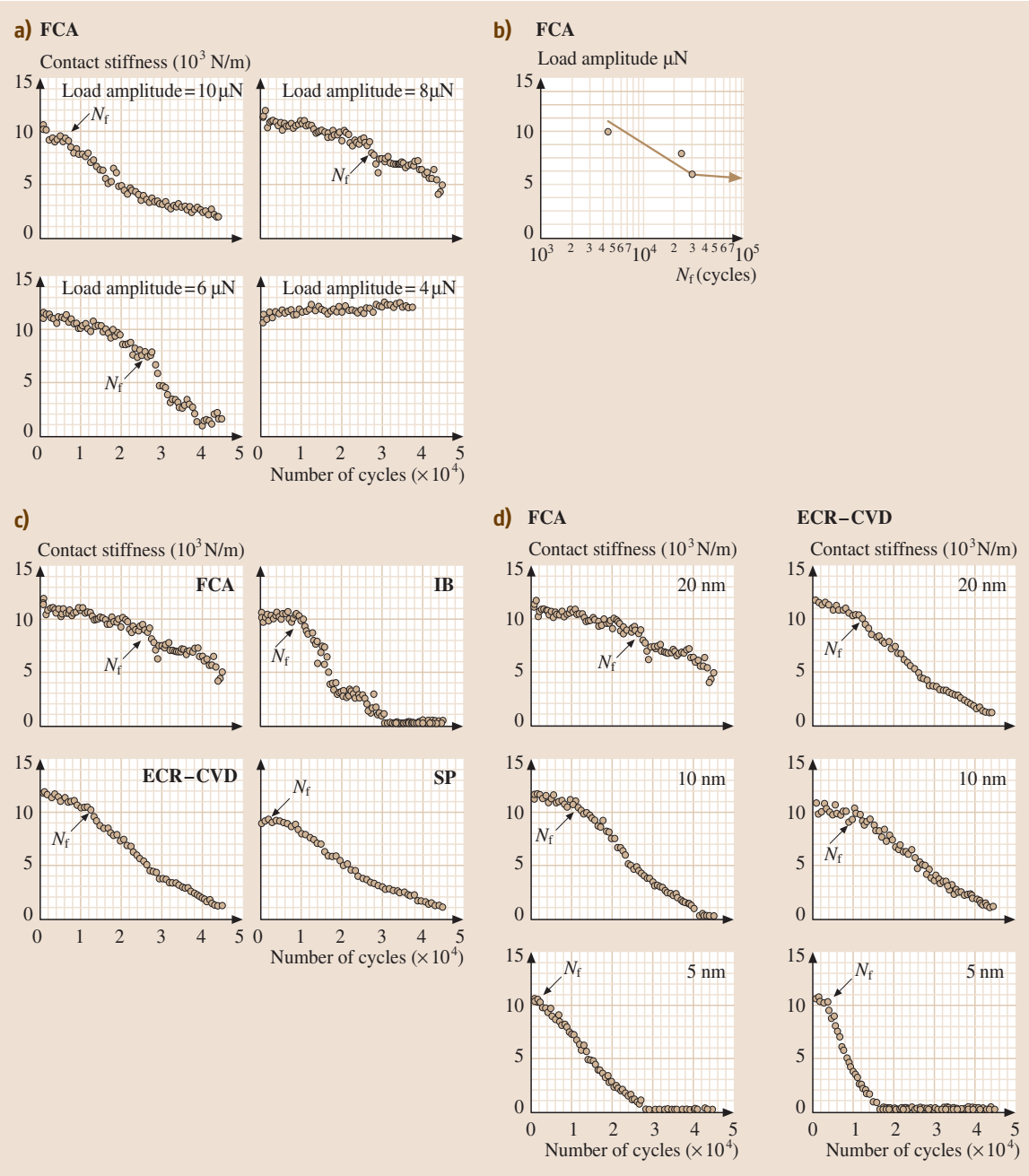


Fig. 26.14 **(a)** Contact stiffness as a function of the number of cycles for 20-nm-thick FCA coatings cyclically deformed by various oscillation load amplitudes with a mean load of 10 μN at a frequency of 45 Hz; **(b)** plot of load amplitude versus N_f; **(c)** contact stiffness as a function of the number of cycles for four different 20-nm-thick coatings with a mean load of 10 μN and load amplitude of 8 μN at a frequency of 45 Hz; and **(d)** contact stiffness as a function of the number of cycles for two coatings of different thicknesses at a mean load of 10 μN and load amplitude of 8 μN at a frequency of 45 Hz

ent thicknesses is compared in Fig. 26.14d. For both coatings, N_f decreases with decreasing coating thickness. At 10 nm, FCA and ECR-CVD have almost the same fatigue life. At 5 nm, ECR-CVD coating shows a slightly longer fatigue life than FCA coating. This indicates that even for nanometer-thick DLC coatings their microstructure and residual stresses are not uniform across the thickness direction. Thinner coatings are more influenced by interfacial stresses than thicker coating.

Figure 26.15a shows the high magnification SEM images of 20-nm-thick FCA coatings before, at, and after N_f. In the SEM images, the net-like structure is the gold film coated on the DLC coating, which should be ignored in analyzing the indentation fatigue damage. Before the N_f, no delamination or buckling was found except the residual indentation mark at magnifications up to 1,200,000× using SEM. This suggests that only plastic deformation occurred before the N_f. At the N_f, the coating around the indenter bulged upwards, indicating delamination and buckling. Therefore, it is believed that the decrease in contact stiffness at the N_f results from the delamination and buckling of the coating from the substrate. After the N_f, the buckled coating was broken down around the edge of the buckled area, forming a ring-like crack. The remaining coating overhung at the edge of the buckled area. It is noted that the indentation size increases with the increasing number of cycles. This indicates that deformation, delamination and buckling, and ring-like crack formation occurred over a period.

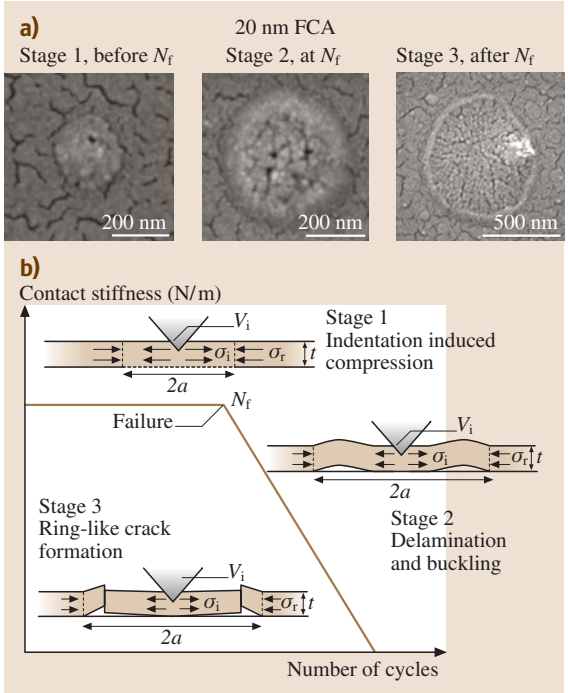


Fig. 26.15 **(a)** High magnification SEM images of a coating before, at, and after N_f, and **(b)** schematic of various stages in the indentation fatigue damage for a coating/substrate system [26.90]

The schematic in Fig. 26.15b shows various stages in the indentation fatigue damage for a coating/substrate system. Based on this study, three stages in the indentation fatigue damage appear to exist: (1) indentation-induced compression; (2) delamination and buckling; (3) ring-like crack formation at the edge of the buckled coating. The deposition process often induces residual stresses in coatings. The model shown in Fig. 26.15b considers a coating with the uniform biaxial residual compression σ_r. In the first stage, indentation induces elastic/plastic deformation, exerting an outward acting pressure on the coating around the indenter. Interfacial defects like voids and impurities act as original cracks. These cracks propagate and link up as the indentation compressive stress increases. At this stage, the coating, which is under the indenter and above the interfacial crack (with a crack length of $2a$), still maintains a solid contact with the substrate; the substrate still fully supports the coating. Therefore, this interfacial crack does not lead to an abrupt decrease in contact stiffness, but gives rise to a slight decrease in contact stiffness, as shown in Fig. 26.14. The coating above the interfacial crack is treated as a rigidly clamped disk. We assume that the crack radius, a, is large compared with the coating thickness t. Since the coating thickness ranges from 5 to 20 nm, this assumption is easily satisfied in this study (radius of the delaminated and buckled area, shown in Fig. 26.15a, is on the order of 100 nm). The compressive stress caused by indentation is given as [26.97]:

$$\sigma_i = \frac{E}{(1-\nu)}\varepsilon_i = \frac{EV_i}{2\pi t a^2(1-\nu)}\,, \qquad (26.3)$$

where ν and E are the Poisson's ratio and elastic modulus of the coating, V_i is the indentation volume, t is the coating thickness, and a is the crack radius. With an increasing number of cycles, the indentation volume V_i increases. Therefore, the indentation compressive stress σ_i increases accordingly. In the second stage, buckling occurs during the unloading segment of the fatigue testing cycle when the sum of indentation compressive stress σ_i and the residual stress σ_r

exceed the critical buckling stress σ_b for the delaminated circular section as given by [26.98]

$$\sigma_b = \frac{\mu^2 E}{12\left(1-\nu^2\right)}\left(\frac{t}{a}\right)^2 , \tag{26.4}$$

where the constant μ equals 42.67 for a circular clamped plate with a constrained center point and 14.68 when the center is unconstrained. The buckled coating acts as a cantilever. In this case, the indenter indents a cantilever rather than a coating/substrate system. This ultrathin coating cantilever has much less contact stiffness than the coating/substrate system. Therefore, the contact stiffness shows an abrupt decrease at the N_f. In the third stage, with an increasing number of cycles, the delaminated and buckled size increases, resulting in a further decrease in contact stiffness since the cantilever beam length increases. On the other hand, a high bending stress acts at the edge of the buckled coating. The larger the buckled size, the higher the bending stress. The cyclically bending stress causes fatigue damage at the end of the buckled coating, forming a ring-like crack. The coating under the indenter is separated from the bulk coating (caused by the ring-like crack at the edge of the buckled coating) and the substrate (caused by the delamination and buckling in the second stage). Therefore, the coating under the indenter is not constrained, but is free to move with the indenter during fatigue testing. At this point, the sharp nature of the indenter is lost, because the coating under the indenter gets stuck on the indenter. The indentation fatigue experiment results in the contact of a relative huge blunt tip with the substrate. This results in a low contact stiffness value.

Compressive residual stresses result in delamination and buckling. A coating with a higher adhesion strength and a less compressive residual stress is required for a higher fatigue life. Interfacial defects should be avoided in the coating deposition process. We know that the ring-like crack formation occurs in the coating. Formation of fatigue cracks in the coating depends upon the hardness and fracture toughness. Cracks are more difficult to form and propagate in the coating with higher strength and fracture toughness.

It is now accepted that long fatigue life in a coating/substrate almost always involves "living with crack", that the threshold or limit condition is associated with the non-propagation of exiting cracks or defects, even though these cracks may be undetectable [26.96]. For all coatings studied, at 4 μN, contact stiffness does not change much. This indicates that delamination and buckling did not occur within the number of cycles tested in this study. This is probably because the indentation-induced compressive stress was not high enough to allow the cracks to propagate and link up under the indenter, or the sum of indentation compressive stress σ_i and the residual stress σ_r did not exceed the critical buckling stress σ_b.

Figure 26.10 and Table 26.5 summarize the hardness, elastic modulus, fracture toughness, and fatigue life of all coatings studied. A good correlation exists between fatigue life and other mechanical properties. Higher mechanical properties result in a longer fatigue life. The mechanical properties of DLC coatings are controlled by the sp^3-to-sp^2 ratio. The sp^3-bonded carbon exhibits the outstanding properties of diamond [26.51]. A higher deposition kinetic energy will result in a larger fraction of sp^3-bonded carbon in an amorphous network. Thus, the higher kinetic energy for the FCA could be responsible for its better carbon structure and higher mechanical properties [26.48–50, 99]. Higher adhesion strength between the FCA coating and substrate makes the FCA coating more difficult to delaminate from the substrate.

26.3.2 Microscratch and Microwear Studies

For microscratch studies, a conical diamond indenter (e.g., having a tip radius of about one micron and included angle of 60°) is drawn over the sample surface, and the load is ramped up (typically from 2 mN to 25 mN) until substantial damage occurs. The coefficient of friction is monitored during scratching. Scratch-induced damage of coating, specifically fracture or delamination, can be monitored by in situ friction force measurements and by optical and SEM imaging of the scratches after tests. A gradual increase in friction is associated with plowing, and an abrupt increase in friction is associated with fracture or catastrophic failure [26.100]. The load, corresponding to the abrupt increase in friction or an increase in friction above a certain value (typically 2× the initial value), provides a measure of scratch resistance or adhesive strength of a coating and is called "critical load". The depth of scratches with increasing scratch length or normal load is measured using an AFM, typically with an area of 10 × 10 μm [26.48, 49, 101].

The microscratch and microwear studies are also conducted using an AFM [26.23, 50, 99, 102, 103]. A square pyramidal diamond tip (tip radius ∼ 100 nm) or a three-sided pyramidal diamond (Berkovich) tip with an apex angle of 60° and a tip radius of about 100 nm mounted on a platinum-coated, rectangular stainless steel cantilever of stiffness of about 40 N/m is scanned

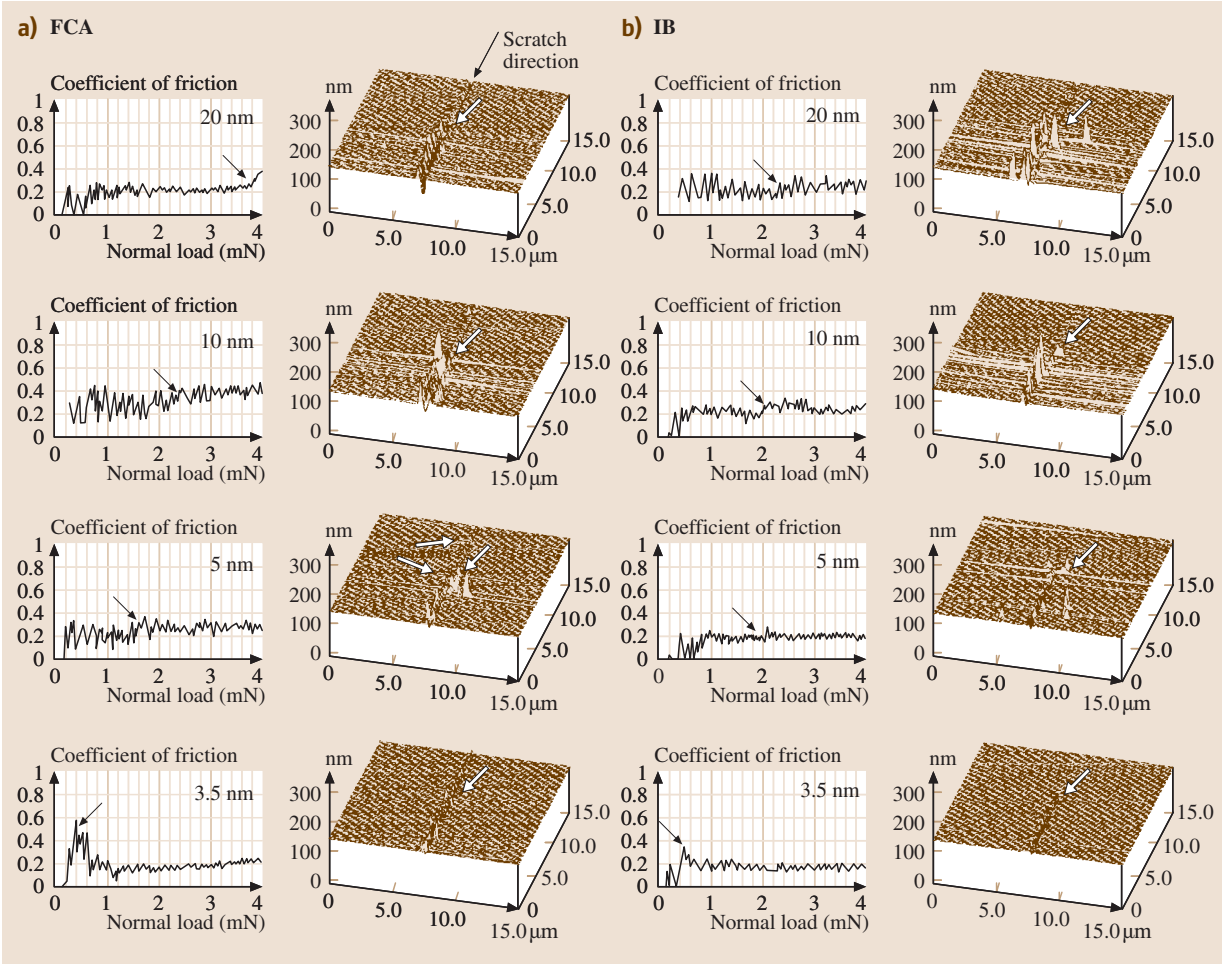


Fig. 26.16a,b

orthogonal to the long axis of the cantilever to generate scratch and wear marks. During the scratch test, the normal load is either kept constant or increased (typically from 0 to 100 μN) until damage occurs. Topography images of the scratch are obtained in situ with the AFM at a low load. By scanning the sample during scratching, wear experiments can be conducted. Wear at a constant load is monitored as a function of the number of cycles. Normal loads ranging from 10–80 μN are typically used.

Microscratch

Scratch tests conducted with a sharp diamond tip simulate a sharp asperity contact. In a scratch test, the cracking or delamination of a hard coating is signaled by a sudden increase in the coefficient of friction [26.23]. The load associated with this event is called the "critical load".

Wu [26.104], *Bhushan* et al. [26.70], *Gupta* and *Bhushan* [26.12,47], and *Li* and *Bhushan* [26.48,49,101] have used a nanoindenter to perform microscratch studies (mechanical durability) of various carbon coatings. The coefficient of friction profiles as a function of increasing normal load and AFM surface height maps of regions over scratches at the respective critical loads (indicated by the arrows in the friction profiles and AFM images) made on the various coatings of different thicknesses and single-crystal silicon substrate using a conical tip are compared in Figs. 26.16 and 26.17. *Bhushan* and *Koinkar* [26.102], *Koinkar* and

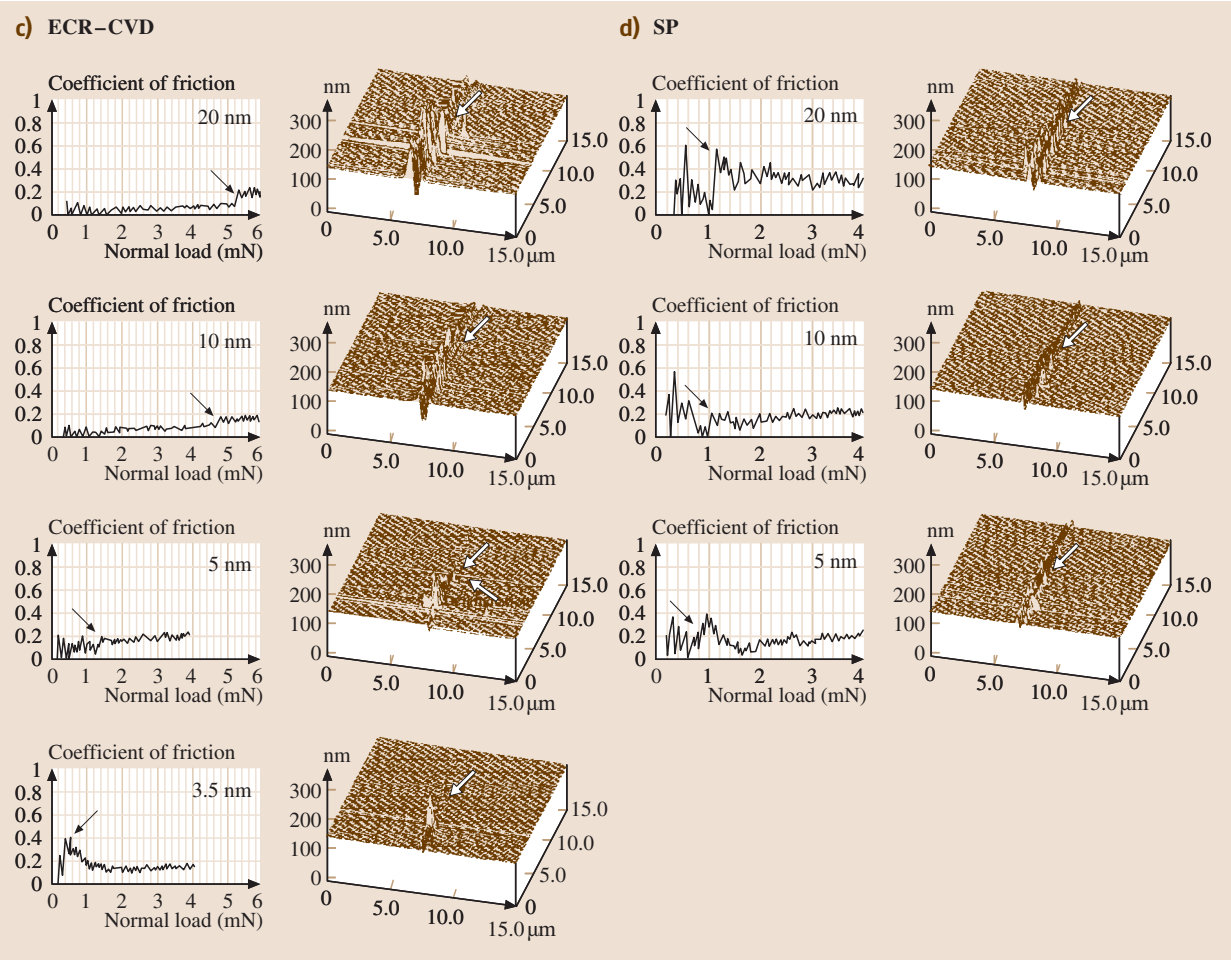


Fig. 26.16a–d Coefficient of friction profiles as a function of normal load and corresponding AFM surface height maps of regions over scratches at the respective critical loads (indicated by the arrows in the friction profiles and AFM images) for coatings of different thicknesses deposited by various deposition techniques: **(a)** FCA, **(b)** IB, **(c)** ECR-CVD, **(d)** SP

Bhushan [26.103], *Bhushan* [26.23], and *Sundararajan* and *Bhushan* [26.50, 99] have used an AFM to perform microscratch studies. Data for various coatings of different thicknesses and silicon substrate using a Berkovich tip are compared in Figs. 26.18 and 26.19. Critical loads for various coatings tested using a nanoindenter and AFM are summarized in Fig. 26.20. The selected data for 20-nm-thick coatings obtained using nanoindenter are also presented in Fig. 26.10 and Table 26.5.

It can be seen that a well-defined critical load exists for each coating. The AFM images clearly show that below the critical loads the coatings were plowed by the scratch tip, associated with the plastic flow of materials.

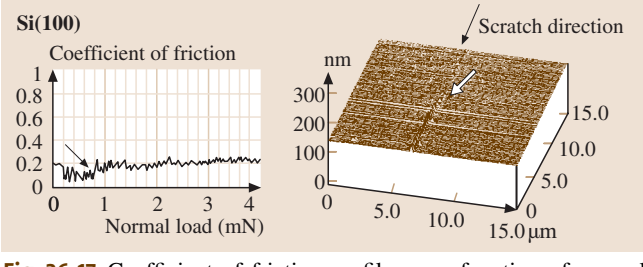
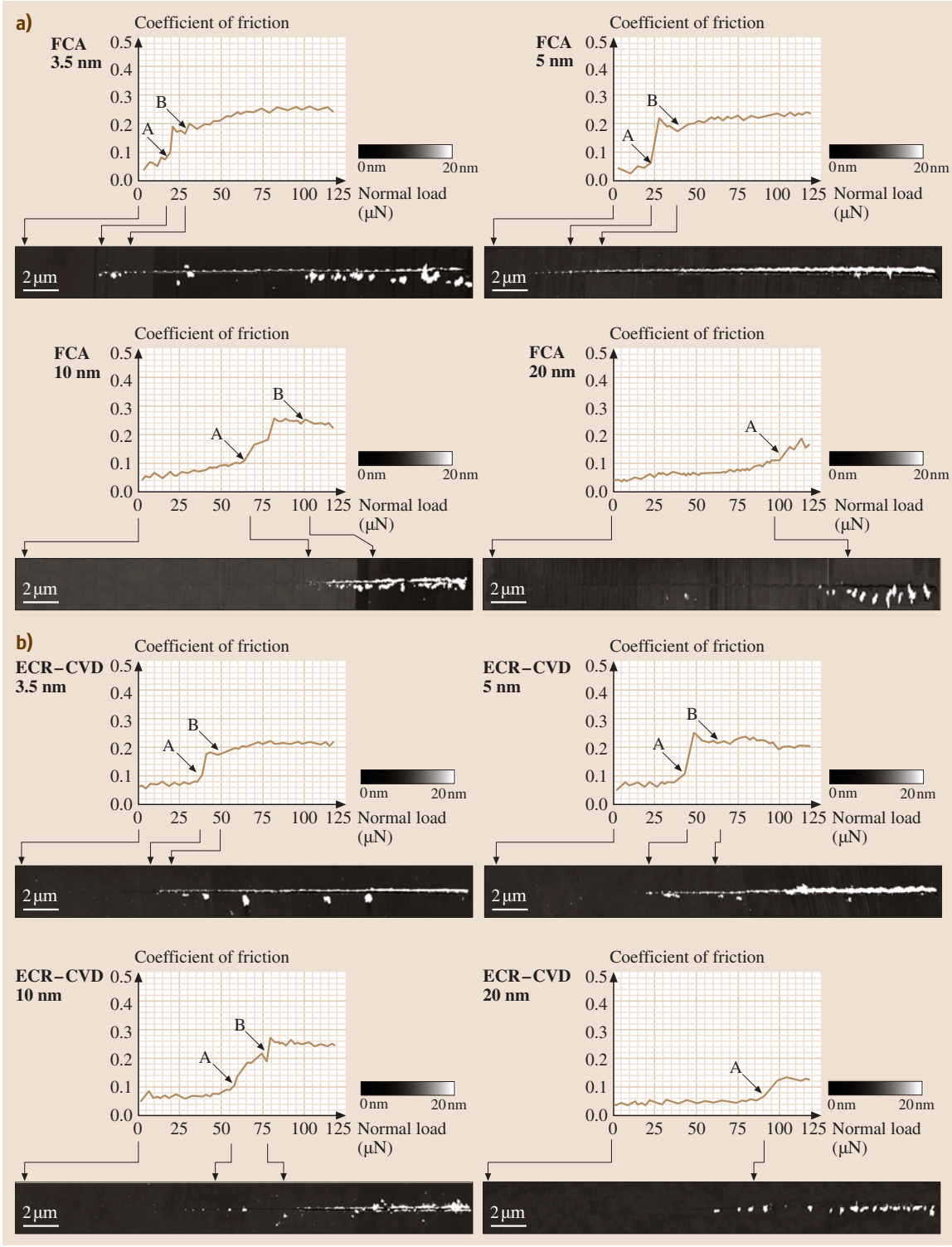


Fig. 26.17 Coefficient of friction profiles as a function of normal load and corresponding AFM surface height maps of regions over scratches at the respective critical loads (indicated by the arrows in the friction profiles and AFM images) for Si(100)


a)
FCA 3.5 nm
FCA 5 nm
FCA 10 nm
FCA 20 nm
b)
ECR–CVD 3.5 nm
ECR–CVD 5 nm
ECR–CVD 10 nm
ECR–CVD 20 nm
Coefficient of friction
0.5
0.4
0.3
0.2
0.1
0.0
0 25 50 75 100 125
Normal load (μN)
A
B
0nm 20nm
2 μm

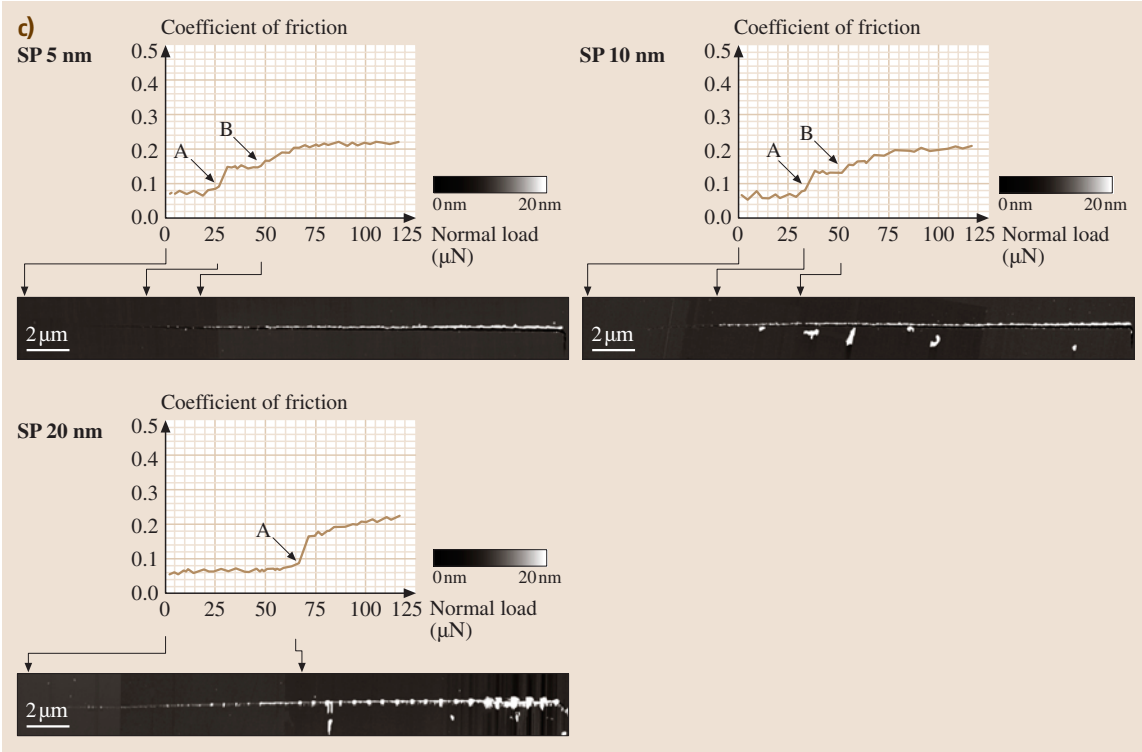


Fig. 26.18a–c Coefficient of friction profiles during scratch as a function of normal load and corresponding AFM surface height maps for **(a)** FCA, **(b)** ECR-CVD, and **(c)** SP coatings [26.99]

At and after the critical loads, debris (chips) or buckling was observed on the sides of scratches. Delamination or buckling can be observed around or after the critical loads, which suggests that the damage starts from delamination and buckling. For the 3.5- and 5-nm-thick FCA coatings, before the critical loads small debris is observed on the sides of scratches. This suggests that the thinner FCA coatings are not so durable. It is obvious that for a given deposition method, the critical loads increase with increasing coating thickness. This indicates that the critical load is determined not only by the adhesion strength to the substrate, but also by the coating thickness. We note that the debris generated on the thicker coatings is larger than that generated on the thinner coatings. For a thicker coating, it is more difficult to be broken; the broken coating chips (debris) for a thicker coating are larger than those for the thinner coatings. The difference in the residual stresses of the coatings of different thicknesses could also affect the size of debris. The AFM image shows that the silicon substrate was damaged by plowing, associated with the plastic flow of materials. At and after the critical load, small and uniform debris is observed and the amount of debris increases with increasing normal load.

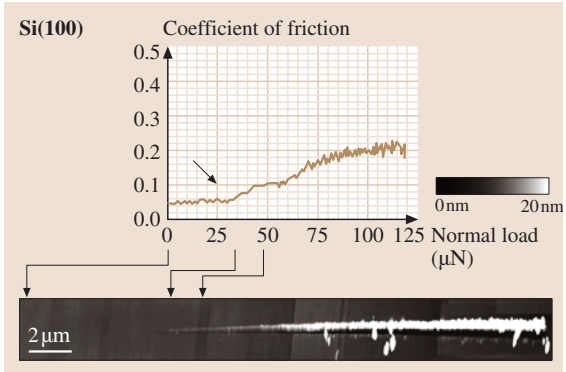


Fig. 26.19 Coefficient of friction profiles during scratch as a function of normal load and corresponding AFM surface height maps for Si(100) [26.99]

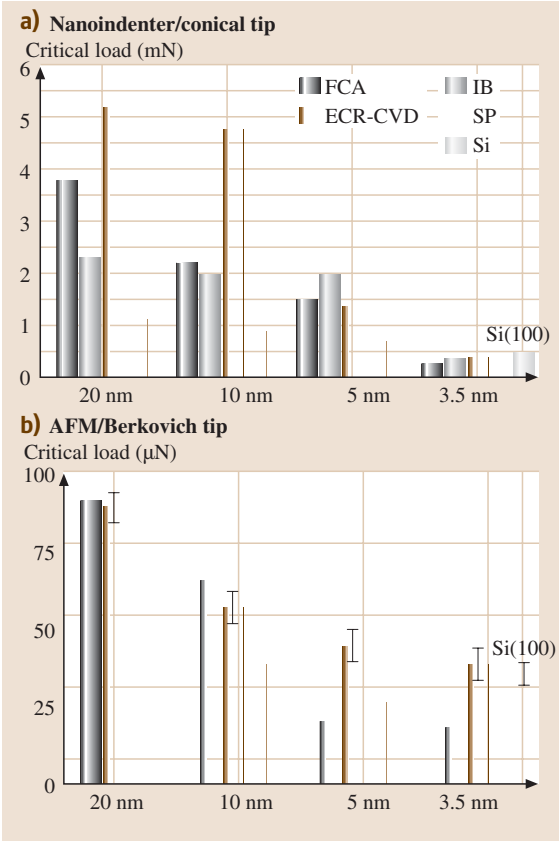


Fig. 26.20a,b Critical loads estimated from the coefficient of friction profiles from **(a)** nanoindenter and **(b)** AFM tests for various coatings of different thicknesses and Si(100) substrate

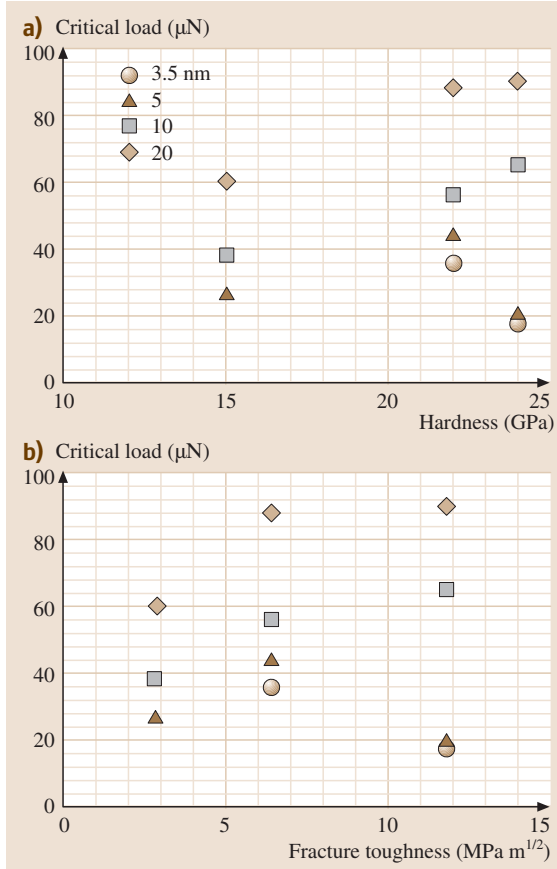


Fig. 26.21a,b Measured critical loads estimated from the coefficient of friction profiles from AFM tests as a function of **(a)** coating hardness and **(b)** fracture toughness. Coating hardness and fracture toughness values were obtained using a nanoindenter on 100-nm-thick coatings (Table 26.5)

Since at the critical load, the damage mechanism appears to be the onset of plowing, higher hardness and fracture toughness of a coating will therefore result in a higher load required for deformation and hence a higher critical load. Figure 26.21 shows critical loads of the various coatings, obtained with AFM tests, as a function of the coating hardness and fracture toughness (from Table 26.5). It can be seen that, in general, higher coating hardness and fracture toughness result in a higher critical load. The only exceptions are the FCA coatings at 5- and 3.5-nm coating thickness, which show the lowest critical loads despite their high hardness and fracture toughness. The brittleness of the thinner FCA coatings may be one reason for their low critical loads. The mechanical properties of coatings that are less than 10 nm thick are unknown. The FCA process may result in coatings with low hardness at such low thickness due to differences in coating stoichiometry and structure compared to the coatings of higher thickness. Also, at these thicknesses stresses at the coating-substrate interface may affect adhesion and load-carrying capacity of the coatings.

Based on the experimental results, a schematic of scratch damage mechanisms of the DLC coatings used in this study is shown in Fig. 26.22. Below the critical load, if a coating has a good combination of strength and fracture toughness, plowing associated with the plastic flow of materials is responsible for the damage of coating (Fig. 26.22a). Whereas if a coating has a lower fracture toughness, cracking could occur during plowing, asso-

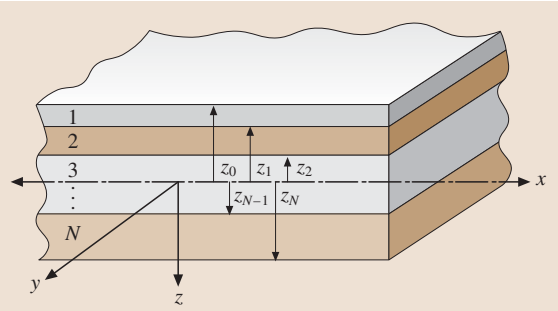


Fig. 35.8 Multilayer film system with a definition of the parameters involved in the analysis

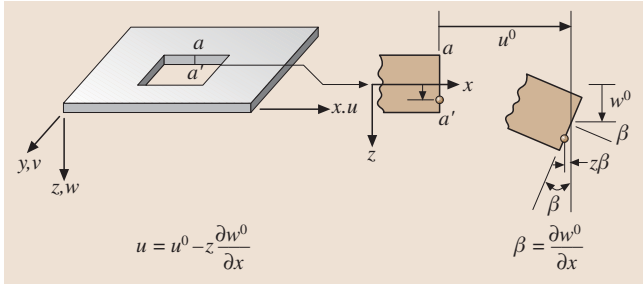


Fig. 35.9 Kinematics of the deformation of the thin-film multilayer

Figure 35.8 shows a schematic of the geometry considered in this study; a multilayer film microstructure with lateral dimensions that far exceed the total thickness. As such we ignore the effects of the free edges and focus on a representative point in the layered film that lies at least a distance of a few film thicknesses from the edge. The layers are numbered $1, 2, \ldots, N$ where 1 is the top layer and N is the bottom layer. Each layer has a thickness t_i $(i = 1 \rightarrow N)$, and the total thickness of the microstructure is t. Each of the film layers consists of an isotropic material with a Young's modulus E_i, Poisson's ratio, ν_i, and may undergo a transformation strain $\varepsilon_i^*(z)$, which may vary with position z through the film thickness. The source of $\varepsilon_i^*(z)$ may be arbitrary, and the effect of multiple sources can be considered as the sum of the individual ones. A common source, and one that we will focus on, is thermal expansion upon a uniform temperature change. In this case $\varepsilon_i^*(z) = \varepsilon_i^* = \alpha_i T$ where α_i is the thermal expansion coefficient and T is the temperature change from a reference state. In addition, each layer may contain an internal stress $\hat{\sigma}(z)$, which also may vary with position z through the thickness. For simplicity we focus our results on isotropic layers, but the framework is applicable to anisotropic behavior of each layer. As a result of the difference in stress-free strains and/or internal stresses between each layer, stresses are developed in the layers and the structure deforms; our intent is to compute these.

Analysis

The analysis presented here is somewhat standard in the treatment of laminated composite plates but has not been often used in the analysis of multilayer thin film problems for microelectronics and/or microelectromechanical systems. Detailed accounts of parts of the theory can be found in texts on composite materials (see for example [35.24,25]), and so only a brief account is presented here.

The displacements u and v at any point z through the thickness of the laminate are assumed to be described by the classical Kirchoff hypothesis. As illustrated in Fig. 35.9, the Kirchoff hypothesis states that after loads are applied to the layered film, the line a–a' which is initially straight and normal to the geometric midplane remains straight, normal to the deformed geometric midplane, and its length is unchanged. In other words, the normal does not deform but only translates and rotates. As a consequence of this assumption, the displacements of any point in the layered film are given by:

$$\begin{aligned} u &= u^0 - z\frac{\partial w^0}{\partial x} \\ v &= v^0 - z\frac{\partial w^0}{\partial y} \\ w &= w^0 \,, \end{aligned} \tag{35.1}$$

where u^0, v^0, and w^0 are the displacements of the geometric midplane at $z = 0$. Note that in addition to this explicit dependence of the displacements on z, they are generally functions of x and y.

The nonzero normal and shear strains are then:

$$\begin{aligned} \varepsilon_x &= \frac{\partial u}{\partial x} = \frac{\partial u^0}{\partial x} - z\frac{\partial^2 w^0}{\partial x^2} = \varepsilon_x^0 + z\kappa_x \\ \varepsilon_y &= \frac{\partial v}{\partial y} = \frac{\partial v^0}{\partial x} - z\frac{\partial^2 w^0}{\partial x^2} = \varepsilon_y^0 + z\kappa_y \\ \gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = \frac{\partial u^0}{\partial y} + \frac{\partial v^0}{\partial x} - 2z\frac{\partial^2 w^0}{\partial x \partial y} \\ &= \gamma_{xy}^0 + z\kappa_{xy} \end{aligned} \tag{35.2a}$$

or in matrix form:

$$\{\varepsilon\} = \left\{\varepsilon^0\right\} + z\{\kappa\} \,, \tag{35.2b}$$

where:

$$\{\varepsilon\} = \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix} \left\{\varepsilon^0\right\} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \gamma_{xy}^0 \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial u^0}{\partial x} \\ \dfrac{\partial v^0}{\partial y} \\ \dfrac{\partial u^0}{\partial y} + \dfrac{\partial v^0}{\partial x} \end{Bmatrix}$$

$$\{\kappa\} = \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix} = \begin{Bmatrix} -z\dfrac{\partial^2 w^0}{\partial x^2} \\ -z\dfrac{\partial^2 w^0}{\partial y^2} \\ -2z\dfrac{\partial^2 w^0}{\partial x \partial y} \end{Bmatrix} . \tag{35.3}$$

In (35.3), $\left\{\varepsilon^0\right\}$ is the midplane strain and $\{\kappa\}$ is the midplane curvature. As with the displacements, $\left\{\varepsilon^0\right\}$ and $\{\kappa\}$ are generally functions of x and y.

The stress-strain relations for each layer are:

$$\begin{Bmatrix} \sigma_x(z) \\ \sigma_y(z) \\ \sigma_{xy}(z) \end{Bmatrix}_k = \begin{bmatrix} Q_{11} & Q_{12} & Q_{16} \\ Q_{12} & Q_{22} & Q_{26} \\ Q_{16} & Q_{26} & Q_{66} \end{bmatrix}_k \times \left[\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix} - \begin{Bmatrix} \varepsilon_x^*(z) \\ \varepsilon_y^*(z) \\ \gamma_{xy}^*(z) \end{Bmatrix}_k \right] + \begin{Bmatrix} \hat{\sigma}_x(z) \\ \hat{\sigma}_y(z) \\ \hat{\sigma}_{xy}(z) \end{Bmatrix}_k \tag{35.4a}$$

or in compact form:

$$\{\sigma(z)\}_k = \left\{\overline{Q}\right\}_k \left[\{\varepsilon\} - \{\varepsilon(z)\}_k^*\right] + \left\{\hat{\sigma}(z)\right\}_k , \tag{35.4b}$$

where $k = 1 \to N$ denotes the layer number. Equation (35.4) is written in a global coordinate system for the layered structure, (x, y, z). As such, $\left[\bar{Q}\right]_k$ are the (generally anisotropic) elastic stiffness coefficients of the kth layer in the global coordinate system. They are obtained by transformation (see [35.24, 25], for example) of the stiffnesses, $[Q]_k$, from a natural coordinate system (x_1, x_2, x_3). For orthotropic materials the nonzero components of $[Q]$ for each layer are,

$$[Q] = \begin{bmatrix} \dfrac{E_1}{1-\nu_{12}\nu_{21}} & \dfrac{\nu_{12}E_2}{1-\nu_{12}\nu_{21}} & 0 \\ \dfrac{\nu_{21}E_1}{1-\nu_{12}\nu_{21}} & \dfrac{E_2}{1-\nu_{12}\nu_{21}} & 0 \\ 0 & 0 & G_{12} \end{bmatrix} , \tag{35.5a}$$

where E_i, ν_{ij}, and G_{ij} are the orthotropic Young's moduli, Poisson's ratios, and the shear modulus, respectively, of the layer, in the local coordinate system. Note that $Q_{12} = Q_{21}$, and so the equality of these expressions in (35.5a) shows the reciprocal relationship between E_i and ν_{ij}; only three of the four are independent. For isotropic materials $\left[\bar{Q}\right] = [Q]$, and the nonzero components for each layer are:

$$\left[\bar{Q}\right] = [Q] = \begin{bmatrix} \dfrac{E}{1-\nu^2} & \dfrac{\nu E}{1-\nu^2} & 0 \\ \dfrac{\nu E}{1-\nu^2} & \dfrac{E}{1-\nu^2} & 0 \\ 0 & 0 & G \end{bmatrix} \tag{35.5b}$$

In (35.4) $\left\{\hat{\sigma}(z)\right\}_k$ is the residual internal stress in the layer and may be an arbitrary function of z, i. e., a stress gradient may exist within each layer; $\{\varepsilon\}$ is the total strain, $\{\varepsilon^*\}_k$ is the inelastic transformation strain, and the term in rectangular brackets is the elastic strain. The stresses and strains are all defined in the (x, y, z) coordinate system. Substituting (35.2) into (35.4) yields the stress distribution through the thickness:

$$\{\sigma(z)\}_k = \left\{\overline{Q}\right\}_k \left[\left\{\varepsilon^0\right\} + z\{\kappa\} - \left\{\varepsilon^*(z)\right\}_k\right] + \left\{\hat{\sigma}(z)\right\}_k . \tag{35.6}$$

For the layered film system we define a resultant force $\{N\}$ and moment $\{M\}$ per unit length as:

$$\begin{Bmatrix} N_x \\ N_y \\ N_{xy} \end{Bmatrix} = \{N\} = \int_{-t/2}^{t/2} \{\sigma(z)\}_k \, \mathrm{d}z$$

$$\begin{Bmatrix} M_x \\ M_y \\ M_{xy} \end{Bmatrix} = \{M\} = \int_{-t/2}^{t/2} \{\sigma(z)\}_k \, z \, \mathrm{d}z . \tag{35.7}$$

For example, N_x is the force in the x-direction per unit length in the y-direction. $\{N\}$ and $\{M\}$ have units of $[F/L]$, $[F \cdot L/L]$, respectively.

Substituting (35.6) into (35.7) and breaking the integrals through the thickness into sums of integrals over each layer yields:

$$\{N\} = [A]\left\{\varepsilon^0\right\} + [B]\{\kappa\} - \left\{N^\varepsilon\right\} + \left\{N^\sigma\right\}$$
$$\{M\} = [B]\left\{\varepsilon^0\right\} + [D]\{\kappa\} - \left\{M^\varepsilon\right\} + \left\{M^\sigma\right\} , \tag{35.8}$$

where:

$$[A] = \int_{-t/2}^{t/2} \{\overline{Q}\}_k \, \mathrm{d}z = \sum_{k=1}^{N} \{\overline{Q}\}_k (z_k - z_{k-1})$$

$$[B] = \int_{-t/2}^{t/2} \{\overline{Q}\}_k \, z \, \mathrm{d}z = \frac{1}{2} \sum_{k=1}^{N} \{\overline{Q}\}_k \left(z_k^2 - z_{k-1}^2\right)$$

$$[D] = \int_{-t/2}^{t/2} \{\overline{Q}\}_k \, z^2 \, \mathrm{d}z = \frac{1}{3} \sum_{k=1}^{N} \{\overline{Q}\}_k \left(z_k^3 - z_{k-1}^3\right)$$

$$\{N^\varepsilon\} = \int_{-t/2}^{t/2} \{\overline{Q}\}_k \{\varepsilon^*(z)\}_k \, \mathrm{d}z$$

$$\{N^\sigma\} = \int_{-t/2}^{t/2} \{\hat{\sigma}(z)\}_k \, \mathrm{d}z$$

$$\{M^\varepsilon\} = \int_{-t/2}^{t/2} \{\overline{Q}\}_k \{\varepsilon^*(z)\}_k \, z \, \mathrm{d}z$$

$$\{M^\sigma\} = \int_{-t/2}^{t/2} \{\hat{\sigma}(z)\}_k \, z \, \mathrm{d}z \, . \tag{35.9}$$

In the terminology of laminated composite materials, $[A]$, $[B]$, and $[D]$ are the extensional, coupling, and bending constants and are functions of the elastic moduli of each layer and the arrangement of the layers through the thickness. $[A]$, $[B]$, and $[D]$ have units of $[F/L]$, $[F]$, and $[F \cdot L]$, respectively. Physically, $[A]$ describes the connection between in-plane forces and straining of the midplane, $[D]$ describes the connection between moments and curvature, and $[B]$ connects moments to midplane strain and forces to curvature. If both the geometry and material properties, including $\varepsilon^*(z)$ and $\hat{\sigma}(z)$, of the layered film system are symmetric about $z = 0$, then $[B] = [0]$ and there is no coupling between in-plane straining and bending. Note that $[A]$, $[B]$, and $[D]$ are 3×3 matrices, as are $\left[\bar{Q}\right]_k$, while all other terms a 3×1 column vectors of the form of (35.3) and (35.4).

Equation (35.8) can be written in compact form as:

$$\begin{Bmatrix} \{N\} \\ \{M\} \end{Bmatrix} = \begin{bmatrix} [A] & [B] \\ [B] & [D] \end{bmatrix} \begin{Bmatrix} \{\varepsilon^0\} \\ \{\kappa\} \end{Bmatrix} - \begin{Bmatrix} \{N^\varepsilon\} \\ \{M^\varepsilon\} \end{Bmatrix} + \begin{Bmatrix} \{N^\sigma\} \\ \{M^\sigma\} \end{Bmatrix} . \tag{35.10}$$

Equation (35.10) is most easily implemented in matrix form as a 6×6 system of equations that hold at each (x, y) coordinate. These are the constitutive equations for the layered film, incorporating the Kirchoff hypothesis and the strain-displacement relations. In order to completely describe the response of a layered film due to external or internal loads, (35.10) must be supplemented with the stress equilibrium equations and appropriate boundary conditions. In general the solution of these equations for a specified multilayer film geometry and loading is a complex undertaking and recourse is often taken to numerical methods such as the finite element method (FEM). Alternatively, the equations can be formulated using an energy approach that will be advantageous when we consider geometric nonlinearity in Section 35.1.3.

Fortunately, using this formulation we can solve a number of important problems simply using (35.10). For example, if there are no externally applied loads then $\{N\} = \{M\} = \{0\}$ and (35.10) can be inverted to yield:

$$\begin{Bmatrix} \{\varepsilon^0\} \\ \{\kappa\} \end{Bmatrix} = \begin{bmatrix} [A] & [B] \\ [B] & [D] \end{bmatrix}^{-1} \times \left\{ \begin{Bmatrix} \{N^\varepsilon\} \\ \{M^\varepsilon\} \end{Bmatrix} - \begin{Bmatrix} \{N^\sigma\} \\ \{M^\sigma\} \end{Bmatrix} \right\} . \tag{35.11}$$

Recall that all terms on the right-hand side are functions of the elastic moduli, the known transformation strains and stresses, and the geometrical arrangement of the layers. Once the midplane strain and curvature are computed using (35.11), the stress distribution in each layer can be computed using (35.6).

Special Cases

Here we apply (35.11) to yield some simple, yet important, results. Before doing so, it is useful to explicitly express the form of the transformation stress distribution $\left\{\hat{\sigma}(z)\right\}_k$. To this end we restrict $\left\{\hat{\sigma}(z)\right\}_k$ to be a linearly varying function through each layer. This can be written as:

$$\left\{\hat{\sigma}(z)\right\}_k = \{S\}_k - \{\beta\}_k \left(z - \frac{z_{k-1} + z_k}{2}\right) , \tag{35.12}$$

where $\{S\}_k$ is the average stress in layer k and $\{\beta\}_k$ is the stress gradient as shown in Fig. 35.10. For this form of the stress, $\{N^\sigma\}$ and $\{M^\sigma\}$ can be expressed

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

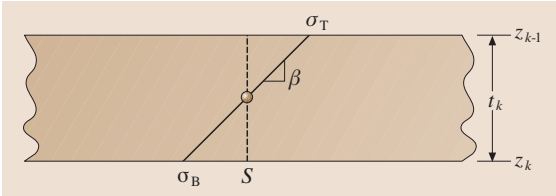


Fig. 35.10 Schematic showing a linear stress distribution characterized by a mean stress S and linear through-thickness stress gradient β

as:

$$\{N^\sigma\} = \sum_{k=1}^{N} \{S\}_k (z_k - z_{k-1})$$

$$\{M^\sigma\} = \frac{1}{2}\sum_{k=1}^{N} \{S\}_k \left(z_k^2 - z_{k-1}^2\right) - \frac{1}{12}\sum_{k=1}^{N} \{\beta\}_k \left(z_k^3 - z_{k-1}^3\right) . \quad (35.13)$$

Furthermore, it is useful to specify the form of $\{\varepsilon^*(z)\}$ when its source is thermal expansion during a uniform temperature change T from a reference temperature. In this case:

$$\left\{\varepsilon^*(z)\right\}_k = \left\{\varepsilon^*\right\}_k = \{\alpha\}_k T . \quad (35.14)$$

For this form of the stress-free strain, $\{N^\varepsilon\}$ and $\{M^\varepsilon\}$ can be expressed as:

$$N^\varepsilon = \int_{-t/2}^{t/2} \left\{\overline{Q}\right\}_k \{\alpha\}_k T \, \mathrm{d}z = \Delta T \sum_{k=1}^{N} \left\{\overline{Q}\right\}_k \{\alpha\}_k (z_k - z_{k-1})$$

$$M^\varepsilon = \int_{-t/2}^{t/2} \left\{\overline{Q}\right\}_k \{\alpha\}_k T z \, \mathrm{d}z = \frac{\Delta T}{2}\sum_{k=1}^{N} \left\{\overline{Q}\right\}_k \{\alpha\}_k \left(z_k^2 - z_{k-1}^2\right) . \quad (35.15)$$

In all of the examples described in this section we will consider only cases in which all layers have elastic constants, thermal expansion properties, and misfit strains that are isotropic. As a result, $A_{16} = A_{26} = B_{16} = B_{26} = D_{16} = D_{26} = 0$. The ramifications of this are that the midplane shear strains $\gamma_{xy}^0 = 0$ and the twist curvatures $\kappa_{xy} = 0$. In addition, the midplane strains and curvatures are equibiaxial, i. e., $\varepsilon_x^0 = \varepsilon_y^0 = \varepsilon^0$ and $\kappa_x = \kappa_y = \kappa$. So in the remainder of this section this equibiaxial deformation state is understood, and we refer to it simply by ε^0 and κ. While this is the case in the examples presented in this section, it is not the case when material anisotropy exists, the film is patterned in lines instead of as a blanket, or large deflections occur which result in geometric nonlinearity. These issues will be discussed briefly in later sections.

Single-Layer Plate. Consider a single-layer film with an internal biaxial mean stress $S = S_{11} = S_{22}$ and stress gradient $\beta = \beta_{11} = \beta_{22}$ that is subjected to a uniform temperature change T. This scenario models the deformation of a film after release from a substrate when subjected to a temperature change where before release the film had an internal stress S, and stress gradient β. In this case, after significant manipulation (35.11), with (35.12) and (35.13), yields:

$$\kappa = \frac{\beta}{M} \quad (35.16)$$

$$\varepsilon^0 = -\frac{S}{M} + \alpha T , \quad (35.17)$$

where $M = \frac{E}{1-v}$ is the biaxial modulus. As previously discussed, after release there is straining due to S and $\varepsilon^* = \alpha T$, and the plate bends with a biaxial curvature κ. The result of (35.16), and especially the simplification that results for beams (simply obtained by setting $\nu = 0$ and interpreting κ and ε^0 as the uniaxial curvature and midplane strain in the direction of the beam length) is commonly used in conjunction with measurements of the curvature after release to determine the stress gradient in a film.

Bilayer Plate. Another simple yet practically important special case is that of two isotropic layers of thicknesses t_1 and t_2, subject only to a uniform temperature change T. In this case, (35.11), with (35.12) and (35.13), yields:

$$\kappa = \frac{6\Delta\alpha T}{t_2} hm \left(\frac{1+h}{1+2hm\left(2+3h+2h^2\right)+h^4m^2}\right) \quad (35.18)$$

$$\varepsilon^0 = T\left(\frac{\alpha_1 mh\left(1+3h+3h^2+mh^3\right)}{1+2hm\left(2+3h+2h^2\right)+h^4m^2} + \frac{\alpha_2\left[1+mh\left(3+3h+h^2\right)\right]}{1+2hm\left(2+3h+2h^2\right)+h^4m^2}\right) , \quad (35.19)$$

where $h = \frac{t_1}{t_2}$, $m = \frac{M_1}{M_2}$, and $M_k = \frac{E_k}{1-\upsilon_k}$ (here $k = 1, 2$ corresponds to layers 1 and 2) and $\Delta\alpha = \alpha_2 - \alpha_1$.

It is often convenient to define a nondimensional curvature $\hat{\kappa}$:

$$\hat{\kappa} = \frac{\kappa\, t_2}{6\Delta\alpha T} \tag{35.20}$$

Equation (35.18) can then be written as

$$\hat{\kappa} = \frac{hm\,(1+h)}{1+2hm\left(2+3h+2h^2\right)+h^4m^2}\,. \tag{35.21}$$

Two limiting values of (35.21) are of interest. The first is the case where there is no elastic mismatch between the layers, i. e., $m = 1$. In this case (35.21) simplifies to:

$$\hat{\kappa} = \frac{h}{(1+h)^3}\,. \tag{35.22}$$

The other is the *thin-film limit* where $t_1 \ll t_2$. Expanding (35.18) in powers of h and retaining only the lowest order term recovers *Stoney's* [35.26] well-known result:

$$\kappa = \frac{6\Delta\alpha T}{t_2} hm$$
$$\hat{\kappa} = hm\,. \tag{35.23}$$

In the thin film limit, (35.19) reduces to $\varepsilon^0 = \alpha_2 T$.

Figure 35.11 shows the normalized curvature $\hat{\kappa}$ as a function of the layer thickness ratio h for various ratios of the biaxial moduli, m. The curves from top to bottom are for $m = 10$, 0.66, and 0.1. $m = 10$ and 0.1 represent a stiff film on a compliant substrate and a compliant film on a stiff substrate, respectively. $m = 0.65$ represents a gold film on a polysilicon substrate. The solid curves are the exact results from (35.21) and the dashed curves are the Stoney's result for the thin-film limit, (35.23). In the latter case, $\hat{\kappa}$ varies linearly with h, with slope m. The results in Fig. 35.10 show how the range of applicability of Stoney's result depends on both h and m; the range decreases for increasing h and decreasing m. For most microstructure applications where the layer thicknesses are comparable ($h > 0.1$), Stoney's result is not accurate unless the film is much more compliant than the substrate. Thus physically, Stoney's result is accurate when the substrate stiffness controls the deformation, either because it is much thicker or has much higher elastic moduli than the film. Over the years many papers have discussed *corrections* to Stoney's result to increase the region of validity in terms of layer thicknesses and modulus mismatch. We emphasize, though, that (35.18) is valid for arbitrary layer thicknesses and modulus mismatch and is quite simple to use itself. We refer the reader interested in more details to [35.7, 18, 22].

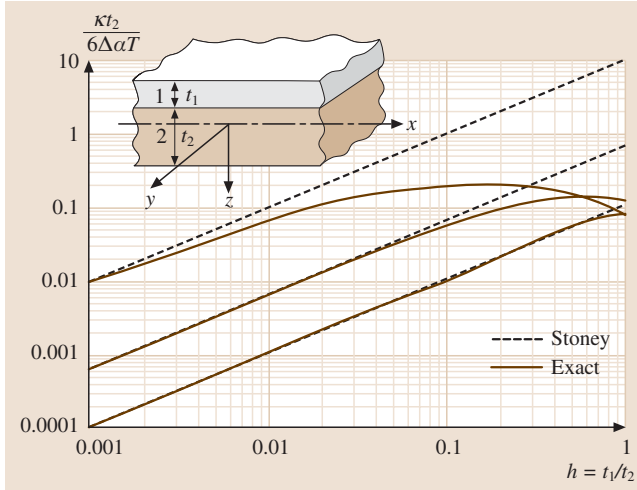


Fig. 35.11 Normalized curvature $\hat{\kappa}$ as a function of the film thickness ratio $h = t_1/t_2$ for three values of the biaxial modulus ratio, $m = M_1/M_2$. From *top* to *bottom*, the curves are for $m = 10$, 0.65 (gold/polysilicon), and 0.1. The *solid lines* are the exact solution, (35.21) and the *dashed lines* are Stoney's result in the thin-film limit, (35.23)

Equally important from a technological viewpoint is the equibiaxial stress distribution through the thickness $\sigma_x(z) = \sigma_y(z) = \sigma(z)$ obtained from (35.6):

$$\sigma(z)_k = M_k\left(\varepsilon^0 + z\kappa - \alpha_k T\right)\,, \tag{35.24}$$

where $k = 1, 2$ denotes the layer. In (35.24), κ and ε^0 are obtained from (35.18) and (35.19), respectively, and M_k and α_k are the values for the kth layer. In the thin-film limit the equibiaxial stress varies so little through the film thickness that it can be taken as constant, i. e., $\sigma_x = \sigma_y = \sigma$. This is the reason one often refers to the *stress* in a thin-film by a single number. In the thin film limit, the substrate (layer 2) thermal expansion dominates the deformation process, and so the film strains by an amount $\Delta\alpha T$, and the biaxial stress in the film is $\sigma_1 = M_1\Delta\alpha T$, or in terms of the curvature:

$$\sigma_1 = \frac{M_2}{6}\frac{t_2^2}{t_1}\kappa\,. \tag{35.25}$$

Note that, as written, the thermoelastic properties of the film do not enter (35.25). We will discuss the practical ramifications of this shortly.

Figure 35.12 shows the stress distribution through the thickness of both layers for a negative unit temperature change, (35.24), as a function of $h = t_1/t_2$ for

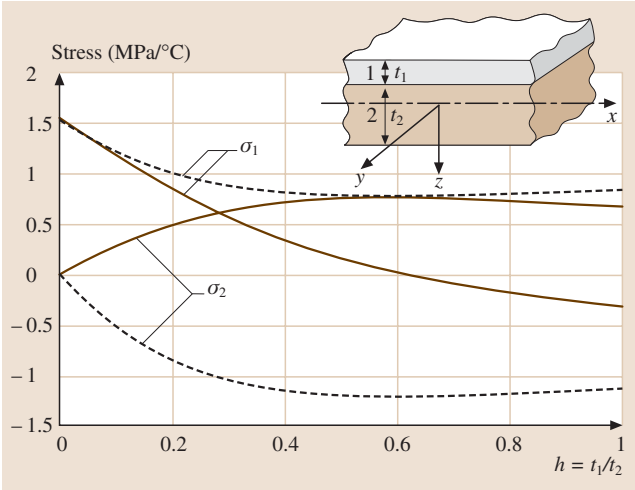


Fig. 35.12 Biaxial stress per unit temperature change, (35.24), for a gold/polysilicon bilayer ($m = 0.67$) for a 0.5-μm-thick gold film as a function of the polysilicon thickness $h = t_1/t_2$. Biaxial stresses are shown at four locations: the interface on the gold and polysilicon sides (*dashed lines*), and the surfaces of the gold and polysilicon (*solid lines*)

$m = 0.67$, and $t_1 = 0.5$ μm, representative of a gold film on a silicon substrate. Equation (35.24) shows, and we have previously discussed, that the biaxial stress varies linearly through each material. Thus we can completely specify the stress distribution by plotting it at four points: the top of material 1, the interface in material 1, the interface in material 2, and the bottom of material 2. Referring to Fig. 35.8, these positions are $z = z_0$, z_1, and z_2. Because of the perfect bonding between the two films, the total strain must be continuous across the interface. The elastic moduli are different for the two films, though, and so the stresses will be discontinuous at the interface. Hence we interpret $z = z_1$ as a position just above, or just below, the interface, and use M_k and α_k in (35.24) as appropriate. As $h \to 0$ the stresses at the top and at the interface in the film (layer 1) converge and take on a maximum value given by $\sigma = M_1 \Delta\alpha T$. The stress in the substrate vanishes in this limit. As the substrate thickness increases (increasing h) the stress in the film decreases with the maximum value occurring at the interface. The stress in the substrate increases at the interface but decreases on the bottom surface. In general this results in a stress distribution in the substrate that is compressive on the bottom surface and tensile at the interface. Depending on h, the stresses in the film are either tensile throughout or tensile on the interface and compressive on the surface. The difference in stress between the interface and the surface is a measure of the stress gradient in each layer; it increases with increasing h. For different values of m, the behavior with varying h is similar to that in Fig. 35.12, with the stress in the film increasing with increasing m.

Equation (35.25) is the theoretical basis of what is perhaps the most common technique to measure thin-film stresses – the wafer curvature method [35.6]. It involves measuring the curvature of a thick (compared to the film thickness) wafer with and without a film, and then using the curvature difference between the two in (35.25) to determine the film stress. It is particularly convenient because the film stress is obtained from the measured curvature regardless of the source of the stress and the thermoelastic moduli of the film. It is necessary, though, to know accurately the thickness of the film and substrate, as well as the elastic properties (biaxial modulus, $E/(1-\nu)$) of the substrate. An appealing feature of the method is that the elastic properties of the film, typically difficult to measure themselves, do not need to be known. This is because the much thicker substrate controls the deformation response.

Three-Layer Symmetric Plate. Considerable simplification also results for the technologically important case of a thin-film multilayer subject to a uniform temperature change T that consists of three layers, but the top and bottom layer are identical in terms of thickness ($t_3 = t_1$) and material properties ($M_3 = M_1, \alpha_3 = \alpha_1$). In this case, from (35.9) $[B] = [0]$, and thus the plate does not bend, i. e., $\kappa = 0$. Equation (35.11), with (35.12) and (35.13), then yields:

$$\varepsilon^0 = \frac{(2\alpha_1 mh + \alpha_2)\,T}{2mh+1}\,. \tag{35.26}$$

In the thin-film limit $h \to 0$, $\varepsilon^0 = \alpha_2 T$.

The equibiaxial stress distribution is such that it does not vary through the thickness within each layer, i. e., $\sigma(z)_3 = \sigma(z)_1 = \sigma_1$, and $\sigma(z)_2 = \sigma_2$. From (35.6) with $\kappa = 0$ and ε^0 from (35.26):

$$\sigma_1 = \frac{M_1 \Delta\alpha T}{2mh+1}$$
$$\sigma_2 = \frac{-2mh M_2 \Delta\alpha T}{2mh+1}\,. \tag{35.27}$$

In the thin-film limit, the stresses of (35.27) reduce to $\sigma_1 = M_1 \Delta\alpha T$ and $\sigma_2 = -2M_1 h \Delta\alpha T$.

Figure 35.13 shows the uniform biaxial stress in each material for a negative unit temperature change as a func-

tion of $h = t_1/t_2$ for $m = 0.65$, and $t_1 = 0.5\,\mu\text{m}$, again representative of a gold film on a silicon substrate. As $h \to 0$, the stresses in the films (layers 1 and 3) are given by $\sigma_1 = M_1 \Delta\alpha T$, and that in the substrate (layer 2) approaches zero. The stress in the films is tensile and decreases with increasing h while that in the substrate is compressive, increasing in magnitude with increasing h.

Beams. Our analysis thus far has focused on multilayer plates where both in-plane dimensions far exceed the thickness and where the transformation strains, and thus the stresses, are equibiaxial. When one in-plane dimension, say x, is much larger than the other, the structure is often referred to as a beam. In the classical treatment of the flexure of beams, it is assumed that only one nonzero normal stress component exists, σ_x. This greatly simplifies the analysis for homogeneous beams, but the situation is not as straightforward for multilayers with biaxial transformation strains. From the results in the previous section, one can obtain results for a beam theory by simply setting $\nu_1 = \nu_2 = 0$ in the formulae developed for biaxial stresses, but these results should be used with caution as the transverse component of the moment that results from the equibiaxial transformation strains cannot be neglected as shown by three-dimensional finite element calculations [35.27, 28]. *Swanson* [35.29] describes how to apply the results from a multilayer plate theory to beams and specifically focuses on the effect of the beam width, giving easily usable results for narrow and wide beams. The main results are an effective structural stiffness $(EI)_{\text{eff}}$ that can be used to replace EI (E is Young's modulus and I is the cross-sectional moment of inertia) in existing beam theory results, such as those that will be presented in Sect. 35.2 for electromechanics analysis. Insight into the behavior can be seen in the results in Table 35.2, which show measurements (using interferometry as described by *Dunn* et al. [35.30]) and predictions based on (35.18) for plates and the simplified version for beams obtained by setting $\nu_1 = \nu_2 = 0$. In all calculations both the gold and polysilicon are modeled as linear thermoelastic with isotropic material properties. Input parameters to the finite element calculations are $E_2 = 163\,\text{GPa}$, $\nu_2 = 0.22$, $E_1 = 78\,\text{GPa}$, $\nu_1 = 0.42$ [35.31]. The thermal expansion coefficients of the materials were assumed to vary linearly with temperature, and values at 100(24) °C used are $\alpha_2 = 3.1(2.6) \times 10^{-6}\,/°\text{C}$, and $\alpha_1 = 14.6(14.2) \times 10^{-6}\,/°\text{C}$ [35.31]. Although some uncertainty exists in the values of these material properties for the gold and polysilicon films, we think they are sufficiently accurate for the purpose of modeling the observed phenomena. Young's modulus and Poisson's ratio of the polysilicon are in line with many measurements over many MUMPS runs [35.32] and agree adequately with bulk polycrystal averages of single-crystal elastic constants. Good agreement exists between the predictions using plate theory and the measurements, while predictions based on beam theory are significantly less accurate, demonstrating the need to use plate, rather than beam, theory.

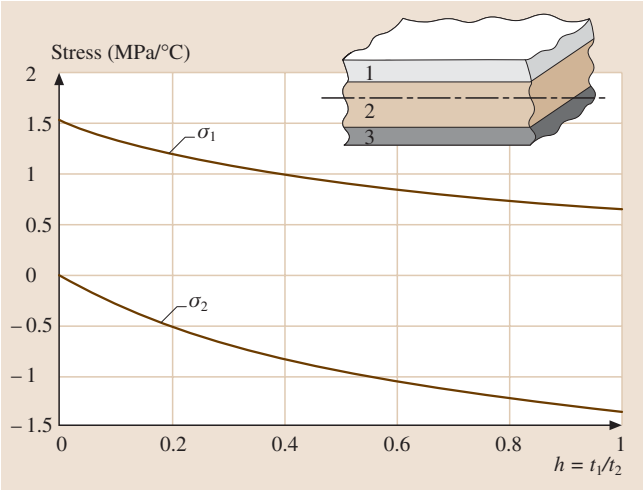


Fig. 35.13 Biaxial stress per unit temperature change for a gold/polysilicon/gold ($m = M_1/M_2 = 0.65$) three-layer film system. The calculations, (35.27), are for a 0.5-μm-thick gold film as a function of the polysilicon thickness $h = t_1/t_2$. Biaxial stresses are shown in the gold (*top curve*) and polysilicon (*bottom curve*)

Patterned Films. In microsystems applications, additional design freedom can be obtained by using multilayer films with patterned, rather than blanketed,

Table 35.2 Predicted and measured curvature per unit negative temperature change for gold/polysilicon microstructures as a function of the polysilicon thickness. In all cases the gold thickness is 0.5 μm. The microstructures are 300 μm long and 50 μm wide and cantilevered from one end

Polysilicon thickness (μm)	$d\kappa/dT$ ($m^{-1}/°C$) Measured	Predicted (plate)	Predicted (beam)
1.5	−5.36	−5.56	−4.86
3.5	−1.42	−1.44	−1.16

films. In general one can use the finite element method with multilayer plate elements to model the deformation and stresses in a multilayer with arbitrary patterned layers. A useful practical example is a beam of width w_2 that is covered with a film in the pattern of a stripe of width w_1 as shown in Fig. 35.14. A simple modification of (35.18) yields the curvature along the length of the patterned beam as a function of the nondimensional linewidth, $c = w_1/w_2$:

$$\kappa_{\text{patterned}} = \frac{6\Delta\alpha T}{t_2} hcm \times \left(\frac{1+h}{1+2hcm\left(2+3h+2h^2\right)+h^4c^2m^2} \right) . \tag{35.28}$$

Predictions from (35.28), normalized as $\eta = \kappa_{\text{patterned}}/T$, are shown in Fig. 35.14 for $m = 0.65$ (the gold/polysilicon bilayer) along with predictions from detailed finite element calculations and measurements over the entire range of $c = w_1/w_2$ [35.33]. Equation (35.28) accurately describes the effect of linewidth on curvature development and can be used to design bilayer beams.

Deflections. Thus far our discussion of deformation has been cast in terms of curvature. An important quantity for many applications where control of the deformation is important is the deflection of the multilayer. Formally this can be obtained by integrating the curvature-displacement relationship and applying the relevant boundary conditions. For example, assuming constant curvature, the tip deflection $w(x = L)$ for a cantilever beam of length L, is:

$$w(x = L) = \frac{\kappa L^2}{2} . \tag{35.29}$$

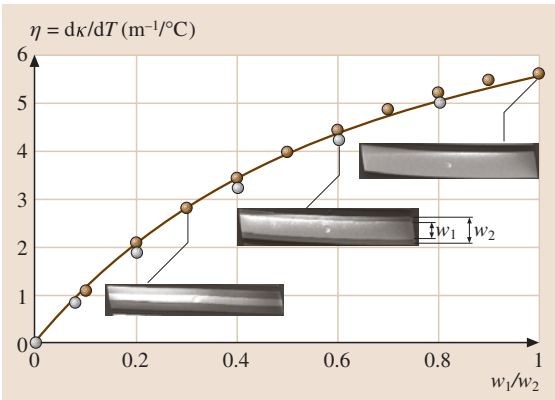


Fig. 35.14 Curvature per unit temperature change as a function of the ratio of gold to polysilicon film width w_1/w_2 for gold (0.5 µm thick)/polysilicon (1.5 µm thick) strip-patterned beams. *Grey circles* are finite element simulation results, the *solid line* is the analytical result of (35.28), and the *brown circles* are measurements

35.1.3 Nonlinear Geometry

In the previous section we noted that in general the deformation of a layered film microstructure subjected to transformation strains consists of three regimes but focused on the first, linear regime. Here we focus on the second and third regimes. To illustrate, consider the seemingly simple case of a plate with total thickness much less than the in-plane dimensions, composed of two isotropic layers with different material properties (elastic moduli and thermal expansion coefficients) subjected to a temperature change. Although we use a temperature change to demonstrate, the results hold for any type of transformation strains. In terms of the average curvature variation as a function of temperature change, three deformation regimes have been identified as illustrated in Fig. 35.15 [35.16–19, 22, 23]. The first regime, I, consists of a linear relation between the average curvature and temperature change where $\kappa_x = \kappa_y$, i. e., the average curvature is spherically symmetric. This deformation regime is characterized by both small transverse displacements and rotations, and so conventional thin-plate theory adequately describes the deformation. In Sect. 35.1.2 we outlined the analysis in this regime. The second regime, II, consists of a nonlinear relation between the average curvature and temperature, but again $\kappa_x = \kappa_y$. The behavior is due to *geometric nonlinearity* that results when the deflections become excessively large relative to the plate thickness, and they contribute significantly to the in-plane strains. It has been shown [35.15–19, 22, 23] that in these two regimes the symmetric deformation modes are stable. The second regime ends at the point when the deformation response bifurcates from a spherical to ellipsoidal deformation, i. e., $\kappa_x \neq \kappa_y$. At this point, the beginning of regime III, it becomes energetically favorable for the plate to assume the ellipsoidal shape because to retain the spherical deformation under an increasing temperature change requires increased midplane straining. After the bifurcation, the curvature in one direction increases while that perpendicular to it decreases; the plate tends toward a state of cylindrical curvature. This observation helps to explain the energetic argument, as unlimited cylindrical curvature can

be obtained with no midplane straining, while spherical curvature cannot. This discussion has been cast in the context of linear material behavior. Additional deformation regimes result if material nonlinearity is present, for example, yielding [35.18].

Analysis

The understanding described above derives from a number of studies with different technological motivations, primarily structural laminated composites and thin films for microelectronics. Most of these studies are analytical [35.15–19, 22, 23, 34] and build upon the original work of Hyer [35.13, 14]. To illustrate the computational approach in a reasonably simple setting, here we consider a two-layer plate with layer thicknesses t_1 and t_2. Each layer is isotropic and characterized by the Young's modulus E_i, Poisson's ratio ν_i, and thermal expansion coefficient α_i $(i = 1, 2)$. We consider a square plate with side length L; a similar analysis can be carried out for circular [35.22] or other plate shapes. The plate is subject to a uniform temperature change, T. In other words, the stress-free strains $\{\varepsilon^*\}_k$ are due to thermal expansion as given by (35.14).

In order to compute the deformed shape when the plate is subject to a temperature change we use the approach of *Hyer* [35.14] as applied by *Masters* and *Salamon* [35.16]. The basic idea is to assume an admissible displacement field $w(x, y)$ in terms of unknown parameters (d_i) that are suitably chosen to be consistent with observed deformation modes. Values of the parameters d_i are then determined via a Ritz procedure so as to minimize the total potential energy of the system. Different choices of the assumed displacement field have been considered, and details of the procedures are given in the above references. Such analyses are sufficient to qualitatively, and in many cases quantitatively, explain the three regimes of deformation shown in Fig. 35.15. In fact, quite simple closed-form expressions result for special cases that provide illuminating descriptions of observed phenomena (see for example [35.22]). A disadvantage of the analytical approaches is that for simplicity a displacement field consistent with a spatially constant curvature deformation mode is usually chosen. Additionally, these formulations are useful for only simple plate shapes. While this may be adequate for the structures demonstrated here, it is not for more complex in-plane shapes, of either or both layers, that arise in microsystems applications. In this more general case the best approach is probably to use the finite element method for the approximate analysis.

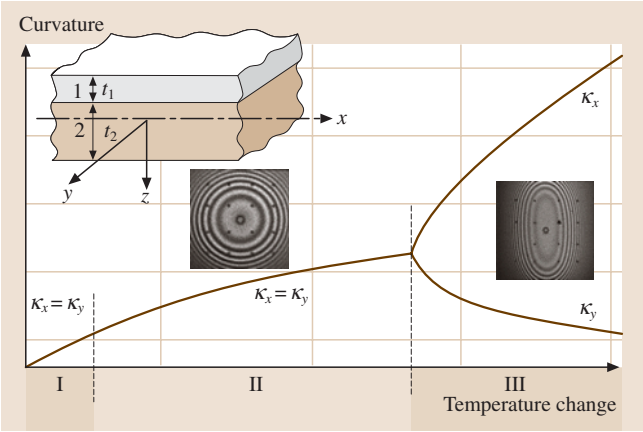


Fig. 35.15 General characteristics of the average curvature-versus-temperature change of a two-layer plate microstructure showing the three deformation regimes. The interferograms in the *inset* show measured displacement contours for gold/polysilicon plates in Regimes II and III

To proceed, we assume the transverse midplane displacement is of the form:

$$w^0(x, y) = d_1 x^2 + d_2 y^2 \,. \tag{35.30}$$

d_1 and d_2 are immediately recognized as one-half the curvature in the x- and y-directions, and the deformation is seen to be a constant curvature mode. The midplane displacements u^0 and v^0 are assumed to be described by third order polynomials in x and y, also with unknown constant coefficients d_3–d_8 to be determined:

$$\begin{aligned} u^0(x, y) &= d_3 x + \frac{d_5}{3} x^3 + d_7 x y^2 - \frac{d_1^2}{6} x^3 \\ v^0(x, y) &= d_4 y + \frac{d_6}{3} y^3 + d_8 x^2 y - \frac{d_2^2}{6} y^3 \,. \end{aligned} \tag{35.31}$$

The midplane strains are then computed from the nonlinear strain-displacement relations of the von Karman plate theory:

$$\begin{aligned} \varepsilon_x^0 &= \frac{\partial u^0}{\partial x} + \frac{1}{2}\left(\frac{\partial w^0}{\partial x}\right)^2 \\ \varepsilon_y^0 &= \frac{\partial v^0}{\partial y} + \frac{1}{2}\left(\frac{\partial w^0}{\partial y}\right)^2 \\ \gamma_{xy}^0 &= \frac{\partial u^0}{\partial y} + \frac{\partial v^0}{\partial x} + \left(\frac{\partial w^0}{\partial x}\right)\left(\frac{\partial w^0}{\partial y}\right) \,. \end{aligned} \tag{35.32}$$

The strains at any point through the thickness are then computed using the standard kinematic relations for thin plates, (35.2), and the stresses are computed from the strains using the conventional linear thermoelastic constitutive relations for each layer, (35.6) with $\{\varepsilon^*\}_k = \{\alpha\}_k T$. The first term in (35.32) is recognized as the conventional small deformation, and the second term arises from the large transverse deflections w^0, which result in straining of the midplane. This is shown schematically in Fig. 35.16 for the strain component ε_x^0; similar results are obtained for ε_y^0 and γ_{xy}^0.

The potential energy density of each layer is computed from the stress and strain in each layer, and the total potential energy of the plate is then computed by integrating it over the volume of the plate. This yields an expression for the potential energy of the plate in terms of the unknown coefficients d_i:

$$U(d_i) = \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} \int_{-t/2}^{t/2} \{\varepsilon\}^T \{\sigma\}_k \, \mathrm{d}z \, \mathrm{d}x \, \mathrm{d}y \,, \tag{35.33}$$

where the superscript T denotes the transpose of the 3×1 column vector. The coefficients d_i are determined by application of the principle of minimum potential energy:

$$\frac{\partial U(d_i)}{\partial d_i} = 0 \,. \tag{35.34}$$

This yields eight equations for the eight unknown d_i and thus solutions for the deformation response of Fig. 35.15. Complete details regarding this analysis can be found elsewhere [35.14, 16]; here we present only the pertinent results in what we hope is an accessible form. When the displacements are small, the solution of (35.34) and (35.35) recovers (35.18) and (35.19).

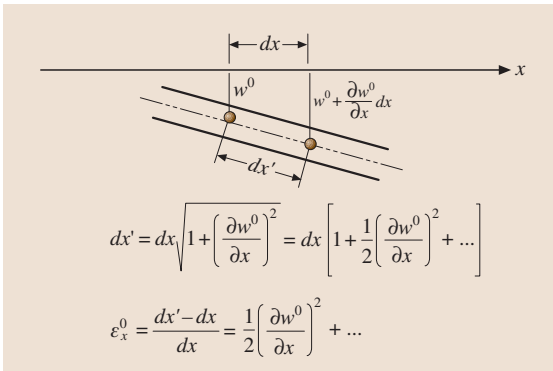


Fig. 35.16 Illustration of the development of midplane strain due to bending when nonlinear geometry effects arise

In the nonlinear but symmetric deformation regime II, the relationship between both the curvature and midplane strain, and the thermal expansion mismatch is too complex to present here. This is also the case in the nonlinear regime III after bifurcation. An important difference between the linear and nonlinear response, though, is that the nonlinear response depends on the plate size. The critical curvature at which bifurcation occurs, though, can be obtained explicitly and is given by:

$$\kappa_{\mathrm{cr}} = \frac{12\sqrt{2}}{L^2} \sqrt{\frac{6 + A_{12}/A_{11}}{1 + A_{12}/A_{11}}} \sqrt{\frac{A_{66} D_{66} - B_{66}^2}{A_{66}^2}} \tag{35.35}$$

for a square plate of side length t, where A_{ij}, B_{ij}, and D_{ij} are the composite moduli of (35.8). In the simplified case where there is no elastic mismatch between the layers and the elastic, response can be expressed in terms of the Young's modulus E and Poisson's ratio ν, (35.35) reduces to:

$$\kappa_{\mathrm{cr}} = 2\sqrt{6} \sqrt{\frac{6+\nu}{1+\nu}} \frac{t}{L^2} \,, \tag{35.36}$$

where $t = t_1 + t_2$. These results have also been obtained by *Finot* and *Suresh* [35.18]; similar results exist for a circular plate [35.22, 30].

The analysis just discussed is made tractable by assuming a displacement field consistent with a constant curvature. In the nonlinear case, this assumption is questionable. While the assumed displacement field used in the Ritz procedure could be modified to incorporate the dependence of curvature on position, perhaps the simplest approach to tackle these more general problems is to use the finite element method to solve the geometrically nonlinear equations over an arbitrary spatial domain. This is also the most viable approach for more complicated in-plane geometries, including patterned films. Complete details regarding the use of the finite element method to carry out such calculations can be found in *Dunn* et al. [35.30]. Here we will simply show the result of some of these calculations. The input parameters used are the same as those given in Sect. 35.1.2.

Deformation Behavior

In order to illustrate the nature of the deformation phenomena in the nonlinear regimes, Fig. 35.17 shows

contour plots of measured (using an interferometric microscope as described by *Dunn* et al. [35.30]) and predicted displacement fields $w(x, y)$ for square plates of four plate sizes that have been cooled from about 100 °C where they are flat to room temperature. Due to the thermal expansion mismatch between the polysilicon and gold, the $L = 150\,\mu\text{m}$ samples deform in a spherically symmetric manner; contours of constant transverse displacement $w(x, y)$ are nearly circles. This is also the case as the size increases to $L = 200$ and $250\,\mu\text{m}$, although the displacements increase as the plate size increases. At $L = 300\,\mu\text{m}$, though, the transverse displacement contours are not circular but elliptical, indicating that the deformation is no longer spherically symmetric. Thus when subjected to the same temperature change, both the magnitude and deformation mode depend on the plate size. As the in-plane dimension of the plate increases with the thickness held constant, the deformation mode changes from one of spherical symmetry to one more like cylindrical symmetry.

A complete picture of the deformation as a function of plate size and temperature change could be obtained by acquiring full-field displacements like those in Fig. 35.17 as a function of temperature. But a reasonable picture of the deformation behavior can be obtained in much simpler form by considering the *average* curvature in the x- and y-directions as a function of temperature. This is shown in Fig. 35.18, in which the average curvature in the x- and y-directions is plotted as a function of the magnitude of the temperature change during cooling. The temperature change is actually negative according to our convention, but its magnitude is plotted for convenience. The average curvatures are determined from the measured and computed $w(x, y)$ by averaging $\kappa_x = -\partial^2 w(x, 0)/\partial x^2$ and $\kappa_y = -\partial^2 w(0, y)/\partial y^2$ along the paths $y = 0$ and $x = 0$, respectively [35.30]. The x- and y-directions are taken to be aligned with the principal curvatures after bifurcation. The use of the average curvature as a measure of the plate deformation seems appropriate if the curvature is, or is close to, spatially uniform. This aspect will not be taken up in detail, but we refer the interested reader to [35.22,30] for details. It is apparent from both the measurements and predictions that in regime I, the curvature-temperature response is

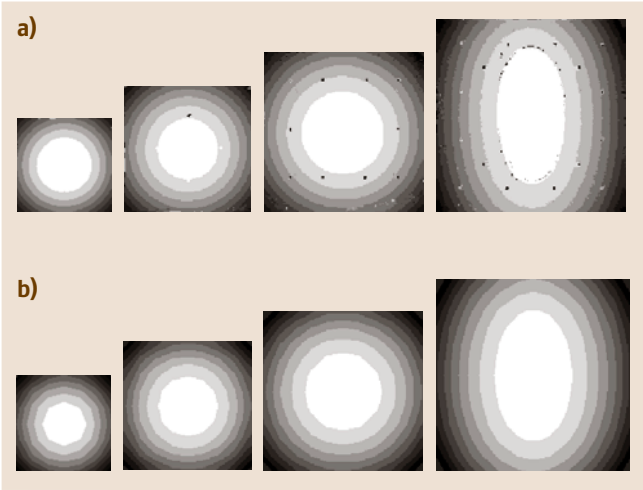


Fig. 35.17a,b Contour plots of the **(a)** measured, and **(b)** predicted transverse displacements $w(x, y)$ at room temperature following cooling from 100 °C for the four gold/polysilicon square plates: $L = 150$, 200, 250, and 300 μm from *left* to *right*. Each contour band represents a displacement of 0.23, 0.35, 0.45, and 0.6 μm for the $L = 150$, 200, 250, and 300 μm plates, respectively

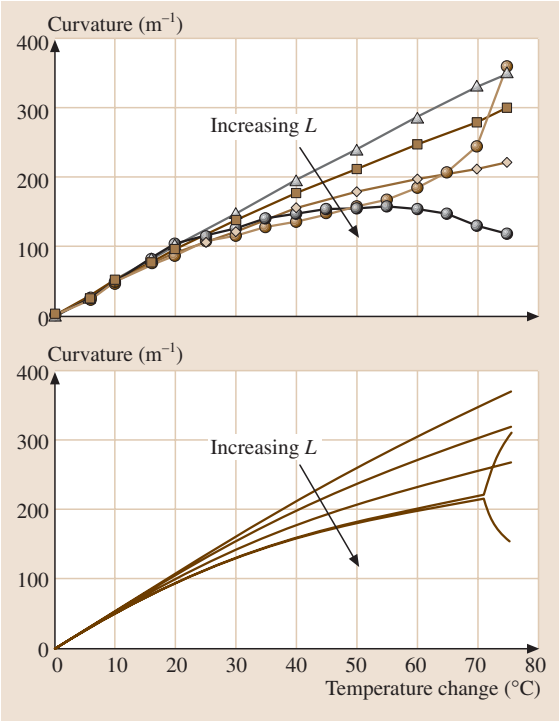


Fig. 35.18 Average measured (*top*) and predicted (*bottom*) curvature as a function of temperature change upon cooling from 100 °C to room temperature. The curves from *top* to *bottom* are for the $L = 150$, 200, 250, and 300 μm structures, respectively

independent of plate size and shape. In regime II, though, there is a strong dependence on plate size, and this is also the case in regime III. The major discrepancy between the measurements and predictions is the sharpness of the bifurcation for the $L = 300\,\mu\text{m}$ plate; it is quite sharp in the predictions but much more gradual in the measurements. To understand this we note that the source of the bifurcation is an *imperfection* of some sort that breaks the ideal symmetry, and in general it is difficult to accurately model the detailed imperfection.

Figure 35.19 demonstrates the connection between the thermomechanical loading (the temperature change), the geometry (plate size, L), and the boundaries between the three deformation regimes for gold/polysilicon plate microstructures. Specifically, it shows the temperature change necessary to initiate nonlinear effects (the transition between regions I and II), and bifurcation (the transition between regions II and III) as a function of polysilicon thickness when the gold film thickness is kept constant at $0.5\,\mu\text{m}$. Despite the fact that the constant curvature approximation becomes questionable for larger plate sizes, (35.35) is a good approximation as seen by the agreement with the finite element calculations, at least for the elastic mismatch and plate sizes considered here. In fact, although not shown, the simplified result of (35.36) for no elastic mismatch is in reasonable agreement with the finite element and the complete analytical results. The measurement for $L = 300\,\mu\text{m}$ is accurately described by both the analytical and finite element results.

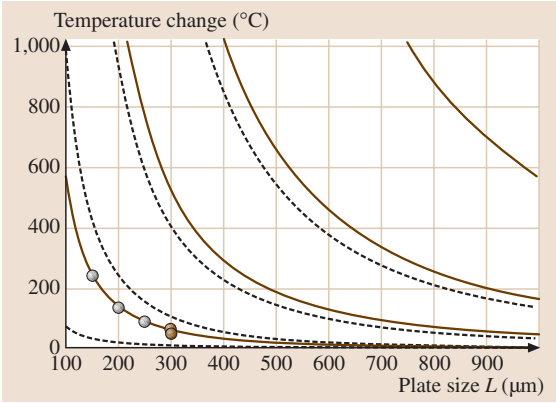


Fig. 35.19 Temperature change required for the initiation of nonlinear geometry effects (*dotted lines*) and critical temperature change for bifurcation (*solid lines*) as a function of square plate size and the thickness of polysilicon for the gold/polysilicon microstructures. Lines in each set from *left* to *right* represent $1.5\,\mu\text{m}$, $3.5\,\mu\text{m}$, $5.5\,\mu\text{m}$, and $8.5\,\mu\text{m}$ thick polysilicon, respectively. *Grey circles* are finite element calculations, and the *brown circle* is a measurement

In the analytical treatments discussed previously, it is assumed that the curvature is spatially uniform. The power of finite element calculations is that this requirement is relaxed and the spatial variation of the curvature can be studied theoretically. Full-field interferometry measurements allow one to study this experimentally as well. Briefly we discuss the resulting behavior for gold/polysilicon plate microstructures but refer the interested reader to [35.22, 30] for more details. In the linear regime, measurements and predictions both show that curvature is essentially uniform across the plate. In the nonlinear regime, though, the curvature varies appreciably with position, increasing by about a factor of two from the center to the periphery of the plate. The spatial nonuniformity of the curvature increases as the plate size increases. The spatial variation of the curvature raises concern regarding the suitableness of an analysis based on constant curvature. As mentioned previously, such an analysis may be adequate to describe the general deformation behavior but not for finer details.

35.1.4 Nonlinear Material Behavior

So far we have discussed in some detail the linear thermoelastic response of multilayer films and to a lesser degree, the effects of geometric nonlinearity. An additional complication is material nonlinearity. It can arise in numerous forms including plasticity, creep, stress relaxation, and evolution of the material microstructure (densification, grain growth, defect annihilation, etc.) during thermomechanical loading. In this section we briefly discuss nonlinear material behavior of multilayer films, focusing on the phenomena most relevant for the realization of reliable devices. The discussion in this section is concerned primarily with multilayer films where one of the films is a metal.

Nonlinear Material Behavior During Initial Thermal Cycles

As deposited, the material microstructure of a metal film is generally not stable. As the film is heated, microstructural evolution leads to the development of internal stress, which then can lead to changes in the curvature of the multilayer film. The general behavior is common to films made of many different metals, such as aluminum, copper, nickel, and gold. It has been studied extensively in the context of microelectronics applications (see for example [35.6, 35, 36]) where the substrate

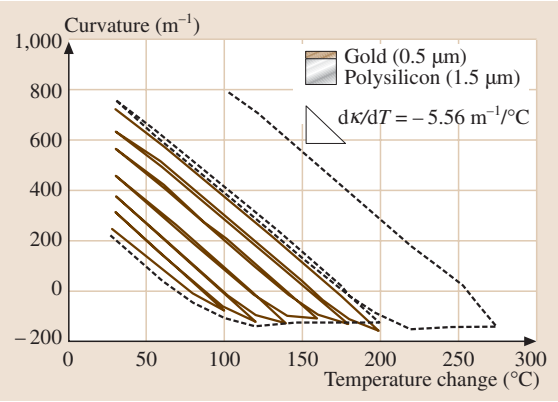


Fig. 35.20 Curvature versus temperature for gold (0.5-μm-thick)/polysilicon (1.5-μm-thick) 300 μm × 50 μm beam microstructures. The *solid lines* are tests with six cycles from room temperature to 200 °C with the maximum temperature in each successive cycle increasing by 20 °C. The *dashed lines* are tests with two cycles: the first from room temperature to 200 °C and then back to 30 °C, and the second to 275 °C and then back to 30 °C

is much thicker than the film and the thin-film limit results, which simplifies the interpretation of the results.

Figure 35.20 shows the general behavior of the curvature vs. temperature for a bilayer film consisting of a 0.5 μm gold film that was evaporated onto a 1.5 μm polysilicon film. During the initial heating the deformation is thermoelastic, characterized by a decrease in curvature with a constant $d\kappa/dT$, until the temperature reaches about 70 °C (the theoretical thermoelastic slope from (35.18) is shown on the figure). Between 70 °C and 100 °C, the curvature decreases at a much lower rate; in fact the curve flattens, i. e., $d\kappa/dT \approx 0$. Interestingly, between room temperature and 100 °C, the curvature changes from positive to negative, our convention being that a positive curvature means a beam with the gold on top is curved upward. The strong departure of the curvature from the thermoelastic behavior results because the microstructure of the evaporated film is not stable in the as-deposited condition. The temperature increase during the first cycle promotes microstructural changes in the film. These result in tensile straining of the film, which competes with the thermoelastic deformation (see for example [35.6, 36]). Upon cooling from 100 °C, the response is again thermoelastic, this time throughout the entire cooling process. Although it does not occur here, if the temperature change is large enough, yielding of the metal film can occur upon cooling. The curvature upon return to room temperature is greater than that initially at room temperature. It is important to understand this behavior because many post-fabrication and packaging processes expose a microstructure to elevated temperature excursions. During these, the film is susceptible to this nonlinear deformation. In the tests in Fig. 35.20, the material has not been taken to a temperature high enough to stabilize fully the gold microstructure in the first cycle. As a result, subsequent cycles going to successively higher temperatures appear to have the effect of continuing to stabilize the microstructure. The response during the second cycle is similar to the first with the exception that the temperature change required to initiate nonlinear behavior is increased. Indeed, over the range of temperature cycles studied, each subsequent cycle is similar, exhibiting thermoelastic response upon heating until a point where nonlinear κ-T behavior commences. The cooling process is again thermoelastic with the subsequent room temperature curvature increasing with each cycle. Also shown in Fig. 35.20 is a result for a second test on a nominally identical microstructure with only two cycles: the first from room temperature to 200 °C and back to 30 °C, and the second to 275 °C and back to 30 °C. The results of the first cycle to 200 °C neatly envelope the results of the set of tests with cycles up to 200 °C. This shows that although there is a path dependence of the deformation behavior, it is controlled by the maximum temperature reached during cycling, independent of how many increments are carried out to reach that temperature. The second cycle to 275 °C shows the same behavior observed in all cycles. The results in Fig. 35.20 have been obtained at a constant heating/cooling rate. In terms of the stabilization of the material microstructure that results in nonlinear material behavior, the rate, along with the time held at the elevated temperature, plays a significant role.

Figure 35.20 suggests that a simple description of the evolution of curvature with temperature can be obtained with parameters as shown in Fig. 35.21: the as-released curvature κ_r (and associated temperature, T_r), a critical temperature at which inelastic mechanisms are activated, T_{cr} (and the associated curvature, κ_{cr}), the slope of the inelastic κ-T behavior (assuming simple linear behavior) for $T > T_{cr}$, ξ, and the thermoelastic slope $\eta = d\kappa/dT$. The as-released curvature is determined primarily by the intrinsic stress development upon film deposition, cooling to room temperature, and release. These also influence T_{cr}. The parameter ξ is physically related to the curvature per unit temperature change due to inelastic mechanism(s) activated beyond T_{cr}. The situation in Fig. 35.20 is much simplified in that $\xi = 0$.

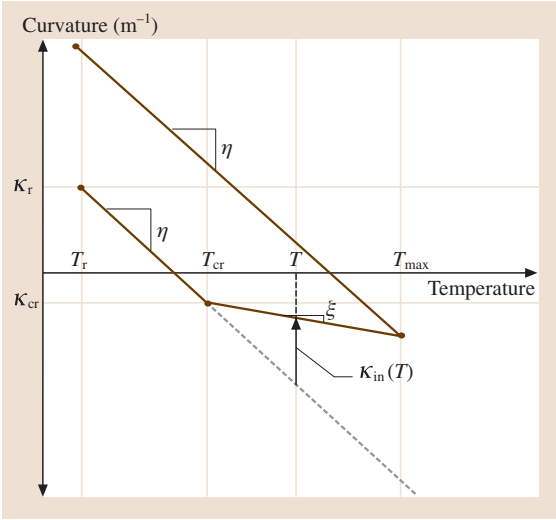


Fig. 35.21 Schematic diagram showing the parameters in the simple model that describes the curvature-versus-temperature behavior of the bilayer beam during thermal cycling after release

With knowledge of these parameters, one can compute the curvature following a defined thermal loading history. The qualitative behavior in Fig. 35.21 also results for a substrate with patterned lines (Sect. 35.1.2), and the same formalism can be used with κ_r, η, and ξ depending on linewidth [35.33]. For example, (35.28) gives the dependence of η on linewidth.

Plasticity

Plasticity in thin-film multilayers has been studied in depth for thin metal films on thick substrates (see, for example [35.6, 37, 38]), but little work has been directed toward freestanding multilayers. This is probably because, at least when subject to thermal loading, the stresses are typically smaller in freestanding multilayer films than in systems with thin films on thick substrates (see Fig. 35.12). Many interesting open issues exist with regard to the general issue of plasticity in microstructures including the understanding of size effects and strain gradient effects (see for example [35.39, 40]).

Creep and Stress Relaxation

For many microsystems applications, it is important to control accurately the deformation of thin-film microstructures over a significant period of time in order to meet performance and reliability requirements. This is especially important for microstructures subjected to thermal loading and/or operated at elevated temperatures. When one or more of the layers consists of a metal or polymer film, creep and stress relaxation in the film can significantly influence deformation and compromise device performance, and so their effects must be carefully considered. For example, *Miller* et al. [35.41] designed and fabricated microrelay switch arrays for RF communications applications using prestressed gold/polysilicon bimaterial beams as electrostatically actuatable switches. They observed a change in the switch shape and position over time and attributed it to stress relaxation in the gold. *Vickers-Kirby* et al. [35.42] report that creep in gold and nickel cantilever beams leads to voltage drops in micromachined tunneling accelerometers over time.

Creep and stress relaxation phenomena have been investigated in some detail for thin film on thick substrate systems, motivated primarily by microelectronics applications (see for example [35.11, 35, 43–46]). These studies have focused on measuring, understanding, and modeling the stress-versus-temperature behavior that occurs when a metal film-substrate system is heated from room temperature, then cooled back to room temperature, over one or a few cycles. The stress-temperature curves in the experimental works are typically obtained using the wafer curvature method that is now widely used to measure stress in thin films. During a typical test thermoelastic and inelastic mechanisms contribute to produce a complex nonlinear stress-temperature curve like that in Fig. 35.20. In most of these studies, the stress-temperature response is studied using wafer curvature measurements at a fixed heating or cooling rate so that rate dependent and independent phenomena are coupled in the response. Stress relaxation during an isothermal hold has been studied by [35.35, 43, 45, 47], among others. These studies all show that significant stress relaxation can occur over periods of only a few hours at modest temperatures of only about 100 °C for many metal thin-film systems. Due to the thin-film/thick-substrate system, the stress in the metal films is quite high, on the order of hundreds of MPa. Models incorporating power law creep of the metal film were successfully able to describe the observed response qualitatively and, to a large degree, quantitatively. The latter, however, sometimes required modification of the power-law exponent. *Shen* and *Suresh* [35.35] showed experimentally that a thin passivation layer on a metal film on thick silicon substrates can significantly reduce creep and stress relaxation. *Thouless* et al. [35.45] also studied the effect of a thin passivation layer on the stress

relaxation of thin films with a focus on the deformation mechanisms. Their experimental results suggest that the presence of a passivation layer on the surface of a film can have a substantial effect on relaxation rates likely by suppressing mechanisms associated with diffusion and dislocation motion.

Thermomechanical Fatigue

It is also important to understand the development of damage and its effect on deformation during cyclic thermomechanical loading. To date, only very limited attention has been directed toward this issue. *Zhang* and *Dunn* [35.33] showed that if gold/polysilicon multilayers were cycled (Fig. 35.20) to an elevated temperature and then cooled to room temperature, they followed the thermoelastic path on subsequent cycling to a temperature below the maximum reached during the initial thermal cycle. Presumably over this range of temperature and time, the gold microstructure has been stabilized by the first cycle, and the polysilicon microstructure is not changing. This was confirmed for only a few, perhaps ten, cycles. *Gall* et al. [35.48] have shown that for thousands of cycles, though, a gradual shift of the thermoelastic curve downward is observed, possibly due to creep and stress relaxation in the gold, although the thermoelastic slope is maintained.

35.1.5 Other Issues

There are a number of additional issues regarding the thermomechanical behavior of thin-film microstructures that warrant discussion. Due to space limitations, however, we will only mention them and refer the reader to appropriate references.

In addition to the deformation itself, the stiffness characteristics of a microstructure, such as for example a beam or plate, can depend significantly on the stress in the beam. *Senturia* [35.49] describes the phenomena and discusses implications for microstructure design. This includes tensile stresses that tend to stiffen beams and plates as well as compressive stresses that can lead to buckling.

An important strategy has emerged to characterize the effects of processing on the residual stress state and thermomechanical behavior of films: the use of on-chip test structures to extract material properties including film stresses. Many test structures, both passive and active, have been developed to measure residual stresses (tensile and compressive), stress gradients, elastic moduli, strength, toughness, thermal expansion, and many others. Nice surveys of many test structures are given by Masters et al. [35.12] and *de Boer* et al. [35.50] where further references to more extensive details can be found.

35.2 Electromechanics of Thin-Film Microstructures

Electromechanics, the coupling of electrostatics and mechanics, is perhaps the most common of the energy couplings used in microsystems technology for sensing and actuation. In this section we summarize the contemporary application of electromechanics, the analysis of fundamental electromechanics problems, and the practical issues that arise in the design of an electromechanical microsystem.

35.2.1 Applications of Electromechanics

Our reference to electromechanics will be limited to the coupling of electrostatics and the mechanics of constrained elastic media such as a cantilever beam. In this case, the electrical energy is transformed into strain energy of the elastic media but with typically negligible energy dissipated due to material damping. An alternative application of electromechanics is the coupling of electrostatics with rigid body mechanics. In this case the mechanical structure is not constrained by a compliant support to the substrate but by a kinematic constraint, which is a pin joint between multiple structural layers or anchor between the substrate and a structural element. An example of these rigid body constraints is the anchor in the electrostatic micromotors. In this case, the electrostatic energy is dissipated by friction (Coulomb type friction between structural elements and pin joints, other structural elements, and/or the substrate) and is stored in the kinetic energy of the rigid body. In this case, the energy stored as strain energy in the structural element is by design negligible. In both electromechanics cases, the energy is dissipated by viscous forces under dynamic conditions. Finally the coupling of electrostatics and mechanics can more broadly; consider the coupling of electrostatics and fluids or electro-thermo-mechanics.

Microsystems have enabled many different commercial applications of electromechanics. These applications have included pressure sensors, accelerometers, gyros, resonators, micropumps, optical mirrors, optical shutters/VOAs, and DC and RF switches. In these

examples, electrostatic forces are produced between a stationary electrode and a movable electrode, which is often the proof mass for an accelerometer or the diaphragm for a pump. The electrostatic forces are imparted by a voltage difference between the movable and stationary electrode. The electromechanics is typically implemented in two ways, with the first characterized by a stable, analog behavior, and the second characterized by an unstable, digital behavior. The stable behavior is implemented as a self-test capability for accelerometers such that a specific, stable displacement is produced with the application of the voltage. The unstable behavior is implemented with a switch so that when the instability point (snap-in or pull-in) is reached, the switch will instantaneously close.

Electrostatic actuation has been used for switches and microrelays [35.51–53], accelerometers [35.54, 55], micropumps [35.56, 57], and micromirror [35.58–65]. The specific case of switches and relays involves multiple regions of electromechanical behavior and represents one of the richer electromechanics design problems. The general behavior is shown in Fig. 35.22. The first region is the stable actuation region. In this region, the switch is actuated to an instability point in the voltage displacement curve called the pull-in or snap-in voltage. Up to the instability, the displacement increases monotonically until the pull-in voltage is reached. At the pull-in voltage, the switch instantaneously closes the gap because of the instability and establishes contact with the substrate. In the case of the switch, this means the closure of a contact. The electromechanics problem continues with a new set of boundary conditions in the next region because the switch is now "simply" constrained at the contact and fixed at its anchors to the substrate. In Fig. 35.22, the switches tip displacement is shown, which is the reason the displacement is unchanged as the voltage increases. Region 2 is then characterized by a constant displacement for increasing voltage. The shape of the beam continues to change as the voltage is increased to the point that the electrodes could short if necessary precautions are not taken.

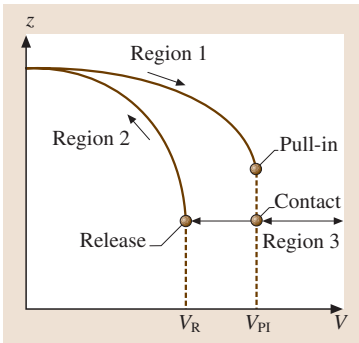


Fig. 35.22 Characteristic electromechanical hysteresis curve showing pull-in, contact, and release voltages

If the first consideration of the switch design is pull-in voltage, the second consideration of the switch design is the development of contact force, which is directly related to the contact resistance. The electromechanical analysis is continued beyond the pull-in voltage once contact is established to develop an increasing contact force. The electrostatic force is distributed between the anchor and the contacts, assuming other portions of the switch are not in contact. The electrostatic force increases as the square of the voltage difference, so the force acting on the contact will increase in proportion to this. Of course other physical limitations must be considered such as gas breakdown or a secondary pull-in that cause shorting between electrodes. The contact areas of the switch that are not the contacts should be minimized, so this does not detract from the contact force.

A third design consideration is the overdrive capability, defined as the maximum voltage the switch control electrodes can support before they touch and short. In some cases, the electrodes may be isolated from each other by a thin dielectric film, whose implications will be discussed in a later section. As suggested in the later section, the insulation material can be patterned to reduce the contacting area between the electrodes and the insulating material, in order to reduce charge accumulation. With the patterned isolation, the switch could be driven by such a large voltage to short the uninsulated regions. This failure mode could be evaluated by the same electromechanics analysis.

A fourth design issue would be the switch's release voltage, which is the voltage when the contact force reaches zero and the switch opens. This is shown as the beginning of region 3, where the deformation is reduced as the voltage is decreased. The release voltage is lower than the pull-in voltage because the electrostatic forces are much greater for the reduced gap than compared to the larger gap just before the pull-in instability. A fifth design consideration is the switch's self-actuation voltage, which is that which occurs across open contact and unintentionally causes the switch to close. By design, the switch's self-actuation voltage should be maximized in the electromechanical analysis. A final design issue is the determination of breakdown limits across the switch's control electrodes. This is not limited to the voltage supported across the equilibrium gap of the control electrodes and needs to be considered for

different cases. The breakdown will be described by the Paschen curve or by vacuum breakdown proportional to the gap.

35.2.2 Electromechanics Analysis

An electromechanics analysis requires modeling of the structural domain, the electrostatic domain, and the coupling between the electrostatic fields and the structural elements. It has been approached at three levels of complexity. The simplest approach is the development of lumped parameter models, which have provided the basic understanding of instability, the impact of nonlinear springs, and the pull-in voltage-gap relationship. The lumped parameter model is comprised of the parallel plate capacitor and a spring (linear or nonlinear). One plate is constrained in space, by a substrate for example. The second plate is elastically constrained to one degree-of-freedom, typically, of motion, which is usually closing the "air" gap. The spring element represents the effective spring stiffness K of the structural element that could be a beam, a plate, a membrane. The parallel plate capacitor represents the electrostatic coupling between the fixed and moving elements.

Increasing in complexity, the second level of analysis considers continuum structural elements with a distributed electrostatic load. For example, the solution for an isotropic homogeneous beam would satisfy the Euler–Bernoulli beam equation:

$$EI\frac{\mathrm{d}^4 w}{\mathrm{d}x^4} = q(x)\,, \tag{35.37}$$

where E is Young's modulus, I is the structure's moment of inertia, $w(x)$ is the transverse deflection of the beam, x is the coordinate along the length of the beam, and $q(x)$ represents the distributed electrostatic load. For a multilayer film, an effective EI can be used, as mentioned in Sect. 35.1.2. The third level of analysis is to model the fully coupled problem as continuum solids. The approach has used different implementations of finite difference (FD), finite element (FE), and boundary element (BE). A typical example would be to implement a finite element code to solve the continuum mechanics problem and a boundary element code to solve the electrostatics problem. Another implementation would use finite element methods to solve both the mechanics and the electrostatics problems. In each case, a similar algorithm is followed for developing a self-consistent solution to the coupled electromechanics analysis.

The coupled electromechanics problem can be described simply as a pair of conductors, where one conductor has a fixed constraint and the second conductor is attached to an elastic structure, such as a beam, a plate, a membrane, or an elastic substrate. The conductors support a voltage difference between them that causes charge to be induced on the surface of the conductors. This charge produces an electrostatic force that attracts the two conductors together. As the elastically supported conductor deforms toward the fixed conductor, the charge redistributes and modifies the field, thereby the force distribution. The process continues until a redistribution of charge is no longer necessary to maintain equilibrium. The typical algorithm begins by calculating the displacements of the undeformed geometry, which is updated before proceeding to the electrostatic analysis. In the electrostatic analysis, the surface charge density distribution is calculated and then used to calculate the electrostatic forces. The electrostatic force distribution is the important parameter because it is passed to the mechanical analysis as the loading condition on the conductors in the undeformed state to deform the structure until a state of equilibrium is reached.

Electromechanical Systems Energy Balance

A general analytical methodology can be applied to many different structures including parallel plate actuators with linear or nonlinear springs and plates supported by torsional suspensions [35.66]. They consider a single input system with charge control or voltage control. To begin, a generalized actuator is considered that consists of two conductors. One conductor is fixed. The second conductor is constrained by an elastic support.

In the charge controlled actuator, the total energy for the generalized actuator is written in terms of the mechanical energy as a function of the generalized coordinate χ and of the electrical energy stored in the capacitance of the actuator. The total energy is:

$$U_\mathrm{T}(\chi, Q) = U_\mathrm{M}(\chi) + \frac{Q^2}{2C(\chi)}\,, \tag{35.38}$$

where U_T is the total energy, U_M is the mechanical energy, and the electrical energy is determined by the charge Q, and the capacitance $C(\chi)$. The first derivative of the total energy with respect to the generalized coordinate set equal to zero determines the equilibrium. The second derivative of the total energy with respect to the generalized coordinate set equal to zero determines the stability. The first and second derivatives are determined and then combined to determine what is referred to as

the charge-controlled pull-in equation:

$$\frac{\partial U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi}\frac{\partial^2}{\partial \chi^2}\left(\frac{1}{C(\chi_{\mathrm{PI}})}\right) - \frac{\partial^2 U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi^2}\frac{\partial}{\partial \chi}\left(\frac{1}{C(\chi_{\mathrm{PI}})}\right) = 0 \qquad (35.39)$$

and the pull-in charge Q_{PI}:

$$Q_{\mathrm{PI}} = \left(\frac{-2\dfrac{\partial U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi}}{\dfrac{\partial C(\chi_{\mathrm{PI}})}{\partial \chi}}\right)^{1/2} . \qquad (35.40)$$

In the voltage controlled case, the total co-energy for the generalized actuator is used because the capacitor is nonlinear, and the actuator is voltage controlled. The co-energy is written in terms of the capacitance between the generalized conductors, the voltage across the conductors, and the mechanical energy. The total co-energy is:

$$U_{\mathrm{T}}^{*}(\chi, Q) = \frac{1}{2}C(\chi)V^2 - U_{\mathrm{M}}(\chi) . \qquad (35.41)$$

The relationship between energy and co-energy is shown in Fig. 35.23.

In a similar process to the charge-controlled system, the voltage-controlled pull-in equation is determined by the first derivative of the total co-energy with respect to the generalized coordinate and by the second derivative of the total co-energy with respect to the generalized coordinate. The equations for the first and second derivative are combined to determine the voltage-controlled pull-in equation

$$\frac{\partial U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi}\frac{\partial^2 C(\chi_{\mathrm{PI}})}{\partial \chi^2} - \frac{\partial^2 U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi^2}\frac{\partial C(\chi_{\mathrm{PI}})}{\partial \chi} = 0 \qquad (35.42)$$

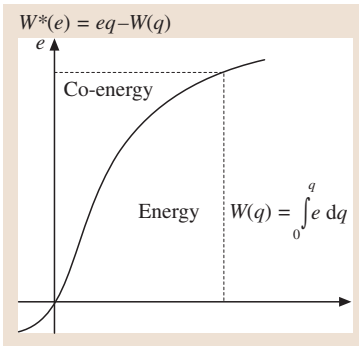


Fig. 35.23 Effort-Displacement curve showing the relationship between energy and co-energy

The pull-in voltage V_{PI} is:

$$V_{\mathrm{PI}} = \left(\frac{2\dfrac{\partial U_{\mathrm{M}}(\chi_{\mathrm{PI}})}{\partial \chi}}{\dfrac{\partial C(\chi_{\mathrm{PI}})}{\partial \chi}}\right)^{1/2} . \qquad (35.43)$$

The application of this approach, to either the voltage-controlled or the charge-controlled case involves determining the mechanical energy and the capacitance. From the mechanical energy and the capacitance, the pull-in parameters and voltage are determined.

Lagrangian Approach to Electromechanics

Li and *Aluru* [35.67] and *Aluru* and *White* [35.68] have proposed a Lagrangian approach for both the electrostatic analysis and the mechanical analysis. Their algorithm begins by calculating the structural displacements for the undeformed geometry. They proceed to the electrostatic analysis to calculate the charge distribution on the conductors in the undeformed state. The charge distribution is used to calculate the electrostatic forces on the undeformed geometry until the system reaches equilibrium. The first advantage of their approach is that the algorithm does not require the structural geometry to be updated. A second advantage is the elimination of the integration error, wich occurs because flat panels are used in other approaches to approximate curved surfaces. The overall algorithm is shortened because the interpolation functions do not have to be updated as the structure shape changes.

35.2.3 Electromechanics – Parallel-Plate Capacitor

The classic element for understanding electromechanics is the simple parallel-plate capacitor model with one plate fixed and the second, moveable, plate, suspended by a spring. This basic electromechanics element is shown in Fig. 35.24. The capacitor is described by an area, A; an initial gap, g_0; and a dielectric constant of ε in the air gap (where ε is the product of the relative dielectric of the media filling the gap, ε_{r}, and the permittivity of free space, $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m). The moveable plate is shown suspended by the spring, with spring constant K, and a single degree-of-freedom represented by the coordinate z. In this case, the equilibrium position is represented by having no charge on the plates (i. e. no electrostatic force) and no displacement of the spring (i. e. no spring force). It should be noted that as the spring length grows by increasing z, the gap decreases. The behavior of this simple model is described in detail in many sources including [35.49, 66].

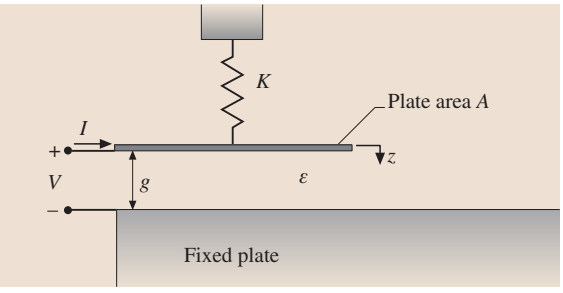


Fig. 35.24 Lumped parameter parallel plate capacitor electromechanical model

The parallel plate capacitor can be analyzed with either a voltage source or a current source. With a voltage source, the charge on the capacitor plates is

$$Q = \frac{\varepsilon A}{g} V \,, \tag{35.44}$$

where Q is the charge on the capacitor plates, ε is the dielectric constant in the gap, A is the capacitor plate area, g is the instantaneous gap between plates, and V is the voltage across the plates.

The electrostatic force on the capacitor plates is determined by the dielectric constant, the capacitor plate area, the instantaneous gap, and the voltage across the capacitor:

$$F = \frac{\varepsilon A V^2}{2g^2} \,. \tag{35.45}$$

The electrostatic force scales as the voltage squared and inversely diminishes as the instantaneous gap squared. This is the force that acts on the movable plate; hence the spring stretches (or is relaxed if the voltage is lowered). The amount of stretch in the spring is:

$$z = \frac{F}{K} \,. \tag{35.46}$$

The instantaneous gap is

$$g = g_0 - z = g_0 - \frac{\varepsilon A V^2}{2Kg^2} \,. \tag{35.47}$$

Electromechanics – Parallel Plate Capacitor with External Forcing

Nemirovsky et al. [35.66] examined the case of constant applied force acting on the suspended capacitor plate of a parallel-plate electrostatic actuator. This situation represents a force-balanced system, in which the electrostatic force is used to balance a pressure load acting on a membrane or an inertial load (e.g. acceleration induced) on a suspended proof mass. With an applied mechanical force, the mechanical energy is:

$$U_{\mathrm{M}}(z) = \frac{1}{2} K z^2 - F_{\mathrm{ext}} z \,, \tag{35.48}$$

where F_{ext} is the externally applied force. The solution to the pull-in equation is

$$\varsigma_{\mathrm{PI}} = \frac{z}{z_{\max}} = \frac{1}{3}\left(1 + 2\frac{F_{\mathrm{ext}}/K}{z_{\max}}\right) \,, \tag{35.49}$$

where ζ_{PI} is the ratio $z/z_{\max}$. It can be seen when the applied force is zero. Equation (35.49) reduces to the familiar result of $z = z_{\max}/3$. The pull-in voltage is

$$V_{\mathrm{PI}} = \left(\frac{8K\left(z_{\max} - \dfrac{F_{\mathrm{ext}}}{K}\right)^3}{27\varepsilon_0 A}\right)^{1/2} \,. \tag{35.50}$$

Electromechanics – Parallel Plate Stabilization

An important practical consideration is the stabilization of the parallel plate electrostatic actuator. *Seeger* et al. [35.69] and Nemirovsky et al. [35.66] showed that the addition of a series capacitance (feedback capacitance) stabilizes the actuator such that it does not pull-in to the substrate. The series capacitance and the variable actuator capacitance combine to form a voltage divider that provides passive control of the voltage across the actuator. In this case, Nemirovsky et al. [35.66] identified three regimes, defined in terms of δ, which is the ratio of the initial actuator capacitance to the series capacitance (Nemirovsky et al. [35.66] had defined the series capacitance in terms of a dielectric layer on one capacitor plate). The first regime is defined for $0 < \delta < 0.5$, when the series capacitance is less than half the nominal actuator capacitance. In this case, the actuator is stable so the pull-in effect is eliminated. The second regime is defined by $\delta > 0.5$ and the pull-in occurs between 1/3 of the gap and the full distance of the gap. In the third regime $\delta \gg 1$, which means the series capacitance is much greater than the nominal capacitance. The pull-in parameter is given by:

$$\zeta_{\mathrm{PI}} = \frac{1}{3}\left(1 + \frac{1}{\delta}\right) \,, \tag{35.51}$$

which reduces to $\zeta_{PI} = \zeta_{max}/3$ when the series capacitance is eliminated. The pull-in voltage is:

$$V_{PI} = \left[\frac{8Kz_{max}^3}{27\varepsilon_0 A} \left(1 + \frac{1}{\delta}\right)^3 \right]^{1/2} . \quad (35.52)$$

Electromechanics – Extending the Range of Motion

Electromechanical systems have two obvious characteristics: nonlinearity and instability. If one desires digital operation or fast switching times, the inherent electromechanical instability is a significant benefit. On the other hand if analog control is desired, the control is faced with both the nonlinearity and the instability. In both cases, the range of motion of the electromechanical device is limited. The range of motion to the instability was extended by the addition of a series capacitor, but this did not address the linearity. *Toshiyoshi* et al. [35.59] addressed the linearity of electrostatically actuated mirrors in an optical scanner. Their's was a two degree-of-freedom scanner, so the increased linearity was necessary to improve the crosstalk between the two different scanner modes. The scanner in question is a square plate suspended by torsion beams within a frame that is itself suspended by another pair of torsion beams. The plate and frame are suspended above the substrate, which has four stationary electrodes defined in the four quadrants of the plate as shown in Fig. 35.25. To actuate the mirror with increased linearity, a bias voltage V_{bias} was applied between each electrode and the plate. Independent differential control voltages are superimposed on the bias voltage supplied between each electrode and the plate. The driving voltages are described by two independent differential voltages such that $V_{diff1} = (V_x + V_y)/2$, $V_{diff2} = (-V_x + V_y)/2$, $V_{diff3} = (-V_x + -V_y)/2$ and $V_{diff4} = (V_x + -V_y)/2$. A second approach taken to improve the linearity was to consider the shape of the mirror, because of the geometric relationship that determines the net electrostatic torque. They considered such simple geometries as square, diamond (square rotated by 45), and a circular plate, as shown in Fig. 35.26, but more complex shapes could be considered. The square plate is presumed to have worse linearity because as the plate rotates the corners deflect toward the control electrodes, which increases the applied torque and further distorts the shape of the mirror. This suggested the removal of the corners could contribute to increased linearity and reduced distortion of the mirror surface. They found the circular plates reduced the distortion by 50% over the rectangular and diamond shaped plates, which is significant for the optical performance. The authors defined distortion as the normalized angle deviation from the ideal value. It is normalized relative to the maximum scan angle. *Nadal-Guardia* et al. [35.70] have described extending the range of travel by using current drive methods instead of voltage drive.

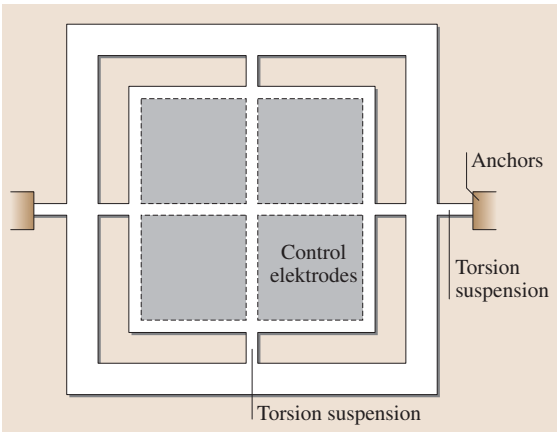


Fig. 35.25 Two degree-of-freedom torsional electromechanical actuator with a rectangular plate and a stationary electrode divided into four sections

35.2.4 Electromechanics of Beams and Plates

The basic understanding of electromechanics and pull-in phenomena has been developed around the simple parallel plate capacitor model developed above. Many MEMS designers innately understand that the electrostatic force is proportional to the square of the voltage, is inversely proportional to the square of the gap, and

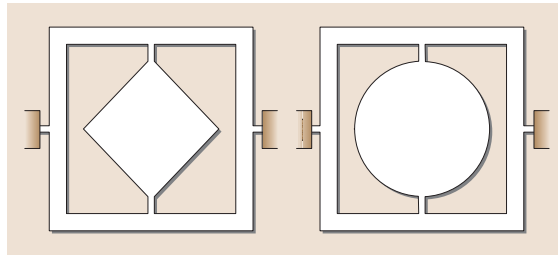


Fig. 35.26 Alternate two degree-of-freedom torsional, electromechanical actuators with diamond and circular plates

that the plate will pull in after displacing 1/3 of the gap. While this is good to first order, it is not adequate for the electromechanical design of other structures such as cantilevered beams/plates, beams/plates with multiple supports, plates with torsional supports, and other structures.

O'Brien et al. [35.71] and *Choi* and *Lovell* [35.72] have studied the electromechanics of cantilever beams by presenting two improved models for the cantilever beam electrostatic actuators. In both cases, the cantilever beam is a homogeneous material anchored to the substrate at one end and otherwise freely suspended along its length, which is supported above a stationary electrode. The new models are compared to the parallel plate capacitance model and a simple extension of the parallel plate model to a cantilever beam. The cantilever beam model uses a simple cantilever of width, W, length, L, thickness, T, and initial air gap of Z_0. The parallel plate capacitance model is based on a capacitance determined by the length and width of the beam. The beam deflection was represented by the displacement function of a beam with a force applied at the beam tip. The beam spring constant was determined by the ratio of the force at the beam tip to the displacement at the beam tip (maximum displacement). The capacitance was determined by integrating the parallel plate capacitance for a differential length element along the length of the beam, where the beam gap is defined by the initial gap in the mechanical displacement function. This showed the beam tip displaced 45% of the initial gap before pull-in occurred. In this case the pull-in voltage can be expressed as a function of the beam's Young's modulus E, cross-sectional moment of inertia I, length L, and width W, the initial gap Z_0, and the dielectric constant in the air gap as

$$V_1 = \sqrt{\frac{18EIZ_0^3}{5\varepsilon_0 L^4 W}} = 0.5477\sqrt{\frac{ET^3Z_0^3}{\varepsilon_0 L^4}} \quad \text{(Volts)}\,. \tag{35.53}$$

In the first improved model, the cantilever beam is defined by the same characteristics, but the electrostatic force acting along the length of the beam is replaced by an equivalent bending moment M applied at the beam tip. The beam displacement was determined for a uniform, homogeneous beam with a bending moment applied at its end. The beam's spring constant was determined by the ratio of the force at the beam tip and the displacement at the beam tip. The capacitance was determined by integrating the parallel plate capacitance for a differential length element along the length of the beam, where the beam gap is defined by the initial gap in the mechanical displacement function. This model showed the beam tip displacement was approximately 46% of the initial gap before pull-in occurred. In this case the pull-in voltage is

$$V_2 = \frac{43}{50}\sqrt{\frac{18EIZ_0^3}{5\varepsilon_0 L^4 W}} = 0.471\sqrt{\frac{ET^3Z_0^3}{\varepsilon_0 L^4}} \quad \text{(Volts)}\,. \tag{35.54}$$

In the second improved model, the cantilever beam was defined by a rigid beam, whose anchor was replaced by a simple hinge and continuum elasticity replaced by an equivalent spring at the beam's free end. Three new variables were introduced to describe the behavior in terms of radial coordinates centered at the hinge. The two radial coordinates describe the beginning of the electrode at r_0 and the end of the beam/electrode at r_1. The angle θ was introduced to describe the angular displacement of the beam. In this model, the tip of the beam displaced 44% of the initial gap before pulling into the substrate. The pull-in voltage is estimated by (35.55) or (35.56) where r_0 and r_1 have been replaced by 0 and L.

$$V_3 = \frac{11}{25r_1^2}\sqrt{\frac{77ET^3Z_0^3}{25\varepsilon_0\left[\frac{11(r_1-r_0)}{r_1-\frac{11}{25}r_0} - 14\ln\left(\frac{25r_1-11r_0}{14r_1}\right)\right]}}\,, \tag{35.55}$$

$$V_3 = \frac{11}{25L^2}\sqrt{\frac{77ET^3Z_0^3}{25\varepsilon_0}} = 0.4548\sqrt{\frac{ET^3Z_0^3}{\varepsilon_0 L^4}} \quad \text{(Volts)}\,. \tag{35.56}$$

As expected, the dependence on parameters, such as Young's modulus, beam thickness, beam length, initial air gap, and the air gap dielectric constant is the same for all cases presented. The difference in the three models is the coefficient of the pull-in voltage function and the tip deflection at pull-in. The coefficients of the pull-in voltage predictions are 0.5479, 0.471 and 0.4548, respectively. The beam tip deflected 45%, 44%, and 46% of the initial gap before pull-in occurred. This is in contrast to the parallel plate model that predicts the plate deflection will be 33% of the initial gap before pull-in. By comparison to experimental test results, the models presented in (35.54) and (35.56) were more accurate at predicting the pull-in voltage for cantilever beams. The parallel plate electromechanics model had greater error and under-predicted the pull-in voltage. The model represented by (35.53) over-predicted the pull-in voltage with greater error than either of (35.54) or (35.55). The beam models predicted more accurately for wider beams, which means they are better applied to beams

with dimensions representative of a parallel plate model used as a basis of the capacitance estimate. In all three cases, the beam models presented in (35.53)–(35.55) predicted poorly the pull-in voltage for narrow beams, which are dominated by the effects of nonuniform fringing fields and not the uniform parallel plate fields. The typical application of the parallel plate capacitor model is to geometries with a width W and length L that are much greater than the capacitor gap Z_0 ($W \gg Z_0$, $L \gg Z_0$; where greater than a factor of 10 is a common rule of thumb). In Fig. 35.27, two capacitors are shown such that in one case the plate dimensions are much greater than the gap, and in the second case the plate dimensions are much smaller than the gap. When the plate dimensions are much larger than the gap, the uniform fields in the gap dominate the capacitive coupling. A small contribution to the capacitive coupling comes from the *fringing* fields at the edge of the plate (or the perimeter of the plate in three dimensions). As shown in Fig. 35.6, the fringing fields could depend on the thickness of the plate. When the plate dimensions are of the same order as the gap, the fringing fields may dominate or be of the same order as the contribution of the uniform field in the gap. This is the nature of the narrow cantilever beams that were not predicted well by a parallel plate capacitor model. A correction can be introduced to the parallel plate model to approximate the contribution of the fringing fields. This approximation is often referred to as the Love approximation.

Segmented Control of Electromechanical Membrane Deformation

Wang and *Hadaegh* [35.62] describe the electrostatic actuation of a deformable membrane mirror with segmented electrodes on the substrate. The membrane mirror, in this case silicon nitride with a conductive film, was suspended above the substrate containing the segmented electrode. The segmented electrodes are defined along angular and radial pattern. A specified voltage can be applied between the membrane and each electrode segment to develop a specific deformation for the mirror. This increases the complexity of the control system but also the flexibility of controlling the shape of the mirror surface. The optical properties could be optimized by individual characterization of membranes or by feedback from the optical output. This basic concept was also applied by *Zhang* and *Dunn* [35.73] who were able to linearize the voltage–deflection relationship by tailoring the shape of the electrodes and the beam structure used to support them.

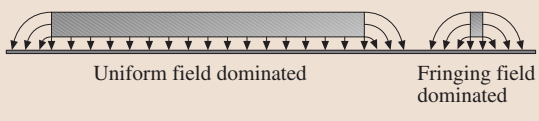


Fig. 35.27 Cross-sections of wide and narrow electrodes showing uniform electric fields in the gap and fringing fields at the edges (perimeter)

35.2.5 Electromechanics of Torsional Plates

The plate suspended by a torsional suspension instead of bending is the final configuration to be considered in terms of electromechanics. The electrostatic torsional plate has many applications, such as optical mirrors [35.58,63] and accelerometers [35.55]. The plate will define the optical surface of the mirror, the moveable plate of a varactor, or the proof mass of an accelerometer. It will be conductive or have a conductive film so that it can couple electrostatically to a stationary plate on the substrate. The moveable plate will be suspended by a torsional suspension such that the bending modes are suppressed and the primary kinematic mode is rotation of the plate. The variable capacitive coupling is used in two ways. First, it is used for actuation by applying a voltage between the moveable plate and a stationary plate. Second, it is used for sensing or use of the variable capacitance. Both actuation function and the variable capacitance function can be present in a device. The key elements of interest are the voltage-displacement function and the pull-in voltage.

Small Angular Deflection Approximation

Degani et al. [35.74], *Xiao* et al. [35.75], and Nemirovsky et al. [35.66] have considered the torsional electrostatic actuator from an analytical perspective with comparison to experimental results and other modeling results. The effective torsional spring constant is determined from the torque-twist angle relationship, where the torque is the electrostatic torque. The electrostatic torque is developed for both small and large angle cases. The pull-in angle and pull-in voltage are determined from the pull-in equations for voltage control (35.42, 43) and charge control (35.39, 40). Their method is an improved approach for modeling torsional electrostatic actuators over approximations involving parallel plate models and effective spring constants as they have shown by comparison to experimental measurements. The torsional plate considered is suspended by a torsional suspension, whose axis of rotation defines a gap

of d. It is assumed the plates extend into the page (z-axis) by a depth b and have a total length a. The torsional plate electrostatic actuator is shown in Fig. 35.28. The solution assumes the tilted plates are semi-infinite, ignoring the effects of fringing fields hence their corrections, and are tilted by angles that satisfy the small angle approximation ($\tan\alpha \approx \sin\alpha \approx \alpha$ and $\cos\alpha \approx 1$). The capacitance is:

$$C(\Theta) = \frac{\varepsilon_0 b}{\alpha_{\max}} \frac{1}{\Theta} \ln\left(\frac{1}{1-\Theta}\right) , \tag{35.57}$$

where $\alpha_{\max} = d/a$ and $\Theta = \alpha/\alpha_{\max}$.

The capacitance function and the linear relationship for the elastic restoring force ($F_M = K\,a$) are used to develop the electrostatic and mechanical energy function to substitute in the pull-in equations. For voltage controlled actuation, the pull-in parameter is developed from (35.42) to yield:

$$4 + \frac{5\Theta_{PI} - 4}{(1-\Theta_{PI})^2} - 3\ln(1-\Theta_{PI}) = 0 . \tag{35.58}$$

Equation (35.58) is solved numerically to show the voltage-controlled pull-in angle is $\Theta_{PI} \approx 0.44$. The pull-in angle is substituted in (35.43) to determine the voltage controlled pull-in voltage:

$$V_{PI} = \left(\frac{0.827 K \alpha_{\max}^3}{\varepsilon_0 b}\right)^{1/2} . \tag{35.59}$$

The charge-controlled pull-in parameter relationship is determined by substituting the electrostatic energy and mechanical energy functions in the charge-controlled pull-in parameter (35.39). This provides the nonlinear relationship for the charge controlled pull-in angle:

$$\ln(1-\Theta_{PI}) - \frac{\Theta_{PI}}{(1-\Theta_{PI})^2} - \frac{2\Theta_{PI}^2}{(1-\Theta_{PI})^2 \ln(1-\Theta_{PI})} = 0 . \tag{35.60}$$

The charge-controlled pull-in angle is given by $\Theta_{PI} \approx 0.71$. The pull-in angle is used in the charge-controlled pull-in equation (35.40) to determine the pull-in charge relationship:

$$Q_{PI} = (1.798 K \varepsilon_0 b \alpha_{\max})^{1/2} . \tag{35.61}$$

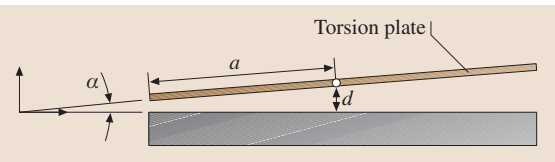


Fig. 35.28 Cross-section of a torsional actuator with a uniform cross section into the page showing the angular degree-of-freedom, α; plate length, a; and initial gap, d

Large Angular Deflection and Plate Geometry

Nemirovsky and *Bochobza-Degani* [35.66] have extended the work to the case of large angle deflections, which included using the trigonometric relationship in the capacitance function and a nonlinear relationship for the mechanical spring constant. These nonlinear relationships were used in the voltage-controlled pull-in equation. In addition, their methodology was extended to planar plates with different shapes (Fig. 35.29). They specifically considered triangular plates (the apex of the triangle is near the axis of rotation), a reversed triangular plate (the triangle apex is at the distal end of the plate), a square plate (with the rotational axis on the edge of the plate), and a circular plate (with the axis of rotation passing through a point on the perimeter and the opposite point on the perimeter defining the most distal point of the plate). Only simple geometries were considered, and their methodology can be applied to other more general geometry. The rectangular plate torsional, electrostatic actuator will provide the lowest actuation voltage relative to a fixed plate length. The reversed triangular plate has proved beneficial with regard to pixel area in mirror application and has been able to provide a higher torque for a specific plate length.

35.2.6 Leveraged Bending

As discussed, when the actuation electrode spans the entire length of the actuator, the stable controllable travel distance of the actuator plate is severely limited due to the pull-in instability. One can get around the problem of

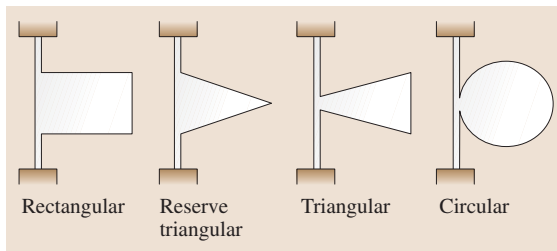


Fig. 35.29 Rectangular, triangular, reverse triangular, and circular plates are shown suspended by torsion beams for application to electromechanical mirrors or other actuators

pull-in by applying an electrostatic force to only part of the structure and using the rest of the structure as a lever to move the structure through a larger range of motion. This concept, referred to as leveraged bending [35.76], is shown schematically in Fig. 35.30 for a cantilever beam. The key idea is that the portions of the beam above the electrodes do not deflect far enough to violate the pull-in limit. The concept can be applied to other microstructure shapes and support conditions, but for fixed-fixed beams, the effect can be diminished by tensile residual stresses [35.76].

35.2.7 Electromechanics of Zipper Actuators

The design requirements for electromechanical actuators usually require minimizing the actuation voltage but maximizing the largest displacement. With the parallel plate actuator, be it a plate, beam, or torsional plate, the pull-in voltage is reduced by minimizing the air gap, but this limits the maximum displacement. Actuators with a zippering behavior have been implemented to achieve greater deflections for a lower actuation voltage. In one example (see [35.77, 78]), a plate is suspended by a torsional suspension. A long slender electrode extends from the plate and terminates with plate-like area. This electrode is supported above the fixed substrate electrode by a largely uniform gap. A voltage is applied between the substrate electrode and the extended electrode, such that the plate-like area is first pulled into the substrate. As the actuation voltage is increased, the long-slender electrode continues to collapse against the substrate electrode until a desired maximum deflection is achieved or until the drive voltage becomes impractical. This is the so-called zippering effect, which has been described for a motion perpendicular to the substrate. This uniform gap scenario is not the best suited for minimizing the pull-in voltage, so other configurations such as the curved electrode actuator have been proposed and applied to such devices as a switch and optical shutters as described by *Jin* et al. [35.79].

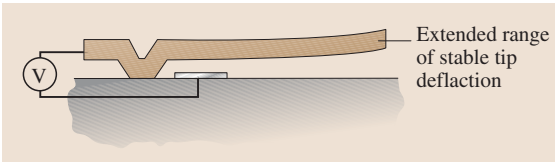


Fig. 35.30 Illustration of the concept of leveraged bending to increase the range of stable motion

Legtenberg et al. [35.80, 81] have described the basic operating features of the curved-electrode actuator. These actuators will provide large deflections in the plane of the substrate. The curved electrode actuator is composed of a stationary electrode and a compliant, moving electrode. The stationary electrode has a shape described by the function $s(x)$ shown in (35.62) and Fig. 35.31, which is written in terms of the maximum gap between electrodes $\delta_{\max}$, the length of the electrode L, the distance along the length of the electrode x, and the polynomial order of the curve n.

$$s(x) = \delta_{\max}\left(\frac{x}{L}\right)^n . \tag{35.62}$$

A necessary feature of the curved electrode actuators and zipper actuators, in general, is the contact between the moving and stationary electrodes. Without due consideration, these electrodes will short upon contact and damage the actuator. Two isolation solutions are implemented to prevent the shorting scenario. First, a continuous insulation film can be added between the electrodes to prevent them from shorting. In this case, the simplest solution is to add the insulation layer to the fixed, curved electrode. The alternative solutions include adding the insulating film to only the moving electrode or to both the moving and stationary electrodes. If the insulation is added to the moving electrode, it will become part of the stiffness of the beam. A second solution would be to add discrete, isolated contact bumpers that are made from the same conducting material as the moving electrode. As it closes the gap, the moving electrode makes contact with the isolated bumpers but not the actuation electrode. These have their operational differences as described by *Legtenberg* et al. [35.81]. Three other issues that must be considered during the design of a curved electrode actuator or zipper actuators are dielectric charging, stiction, and dielectric breakdown.

Legtenberg et al. [35.81] found that as the polynomial order increased for the electrode curve, the pull-in voltage will decrease for about the same maximum displacement. This means larger displacements can be achieved for lower voltages when compared to the parallel plate actuator. Once the moving electrode is pulled in to the curved electrode, the moving electrode will collapse to the maximum displacement. It was shown for polynomial curves of order above $n = 2$ that the maximum tip deflection becomes stable because of the geometric constraints provided by the curved electrode. This means to achieve the maximum displacement with the lowest pull-in voltage, a polynomial curve of or-

der $n = 2$ should be specified. For polynomial curves with $n > 2$, stable continuous deflection along the curved electrode can be implemented with the maximum deflection occurring at lower voltages for polynomial orders just above two. If an approximate linear range of operation is desired (i. e. the tip deflection versus voltage is approximately linear), the higher order curves ($n > 2$) are desired. *Legtenberg* et al. [35.81] analyzed the curved electrode problem by application of analytical energy methods, specifically the Rayleigh–Ritz method and of three-dimensional self-consistent electromechanical numerical simulations using CoSolve-EM in MEMCAD (now integrated into CoventorWare). From an energy perspective, the total potential energy U_T is determined in terms of the mechanical energy associated with the strain energy of bending and the electrostatic energy stored in the electric field. For the curved electrode of Fig. 35.31, the total potential energy is:

$$\begin{aligned} U_T &= U_M + U_E \\ &= \frac{1}{2}\int_0^L EI\left(\frac{d^2 w(x)}{dx^2}\right)^2 dx \\ &\quad - \frac{1}{2}\int_0^L \frac{\varepsilon_0 h V^2}{\frac{d}{\varepsilon_r} + s(x) - w(x)}\, dx\,, \end{aligned} \tag{35.63}$$

where L is the beam length, E is the beam's Young's modulus, I is the beam's cross-sectional moment of inertia, h is the thickness of the beam, V is the voltage applied between the beam and electrode, d is the initial gap at the start of the electrode, ε_0 is the permittivity of free space, and $w(x)$ is the beam's deflection curve. The next step in the solution of this problem is to determine an admissible trial function that satisfies the beam boundary conditions and can adequately approximate the solution. In this case the deflection curve for a uniformly loaded cantilever beam is chosen as the trial function:

$$\omega(x) = cx^2\left(6L^2 - 4Lx + x^2\right) = c\,f(x)\,, \tag{35.64}$$

where c is a constant to be determined for specific cases.

The approach of Nemirovsky et al. [35.66] can be taken by substituting the potential energy functions in the voltage-controlled pull-in equations. The simultaneous solution of the pull-in equations will result in the pull-in voltage and an implicit equation in terms of the constant c at pull-in. The pull-in voltage relationship is

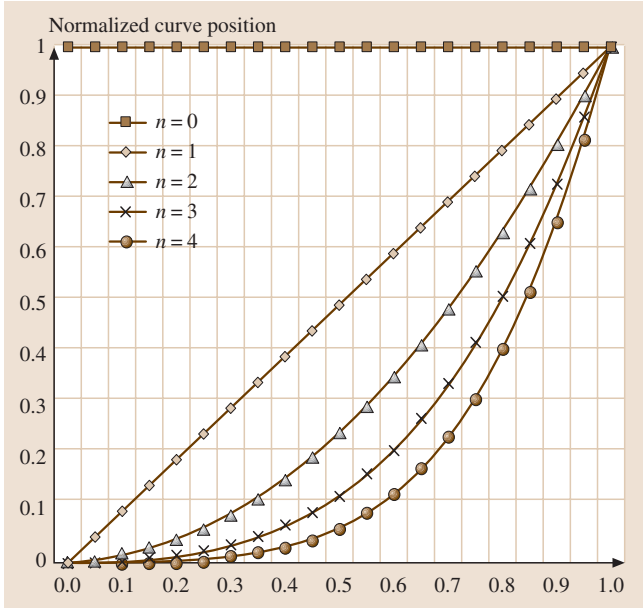


Fig. 35.31 Curved electrode shape curve shown as a function of order of the curve n

$$V_{PI}^2 = \frac{EI}{\varepsilon_0 h}\frac{\int_0^L \left(\frac{d^2 f(x)}{dx^2}\right)^2 dx}{\int_0^L \frac{f^2(x)}{[d/\varepsilon_r + s(x) - c_{PI} f(x)]^3}\, dx}\,. \tag{35.65}$$

The implicit equation to determine the coefficient of the beam deformation function is

$$\begin{aligned} &\int_0^L \frac{c_{PI} f^2(x)\, dx}{[d/\varepsilon_r + s(x) - c_{PI} f(x)]^3} \\ &\quad = \int_0^L \frac{f(x)\, dx}{[d/\varepsilon_r + s(x) - c_{PI} f(x)]^2}\,. \end{aligned} \tag{35.66}$$

DeReus [35.82] designed the curved electrode actuator to convert the curvilinear motion into a translational motion at the tip of the moving electrode. The transformation is accomplished by a folded spring that connects the electrode to the body that will be translating. A solid model of this actuator design is shown in Fig. 35.32. A pull-in analysis was performed with the results before pull-in and after pull-in presented in Fig. 35.33. This shows that a 60 μm translation was accomplished

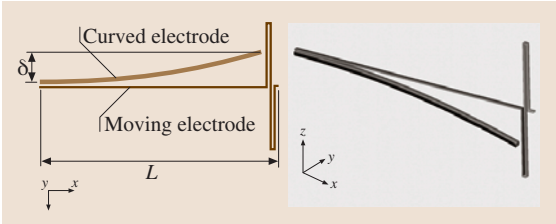


Fig. 35.32 Solid model of a curved electrode actuator with a folded suspension showing the curved electrode δ and the actuator beam length L

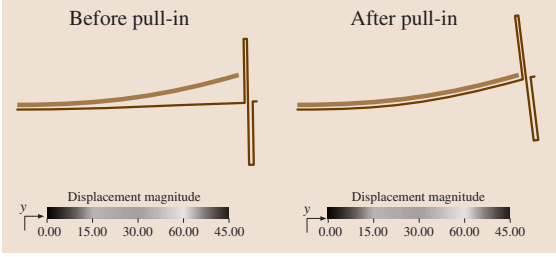


Fig. 35.33 Results of electromechanical pull-in analysis showing curved electrode actuator before and after pull-in has occurred

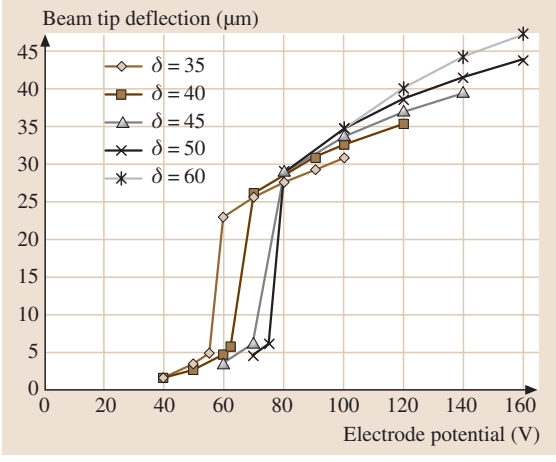


Fig. 35.34 Beam tip deformation as a function of electrode potential and δ for a suspension with a single fold

with this particular design, which has a twofold flexure and could have application as a optical shutter or VOA. In Fig. 35.34, electrode tip displacement is shown as a function of electrode voltage and electrode curve δ for a onefold flexure design. The displacement as a function of electrode potential and curve δ is shown for a twofold flexure in Fig. 35.35.

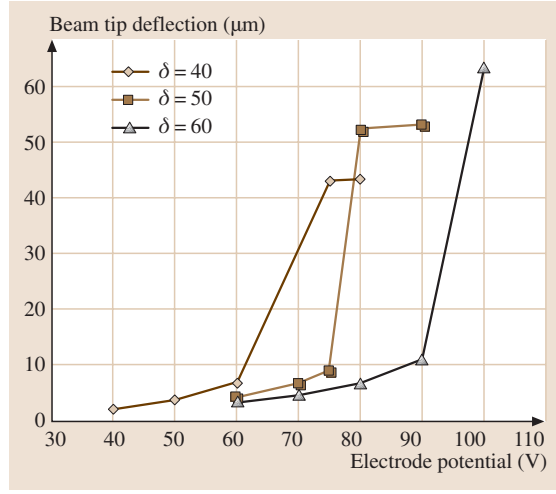


Fig. 35.35 Beam tip deformation as a function of electrode potential and δ for a suspension with a doubly folded suspension

35.2.8 Electromechanics for Test Structures

Electromechanics test structures are an important contributor to the development of microsystems products as they are used to extract information about material properties of thin films for specific process parameters. They are also used to monitor the material property variation during the fabrication process. In this case, they are distributed around the wafer on specific test die or they may be included in the dicing streets of more mature processes. These distributed test structures enable the manufacturer to track the variation in material properties with its associated within-wafer variation, wafer-to-wafer variation, and lot-to-lot variation. Another application of test structures is to characterize the impact of the first level packaging on the performance of the product. With all the critical information being extracted from the test structures, it is important for the information to be easy to interpret. Test structures have been developed that involve electrical, optical, and mechanical interrogation of the structure, but these are not easily instituted at all levels of characterization. The purely electrical test structures offer simple evaluation at the wafer, die, and package level so manufacturers often prefer them. The most common electrical test structures used for extracting information on mechanical properties involve electromechanics, of course. This section has provided a basis for interpreting and modeling the behavior of specific test structures, such as a homoge-

neous cantilever beam, and can be extended to other structures.

Osterberg and *Senturia* [35.83] describe in detail the application of electromechanics to the extraction of mechanical properties for microsystems design. They labeled their test structures as M-TEST for mechanical test structure. Their test structures comprised arrays of cantilever beams, fixed-fixed beams, and clamped, circular plates of various lengths, widths, or diameters. With these test structures, they demonstrated the extraction of important mechanical properties such as Young's modulus, plate modulus, and the residual film stress. Why choose an electromechanics test structure unless for the breadth and ease of application? An additional reason is the inherent instability that has been described as the pull-in of an electromechanical structure. The pull-in voltage is very distinct and can be measured in terms of a single parameter, the pull-in voltage. Given a control range of pull-in voltages for different structural layers in the manufacturing process, the pull-in parameter can be used as a simple control parameter.

Osterberg and *Senturia* [35.83] describe the methodology for extracting the mechanical properties. Their methodology includes the measurement of plan-view dimensions (i. e., width and length), out-of-plane curvature, pull-in voltage, film thickness, and air gap. With these measured parameters, the mechanical properties can be extracted. Their extraction methodology includes general closed form expressions for pull-in voltage that depend on the material properties and geometry of a homogeneous, isotropic material.

35.2.9 Electromechanical Dynamics: Switching Time

Switching time is an important factor in the design of such electromechanical devices as electrical or optical switches. The switching time analysis requires consideration of the dynamic equations representing the device and the electrical source. *Castaner* et al. [35.84–86] have considered the optimization of the speed-energy product. In this case, a parallel plate electrostatic actuator is described by an equivalent mechanical circuit that includes the mass m, the damping b, and the spring k, and a variable gap g. The variable gap depends on the initial gap g_0 and the displacement of the moving electrode x, such that $g = g_0 - x$. In the parallel plate, variable gap model the damping is described by squeeze film damping. (Other electrostatic actuators that move parallel to the substrate with variable gaps defined by groups of moving and stationary fingers.) In these cases, the damping model will need to include squeeze film damping and shear flow damping (Couette or Stokes Flow models). The electrical source is modeled as a voltage source and a source series resistance. A schematic of the model is shown in Fig. 35.36.

The equivalent electrical circuit of the parallel plate electrostatic actuator is modeled by (35.67), where V_s is the source voltage, V is the voltage across the variable capacitance, R is the series source resistance, and Q is the charge on the capacitor plates

$$\frac{V_s}{R} = \frac{V}{R} + \frac{dQ}{dT} \,. \tag{35.67}$$

The equivalent mechanical circuit is described by:

$$m\frac{d^2X}{dT^2} + b\frac{dX}{dT} + kX = \frac{Q^2}{2C_0G_0} \,. \tag{35.68}$$

Equation (35.68) includes the sum of the inertial force, the damping force, the spring force, and the electrostatic force. Equation (35.68) is written in terms of the mass of the moving capacitor plate, m; the damping coefficient due to squeeze film damping, b; the elastic constant of the spring by which the capacitor plate is suspended, k; the charge on the capacitor plate, Q; the initial gap, G_0; and the initial capacitance, C_0. The initial capacitance is written in terms of the area of the capacitor A, the dielectric constant of the capacitor air gap ε, and the initial gap:

$$C_0 = \frac{\varepsilon A}{G_0} \,. \tag{35.69}$$

The dynamics of the parallel plate electrostatic actuation were divided into four events. The first occurrence is a short charging time for the capacitor. The second phase is the acceleration phase, which defines the rapidly increasing velocity of the mass, but the

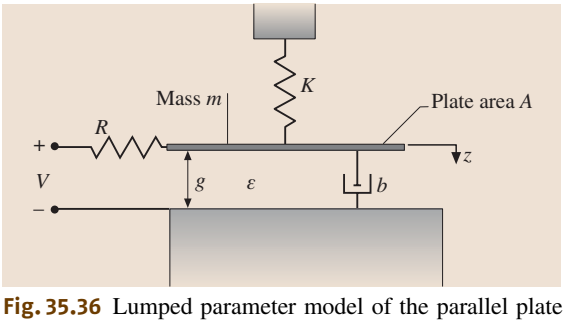


Fig. 35.36 Lumped parameter model of the parallel plate electromechanical actuator including damping and series resistance for dynamic analysis

position is still close to the initial equilibrium position. The third phase is the damping phase, defined by the velocity reaching a maximum, then decreasing due to the increased damping, and then finally increasing due to the increasing electrostatic forces as the gap closes. The fourth and final phase is the pull-in phase, characterized by a rapidly increasing velocity. *Hung* and *Senturia* [35.76] have described generating efficient dynamical models from a minimal set of finite element simulations. *Pons-Nin* et al. [35.87] have described the behavior of pull-in voltage and pull-in time for electromechanical actuators with current drives.

35.2.10 Electromechanics Issues: Dielectric Charging

Electrostatic actuators are operated in three different regimes depending on the application. In the first regime, the electrostatic actuator is operated over the voltage range that does not include the pull-in voltage. In this case, the actuated element does not contact the opposing substrate. In the second regime, the electrostatic actuator is operated through a voltage range that does include the pull-in voltage ($V_{act} \geq V_{PI}$). In this case, the desired response is to establish contact with or without some overdrive to achieve greater contact force or other response parameter. In the third regime, the electrostatic actuator is operated through a voltage range beyond the pull-in voltage. This is characteristic of electrostatic zipper actuators. In this case, the operating voltage is programmed to quickly establish pull-in, thereby contact, then the actuator operates in a range above the pull-in voltage ($V_{PI} < V_{act}$). In all three regimes, dielectric isolation layers are considered to prevent shorting between actuation electrodes. For the first and second case, the use of the insulating material is a preventative measure to protect against electrode shorting during an unintentional overdrive of the actuator. It is preventative because the actuation electrodes are not intended to touch. In the third case, the gap separating the moving and stationary actuation electrodes is reduced to zero so that an insulating layer is required. The insulating layer may be patterned or unpatterned.

The third regime defines the touch mode electrostatic actuators. *Cabuz* et al. [35.88] have described many of the failure modes that can be observed with the touch mode electrostatic actuators and that should be addressed during the design of the actuator and process. The observations and failure modes included: 1) a significant difference between the experimental actuation voltage (DC) and the theoretical actuation voltage, 2) a closed state electrostatic pressure lower than predicted, 3) an actuation voltage that increases with each subsequent cycle until electrical breakdown occurs, 4) actuator vibration observed between an open and a closed state, 5) actual release voltage is higher than theoretical, but the electrostatic forces should be much higher in the closed state such that the release voltage is always lower than the pull-in voltage, 6) the actuator remains closed temporarily after shorting, 7) permanent stiction will occur when in the closed state for a long period of time, 8) permanent stiction occurs after many actuation cycles.

Wibbeler et al. [35.89] have analyzed this problem in terms of one of the contributing factors, the accumulation and storage of localized charge in dielectrics. The common dielectrics in microsystems devices are SiO_2 and Si_3N_4, which show low mobility for surface charges and provide trap sites for positive or negative charge within their volume and at interfaces. *Kubo* et al. [35.90] showed this charge can be produced by the fabrication process by contact electrification, and by gas discharge in the air gap (the case focused on that by *Wibbeler* et al. [35.89]), and static electrification during handling (ESD). The contact electrification created electrical charge through the mechanical contact of two materials with different work functions and the separation of the two materials.

To understand the impact of the accumulated charge, *Wibbler* et al. [35.89] considered the impact of stored surface charge on the electrostatic force acting on a simple parallel plate capacitor. In Fig. 35.37, a simple capacitor is considered with a dielectric (ε_d) filling part of the air gap (ε). The dielectric has a surface charge of $\sigma_s(x, y)$. The capacitor plates carry charge shown as σ_1 and σ_2, which are provided by the battery, of voltage V, connected across the capacitor plates. The thicknesses of the dielectric and air gap are d_d and d, respectively. The force was estimated from the applied voltage and from the offset voltage due to the accumulated surface charge. The force acting on the capacitor plate as a function of the applied voltage, V; the offset

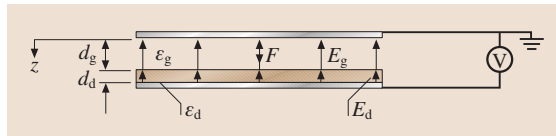


Fig. 35.37 Parallel plate capacitor with dielectric and parasitic charges

voltage, V_s; and the capacitance, C, is:

$$F = \frac{dC}{dz} \int_A \frac{[V - V_s(x, y)]^2}{2} dA \,. \tag{35.70}$$

The offset voltage is shown in (35.71) as a function of the surface charge; the thickness of the dielectric, d_d; the dielectric constant, ε_d; or the dielectric's contribution to the total capacitance, C_d. The capacitance of the dielectric layer, per unit area, being defined as $C_d = \varepsilon_d/d_d$.

$$V_s = -\sigma_s(x, y)\frac{d_d}{\varepsilon_d} = -\frac{\sigma_s(x, y)}{C_d} \,. \tag{35.71}$$

Electromechanical Pull-In with Dielectric Charging

Nemirovsky et al. [35.66] and *Chan* et al. [35.91] studied the effect of residual charge at the interface of a dielectric and the air in the capacitor gap. With the dielectric in the capacitor gap and the residual charge, the total energy for the actuator is:

$$U_T = \frac{1}{2}Kz^2 + \frac{Q_{res}^2}{2(C(z) + C_d)} - \frac{C(z)C_d}{2(C(z) + C_d)}V^2 \,. \tag{35.72}$$

In (35.72), Q_{res} is the residual charge, $C(z)$ is the air gap capacitance, and C_d is the dielectric capacitance. From the total energy, the pull-in parameter and pull-in equations (35.42) and (35.43) can be solved for the pull-in parameter ζ_{PI} and pull-in voltage V_{PI}:

$$\varsigma_{PI} = \frac{1}{3}\left(1 + \frac{1}{\delta}\right) \,, \tag{35.73}$$

$$V_{PI} = \left[\frac{8Kz_{max}^3}{27\varepsilon_0 A}\left(1 + \frac{1}{\delta}\right)^3 - \frac{Q_{res}^2}{C_d^2}\right]^{1/2} \,. \tag{35.74}$$

In (35.73), ζ_{PI} is the ratio z/z_{max} and δ is the ratio $\varepsilon_d z_{max}/z_d$. Equation (35.74) shows that the dielectric layer will increase the pull-in voltage and the residual charge will lower the pull-in voltage. It can be seen that the residual charge does not impact the normalized pull-in coordinate shown in (35.73).

Test structures were fabricated to verify the shift predicted by (35.70) that would result with a dielectric in the air gap. Test structures with a dielectric did show a shift in the voltage–deflection curve, whereas a shift was not observed when a dielectric was not present. A shift of −25 V was observed. Following electrostatic discharge events, the offset voltages of −210 V and 145 V were observed during different phases of a charge-discharge cycle. This demonstrates the significant bias the accumulated surface charge can create. As a result, *Wibbler* et al. [35.89] did not recommend trying the fix shorting by adding dielectric films over the electrodes. They did recommend estimating the influence of accumulated charge on the response of the device being designed. Instead of complete dielectric film coverage of an electrode, they recommended patterning the dielectric such that it was limited to the edges of the electrodes or patterned at locations that will touch first. By patterning small areas of insulation, the surface area available to store charge is smaller and the distance to the electrode is shorter for more timely dissipation of stored charge.

Cabuz et al. [35.88] identified the effects of humidity, DC or AC driving voltage, and SAM coatings on the performance of electromechanical actuators. With a DC operating voltage, the devices operated as expected below 35% RH except in dry air or dry N_2. In the dry environment the actuator would temporarily remain closed, would remain closed after high fields are applied, or would remain closed after a high number of cycles. When operated above 35% RH, the DC voltage was higher than expected, the closed state electrostatic pressure was lower, the actuator did not release when expected, the actuator vibrated, and the voltage increased with subsequent cycles. When an AC actuation voltage is used the actuators operated up to 55% RH. At higher RH, the actuation voltage increased. The dry condition operation was improved with AC driving compared to DC operation. AC voltage control is recommended for eliminating RC effects but cannot overcome the effects of humidity. To avoid the effects of humidity, two SAM coatings (octadecyltrichlorosilane (OTS) and perfluorodecyltrichlorosilane (PFTS)) were tried to influence the hydration of the surfaces. The OTS eliminated the effects observed for uncoated samples up to a 95% RH level for both DC and AC voltages. The PTFS improved operation up to 95% RH for AC voltages and 70% RH for DC control voltages. The residual in-use stiction was determined to be associated with accumulated and trapped charge on the surfaces or at interfaces.

35.2.11 Electromechanics Issues: Gas Discharge

As described in previous sections, the electromechanical actuation of a microstructure is performed by applying a voltage across two plates. The two plates can be defined by two conductors (e.g. Al to Al), by semiconductors

(e.g. polysilicon), or by a combination of a conductor and a semiconductor (e.g. Al and polysilicon). When these two plates (or other similarly described structures) are fabricated in proximity with each other, they are usually fabricated with an "air" gap less than 10 μm. With the electrostatic actuators, a small gap is desirable to minimize the voltage required for actuation.

The presence of small gaps in microsystems devices has prompted the further investigation of electrical breakdown in air and other ambient conditions at small gaps (i. e. less than 50 μm). *Slade* and *Taylor* [35.92] have demonstrated that breakdown in small gaps is defined by three regimes: less than 4 μm, 4 to 6 μm, and greater than 6 μm. In the first case, the breakdown event follows the behavior for a vacuum breakdown process at a larger gap. In the last case, the breakdown follows the behavior of Paschen's law for breakdown in air. The middle case of gaps between 4 and 6 μm, the breakdown process is in a transition from following the vacuum breakdown process to following the Paschen curve. In this case, the breakdown voltage is lower than predicted in either case, which indicates significant contributions from factors in both cases. The deviation from Paschen's curve was observed by *Torres* and *Dhariwal* [35.93] for gaps less than 4 μm.

The goal of this section is to describe some basic breakdown principles that can be used to determine whether or not a particular design will break down during an actuation event. Different processes will be considered for different gap regimes so that breakdown estimates can be made.

The vacuum breakdown region is characterized by a breakdown voltage that is a function of the gap separating the two electrodes. This contrasts with the breakdown behavior that follows the Paschen curve, which shows electrode gap-pressure product dependence. The breakdown voltage in vacuum is

$$V_B = K_V t_g \,. \tag{35.75}$$

Equation (35.75) shows the breakdown voltage V_B is proportional to the gap t_g where the proportionality constant is $K_v = 97\,\mathrm{V/\mu m}$ for data from many sources as indicated by *Slade* and *Taylor* [35.92]. The breakdown data for vacuum was determined for electrode gaps between 35 and 200 μm. For electrode gaps between 0.2 and 40 μm in air, the proportionality constant was between 65 and 110 V/μm.

The electric field at the surface of the electrode is a critical parameter for determining breakdown in vacuum. This is estimated by the ratio of the voltage between the electrodes V and the gap between the electrodes t_g with a modification factors β_g. β_g is a geometrical enhancement factor that captures the field enhancement due to the macroscopic geometry:

$$E_g = \beta_g V/t_g \,. \tag{35.76}$$

The geometrical enhancement factor can be determined by performing finite element analysis to determine the macroscopic electric field as a function of the electrode gap for the electrode geometry under consideration. The geometric enhancement factor is determined from the ratio of the macroscopic electric field E_g, from finite element simulations, to the electric field estimated by V/t_g [35.92, 94]. The estimated electric field, V/t_g, can be used as a reasonable approximation when the electrode gap is much smaller than that radius of curvature of the cathode electrode. If this is not the case, the geometric enhancement factor needs to be estimated from numerical analysis such as finite element or boundary element methods. The geometric enhancement factor is shown, by *Slade* and *Taylor* [35.92], to be between 1.01 and 1.57 for gaps ranging from 0.2 to 40 μm, respectively, for the geometry of a needle cathode and a plate anode.

A second enhancement factor is attributed to the microscopic field enhancement due to the surface roughness of the electrodes. The electric field containing the effects of the geometric enhancement and the microscopic enhancement is:

$$E = \beta_m \beta_g V/t_g \,. \tag{35.77}$$

In (35.77) β_m is the microscopic enhancement factor. The total enhancement factor is $\beta_g \beta_m$, which can have a range of 100 to 250 for polished electrodes [35.95].

Following the estimate of the electric field at the surface of the electrodes, the current density for the field emission of electrons is calculated. This current density j_{FE} is described by the Fowler–Nordheim equation, which depends on the electric field E, the work function of the electrode material ϕ, and the dimensionless parameters of $t(y)$ and $\nu(y)$, where y is a parameter defined as $3.79\times10^{-5}\sqrt{E}/\phi$. Typically, $t(y)$ is one. $\nu(y)$ is given by $0.956 - 1.06y^2$. The Fowler–Nordheim equation is presented in (35.78) with j_{FE} in $\mathrm{A/m^2}$, E in V/m, and ϕ in eV.

$$j_{FE} = \frac{1.54\times10^{-6}E^2}{\phi t(y)^2} \exp\left(\frac{-6.83\times10^9\phi^{3/2}\nu(y)}{E}\right) \,. \tag{35.78}$$

Equation (35.78) can be rewritten with E substituted from (35.77) and the current density replaced by an emission current I_e and the area A_e of a microprojection where the emission occurs. Equation (35.78) is shown rewritten in (35.79), such that the slope from the Fowler–Nordheim plot will provide the field enhancement factor as shown in (35.80) where m is the slope. The Fowler–Nordheim plot is the plot of I_e/V^2 versus $1/V$, which is constructed by measuring the voltage across the electrode gap and measuring I_e.

$$\mathrm{Log}_{10}\left(\frac{I_e}{V^2}\right) = \mathrm{Log}_{10}\left(\frac{1.54\times 10^{-6} A_e \beta^2}{\phi\, t(y)^2 t_g^2}\right) - \frac{1}{2.303}\left(\frac{6.83\times 10^9 \phi^{3/2} t_g \nu(y)}{\beta}\right)\frac{1}{V}, \tag{35.79}$$

$$\beta = -\frac{2.303m}{6.83\times 10^9 \phi^{3/2} t_g \nu(y)}. \tag{35.80}$$

For gaps larger than 6 μm, the electrical breakdown as a function of gap does follow the Paschen curve for air and does not follow the vacuum breakdown curve for these larger gaps. The electrode gap is now larger than the mean free path of the air molecules so it is expected that emitted electrons will collide with the air molecules in the gap. For these larger electrode gaps, it is assumed the electrical breakdown will follow the Townsend electron avalanche theory. A similar process is followed to estimate the breakdown. First, the electric field at the surface of the cathode electrode is estimated including the enhancement factors for geometry and surface microstructure. An estimate of the total field enhancement factor is required for this. The electric field is substituted into the Fowler–Nordheim equation (35.78) to estimate the field emission current. In this case, the emitted electrons will interact with the background gas molecules causing an avalanche in the current. The current avalanche process is described by (35.79) and (35.80), where t_g is the electrode gap, α is the first Townsend coefficient, and γ is the second Townsend coefficient. The first Townsend coefficient represents the number of electrons produced per unit distance along the direction of the electric field. The second Townsend coefficient is the number of electrons generated by secondary processes per each primary avalanche. Breakdown will occur when the denominator of the expression for current approaches zero: $\gamma \exp(\alpha t_g)$ approaches unity.

In the transition region, the breakdown voltage can be lower than predicted by the Paschen curve or vacuum breakdown. It is suggested that in this regime, the partial electron avalanche process enhances the vacuum breakdown process to lead to a lower breakdown voltage [35.92].

The choice of materials is another consideration for the design of an electrostatic microactuator. As mentioned, the electrode pairs could comprise the following combinations: conductor-conductor, conductor-semiconductor, or semiconductor-semiconductor. In *Torres* and *Dhariwal* [35.93], the breakdown voltage was measured for electrode pairs comprised of nickel, brass, or aluminum. A significant difference was not observed between the breakdown voltages of the different conductors or for the size of the electrode. At larger gaps, the breakdown voltage was different for different electrode shapes (cylindrical versus spherical) but this can be explained by the relative difference in the geometric enhancement factor for the electric field for the different electrodes. *Ono* et al. [35.96] made a comparison between electrode combinations where both a semiconductor (silicon) and a silicon–metal combination were used. In the case of a silicon electrode with a metal electrode, the breakdown voltage deviated from the Paschen curve at approximately 6 μm. It was lower than the breakdown voltage predicted by the Paschen curve and trailed off with an approximate linear dependence on the electrode gap. With the silicon as one electrode, Au/Cr or Pt/Cr was used as the second electrode but little difference was found for the different metals. With the silicon-silicon electrode combination, the breakdown voltage followed the Paschen curve for air. For similar gaps, the breakdown voltage was significantly higher for the silicon-silicon electrodes than for the silicon-metal electrode. This was true for a thin silicon film over a Cr/Pt/Cr electrode.

Another option for reducing the electric breakdown in the air gap is to introduce an insulating material over the electrode. But *Wibbeler* et al. [35.89] have demonstrated this does not solve the problem of breakdown by itself and can lead to parasitic charging of the insulator. An electric breakdown will produce a large current, as described previously, that is destructive to the electrodes without insulation. The dielectric limits the current so that the electrical breakdown event is not so destructive. *Wibbeler* et al. [35.89] demonstrated through highly chaotic behavior in the voltage–displacement response of an electrostatic actuator. It is reasoned that the local field strength at the electrode reached a magnitude

to cause an avalanche of free electrons and ions, as described here. These free charges travel the length of the air gap, but they are stopped and neutralized at the dielectric surface. This reduces the local electric field at the location of the surface charge so that the force on the cantilever is reduced and the process repeats itself. Throughout this process, a large amount of charge can be introduced onto the dielectric surface, which will typically have a very long dissipation time constant.

35.3 Summary and Mention of Topics not Covered

This chapter has focused on the thermomechanics and electromechanics of thin-film microstructures. Emphasis has been placed on the basic physical phenomena, approaches to analyze the phenomena, and the presentation of accessible results for special cases of practical importance in a variety of applications. We have not discussed many other energy domain couplings that play major roles in microsystems technology such as electrothermomechanics and magnetomechanics. We have also not discussed the role and behavior of materials such as piezoelectrics or shape memory alloys, which play important roles in electromechanics and thermomechanics, respectively. We refer the interested reader to *Senturia* [35.49] for an introduction and further references.

References

35.1 M. Madou: *Fundamentals of Microfabrication* (CRC, Boca Raton 1997)

35.2 M. Gad-el-Hak: *The MEMS Handbook* (CRC, Boca Raton 2001)

35.3 J. J. Sniegowski, M. P. de Boer: IC-compatible polysilicon surface micromachining, Annu. Rev. Mater. Sci. **30** (2000) 299–333

35.4 C. H. Mastrangelo: Adhesion-related failure mechanisms in microelectromechanical devices, Tribol. Lett. **3** (1997) 223–238

35.5 M. P. de Boer, T. M. Mayer: Tribology of MEMS, MRS Bull. **26** (2001) 302–304

35.6 W. D. Nix: Mechanical properties of thin films, Metall. Trans. A **20A** (1989) 2217–2245

35.7 C. Y. Hui, H. D. Conway, Y. Y. Lin: A reexamination of residual stresses in thin films and of the validity of Stoney's estimate, J. Electron. Packaging **122** (2000) 267–273

35.8 D. B. Bogy, K. C. Wang: Stress singularities at interface corners in bonded dissimilar materials, Int. J. Solids Struct. **7** (1971) 993–1005

35.9 V. L. Hein, F. Erdogan: Stress singularities in a two material wedge, Int. J. Fract. Mech. **7** (1971) 317–330

35.10 E. D. Reedy, T. R. Guess: Nucleation and propagation of an edge crack in a uniformly cooled epoxy/glass bimaterial, Int. J. Solids Struct. **39** (2002) 325–340

35.11 M. D. Thouless: Modeling the development and relaxation of stresses in films, Annu. Rev. Mater. Sci. **25** (1995) 69–96

35.12 N. D. Masters, M. P. de Boer, B. D. Jensen, M. S. Baker, D. Koester: Side-by-side comparison of passive MEMS residual strain test structures under residual compression. In: *Mechanical Properties of Structural Films*, ASTM STP **1413**, ed. by C. L. Muhlstein, S. B. Brown (2001) 168–200

35.13 M. W. Hyer: Calculation of the room-temperature shapes of unsymmetric laminates, J. Compos. Mater. **15** (1981) 296–310

35.14 M. W. Hyer: The room-temperature shape of four-layer unsymmetric cross-ply laminates, J. Compos. Mater. **16** (1982) 318–340

35.15 D. E. Fahnline, C. B. Masters, N. J. Salamon: Thin film stress from nonspherical substrate bending measurements, J. Vac. Sci. Technol. A **9** (1991) 2483–2487

35.16 C. B. Masters, N. J. Salamon: Geometrically nonlinear stress-deflection relations for thin film/substrate systems, Int. J. Eng. Sci. **31** (1993) 915–925

35.17 C. B. Masters, N. J. Salamon: Geometrically nonlinear stress-deflection relations for thin film/substrate systems with a finite element comparison, J. Appl. Mech. **61** (1994) 872–878

35.18 M. Finot, S. Suresh: Small and large deformation of thick and thin film multilayers: effects of layer geometry, plasticity and compositional gradients, J. Mech. Phys. Solids **44** (1996) 683–721

35.19 M. Finot, I. A. Blech, S. Suresh, H. Fujimoto: Large deformation and geometric instability of substrates with thin film deposits, J. Appl. Phys. **81** (1997) 3457–3464

35.20 L. B. Freund: The stress distribution and curvature of a general compositionally graded semiconductor layer, J. Cryst. Growth **132** (1993) 341–344

35.21 L. B. Freund: Some elementary connections between curvature and mismatch strain in compositionally graded thin films, J. Mech. Phys. Solids **44** (1996) 723–736

35.22 L. B. Freund: Substrate curvature due to thin film mismatch strain in the nonlinear deformation range, J. Mech. Phys. Solids **48** (2000) 1159–1174

35.23 N. J. Salamon, C. B. Masters: Bifurcation in isotropic thin film/substrate plates, Int. J. Solids Struct. **32** (1995) 473–481

35.24 R. M. Jones: *Mechanics of Composite Materials*, 2nd edn. (Taylor & Francis, London 1999)

35.25 M. W. Hyer: *Stress Analysis of Fiber-reinforced Composite Materials* (McGraw-Hill, Boston 1998)

35.26 G. G. Stoney: The tension of metallic films deposited by electrolysis, Proc. R. Soc. Lond. A **82** (1909) 172–175

35.27 C. D. Pionke, G. Wempner: The various approximations of the bimetallic thermostatic strip, J. Appl. Mech. **58** (1991) 1015–1020

35.28 P. Krulevitch, G. C. Johnson: Curvature of a cantilever beam subjected to an equibiaxial bending moment, Mater. Res. Symp. Proc. **518** (1998) 67–72

35.29 S. R. Swanson: *Introduction to Design and Analysis with Advanced Composite Materials* (Prentice-Hall, Upper Siddel River 1997)

35.30 M. L. Dunn, Y. Zhang, V. Bright: Deformation and structural stability of layered plate microstructures subjected to thermal loading, J. Microelectromech. Syst. **11** (2002) 372–383

35.31 J. A. King: *Materials Handbook for Hybrid Microelectronics* (Teledyne Microelectronics, Los Angeles 1988)

35.32 W. N. Sharpe: Mechanical properties of MEMS materials. In: *The CRC Handbook of MEMS*, ed. by M. Gad el Hak (CRC, Boca Raton 2001) Chap. 3

35.33 Y. Zhang, M. L. Dunn: Deformation of blanketed and patterned bilayer thin film microstructures during post-release thermal and cyclic thermal loading, J. Microelectromech. Syst. (2003) in press

35.34 B. D. Harper, C.-P. Wu: A geometrically nonlinear model for predicting the intrinsic film stress by the bending plate method, Int. J. Solids Struct. **26** (1990) 511–525

35.35 Y. L. Shen, S. Suresh: Thermal cycling and stress relaxation response of Si-Al and Si-Al-SiO_2 layered thin films, Acta. Metall. Mater. **43** (1995) 3915–3926

35.36 S. P. Baker, A. Kretschmann, E. Arzt: Thermomechanical behavior of different texture components in Cu thin films, Acta Mater. **49** (2001) 2145–2160

35.37 W. D. Nix: Elastic and plastic properties of thin films on substrates, Mater. Sci. Eng. A **234–236** (1997) 37–44

35.38 C. Thompson: The yield stress of polycrystalline thin films, J. Mater. Res. **8** (1993) 237–238

35.39 H. Gao, Y. Huang, W. D. Nix, J. W. Hutchinson: Mechanism-based strain gradient plasticity – I. Theory, J. Mech. Phys. Solids **47** (1999) 1239–1263

35.40 H. D. Espinosa, B. C. Prorok, M. Fischer: A novel method for measuring elasticity, plasticity, and fracture of thin films and MEMS materials, J. Mech. Phys. Solids **51** (2003) 47–67

35.41 D. C. Miller, M. L. Dunn, V. M. Bright: Thermally induced change in deformation of multimorph MEMS structures, Proc. SPIE **4558** (2001) 32–44

35.42 D. J. Vickers-Kirby, R. L. Kubena, F. P. Stratton, R. J. Joyce, D. T. Chang, J. Kim: Anelastic creep phenomena in thin film metal plated cantilevers for MEMS, Mater. Res. Soc. Symp. **657** (2001) EE2.5.1–EE2.5.6

35.43 R. M. Keller, S. P. Baker, E. Arzt: Stress-temperature behavior of unpassivated thin copper films, Acta Mater. **47** (1999) 415–426

35.44 M. D. Thouless, J. Cupta, J. M. E. Harper: Stress development and relaxation in copper films during thermal cycling, J. Mater. Res. **8** (1993) 1845–1852

35.45 M. D. Thouless, K. P. Rodbell, C. Cabral Jr.: Effect of a surface layer on the stress relaxation of thin films, J. Vac. Sci. Technol. **14** (1996) 2454–2461

35.46 R. P. Vinci, E. M. Zielinski, J. C. Bravman: Thermal stress and strain in copper thin films, Thin Solid Films **262** (1995) 142–153

35.47 J. Koike, S. Utsunomiya, Y. Shimoyama, K. Maruyama, H. Oikawa: Thermal cycling fatigue and deformation mechanism in aluminum alloy thin films on silicon, J. Mater. Res. **13** (1998) 3256–3264

35.48 K. Gall, M. L. Dunn, Y. Zhang, B. Corff: Thermal cycling response of layered gold/polysilicon MEMS structures, Mech. Mater. (2003) in press

35.49 S. D. Senturia: *Microsystems Design* (Kluwer, Dordrecht 2001)

35.50 M. P. de Boer, N. F. Smith, N. D. Masters, M. B. Sinclair, E. J. Pryputniewicz: Integrated platform for testing MEMS mechanical properties at the wafer scale by the IMaP methodology. In: *Mechanical Properties of Structural Films*, ASTM STP **1413** (2001) 85–95

35.51 Z. J. Yao, S. Chen, S. Eshelman, D. Denniston, C. Goldsmith: Micromachined low-loss microwave switches, J. Microelectromech. Syst. **8** (1999) 129–134

35.52 B. McCarthy, G. G. Adams, N. E. McGruer, D. Potter: A dynamic model, including contact bounce of an electrostatically actuated microswitch, J. Microelectromech. Syst. **11** (2002) 276–283

35.53 H.-S. Lee, C. H. Leung, J. Shih, S.-C. Chang, S. Lorincz, I. Nedelescu: Integrated microrelays: Concept and initial results, J. Microelectromech. Syst. **11** (2002) 147–153

35.54 S. J. Cunningham, S. Tatic-Lucic, J. Carper, J. Lindsey, L. Spangler: A high aspect ratio accelerometer fabricated using anodic bonding, dissolved wafer, and deep RIE processes, Proc. Transducers '99, the 10th Intl. Conf. Solid-State Sensors and Actuators, Sendai 1999, ed. by M. Esashi (IEE Jpn. 1999) 1522–1525

35.55 L. Spangler, C. Kemp: ISAAC: integrated silicon automotive acceleromter, Sens. Actuators A **54** (1996) 523–529

35.56 M.T.A. Saif, B.E. Alaca, H. Sehitoglu: Analytical modeling of electrostatic membrane actuator for micropumps, J. Microelectromech. Syst. **8** (1999) 335–345

35.57 C. Huang, C. Christophorou, K. Najafi, A. Naguib, H.M. Naguib: An electrostatic microactuator system for application in high-speed jets, J. Microelectromech. Syst. **11** (2002) 222–235

35.58 H. Toshiyoshi, H. Fujita: Electrostatic micro torsion mirrors for optical switch matrix, J. Microelectromech. Syst. **5** (1996) 231–237

35.59 H. Toshiyoshi, W. Piyawattanametha, C.-T. Chan, M.C. Wu: Linearization of electrostatically actuated surface micromachined 2-D optical scanner, J. Microelectromech. Syst. **10** (2001) 205–214

35.60 M. Fischer, M. Giousouf, J. Schaepperle, D. Eichner, M. Weinmann, W. von Münch, F. Assmus: Electrostatically deflectable polysilicon micromirrors-dynamic behaviour and comparison with the results from FEM modelling with Ansys, Sens. Actuators A **67** (1998) 89–95

35.61 P.K.C. Wang, R.C. Gutierrez, R.K. Bartman: A method for designing electrostatic-actuator electrode pattern in micromachined deformable mirrors, Sens. Actuators A **55** (1996) 211–217

35.62 P.K.C. Wang, F.Y. Hadaegh: Computation of static shapes and voltages for micromachined deformable mirrors with nonlinear electrostatic actuators, J. Microelectromech. Syst. **5** (1996) 205–220

35.63 J. Bühler, J. Funk, J.G. Korvink, F.-P. Steiner, P.M. Sarro, H. Baltes: Electrostatic aluminum micromirrors using double-pass metallization, J. Microelectromech. Syst. **6** (1997) 126–135

35.64 M. Fischer, H. Graef, W. von Münch: Electrostatically deflectable polysilicon torsional mirrors, Sens. Actuators A **44** (1994) 83–89

35.65 H. Schenk, P. Dürr, D. Kunze, H. Lakner, H. Kück: An electrostatically excited 2D-micro-scanning-mirror with an in-plane configuration of the driving electrodes, Proc. IEEE MEMS 2000, The 13th Ann. Intl. Conf. Micro Electro Mechanical Systems, Miyazaki 2000, ed. by I. Shimoyama, H. Kuwano (IEEE, Piscataway 2000)

35.66 Y. Nemirovsky, O. Bochobza-Degani: Methodology and model for the pull-in parameters of electrostatic actuators, J. Microelectromech. Syst. **10** (2001) 601–615

35.67 G. Li, N.R. Aluru: A Lagrangian approach for electrostatics analysis of deformable conductors, J. Microelectromech. Syst. **11** (2002) 245–254

35.68 N.R. Aluru, J. White: A multilevel Newton method for mixed-energy domain simulation of MEMS, J. Microelectromech. Syst. **8** (1999) 299–308

35.69 J.I. Seeger, S. B. Crary: Stabilization of electrostatically actuated mechanical devices, Proc. Transducers '97, 1997 Intl. Conf. Solid-State Sensors and Actuators, Chicago 1997, ed. by K. Wise (IEEE, Piscataway 1997) 1133–1136

35.70 R. Nadal-Guardia, A. Dehe, R. Aigner, L.M. Castaner: Current drive methods to extend the range of travel of electrostatic microactuators beyond the voltage pull-in point, J. Microelectromech. Syst. **11** (2002) 255–263

35.71 G.J. O'Brien, D.J. Monk, L. Lin: Electrostatic latch and release; a theoretical and empirical study, Proc. Micro-Electro-Mechanical Syst. (MEMS) 2000, MEMS-Vol. 2, The 2000 ASME Intl. Mech. Eng. Cong. and Expo., Orlando 2000, ed. by A.J. Malshe, Q. Tan, A.¶. Lee, F.R. Forster, R.S. Kenten (ASME, New York 2000) 19–26

35.72 B. Choi, E.G. Lovell: Improved analysis of microbeams under mechanical and electrostatic loads, J. Micromech. Microeng. **7** (1997) 24–29

35.73 Y. Zhang, M.L. Dunn: A vertical electrostatic actuator with extended digital range via tailored topology, Proc. SPIE **4700** (2002) 147–156

35.74 O. Degani, E. Socher, A. Lipson, T. Leitner, D.J. Setter, S. Kaldor, Y. Nemirovsky: Pull-in study of an electrostatic torsion microactuator, J. Microelectromech. Syst. **7** (1998) 373–379

35.75 Z. Xiao, X. Wu, W. Peng, K.R. Farmer: An angled-based design approach for rectangular electrostatic actuators, J. Microelectromech. Syst. **10** (2001) 561–568

35.76 E.S. Hung, S.D. Senturia: Generating efficient dynamical models for microelectromechanical systems from a few finite-element simulation runs, J. Microelectromech. Syst. **8** (1999) 280–289

35.77 J.R. Gilbert, R. Legtenberg, S.D. Senturia: 3D coupled electro-mechanics for MEMS: Application of CoSolve-EM, Proc. IEEE MEMS Conference, Amsterdam 1995, ed. by M. Elwenspoek, N. de Rooij (IEEE, Piscataway 1995) 122–127

35.78 J.R. Gilbert, G.K. Ananthasuresh: 3D modeling of contact problems and hysteresis in coupled electro-mechanics, Proc. IEEE MEMS 1996, The ninth Ann. Intl. Workshop on Micro Electro Mechanical Syst., San Diego 1996, ed. by M.G. Allen, M.L. Reed (IEEE, Piscataway 1996) 127–132

35.79 Y.-H. Jin, K.-S. Seo, Y.-H. Cho, S.-S. Lee, K.-C. Song, J.-U. Bu: An integrated SOI optical microswitch using electrostatic micromirror actuators with insulated touch-down beams and curved electrodes, Proc. Micro-Electro-Mechanical Syst. (MEMS) 2000, MEMS-Vol. 2. The 2000 ASME Intl. Mech. Eng. Cong. and Expo., Orlando 2000, ed. by A.J. Malshe, Q. Tan, A.P. Lee, F.R. Forster, R.S. Kenten (ASME, New York 2000) 177–181

35.80 R. Legtenberg, E. Berenschot, M. Elwenspoeke, J. Fluitman: Electrostatic curved electrode actuators, Proc. IEEE MEMS Conference, Amsterdam 1995 (IEEE, Piscataway 1995) 37–42

35.81 R. Legtenberg, J. Gilbert, S. D. Senturia, M. Elwenspoek: Electrostatic curved electrode actuators, J. Microelectromech. Syst. **6** (1997) 257–265
35.82 D. DeReus: Personal communications and internal reports (2002)
35.83 P. M. Osterberg, S. D. Senturia: M-TEST: A test chip for MEMS material property measurement using electrostatically actuated test structures, J. Microelectromech. Syst. **6** (1997) 107–118
35.84 L. M. Castaner, S. D. Senturia: Speed-energy optimization of electrostatic actuators based on pull-in, IEEE J. Microelectromech. Syst. **8** (1999) 290–298
35.85 L. Castaner, A. Rodriguez, J. Pons, S. D. Senturia: Measurement of power-speed product of electrostatic actuators, Proc. Transducers '99, the 10th Intl. Conf. Solid-State Sensors and Actuators, Sendai 1999, ed. by M. Esashi (IEE Jpn., Tokyo 1999) 1772–1775
35.86 L. Castaner, A. Rodriguez, J. Pons, S. D. Senturia: Pull-in time-energy product of electrostatic actuators: Comparison of experiments and simulation, Sens. Actuators **83** (2000) 263–269
35.87 J. Pons-Nin, A. Rodriguez, L. M. Castaner: Voltage and pull-in time in current drive of electrostatic actuators, J. Microelectromech. Syst. **11** (2002) 196–205
35.88 C. Cabuz, E. I. Cabuz, T. R. Ohnstein, J. Neus, R. Maboudian: Factors enhancing the reliability of touch-mode electrostatic actuators, Sens. Actuators **79** (2000) 245–250
35.89 J. Wibbeler, G. Pfeifer, M. Hietschold: Parasitic charging of dielectric surfaces in capacitive microelectromechanical systems (MEMS), Sens. Actuators A **71** (1998) 74–80
35.90 H. Kubo, T. Namura, K. Yoneda, H. Ohishi, Y. Todokoro: Evaluation of charge build-up in wafer processing by using MOS capacitors with charge collecting electrodes, ICMTS 1995 Proc. IEEE Intl. Conf. Microelectronic Test Structures, Vol. 8, Nara 1995, ed. by T. Sugano, K. Asada (IEEE, Piscataway 1995) 5–9
35.91 E. K. Chan, K. Garikipati, R. W. Dutton: Characterization of contact electromechanics through capacitance-voltage measurement and simulations, IEEE J. Microelectromech. Syst. **8** (1999) 208–217
35.92 P. G. Slade, E. D. Taylor: Electrical breakdown in atmospheric air between closely spaced (0.2 μm–40 μm) electrical contacts, Proc. 47th IEEE Holm Conf. Electrical Contacts, Montreal 2001, ed. by K. Leung (IEEE, Piscataway 2001) 245–250
35.93 J-M. Torres, R. S. Dhariwal: Electric field breakdown at micrometer separations, Nanotechnology **10** (1999) 102–107
35.94 R. Longwitz, H. van Lintel, R. Carr, C. Hollenstein, P. Renaud: Study of gas ionization schemes for microdevices, Proc. Transducers '01, Eurosensors XV, The 11th Intl. Conf. Solid-State Sensors and Actuators, Munich 2001, ed. by O. Obermeier (Springer, Berlin, Heidelberg 2001) 1258–1261
35.95 D. K. Davies, M. F. Biondi: Vacuum breakdown between plane-parallel copper plates, J. Appl. Phys. **37** (1966) 2969–2977
35.96 T. Ono, D. Y. Sim, M. Esashi: Micro-discharge and electric breakdown in a micro-gap, J. Micromech. Microeng. **10** (2000) 445–451

36. High Volume Manufacturing and Field Stability of MEMS Products

Low volume MEMS/NEMS production is practical when an attractive concept is implemented with business, manufacturing, packaging, and test support. Moving beyond this to high volume production adds requirements on design, process control, quality, product stability, market size, market maturity, capital investment, and business systems. In a broad sense, this chapter uses a case study approach: It describes and compares the silicon-based MEMS accelerometers, pressure sensors, image projection systems, and gyroscopes that are in high volume production. Although they serve several markets, these businesses have common characteristics. For example, the manufacturing lines use automated semiconductor equipment and standard material sets to make consistent products in large quantities. Standard, well controlled processes are sometimes modified for a MEMS product. However, novel processes that cannot run with standard equipment and material sets are avoided when possible. This reliance on semiconductor tools, as well as the organizational practices required to manufacture clean, particle-free products partially explains why the MEMS market leaders are integrated circuit manufacturers. There are other factors. MEMS and NEMS are enabling technologies, so it can take several years for high volume applications to develop. Indeed, market size is usually a strong function of price. This becomes a vicious circle, because low price requires low cost – a result that is normally achieved only after a product is in high volume production. During the early years, IC companies reduced cost and financial risk by using existing facilities for low volume MEMS production. As a result, product architectures are partially determined by capabilities developed for previous products. This chapter includes a discussion of MEMS product architecture with particular attention to the impact of electronic integration, packaging, and surfaces. Packaging and testing are critical, because they are significant factors in MEMS product cost. These devices have extremely high surface/volume ratios, so performance and stability may depend on the control of surface characteristics after packaging. Looking into the future, the competitive advantage of IC suppliers will decrease as small companies learn to integrate MEMS/NEMS devices on CMOS foundry wafers. Packaging challenges still remain, because most MEMS/NEMS products must interact with the environment without degrading stability or reliability. Generic packaging solutions are unlikely. However, packaging subcontractors recognize that MEMS/NEMS is a growth opportunity. They will spread the overhead burden of high-capital-cost-facilities by developing flexible processes in order to package several types of moderate volume integrated MEMS/NEMS products on the same equipment.

36.1 **Manufacturing Strategy** 1086
36.1.1 Volume 1086
36.1.2 Standardization 1086
36.1.3 Production Facilities 1086
36.1.4 Quality 1087
36.1.5 Environmental Shield 1087

36.2 **Robust Manufacturing** 1087
36.2.1 Design for Manufacturability 1087
36.2.2 Process Flow and Its Interaction with Product Architecture 1088
36.2.3 Microstructure Release 1095
36.2.4 Wafer Bonding 1095
36.2.5 Wafer Singulation 1097
36.2.6 Particles 1098
36.2.7 Electrostatic Discharge and Static Charges 1098
36.2.8 Package and Test 1099
36.2.9 Quality Systems 1101

36.3 **Stable Field Performance** 1102
36.3.1 Surface Passivation 1102
36.3.2 System Interface 1105

References 1106

Solid-state pressure sensors were first reported in the 1960s and commercialized by a number of start-up companies in the 1970s. By 1983, a *Scientific American* cover story [36.1] described pressure sensors, accelerometers, and ink-jet print heads and featured a gas chromatograph that combined an injection valve, a detector, and a chromatographic column on a silicon wafer. The future of the budding micromechanical device industry looked bright, but commercial reality has seriously lagged the rosy market projections. The reason was simple: Market forecasters failed to recognize the difference between lab prototypes and high volume production of stable products. Many companies have demonstrated that small quantities can be produced for niche markets. However, routine high volume production of reliable packaged devices is much more challenging. It requires the production prototype skills plus the manufacturing disciplines that are critical to long-term stable production.

Successful companies understand and apply these lessons. Indeed, over a hundred million MEMS pressure sensor, accelerometer, gyro, gas flow, and optical projection devices are shipped to customers annually. There is no universally accepted definition of MEMS. This chapter focuses on silicon-based products with movable micromechanical elements. It excludes disk drive heads, ink-jet print heads, hearing aids, microscale plastic, ceramic, quartz and metal components, and test strips for in vitro diagnostics that are sometimes included within the definition of MEMS. Many MEMS products do not integrate support electronics with the MEMS element on the chip even though they are produced with IC equipment, materials, and processes. However, integrated MEMS products are becoming ubiquitous. By 2002 Analog Devices had shipped over a hundred million surface micromachined MEMS accelerometers with integrated electronics to customers.

The literature related to high volume production of MEMS products is quite uneven. Some suppliers have published detailed process and design descriptions, while others limit disclosures to market-focused publications. The two major suppliers of integrated MEMS products have discussed their designs and processes in detail. *Core* et al. [36.2], *Kuehnel* et al. [36.3], *Chau* et al. [36.4], and *Sulouff* [36.5] describe the Analog Devices *i*MEMS accelerometers and gyros. *Mignardi* et al. [36.6] give an in-depth review of the Texas Instruments DLP (Digital Light Processing) products.

MEMS devices are components in larger systems, so every MEMS product must define its system interface in a way that adds value from the perspective of the system designer. Some MEMS pressure sensor products are simply Wheatstone bridge elements in a circuit (Fig. 36.1). The image projection systems that use the Texas Instruments Digital Light Processing Technology™ could never be designed in this manner, because up to 1.3 million mirrors are driven on each chip (Fig. 36.2). Such a system would be impractical unless electronics are integrated with the MEMS mirrors on the chip. Air bag sensors also illustrate how integration adds value. Non-MEMS ball-and-spring air bag sensors were often used in the early 1990s. Upon impact, a ball would be released and detected as it passed through the sensor tube. Unfortunately, these binary sensors could not always distinguish a frontal collision from the jolt of a pothole or a side collision. Therefore, multiple ball-and-spring sensors were placed at different locations in a vehicle and wired together in order to reliably identify accidents that should result in air bag deployment. High cost limited the use of these distributed systems. Analog Devices changed the air bag market when it introduced the ADXL50 in 1993 (Fig. 36.3). The ADXL50 integrated electronics and an accelerometer on one chip to provide an analog output of deceleration versus time when an impact event occurred. This allowed automobile manufacturers to design air bag modules that compared the ADXL50 output to the known signature of a frontal collision. The result was greater reliability

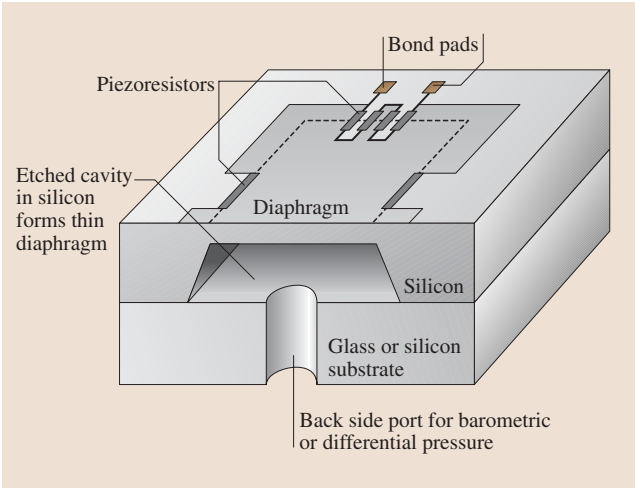


Fig. 36.1 Cross section of a piezoresistive silicon sensor after it is bonded to a silicon or glass substrate. When pressure is applied, the thin silicon diaphragm deforms. This causes changes in four implanted resistors that form a Wheatstone bridge. The resistors are located and oriented to cause resistance to increase in two resistors and decrease in the other two

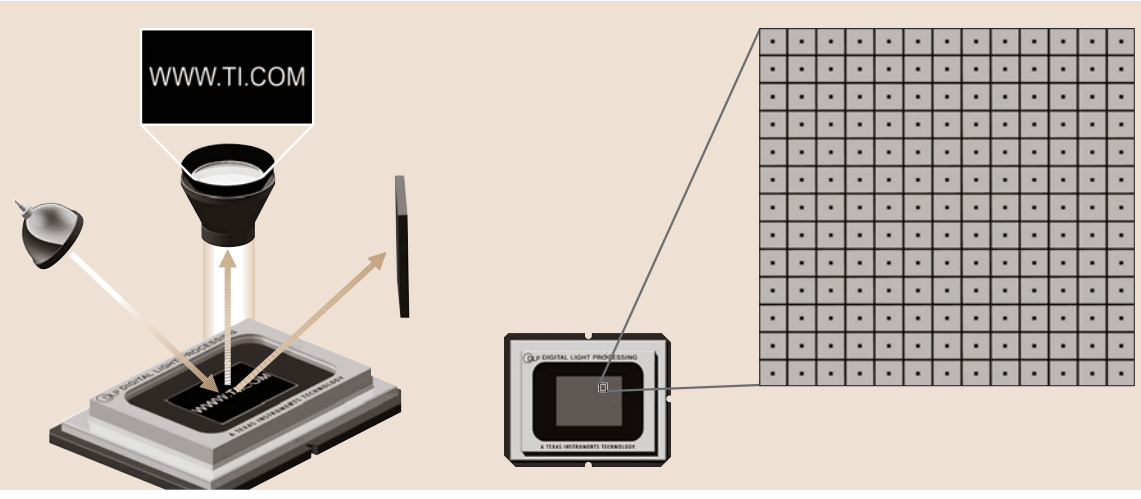


Fig. 36.2 Digital Light Processing technology™ is based on Digital Micromirror Device (DMD) chips that address and drive a "sea of mirrors." The DMD is the key element in small computer-based projection systems. Each of the 14 or 17 μm mirrors represents one pixel. The diagram on the *left* represents light reflecting from a mirror to the system lens. The illustration on the *right* is an exploded view of a small region of mirrors in a packaged DMD. (Courtesy of Digital Light Processing™, Texas Instruments, Inc.)

and lower cost, which led to the installation of air bags in almost every new automobile.

This chapter will not recite detailed descriptions of particular commercial products. Instead, brief descriptions will be used to illustrate basic principles and concepts. Part of the chapter is organized according to MEMS-specific issues like wafer singulation. This is useful because some designs and unit processes are well suited for lab scale or low volume production, but impractical in a high volume manufacturing environment. Complete reliance on such an organization would be artificial, because partitioning omits the greatest challenge – combining these individual topics into a manufacturing flow that routinely produces products that meet cost, performance, and reliability expectations. The importance of process integration cannot be

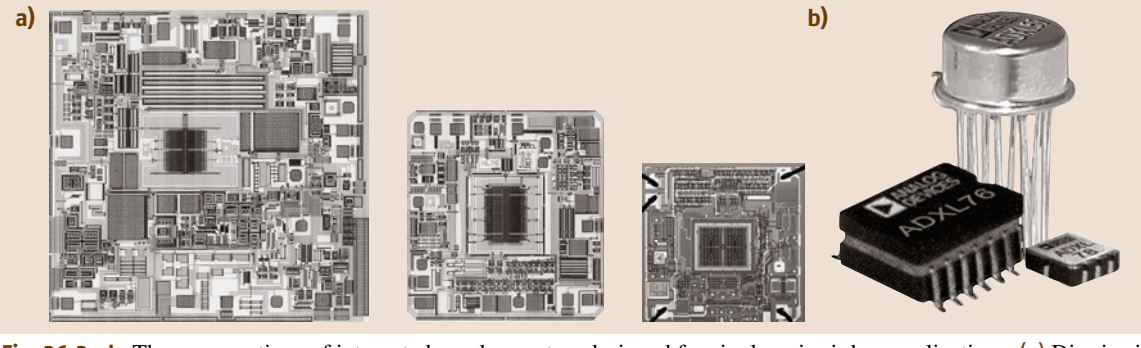


Fig. 36.3a,b Three generations of integrated accelerometers designed for single-axis air bag applications. **(a)** Die size is reduced by a factor of two about every four years. **(b)** Package technology also evolved to meet size, cost, and automated handling requirements. The ADXL50 was packaged in TO-100 seam-sealed metal headers. ADXL76 cerpacs (*bottom left*) are being replaced by the ADXL78 in 5 mm × 5 mm LCC packages (*bottom right*). (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

overemphasized, because changes in one step almost always affect other steps. Solutions may involve design modifications, so design-for-manufacturability is an important topic. Manufacturing and business skills are included to the extent that they apply to stable production of reliable MEMS products.

36.1 Manufacturing Strategy

MEMS manufacturing is based on existing IC technology. However, there are only a few examples of routine MEMS and IC production intermixed on the same equipment (or of MEMS devices with electronics integrated on the chip). Even when a MEMS process step is nominally the same as a standard IC process, the specific conditions or end point or some quality such as film stress will differ. Technologists from Analog Devices and Motorola have discussed these thin film issues in depth [36.7–10].

Manufacturers prefer to use dry etch processes when possible in order to achieve better control of features, minimize waste disposal, and utilize automated equipment developed for the IC industry. However, anisotropic wet etching provides unique value in the production of some MEMS products like pressure, mass flow, and yaw rate sensors [36.11].

Although MEMS production economics is closely tied to the semiconductor model, useful lessons can be drawn from many industries. Some of the key characteristics that affect success are as follows:

36.1.1 Volume

Semiconductor economics is rooted in the principle that expensive processes are cost-effective when many devices are made simultaneously. Obviously, a MEMS product that is sold in low quantities loses this critical advantage. MEMS components are "disruptive technologies" – they are the key element in higher-level products that were previously impractical. It can take several years to develop a higher-level product and build its customer base. Therefore, it is prudent to include an incubation phase in the business plan for a novel MEMS device. Low sales volume during this phase causes high unit costs – batch process cost benefits are achieved only after the market develops.

Start-up companies often do not understand the implications of design-for-volume. It is a major challenge to transfer products made on lab tools to manufacturing equipment. The challenge is sufficient to justify the use of manufacturing people and equipment for product development.

Realistic price-volume estimates are also essential to a good business plan. The highest volume MEMS products (accelerometers and pressure sensors) all represent a small part of the total system cost, even though the functions they provide are absolutely critical to system performance.

Market-pull business models that respond to the needs of a specific large application in a particular industry are usually more successful than technology-driven models that seek applications for new devices. Given this reality, it is prudent to thoroughly understand the target industry and its quirks. Biotech, wireless, image projection, optical communications, and consumer markets offer tremendous growth potential. However, the past and present growth in high volume MEMS products have been driven primarily by the automotive industry, with pressure sensors also benefiting from medical applications. *Sulouff* [36.12], *Weinberg* [36.13], *Marek* et al. [36.14], *Eddy* et al. [36.15], and *Verma* et al. [36.16] discuss automotive applications and the challenges a successful supplier must master.

36.1.2 Standardization

Mature industries develop standards in response to equipment vendors, material suppliers, and customer interface requirements. The MEMS industry is fragmented, so even after 30 years it is not mature. Deviations from semiconductor industry standards are inevitable because MEMS products impose unique requirements on wafer processing, packaging, handling, and testing. For example, KOH wet etching raises mobile ion concerns. Gold used in some optical products is incompatible with CMOS. Any deviation from standard process and material sets is serious, because the price-volume relationship is a steep curve in high volume applications.

36.1.3 Production Facilities

The semiconductor industry is divided into front-end (wafer fabrication) and back-end (product assembly and test) operations. IC design rules are based on pre-

qualified fab, assembly, and test capabilities at specific sites. Consider how each of these functions matches the needs of MEMS production:

1. An IC product might be produced at an internal wafer fab, or designed to run on a standard foundry process. In contrast, production challenges and "know-how" have caused MEMS suppliers to retain wafer fab production within their internal facilities. High volume production in MEMS foundries was limited to pressure sensors. However, Motorola is now outsourcing accelerometer production. This suggests that the MEMS industry will follow the IC model, with foundry services becoming more substantial in the future.
2. IC suppliers have shifted assembly operations to subcontractor facilities located in Southeast Asia. Some MEMS suppliers have followed this model. It is common to assemble low volume products in North America, Europe, or Japan. However, semiconductor assembly and package technology are evolving rapidly and are driven by subcontractors based in Southeast Asia. MEMS companies that do not utilize these resources pay a significant cost penalty. Obviously, the break point at which offshore assembly is attractive differs from one product to the next. Products that have unique assembly requirements and support high market prices may never reach this point.
3. IC testing is often carried out at the assembly subcontractor site. MEMS package and testing comprise a substantial portion of the total product cost, so a high volume manufacturing strategy should consider Southeast Asia packaging and testing. This is complicated by the fact that MEMS testing requires unique stimuli that are not in the standard IC portfolio.

36.1.4 Quality

The goal in lab and low volume production environments is to make functioning devices. As quantities increase, other factors become more important: lot-to-lot repeatability, yield, device performance that meets well-defined specifications, process stability to ensure predictable on-time delivery, documentation and procedures that enable traceability and corrective actions when problems arise, etc. Methodologies that promote stable, high-quality manufacturing practices have been established in several industries. These standard procedures such as ISO9000, QS9000, and cGMP overlap to a large extent. QS9000 was developed to help automotive suppliers implement effective, well controlled processes to produce reliable products. It has several unique features, so it is used as the model in this chapter.

36.1.5 Environmental Shield

IC wafers are completely passivated before packaging in order to meet product stability and reliability requirements. Aside from ink-jet print heads, the active regions of all MEMS devices currently in high volume production replicate this hermetic barrier in some manner. This passivation requirement is noteworthy because it has been a primary limitation on the growth of the MEMS industry. For example, it limits electrochemical sensors to benign or "throw-away" applications.

36.2 Robust Manufacturing

Stable production requires a manufacturing flow that is well controlled when measured against the product performance specifications. This has several implications.

36.2.1 Design for Manufacturability

Low maintenance products and processes must be designed to run on standard equipment. The semiconductor industry has invested billions of dollars to develop equipment that is automated, maintainable, reliable, and capable of supporting well controlled processes. Processes that are implemented on custom-designed equipment put this experience base aside and invariably have a long learning curve.

Equipment

Microstructure release illustrates this issue and how it changes over time. Virtually all semiconductor processes operate in vacuum or at atmospheric pressure. In a brief, unpublished 1992 study related to ADXL50 development, the author demonstrated the use of supercritical CO_2 mixtures for microstructure release and particle cleaning. Unfortunately, supercritical processes were incompatible with the industry infra-

structure, because they require pressures in excess of 7.38 MPa (72.8 atm; 1,070 psi). Implementation would have required process and equipment development, so a release process that ran on existing equipment was selected.

Supercritical equipment designed for MEMS release was introduced a few years later. However, maintenance and cycle time concerns, as well as single wafer capabilities limited its adoption to university and low volume manufacturing.

The intrinsic limitations of non-standard equipment have caused high volume MEMS suppliers to avoid supercritical processes. However supercritical equipment will continue to evolve and become competitive, particularly in applications where the design requirements make alternative processes unusable.

Materials

Process control can be no better than the materials used in the process. SAM (*s*elf-*a*ssembled *m*onolayer) coatings illustrate this limitation. These materials suppress stiction following aqueous release of microstructures by treating the wafers in a SAM solution before drying. Although straightforward in principle, high volume manufacturers have not (yet) adopted SAM processes (Sect. 36.2.3 reviews the processes that are used commercially). There are several reasons for this reticence. Classical SAM materials are chlorosilanes that have at least one organic substituent. Organosilanes have been used to treat the surface of inorganic materials for many years. Applications include coupling agents on the surface of fillers and reinforcing agents in polymers, as well as agglomeration control of particles. The chlorine sites hydrolyze when dissolved in solutions that contain a small amount of water. The resulting hydroxyl groups react with the microstructure surface oxide to produce a chemically bonded layer that reduces stiction due to the low surface energy and hydrophobic nature of the organic-rich surface. The concerns that cause manufacturers to use alternative stiction solutions include:

1. High variability due to the SAM chemical reactivity
2. Particles are common by-products of the reactivity
3. Many SAM process flows require organic baths. This raises health, safety, and waste disposal concerns
4. Possibility of chloride corrosion on aluminum interconnects

Maboudian et al. [36.17], *Srinivasan* et al. [36.18], *Kim* et al. [36.19], and *Pamidighantam* et al. [36.20] describe the use of SAM solutions to suppress stiction. There is also ongoing research focused on solving the manufacturing concerns [36.21–25].

Note that stiction also arises in packaged MEMS products. This yield and reliability problem is discussed in Sect. 36.3.1.

36.2.2 Process Flow and Its Interaction with Product Architecture

MEMS products have structures and functions that do not exist in standard IC devices, so it is unrealistic to expect that every fab, assembly, and test step will reapply a previously qualified IC process. The challenge is to maximize utilization of available IC technology within the constraints of the product function and cost requirements. This leads to fundamental choices in product architecture, product design, and process flow, as discussed below.

Integration of MEMS and Circuits

The integration of MEMS and electronics onto one chip has proven difficult. It adds little value in some MEMS products. However, as noted earlier, integration is essential to products based on the Texas Instruments (TI) DLP (Digital Light Processing™). The third (and largest) category consists of applications where integrated and non-integrated products compete for market share. Pressure sensors and accelerometers are examples of this group.

Image Projection. The first projectors based on the DLP technology were shipped in 1996. DLP chips are produced on CMOS wafers in a mature TI process. The circuitry addresses and drives aluminum mirror arrays that function as on-off pixels – 1.3 million mirrors in the latest SXGA products (Fig. 36.2). *Mignardi* et al. [36.6] describe the DMD (Digital Mirror Device) fabrication process and illustrate some of the factors that must be considered when manufacturing flows are changed. The manufacturing flow patterns three aluminum depositions over SRAM cells that are positioned with 14 or 17 μm center-to-center spacing (Figs. 36.4 and 36.5). The first of these metal films form electrostatic drive electrodes and landing pads. The yoke and hinge layer has spring tips that land on the first metal layer when mirrors are rotated (Fig. 36.6). Mirrors are mechanically connected to the yokes through center pedestals. Sacrificial organic films separate the metal layers. These organics, along with a protective organic cover film, remain on the device until after the wafer is sawn and the chips are mounted in a ceramic package. They are removed by dry etching.

Fig. 36.4 DMD pixel array with tilted and non-tilted mirrors. (Courtesy of Digital Light Processing™, Texas Instruments, Inc.)

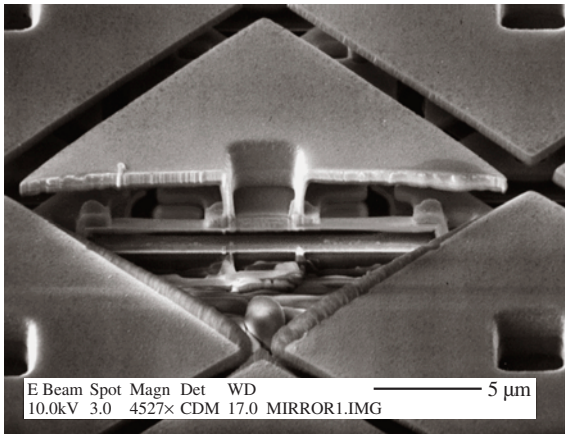


Fig. 36.5 Ion milled section of DMD pixel showing one mirror on its center support and its relation to the underlying layers. Note the close mirror spacing. (Courtesy of Digital Light Processing™, Texas Instruments, Inc.)

Following surface passivation, inspection, and testing, an optical glass sub-assembly is hermetically sealed to the ceramic package (Fig. 36.7). At this point, the device is ready for burn-in and final test – a significant task because each mirror is actuated on chips with as many as 1.3 million mirrors.

In operation, a (nominal) 26 V electrode bias generates an electrostatic force that causes the mirror/yoke assembly to rotate either plus or minus 10 degrees around the hinge. Over one million DMD devices had been shipped by 2002 [36.26].

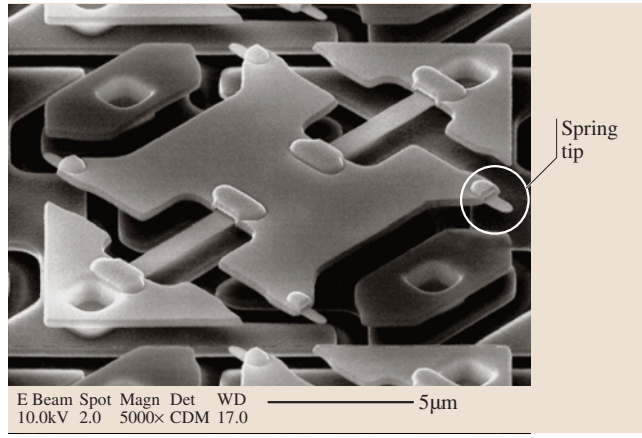


Fig. 36.6 DMD yoke and hinge layer. The spring tip touches landing pads when the drive electrode tilts the mirror. Elastic energy stored in the deformed spring is part of the restoring force that overcomes stiction when the drive electrode voltage is removed. (Courtesy of Digital Light Processing™, Texas Instruments, Inc.)

Unlike the pressure sensors and accelerometers discussed below, TI had no choice but to integrate. It would be difficult, if not impossible, to devise a cost-effective multi-chip solution to control and drive a matrix of over a million mirrors. Successful integration of circuits and mirrors was only one step in the commercialization process. To remain competitive against lower cost LCD products, TI continues to develop their package and test technologies, because package and test comprise a major part of the total product cost. This is discussed further in Sect. 36.2.8.

Pressure Sensors. Early (1970s-era) bulk micromachined piezoresistive pressure sensors from Honeywell, ICT (later Foxboro-ICT), and Kulite were not integrated. However, in the 1980s, Motorola commercialized the

Fig. 36.7 DMD products in hermetic packages. (Courtesy of Digital Light Processing™, Texas Instruments, Inc.)

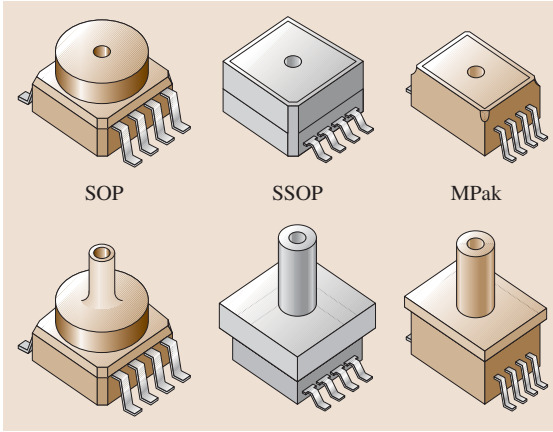


Fig. 36.8 Pressure sensor surface mount packages. (Courtesy of Semiconductor Products Sector, Motorola, Inc.)

first of its MPX5100 series piezoresistive pressure sensor products. These products had bipolar signal conditioning electronics and temperature compensation on the sensor chip. Current MPX products are offered in various plastic packages (Figs. 36.8 and 36.9) with maximum pressure ranges between 10 and 300 kPa (1.45–44 psi). Medical and tire pressure products are also available.

Integrated Motorola piezoresistive pressure sensor products are cited here for illustrative purposes, but capacitive and piezoresistive silicon pressure sensors are available from many companies. Capacitive designs are less temperature sensitive and require less power than piezoresistive designs. However, die size tends to be larger. The fabrication process and interface electronics are also more involved. Piezoresistive products (sensor plus signal processing) are less expensive, smaller, easier to manufacture and suitable for most applications. They rely on the fact that the resistance of silicon changes in response to strain. This effect is very temperature sensitive. However, configuring four piezoresistors as a Wheatstone bridge, judicious selection of the doping level and junction depth and the use of a temperature compensating resistor network minimizes temperature error.

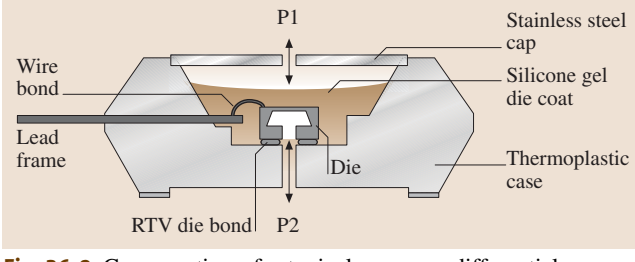


Fig. 36.9 Cross section of a typical gauge or differential pressure sensor. The silicon pressure sensor chip is bonded to a silicon substrate. This bonded unit is mounted in a pre-molded plastic package and protected from the ambient environment by silicone gel. (Courtesy of Semiconductor Products Sector, Motorola, Inc.)

Piezoresistive pressure sensors (Fig. 36.1) require double-sided polished wafers and front-to-back alignment. The piezoresistors are implanted into the front side of (100) wafers and precisely located with respect to the edges of thin diaphragms. The diaphragms are produced by anisotropically wet etching cavities into the back surface of the wafers. KOH etch solutions are usually used to avoid the safety, toxicity, and waste disposal concerns that arise when large quantities of organic etchants like EDP (Ethylenediamine-pyrocatechol-water) are used. Front-to-back side wafer alignment is critical in order to achieve precise resistor location with respect to the cavity edges. Alignment to the wafer crystal axis is also important, because the slow-etching (111) planes determine the cavity wall, and thus the location of the diaphragm edges. The piezoresistors are commonly placed near these edges to maximize the signal (the effect of pressure on diaphragm stress is greatest at the edges). An etch stop is sometimes used at additional cost because diaphragm thickness variations also have a large effect on sensitivity.

Control of mount stress is critical to achieving predictable performance. After wafer fab, some suppliers hermetically bond pressure-sensor wafers to a backup silicon wafer using glass frit. Others anodically bond them to a borosilicate glass wafer that has a thermal expansion coefficient close to silicon (Pyrex 7740 and Schott 8330 are two glasses used for this purpose). Absolute pressure sensors are produced when this bonding step is done in vacuum. As shown in Fig. 36.1 and Fig. 36.9, differential pressure and gauge pressure products incorporate through holes in the backup wafers. These ports allow fluid pressure to be applied through the cavities to the back side of the diaphragms.

After wafer singulation, the chips are mounted and sealed in cavity packages. Soft die attach materials, often silicone-based, are used to decouple the sensor from package and substrate stresses. In some products, the sensor chip is directly mounted to a substrate with a soft die attach without the intermediate backup wafer. Many package variations of MEMS pressure sensors are available, as summarized in Sect. 36.2.8.

The MEMS pressure sensor market is believed to exceed a hundred million pressure sensors annually [36.27]. Automotive applications like manifold absolute pressure (MAP) sensors and a wide range of gas and liquid pressure sensors form the largest market segment. Health care and medical uses such as disposable blood pressure transducers and sensors to measure pressure in angioplasty catheters, infusion pumps, and intrauterine products are the second largest market. Industrial products such as process control pressure and differential pressure transmitters, household appliance, and aeronautical products are also significant. *Maudie* et al. [36.28] review the performance and reliability issues related to appliance applications like household washing machines.

Accelerometers. Most MEMS accelerometer designs apply some variant of Newton's Law, $F = ma$, to sense the response of a proof mass. Many sensing principles, including piezoresistive, capacitive, piezoelectric, and resonant, have been examined [36.29]. Figure 36.10 has a conceptual view of the capacitive design used in Analog Devices ADXL accelerometers.

SensorNor was an early pioneer in silicon accelerometer manufacturing. Their product had a mass bonded to the end of a silicon piezoresistive element in an oil-filled package. However, the first MEMS accelerometer product to achieve large-scale market acceptance was the integrated ADXL50 air bag sensor from Analog Devices. When it was fully qualified in 1993, the ADXL50 was the first integrated surface-micromachined MEMS device of any type in production. The MEMS element in the ADXL50 was part of a closed loop differential capacitance circuit, but was designed to allow routine self-testing after the air bag module was installed in vehicles. This self-test capability eased concerns related to the adoption of a new technology in a critical safety application. The analog output feature also allowed the industry to design single point sensors – a significant system cost savings. *Core* et al. [36.2] and *Sulouff* [36.5] describe the ADXL wafer fabrication process. In essence, circuits are fabricated with a well-established BiCMOS process. The sensor polysilicon is deposited after these high temperature steps, followed by the lower temperature metal and passivation processes. The in situ doped polysilicon is connected to the circuit through N+ doped runners. Following wafer fabrication, thin film resistors are laser trimmed in an automated trim system to meet the specific end user's requirements. After wafer singulation (Sect. 36.2.5), the chips are assembled in hermetic cavity packages and screened in an automated test system to ensure compliance to the performance specification (Sect. 36.2.8).

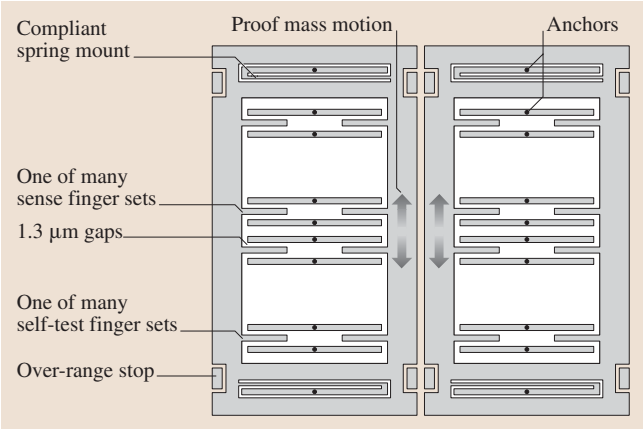


Fig. 36.10 Block diagram of the MEMS element in an ADXL78 single-axis accelerometer. The detailed design applies design-for-manufacturability principles to ensure close control of critical dimensions and residue-free removal of the sacrificial layer. (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

Front air bag sensors have an output of 2 V at a full scale acceleration that is specified by the user (typically 35–50 g). Satellite air bag sensors are located in door pillars and near the front of the automobile and typically have higher ranges (250 g). Regardless of range, all air bag sensors must meet stringent cross-axis specifications, i. e., a front air bag sensor must not react to a side collision or to potholes in the road. This cross-axis requirement also applies to products that are designed to sense in two axes (Fig. 36.11).

Progressive reductions in chip and package sizes (Fig. 36.3) led to cost reductions and penetration into new markets. One example is the ADXL202 (Fig. 36.11). This two-axis, low-g accelerometer has a full scale of 2 g in each axis. It is widely used in industrial, consumer, and automotive applications. Production of integrated MEMS accelerometers at Analog Devices has grown rapidly in recent years (several million devices per month by 2001). It is worth noting that these products are developed using an organizational philosophy that emphasizes design-for-manufacturability. For example, development is conducted on production equipment and directly involves production personnel. This may appear to be expensive and unwieldy, however, it ensures that product designs remain within the bounds of practical man-

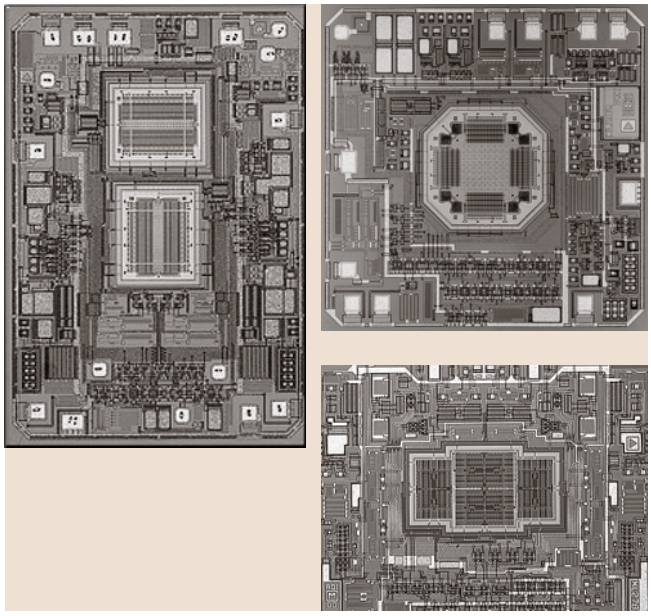


Fig. 36.11 Three two-axis accelerometers. *Top left:* ADXL276 was the first two-axis air bag accelerometer. The ADXL278 family (*bottom right*) is replacing it. *Top right:* 2-g ADXL202. The ADXL202 and ADXL278 are packaged in the LCC shown in Fig. 36.3. (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

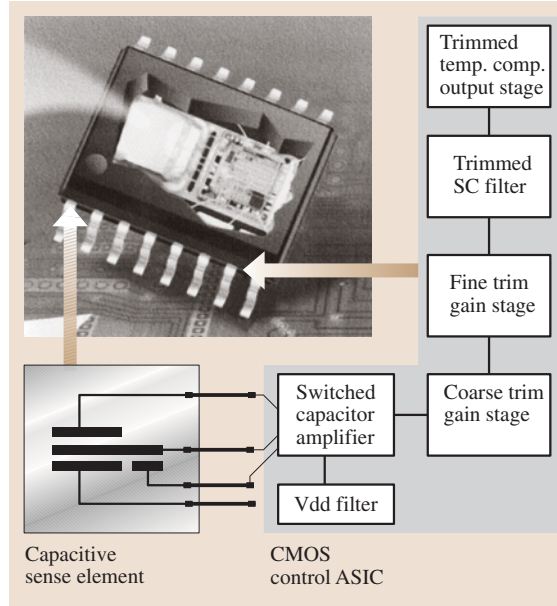


Fig. 36.12 Two-chip accelerometer in 16 lead plastic SOIC package. (Courtesy of Semiconductor Products Sector, Motorola, Inc.)

ufacturing. It also minimizes the difficulties that are normally encountered when new products are transferred to production.

Prior to the introduction of the ADXL50, most of the MEMS air bag candidates were nonintegrated, bulk-micromachined piezoresistive sensors [36.3]. However, Motorola and Bosch soon followed with an integrate-in-the-package strategy. Denso, Delphi, and VTI Hamlin are also active in some market sectors. The Motorola and Bosch products are nonintegrated capacitive MEMS products that are hermetically sealed at the wafer level (Sect. 36.2.4). By interconnecting a separate circuit chip and a sealed accelerometer chip in one package, they are able to use low cost CMOS for signal processing and near-standard plastic packaging (Sect. 36.2.8).

In contrast to the ADXL sensors, Motorola's MMA air bag products measure acceleration normal to the plane of the chip (z-axis). These differential capacitance sensors are formed from three layers of polysilicon. The middle layer is a "proof mass" that is suspended with ligaments and is free to move in response to an accelerating force. Ranges and outputs are similar to the ADXL products. Recent Motorola expansions add x-axis air bag products and a new line of low-g z-axis products (Fig. 36.12).

The Bosch x-axis sensor has interdigitated fingers that are patterned in a thick (11 μm) polysilicon film that is deposited in an epi reactor. Tight control of stress and stress gradients in these films is essential in order to meet sensor performance requirements [36.30, 31]. After patterning, the polysilicon is dry etched using fluorine-based chemistry [36.32] to form the differential capacitance sensor. An HF vapor process is used to remove the sacrificial oxide and release the microstructures. These thick structures have working capacitances near 1 pF [36.33]. *Laermer* et al. [36.34] explain how Bosch applies deep silicon etch processing to several production and experimental inertial sensors. Better known as the "Bosch etch," this deep reactive ion etching technique has pioneered the field of high-aspect-ratio silicon-based MEMS.

The Future of Integration. The high growth rate of integrated MEMS products demonstrates the value of integration in MEMS applications. To date, integration has been limited to large semiconductor companies that have in-house signal processing expertise and the willingness to invest substantial resources over a number

of years. Each company identified a large volume market that could support a profitable integrated product. This "large semiconductor company" barrier-to-entry for integrated MEMS products will diminish in the future due to new SOI wafer technology (Sect. 36.2.4), new flip-chip equipment, and the development of low temperature MEMS processes. With these capabilities, MEMS-only companies will be able to integrate microstructures on circuit wafers that are preprocessed in CMOS foundries. To be successful, they will also have to develop cost-effective manufacturing and quality expertise (Sect. 36.2.9).

Risk

High risk is normal in research programs. However, the tremendous financial investments required to manufacture new products make the risk of failure a key consideration. This section discusses how the perception of manufacturing risk changes with time and is influenced by available knowledge and expertise.

Every development program makes fundamental product architecture decisions. These decisions can lead competing companies down different paths. For example, Analog Devices pursued chip-level integration, while Bosch and Motorola chose package-level integration. Why did these companies choose different paths? The answer is related to available resources, cost, time-to-market, and technical maturity. Analog Devices chose to integrate because of its expertise in process innovation and signal processing. This path did have risks. When the ADXL50 was designed, the fatigue life of polysilicon MEMS devices was unknown. Therefore, a closed loop circuit architecture was chosen in order to keep the proof mass motion under 100 Å. Polysilicon actually has excellent fatigue life. Once this was established, Analog changed their ADXL products to an open loop architecture that can detect capacitance changes as low as 20 zF (10^{-21} F). Most companies do not have the expertise to work in this regime, but the decision to integrate on one chip allowed Analog to make full use of its signal processing knowledge.

The two-chip Motorola and Bosch designs allow the circuit chip to be fabricated on any IC process. This product concept also removed the concern that integrating MEMS and circuits onto one chip might reduce overall product yield. However, it does require higher output signals from the MEMS chip in order to overcome parasitics that arise in the bond wires between the circuit and sensor chips. This requirement led Motorola toward large-area z-axis designs, while Bosch utilized its deep-etch process capability. To minimize cost and ensure reliability, both companies seal the sensors with glass at the wafer level. Sensor and circuit chips are then molded together in one plastic package. Each company built on expertise that it already had in order to minimize risk and cost. Motorola had in-house manufacturing expertise in wafer-level glass frit sealing of pressure sensors that are subsequently molded in plastic packages. Bosch had deep-etch process capability and the world's largest hybrid manufacturing plant, so screen printing and glass sealing were well-established. These considerations plus the concern that MEMS yield loss would undermine product yield contributed to their two-chip decision.

Semiconductor companies continually drive down cost and serve evolving markets by shrinking product size. Analog Devices believes that this long-term roadmap requires an integrated product. They chose to deal with one-chip risks, rather than package uncertainties. ADXL50 chips had been capped and molded in plastic long before the first qualified products were shipped in 1993 [36.35]. However, production-worthy capping equipment did not exist at that time. The effects of plastic package stress on yield and long-term device parametrics were also uncertain. By introducing the initial product in TO-100 hermetic metal packages, the risks associated with custom equipment and package stress were removed.

Reusable Engineering and Facilities to Achieve Economies of Scale

The IC industry produces many products on each process flow. Foundries and packaging subcontractors routinely apply this "economy-of-scale" principle. Only basic items like mask sets and bond pad diagrams are changed. The MEMS industry has been less successful in this area, although techniques like wafer-level laser trim and blowing poly fuses allow accelerometer suppliers to offer "specialty" products without losing the advantages of high volume production.

If only a few devices are produced in a capital-intensive facility, unit costs are high. This issue is particularly critical during the first years of product life when sales volume is normally low. One solution is to build the market by manufacturing the new product in an existing facility. The two-axis low-g ADXL202 (Fig. 36.11) was built using a new mask set with an existing air bag sensor process. Many ADXL202 consumer, industrial, and automotive applications are too small to justify the capital investment risk required to

develop and produce a two-axis accelerometer, but use of an existing design and manufacturing base removed this constraint.

Market Dynamics Drive New Product Opportunities

In the 1980s, low cost and high reliability allowed MEMS pressure sensors to penetrate automotive applications like the measurement of manifold vacuum (MAP sensors). Once a technology is accepted, new opportunities usually develop. For example, tire pressure monitors with RF communication capability are now available in some car models. In the future, this type of monitor will be a safety feature expected by all consumers. Indeed, wireless communication capabilities will also be a common MEMS product requirement.

This example illustrates the fact that MEMS products are merely system components. They must support seamless integration into the larger system. MEMS opportunities in automobile safety systems will be affected by both communication protocols and interconnect technology (copper, fiber optic, and wireless). The growth spurt of the 1990s started with front-air-bag crash sensors located in the center module of automobiles. It has evolved to include satellite crash sensors behind bumpers, side-impact sensors in the B pillar (the vertical post between doors), impact sensors inside doors, rollover sensors, and vehicle dynamic control sensors to improve ride and handling. Each of these applications has unique characteristics that can favor one technology over another. For example, side-impact sensing systems must activate the air bags faster than front-impact systems. This response time characteristic can favor the use of pressure sensors that monitor air pressure inside door cavities – a good solution as long as the cavity remains sealed. An alternative is to place an accelerometer in the B pillar. These thin structural members have very little space, so z-axis accelerometers are sometimes easier to implement. Delphi and Motorola products have this characteristic. Analog's small x-axis accelerometers also meet this space requirement.

Gyros – New Products Produced with Preexisting Manufacturing Capabilities

Gyroscopes are transducers for rotational motion. They have a wide range of uses in platform stabilization and robotics, as well as automotive applications (dead reckoning backup for GPS navigation systems when satellite communication links are lost, skid control systems, and rollover detection). Unfortunately, the non-MEMS and resonant quartz gyros used in military and aerospace applications are rather expensive for many high volume applications. BEI Systron Donner tailored their high performance quartz gyro to meet automotive needs and has shipped several million gyros. Other lower cost gyro and angular rate products, like the Murata vibrating ceramic bimorph products, are used in cameras. However, as a group, these low cost products do not meet the performance required for many applications.

Analog Devices and Bosch reapplied their accelerometer capabilities to develop and manufacture low cost gyroscopes. *Geen* et al. [36.36] describe the Analog Devices product. *Funk* et al. [36.37] and *Lutz* et al. [36.38] describe the Bosch product. These three publications reference previous work and discuss the challenges of producing gyros that meet performance, reliability, and cost targets.

The Analog Devices ADXRS150 gyro (Fig. 36.13) illustrates the technology. It has a resonant microstructure that is integrated with two electronic systems on one chip. One of the electronic systems drives the MEMS resonator. When the ADXRS150 is rotated, Coriolis ac-

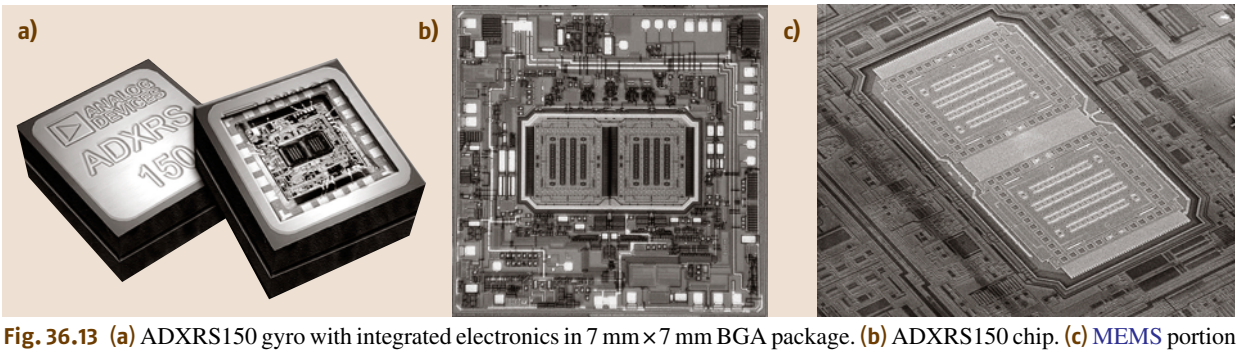


Fig. 36.13 **(a)** ADXRS150 gyro with integrated electronics in 7 mm × 7 mm BGA package. **(b)** ADXRS150 chip. **(c)** MEMS portion of the ADXRS150 chip. (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

celeration is generated in a direction that is perpendicular to the vibration axis of the resonator. This motion is detected by the second accelerometer system. Full scale is rated at 150 deg/s, although it can sustain overloads up to 10,000 deg/s. Integration of electronics on the chip, microstructure modeling, and predictable manufacturing were essential to success, because the full-scale Coriolis motion is only about 1 Å. The combination of differential capacitance detection, area averaging, and correlation techniques allow this motion to be resolved from thermal noise down to about 0.00016 Å [36.36].

Resonant gyros are not new. However, as noted above, their high cost restricts them to specialty applications. Many are non-MEMS, or utilize the piezoelectric properties of quartz, so chip-level integration is not an option. Most require expensive vacuum packaging to achieve high mechanical Q at resonance. Unfortunately, vacuum packaging also makes them susceptible to mechanical damage. Viscous damping in gas-filled devices like the ADXRS150 suppresses the destructive impact events that occur in vacuum-packaged MEMS products during shipping and handling and in-use conditions. Equally important are ADXRS150 design features that allow the measurement to reject mechanical shock, vibration, and other environmental noise sources. As a result, under-the-hood location is practical in automotive applications, an unmatched characteristic that significantly reduces system cost.

Analog Devices and Bosch both needed several years to develop their gyro products. Use of existing accelerometer infrastructure during the prototype and early production phases significantly reduced development costs and manufacturing risks. Early ADXRS150 gyros were sold as engineering grade products in socketable side braze packages. This allowed the development team and customers to understand the design and process parametrics that affect performance, reliability, and yield. Side braze packages are unacceptable in high volume applications, so the ADXRS150 is sold in a hermetic, ceramic, surface mount package.

36.2.3 Microstructure Release

Release stiction occurs when microstructures stick together after the sacrificial material is removed in a wet etch. It can cause considerable yield loss and must be considered in devices like capacitive pressure sensors and accelerometers that have closely spaced, mechanically compliant microstructures. Piezoresistive pressure sensors are not susceptible to release stiction, because these sensing diaphragms are not near other surfaces.

Release stiction is most frequently observed in silicon MEMS devices that use a sacrificial oxide. The oxide is removed by etching in wet HF, rinsed in deionized water, and then dried. Water promotes growth of a hydrophilic surface oxide. It also has a high surface tension. To minimize energy, the high surface tension in the shrinking water droplets causes the microstructures to be pulled together. When they touch, clean oxide surfaces stick, thus destroying device functionality. *Maboudian* et al. [36.17] review release stiction mechanisms, as well as the solutions that have been reported in the literature.

Some manufacturers do not publicly discuss their fab process. Those that do avoid release stiction by using gas phase processes to remove the sacrificial material. Texas Instruments uses photoresist, rather than silicon dioxide as the sacrificial material in their DMD manufacturing process. Two photoresist layers are under the aluminum mirrors. The wafer is also covered with a third organic layer to provide mechanical damage protection and allow for particle cleaning after the wafer is singulated into individual devices. By removing these organics in a dry etch process, the stiction problems that characterize wet etch processes are avoided.

Analog Devices avoids release stiction by dividing its accelerometer process flow into several steps, each of which is well controlled with standard manufacturing equipment. After the microstructure is formed, a few small channels are etched in the sacrificial oxide and filled with photoresist. These photoresist pedestals hold the microstructures in place when the remaining sacrificial oxide is etched, rinsed, and dried. The pedestals are then removed in a dry etch process. *Core* et al. [36.39] and *Sulouff* [36.5] describe this technique in more detail.

Bosch avoids release stiction by holding accelerometer wafers above an HF solution. HF vapors from the solution react with the sacrificial oxide, converting it to gaseous SiF_4 and water. The wafers are heated to prevent water from condensing. This process is noted in *Offenberg* et al. [36.40]. *Anguita* et al. [36.41] describe a similar process.

36.2.4 Wafer Bonding

To date, the growth of MEMS wafer bonding has been driven by three factors:

- Sealing microstructures by bonding wafers together addresses two sources of yield loss in MEMS manufacturing: wafer singulation and particles.

- Wafer-level mounting can be a cost-effective way to control stress in products that are sensitive to variations in mount stress.
- Fabrication of 3-D microstructures from elements that are formed on multiple wafers.

Piezoresistive pressure sensor wafers have been anodically bonded to borosilicate glass wafers for about 25 years. The earliest products were manufactured using little more than a hot plate and a high-voltage power supply. The era of custom built manual equipment has passed, because both Electronic Visions and Karl Suss now offer automated production-worthy tools for anodic, glass frit, organic, and silicon-silicon wafer bonding. *Schmidt* [36.42] reviews wafer-wafer bonding processes, while *Mirza* [36.43,44] addresses MEMS wafer-level bonding applications and equipment.

Anodic Bonding

Anodic bonding is commonly used to seal glass wafers to the back side of bulk micromachined pressure sensor wafers. The process requires that a silicon wafer be placed in intimate contact with a glass wafer that contains a mobile ion at elevated temperature. An applied electric field causes mobile ions (usually sodium) in the glass to move away from the silicon interface toward the cathode at the other side of the glass. Bound negative charges remain in the glass near the silicon interface. These charges produce an electric field that pulls the wafers together and drives oxygen across the interface to anodically oxidize the silicon surface. Hermetic seals are routinely produced between flat wafers when particles are rigorously excluded. Process conditions depend on the glass composition and thickness, but 500–1,000 V at 400 °C with 10 min cycle times is typical. Borosilicate glasses with thermal expansion coefficients close to silicon are used to minimize stress. Anodic bonding promotes surface conformation so hermetic glass-silicon seals can be formed with surface grooves as deep as 50 nm [36.45]. This process has been demonstrated to work with many material combinations [36.46].

Glass Frit Bonding

Glass frit bonding reapplies techniques and materials developed for hybrid processes and cerdip sealing. It is often used to seal pressure sensor wafers to silicon substrate wafers. A second common application is to glass-frit seal bulk-micromachined silicon cap wafers over the microstructures in accelerometer wafers. In this process, fine glass powder is dispersed in an organic binder to form a paste. This paste is screen printed or stenciled in the desired pattern on a wafer. The wafer is then passed through a furnace (typically 430–500 °C). This causes the organics to burn off and the individual glass particles to coalesce into the desired pattern.

The wafer bonding process starts with a wafer prepared as described in the preceding paragraph. This wafer is aligned to a second wafer, where they are held in intimate contact and heated until the glass softens. Liquid glass readily wets the surface of the adjacent wafer when at least trace amounts of oxygen are present. The liquid-wetting mechanism produces hermetic seals even when the wafer surfaces are relatively rough. However, seal widths less than 100 μm are difficult to achieve. Fixtures that maintain wafer alignment through the bonding process are also required. Low-melting lead oxide and boron oxide glasses are commonly used in processes that typically run at 450 °C. Glass wet-out occurs quickly, but process cycle times are limited by the cooling rate (rapid cooling can cause stress gradients and cracking in the glass if proper design practices are not followed). Glass frit bonding to semiconductor wafers requires a good understanding of glass composition effects and process conditions [36.47].

Silicon–Silicon Wafer Bonding

Silicon-silicon bonding has been demonstrated in many IC and MEMS applications over the last 30 years. These processes bring highly polished wafers into intimate contact and are often promoted by a thin hydrophilic oxide that is left on the surfaces by the aqueous cleaning process. Clean, well polished wafers will bond at room temperature. However, heating (800–1,200 °C) increases bond strength by an order of magnitude. A fundamental difference between MEMS silicon–silicon bonding and either anodic or glass frit bonding is that it is used to produce complex microstructures that cannot be made in one wafer. In contrast, most anodic and glass frit bonding applications address assembly and packaging issues.

Recent silicon-silicon advances have solved technical factors like stress control that had limited MEMS applications. For example, Analog Devices balances the stresses at oxide interfaces and the handle wafer to produce flat, three- and four-layer SOI wafers [36.48,49]. The thickness of each layer is precisely controlled using grinding and CMP (chemical-mechanical polishing). Intermediate layers can be patterned, electrically isolated, electrically interconnected and locally etched to produce single-crystal silicon microstructures that are hundreds of microns thick on CMOS compatible wafers. With this new capability, several organizations are fab-

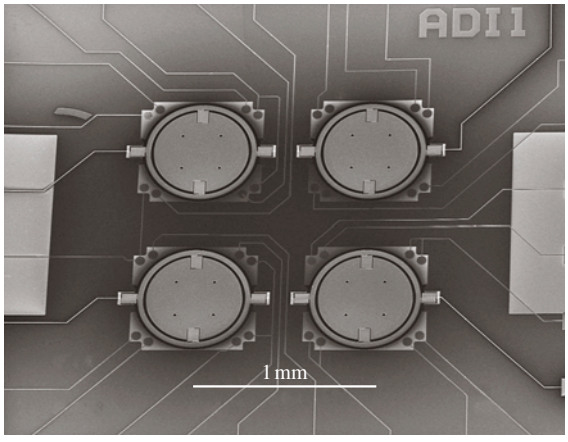


Fig. 36.14 Array of steerable mirrors designed for fiber optic network switching applications. The gimbal-mounted mirrors in this experimental product are electrostatically driven. Closed loop control based on capacitive sense electrodes ensures that the proper angle is maintained. High and low voltage circuits are integrated with mirrors on one chip. (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

ricating MEMS devices for biotech, automotive, and communications applications. One example is the array of gimbal-mounted mirrors shown in Figs. 36.14 and 36.15. These mirrors can be tilted in any axis to steer light between fibers in optical fiber bundles. The mirrors are driven by electrostatic force applied from electrodes that are patterned under the mirrors. Capacitance measurements based on a second set of electrodes are also part of a closed loop control system that maintains position control. Integration of high voltage drive electronics, low voltage control electronics, and MEMS mirrors would be impractical without this SOI capability and related MEMS processes.

36.2.5 Wafer Singulation

The cooling water and the particles generated in a standard IC diamond saw will destroy MEMS wafers unless some form of protection is used. Manufacturers solve this problem in several ways. Bosch and Motorola bulk micromachine cap wafers that are glass frit bonded to surface micromachined wafers. The caps protect the microstructures from water, particles, and mechanical damage so the wafers can be sawn with standard equipment.

The optical function of DMD chips required Texas Instruments to find a different solution for DMD wafers. Originally, they used a partial saw process on wafers that were protected with an organic film. This allowed wafer-level testing, but caused particle contamination and die loss when wafers were broken along the partial saw cuts. They now use a standard saw process, but do it before the mirrors are released. The wafers have the protective organic film over the mirrors, so normal cleaning processes are used. The singulated chips are

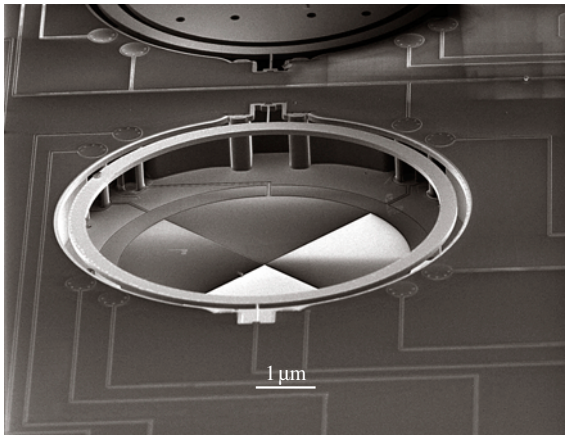


Fig. 36.15 View of the deep etched trenches, interlayer connections, and electrodes under a mirror of the SOI MEMS device illustrated in Fig. 36.14. (Courtesy of Micromachined Products Division, Analog Devices, Inc.)

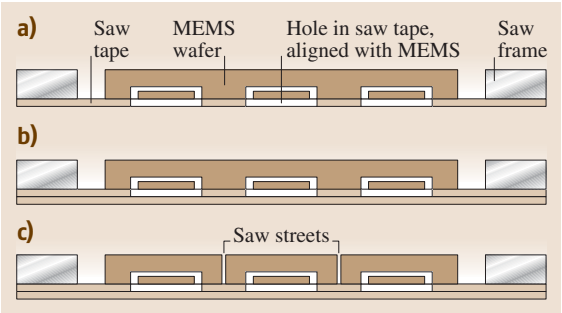


Fig. 36.16a–c Analog Devices "upside-down" saw process flow. **(a)** Saw tape is mounted on a saw frame. Holes are punched in the tape to match the microstructures on the wafer. The wafer is mounted-upside down on the saw tape and aligned so that the microstructures are in the holes. **(b)** A second layer of tape is placed over the first layer to form watertight pockets. **(c)** Wafer is aligned and sawn using standard equipment (adapted from [36.50])

mounted in ceramic packages before the organic layers are dry etched. This flow eliminates a primary source of particle contamination and problems associated with handling partially sawn wafers [36.6].

Analog Devices singulates uncapped MEMS wafers using an upside-down saw process and standard equipment and fixtures that are slightly modified (Fig. 36.16) (*Roberts* et al. [36.50]). A major attraction is that it allows wafer-level test and trim, so only chips that meet the product specification are assembled. It also avoids the added cost and yield loss associated with wafer capping. Although originally developed for accelerometers, this technology was extended to gyro wafers and to wafers that contain optical mirror arrays (see Figs. 36.13 and 36.14).

36.2.6 Particles

Particles cause yield loss in every product that has closely spaced movable microstructures. They also cause large area defects in anodic and silicon-silicon wafer bonded processes. Commercial MEMS suppliers use several solutions:

1. Cap the MEMS wafers before they leave the wafer fab clean area. Bosch and Motorola use this approach to produce accelerometer products (Fig. 36.12). It solves the wafer saw dilemma faced by most MEMS suppliers and allows the use of plastic packaging. Disadvantages include yield loss due to cap misalignment and stress, as well as the cost of the added capping steps. Hermeticity testing is also difficult because capped devices require sensitivity levels beyond the capabilities of standard analytical test equipment.
   Capping protects MEMS from particle contamination during assembly, but does not prevent wafer fab contamination. A greater concern is that glass-frit seal processes and equipment do not usually meet fab cleanliness standards. Once capped, inspection is essentially impossible.
2. Analog Devices and Texas Instruments assemble their MEMS products in clean rooms. This option increases capital costs. TI invested in environmental and process control, rather than develop optical quality capping for the DMD. Analog Devices had solved the wafer saw dilemma, so environmental and process control was the low risk path for particle control.

Particle control is central to high yield and reliability. The first control level is to minimize particle contamination through proper handling procedures, equipment maintenance, environment controls, and process design. The DMD protective organic film illustrates how process design can be used to control particles. This film allows the wafer to be cleaned after wafer saw and die attach.

Even with the best of controls, some particle contamination occurs in every clean room. Electrical tests in the automated wafer- and package-level test programs are designed to detect particles in Analog Devices products. Texas Instruments also finds particles in DMD products by driving each mirror in an array to ensure proper operation. Such screens are good but imperfect. Attempts have been made to replace visual inspection with automated particle inspection based on pattern recognition software. To date, however, machine vision systems have not been able to match the human eye and mind aided by a high quality microscope.

36.2.7 Electrostatic Discharge and Static Charges

Susceptibility to electrostatic discharges (ESD) led semiconductor manufacturers to make ESD avoidance a central criterion in equipment and fixture design, as well as handling procedures. Control of ESD events on the manufacturing floor is only the first step. Analog Devices and Texas Instruments incorporate ESD protective circuitry into their MEMS products as a standard practice to avoid failures in their customer assembly lines, as well as in the final system.

Most MEMS products use electrostatic force to actuate or control suspended microstructures. This design characteristic introduces performance parasitics and failure mechanisms that are not encountered in normal IC devices. For example, charges that build up on a dielectric surface can deflect unshielded microstructures, even when they are several millimeters apart. Such effects are insidious, because closed loop control systems do not always correct for electrostatic-induced errors.

Section 36.3.1 discusses surface treatments designed to suppress stiction caused by surface forces. These treatments often produce a dielectric surface. If the dielectric properties of the treatment and the device design cause the surface to hold a charge, that anti-stiction coating may actually promote electrostatic stiction! Thus, any "solution" must be critically evaluated in order to identify and remove undesirable side effects.

36.2.8 Package and Test

IC packages are environmental (mechanical, chemical, electromagnetic, optical) barriers that protect the chips from surrounding media. Only power and electrical signals and inertial forces pass unimpeded, so most MEMS sensors and actuators require that this barrier be selectively penetrated. Package and test functions are unique to each company's design and MEMS product type. Standard semiconductor packaging equipment and processes are modified when possible. However, even the line of demarcation that separates wafer fab from packaging is unique to each product.

High volume MEMS testing is even more challenging than MEMS packaging. In essence, the task is to measure and trim the response of products mounted in non-standard packages to calibrated stimuli over a range of temperatures. Automated IC test equipment does not have the calibrated pressure, acceleration, or optical test functionality required for this task. *Maudie* et al. [36.51] detail the test system elements that are required to support volume production of MEMS products.

Pressure Sensors

Pressure sensors are extremely susceptible to stress, so die-attach stress variations broaden performance distributions. The uniformity produced by wafer-level bonding is a primary reason why these processes are used to make the first-level package in many pressure sensor products. Soft silicone and fluorosilicone rubber products are often used to attach the sensor chip to the package to maximize isolation from mechanical stresses transmitted through the mount.

Most semiconductor products are assembled in plastic packages. The automated plastic presses transfer viscous epoxy compounds into multi-cavity molds at 175 °C and pressures near 6,900 kPa (1,000 psi). The hot liquid plastic creates high shear stress as it flows around devices in the mold cavities. Further stresses develop as the plastic hardens and cools. Even if this technology were adapted to create pressure ports in packages, the mold stresses would substantially affect pressure sensor performance.

Pre-molded plastic packages avoid the stress problem and incorporate pressure ports (Figs. 36.8 and 36.9). Polyphenylene sulfide (a high temperature thermoplastic) and medical-grade polysulfone are commonly used to make pre-molded pressure sensor packages.

The liquid or gas that is being measured can cause corrosion of interconnects and chip metallurgy, as well as parasitic leakage paths in the sensor. For that reason, a barrier such as silicone or fluorosilicone gel is often applied over the sensor. The soft gel transmits pressure with high fidelity over a wide temperature range and passivates the chip against many types of chemical attack. Parylene is also used for this purpose. This vapor-deposited organic coating is much stiffer than the gels, so thickness is typically controlled near 1 μm.

Petrovic et al. [36.52] tested the ability of fluorosilicone gels and several Parylene C thicknesses to protect powered pressure sensors against automotive and white goods benchmark liquids. One purpose of their study was to propose and demonstrate a formal media compatibility test protocol for pressure sensors similar to the IC industry standard tests. They found that the lifetime of sensors coated with both gel and parylene was considerably longer than sensors coated with either gel or parylene alone. *Petrovic* [36.53] later summarized the advantages and limitations of gels, parylene, and other techniques.

Figure 36.8 has examples of commercial plastic packaged pressure sensors. Pressure calibration equipment and software must be custom designed and built for these products. However, most pressure products are dimensioned to match IC packages in order to maximize compatibility with standard IC test handlers.

The harsh environment of some under-the-hood applications has been used to justify the high cost of metal cavity packages. For example, Bosch introduced a piezoresistive MAP sensor with signal conditioning that was hermetically sealed in an evacuated T08 header [36.54]. In-package trim was based on planar thyristors that were zapped as needed to bring the sensor within spec. The design avoided chip metallurgy corrosion and surface electrical parasitics by using the back surface of the silicon diaphragm as the fluid interface.

Much more expensive packaging is used in the pressure and differential pressure transmitters that are designed for industrial process control. These products seal the sensor in silicone oil behind thin metal diaphragms. Such products can measure pressure differences of 20 kPa superimposed on a pressure of 20,000 kPa (3,000 psi) over wide temperature ranges with high accuracy and stability. Soft die attach materials decouple the sensor from thermal stresses that arise when they are mounted in metal housings. Note, however, that silicone die attach materials cannot be used if the transmitter is filled with silicone oil. Fluorosilicones are one alternative. Piezoresistive sensors have been used in pressure and differential pressure transmitters since the 1980s. *Fung* et al. [36.55] describe a relatively new version. This product uses a piezoresistive polysil-

icon sensor to measure both absolute and differential pressure on the same chip. See *Chau* et al. [36.56] for early work on a multirange version of this technology.

Image Projection

The Texas Instruments DMD poses unusual challenges because the package lid must be an optical quality glass with anti-reflection coatings to improve optical performance and reduce heat load. The glass lid has opaque borders around the image area to minimize stray light effects and create a sharp edge on the projected image. It is fused to Kovar frames to produce lid sub-assemblies. The package base is a multilayer alumina substrate with co-fired tungsten to provide the electrical interconnects. The package sidewalls are formed by brazing a Kovar seal ring to the substrate. Heat dissipation through this substrate is a significant consideration because the service life of DMD mirror hinges (Fig. 36.6) is largely determined by operating temperature [36.57]. Excessive high temperature creep results in "hinge memory" and would cause a gradual drift in mirror orientation.

The DMD chips are die attached in the package cavity before the mirrors are released. A dry etch process is used to remove the organic sacrificial layers. The die are then passivated and tested. Getter strips are attached on the inside surface of the glass at the sides to control vapor composition in the package cavity before the lid is aligned and seam sealed to the package seal ring. Image quality requirements place stringent requirements on handling, alignment, and spacing tolerances. Thermal stresses also arise when dissimilar materials are joined in high temperature processes to produce the final hermetic package. The DMD packages (Fig. 36.7) and the related processes are discussed further in *O'Connor* [36.58], *Mignardi* et al. [36.6], *Bang* et al. [36.59], and *Poradish* et al. [36.60].

DMD mirror release does not occur until after die attach. This has advantages with respect to wafer saw and particle suppression, but it sacrifices the economic attraction of wafer-level testing. There is no commercial test system capable of combining CMOS electrical testing with 100% testing of optical mirrors. Therefore, a custom electro-optic test system was built around an x-y-θ translation stage with a CCD camera, light source, control hardware and software, and test programs [36.6]. A characteristic of the papers published by the DMD group is their effective use of this test system to examine problems and statistically validate the solutions.

Assembly of the DMD on the projector electronics board requires alignment with the system optics, in addition to electrical connections. Initially, the DMD was held with a plastic clamp and electrically connected through elastomer pads that had alternating layers of conductive and nonconductive material. This allowed easy replacement if a DMD was damaged during assembly. However, the impedance of this connector system was too high for new, higher speed products. Electrical intermittents were also observed. Therefore, it was replaced [36.61] by a grid of c-shaped springs (cLGA^{TM}, Intercon Systems, Inc., Harrisburg, PA).

The DMD has about 40% of the business and entertainment image projection market [36.62]. Products based on liquid crystal technology serve the balance of the market. Market share is largely driven by price, so considerable effort has been placed on reducing the cost of DMD package and board-level assembly. *Migl* [36.61] reports on the assembly benefits achieved by replacing the epoxied heat sink with a mechanically attached heat sink. *Jacobs* et al. [36.62] describes the effort to replace the seam sealed window mount with a lower cost epoxy-bonded design. In theory, such a bond is not hermetic. However, the team realized that proper material selection, design, analysis, and use of moisture getters would produce a low humidity package through the product life. This change was not released to production. If it is implemented, success will require that adhesive bond integrity also be maintained, because the adhesive joint sustains thermal expansion and mechanical clamp stress cycling each time the projector is used.

Accelerometers and Gyros

The Bosch and Motorola capping processes are a first-level package that is applied at the end of wafer fab. After capping, the wafers can be sawn using standard equipment and do not require clean room assembly conditions. In principle, capped devices are compatible with standard plastic packaging. However, molding stresses can be a serious problem. To minimize stress, capped sensor chips are mounted on lead frames with a soft elastomer and often coated with a silicone gel to isolate the sensors from the package stress. Motorola products are calibrated by burning EPROMS after molding. Their early air bag accelerometers were molded in DIP and SIP packages for assembly on through-hole circuit boards. The shift to smaller surface mount SOIC packages reduces the quantity of plastic, so package stresses are reduced. To maintain compatibility with standard tooling in small thin plastic packages, either the cap or the MEMS wafers must be backlapped.

Analog Devices uses a near-standard saw process (Sect. 36.2.5 and Fig. 36.16), so capping is not required

if hermetic cavity packages are used. Such packages are standard, but more expensive than molded plastic. A major attraction of cavity packages is that they eliminate plastic package stress (see page 1099), because the only mechanical connection between the chip and the package is through the die attach and bond wires. As a result, ADXL accelerometers and ADXRS gyro products are fully tested and trimmed on automated systems before the wafers are sawn. A lot-tracking system transfers this data to automated die attach systems. These systems are programmed to pick only chips that meet performance specifications from the wafer saw film frames. Thus, reject die are not assembled into packages. After packaging, devices are tested to ensure conformance to specification. However, package-level trim is not required, because the assembly process and cavity packages do not appreciably shift device parametrics.

The evolution of ADXL cavity packages is illustrated in Fig. 36.3. Initial products were packaged in TO-100 metal packages. This seam-sealed package is useful for development purposes. However, it is expensive and incompatible with the automated equipment used to assemble electronic circuit boards. Therefore, the early air bag sensors were soon switched to cerdips. Cerdips are made from low cost, molded, ceramic bases that have glass seal surfaces. The high process temperature requires use of a silver-glass die attach product. The full assembly flow includes two or three furnace passes near 450 °C to produce hermetic cavity packages.

IC products have been packaged in cerdips for decades, but through-hole circuit boards are seldom used today. A simple change in lead forming allows cerdips to be fully compatible with standard surface mount boards ("cerpacs"), so cerdips and cerpacs are assembled on the same equipment. Most of the ADXL products shipped in the late 1990s were packaged in cerpacs. However, small LCCs have become the package of choice for one- and two-axis accelerometers. The LCC (Leadless Chip Carrier) uses solder-sealed ceramic bases and metal lids. Organic die attach materials are practical, because the furnace gas is nitrogen, rather than air, and temperatures are about a hundred degrees lower than cerpac furnace temperatures. ADXRS gyros are assembled in a solder-sealed 7 mm × 7 mm ceramic ball grid array package (Fig. 36.13).

The fundamental message of this section is that IC package technology – both plastic and hermetic – is evolving very rapidly. Cost-competitive suppliers must remain cognizant of these trends in order to use them to best advantage. MEMS products that are packaged in a way that is not compatible with standard IC equipment and interfacing standards bear a significant cost premium. High package cost seriously limits market size because the price-volume curve is steep for most products. The package size trend is also critical to new market penetration. In summary, customer interface and new application requirements will continue to drive down both the cost and the size of most MEMS products.

The ADXL and ADXRS automated testers and handlers measure multiaxial linear and rotational acceleration. Analog Devices has a division that designs and manufactures automated test systems (many of the corporation's IC products have unique test requirements), so it avoided some of the problems associated with the procurement of custom-automated test equipment. MEMS suppliers have relied on custom-designed test equipment, but this situation is gradually changing. For example, Multitest GmbH now makes handlers with integrated shakers that are specifically designed for accelerometer testing.

36.2.9 Quality Systems

Quality systems are an intrinsic part of stable MEMS production. Management must drive a systemic approach to quality and set continuous improvement as a high priority goal.

Several quality systems define continuous improvement methodologies. In general, they formalize the process used to minimize defects and variations in products and in the processes used to manufacture them. Automotive supply companies must implement quality systems that meet QS-9000 [36.63] before product volume ramps up and continue them through the product life cycle. Four elements that are worthy of note:

1. *Failure Mode Effect Analysis* (*FMEA*). Early in the development phase, a team with representatives from several disciplines reviews the design (or process) in order to identify possible causes of failure. Each potential failure mode is given three numerical rankings. One ranking represents the likelihood of an occurrence, while the others rank the severity of that result and detectivity. These numerical scores are combined for each potential failure mode in order to identify the issues that merit the most attention before they become problems. FMEA spreadsheets are periodically updated as the product or process moves into production. More targeted versions are used in products or processes that are in stable production.

2. *Process Control*. Every significant process must have a "short loop" monitor to ensure stability. This measurement is tracked and statistically analyzed with respect to the control limits to identify changes before the process strays beyond the control limits.
3. *Review Boards*. Every change, unusual occurrence, and proposed solution is assessed by boards that meet on a regular basis to ensure that it does not put product quality at risk. Affected production material is put aside until the appropriate board approves its release. This may appear to be expensive and bureaucratic. However, the cost of a scraped wafer lot is insignificant when compared to the cost of a field replacement program.
4. *Procedures and Specifications*. Each step must be fully specified and identified in the process flow, along with the appropriate metrics for each lot. The compilation of this data in a retrievable form is an essential part of every continuous improvement program. By combining these product and process databases with lot tracking software, Pareto charts linking yield loss and test failures to process variations can be identified and eliminated. Experience has shown that unexpected second-order effects are present in every production line – but data-driven decisions on indirect effects cannot be made on small test populations.

Implementation of a continuous improvement program requires well-informed failure analysis teams, methodologies, and programs. Such analyses start with gathering facts and making relatively simple tests. Often this is sufficient. However, MEMS products are susceptible to uncommon failure modes, so *Walraven* et al. [36.64] gathered examples of more powerful analytical techniques and showed how they are applied to MEMS devices.

Each supplier uses the information and insights gathered from product performance, as well as control and yield data at different points in the process flow to refine their operations. *Douglass* [36.65], for example, outlined the yield loss and failure mechanisms observed in early Texas Instruments DMD products. Concerns like hinge memory, hinge fatigue, particles, stiction, and environmental robustness were each addressed and mitigated by focused teams. For example, *Mignardi* et al. [36.6] describe the partial wafer saw process used in the original DMD manufacturing flow process. Breaking these delicate wafers into individual product chips generated particles that caused yield loss and were potential sources of field failures. A new flow based on a standard full saw process eliminates this particle source. The result was an increase in both yield and reliability. Hinge fatigue characterization showed that bulk metal fatigue models do not properly describe thin film behavior. The hinge memory effort required an understanding of how thin film metal creep is affected by alloy composition and the environment [36.57]. Stiction control involves surface passivation, spring design, mirror dynamics, and moisture level [36.26, 65]. These publications give an insight into the quantities of data and the time required to bring robust products to market. They also illustrate how data-based evaluations can uncover unexpected effects like the acceleration of hinge creep by adsorbed moisture [36.26].

The second continuous improvement example applies to MEMS integration. High yield loss is perceived to be a serious risk when electronics and MEMS are integrated on one chip. Indeed, the initial yields on the Analog Devices integrated MEMS accelerometers were not impressive. However, it increased each year due to the work of many mission-oriented teams. Yield has reached defect density limited levels because the teams eliminated all significant failure mechanisms.

Coincident with the annual yield increases were reductions in customer failure rates. Failure rates in Analog's accelerometer products are in the low single digit ppm range. Quality does pay.

36.3 Stable Field Performance

Some topics discussed in Sect. 36.2 like particles and ESD are equally relevant to long-term stability. Most microstructure products have elements that are in close proximity, so stability is also affected by surface characteristics and mechanical shock.

36.3.1 Surface Passivation

Early, 1970s-era piezoresistive pressure sensors were not passivated. Like early ICs, these wafers only had oxide over the piezoresistors. Performance was

inconsistent and drifted over time, because surface interactions with moisture and other atmospheric gases created a variety of shunts, parasitics, and charging issues.

Electrical Surface Passivation

The introduction of silicon nitride passivation and conductive field plates led to stable products that were hitherto unobtainable. The ability to achieve measurement stability was absolutely critical to growth of the MEMS pressure sensor industry.

Aside from pressure sensors, few MEMS products are passivated. Indeed, standard plasma nitride processes are usually impractical, because they produce dielectric coatings that support static charges. Such charges can cause electrical drift, or be the source of electrostatic forces that cause stiction.

There have been published reports attributing instability to the lack of passivation on microstructures. For example, Analog Devices implemented a special process in ADXL50 accelerometers before the product was released in order to suppress high temperature electrical drift [36.66]. This was not a moisture effect, because it occurred in dry nitrogen packages. Later ADXL designs eliminated the root cause.

Scientists at Lucent Technologies [36.67] observed anodic oxidation in polysilicon electrodes used to electrostatically drive mirrors in optical cross-connect products. If allowed to occur, this corrosion mechanism would cause the mirror position to drift and the product to eventually fail. The study concluded that, within the limits of the test, maintaining low moisture in these optical packages eliminates anodic oxidation.

The preceding paragraphs suggest that lack of MEMS passivation can affect long-term product stability. This issue is very design-related and can be driven by factors other than humidity. Even when low package humidity is determined to be adequate, design reviews should consider how normal manufacturing variations, outgassing and diffusion over the product life affect the moisture level in the gas adjacent to the microstructures.

Mechanical – Stiction and Wear

In-use stiction is difficult to predict and may not become an issue until manufacturing volumes increase, or when endusers handle the part in ways that are not anticipated [36.68]. The resulting product liability and field replacement programs carry substantial financial costs and have caused MEMS suppliers to withdraw from the market. End users have little tolerance for failure. Since in-use stiction is a failure mechanism in MEMS devices, a company that solves this problem creates an effective market barrier against competitors.

The scale of this problem varies between product areas. For example, the Texas Instruments DMD has moving mirrors that are designed to touch down on a substrate and later release. This product must address both stiction and wear. In contrast, the accelerometer products from Analog Devices, Bosch, and Motorola have proof masses that are suspended on compliant springs. By design, an automobile collision only displaces the proof mass by a few percent of the gap that separates it from adjacent surfaces. It is difficult to generate a shock wave in an air bag module that is sufficient to cause these proof masses to contact adjacent surfaces (acceleration levels of several thousand g are required). However, handling of the discrete packaged parts during test, shipment, and module assembly frequently cause shock events that bring MEMS surfaces into contact. For example, *Li* et al. [36.69] used both math modeling and tests to show that dropping a packaged MEMS accelerometer from the height of a table top generates several tens of thousands of g in the device when it lands on the floor. Thus, noncontacting MEMS products must be designed to withstand at least transient stiction.

Design. When shock, vibration, or functional operation causes MEMS elements to touch, the mounting springs are designed to pull them apart (Fig. 36.10). High stiffness springs are desirable to suppress stiction, but low stiffness springs increase measurement sensitivity and signal-to-noise ratio. Thus stiction is a fundamental MEMS design constraint that often limits product performance. Each manufacturer has proprietary design practices to address this performance versus reliability trade-off. In general, designers suppress stiction by minimizing the sources and applying supplementary techniques to ensure recovery when contact occurs:

1. Thorough analysis and testing to move harmful resonances beyond the range where they might be excited. Note that MEMS component-level analysis is insufficient – this is a system issue based on the packaged part as it is handled through manufacturing, system assembly, and end use.
2. Surface modification (discussed in the next section).
3. Minimizing contact area by integrating bumps and "stoppers" into the design [36.70].
4. Elimination of dielectric surfaces that may accumulate surface charges and result in electrostatic attraction.

5. Release of DMD mirrors from the touchdown position is assisted by pulsing the reset voltage to excite a mirror resonant frequency [36.65, 71].
6. Use of gas-filled packages to reduce contact velocity, as discussed later in this chapter.
7. ESD-protected designs that prevent voltage transients that may cause electrostatic attraction.

The only product in significant production that contains MEMS elements that are designed to touch is the Texas Instruments DMD. The proprietary aluminum alloy springs used in the mirror contact points touch and release millions of times during the product life. Considerable development was required to overcome metal creep that occurred in early designs.

Surface Modification. In addition to the touchdown springs, early DMD die were treated with perfluorodecanoic acid vapor before the glass lid was sealed onto the package. This treatment [36.72] created a low energy monolayer coating that suppressed initial stiction. However, improvements were required to achieve longer wear life, higher production rates, and lower contamination levels. A new process [36.73] places capsules inside DMD packages. After cleaning, the packages are immediately sealed. An oven bake releases anti-stiction vapor from the capsules to create a monolayer organic coating on surfaces inside these sealed packages. Excess anti-stiction material in the capsules maintains a low vapor pressure in the package cavities throughout the product life. Therefore, wear damage at the mirror contact springs is continuously repaired to prevent stiction failures before they occur.

Maintaining a low level of organic vapor inside the DMD reapplies a concept used with other organics to prevent corrosion of electronic metallurgy and contact fretting of separable connectors (see, for example, [36.74, 75]). These treatments have been shown to add years to the useful life of electronic systems that are in enclosures but also exposed to aggressive environments.

Stiction can be minimized by design and process spin-offs – stiff springs, control of contact area, etch residues, etc. Some suppliers design within these limits; others seek to modify the MEMS surfaces. The requirements of any surface modification technology are quite substantial. In addition to normal manufacturing requirements (stable, scalable process and materials), the process must be conformal and must uniformly treat all areas, including microstructure bottom and side surfaces. If a conductive material is used, it must somehow be patterned in order to avoid electrical shorting. If a dielectric surface is formed, it must be extremely thin because thick dielectrics support surface charges. Such charges cause performance shifts and induce electrostatic stiction.

Since devices with movable microstructures are particularly susceptible to stiction, most suppliers who develop production-worthy surface modification processes consider this information a trade secret. Analog Devices chose to patent portions of its technology. The following paragraphs give an overview of the published Analog Devices surface treatment technologies. *Martin* [36.66] describes the evolution of these technologies from 1992 to 1999. Further information is contained in *Martin* et al. [36.76] and *Martin* [36.77].

The earliest of the published Analog surface treatment technologies was implemented after stiction was observed during automated testing and module assembly of air bag sensors that were packaged in cerdips. Cerdip assembly typically includes at least two furnace processes in air at 440–500 °C, so they seldom contain organic materials. The high temperature furnace air thermo-oxidizes adsorbed materials and removes them from the MEMS surface. The resulting surfaces are extremely clean. Unfortunately, clean inorganics have high surface energies. Therefore, adjacent microstructures readily stick together if shock and vibration during handling cause them to touch. Cerdips (and the surface mount cerpacs) have very low moisture levels (Mil specs allow up to 5,000 ppm moisture, but < 100 ppm moisture is common in a well controlled manufacturing line). Deliberately raising the moisture level was found to be a potential solution because, within limits, moisture adsorption reduces surface energy. However, the solution adopted by Analog Devices in the mid-1990s was the anti-stiction agent (ASA) process. This process involved dispensing a controlled amount of a pure siloxane liquid (ASA) into each package immediately before sealing. ASA selection was based on thermo-oxidative stability, low volatility, purity, and the requirement that it be a liquid at the dispensing temperature. As packages heat in the furnace, adsorbed material on the microstructures is removed, leaving chemically active, high energy surfaces. Further heating volatilizes the ASA. Before escaping into the furnace, the ASA vapor surrounds the reactive MEMS surfaces and chemically bonds to them. The result is infusion of organic groups into the native oxide surface. Organic-rich surfaces are unreactive and have low surface energy. Thus the treatment is self-limiting after the first vapor molecules react. Obviously, ASA volatilization must be complete before the glass softens and seals the package cavity. Some

ASA degrades in this process. However, silicone thermo-oxidation simply adds a few angstroms of silicon oxide to the native oxide, so the ASA process, by design, is non-contaminating.

Although variations and refinements of the ASA process were developed, this technology required that each individual package be treated. Automated dispense equipment made ASA acceptable when product volumes were moderate. However, as production approached several million per month, it became evident that a wafer-level anti-stiction process was desirable.

Assessment of candidate processes and materials led to a vapor treatment based on a custom-synthesized solid polymeric siloxane. This polymer is volatilized in a standard CVD furnace that treats up to a hundred wafers at a time. Repeatability is extraordinary because thickness variation within a furnace run, or from run to run, is only about one angstrom.

Package Environment Effects. The viscosity and density of the gas that contacts MEMS resonators limit the device Q, so resonant devices are often designed to operate in vacuum. Unfortunately, this makes them susceptible to handling damage. Viscous fill gases reduce destructive excursions of high-Q microstructures when they are mechanically shocked. They also provide squeeze film damping to cushion the contact event as surfaces approach each other. None of the high volume MEMS products on the market are evacuated. Presumably, this is due to the high cost of vacuum packaging and the susceptibility to handling damage.

Many investigators believe that suppression of in-use stiction requires low humidity in the package. This view is not surprising because most MEMS research groups have observed stiction caused by aqueous surface tension during microstructure release. However, in-use stiction and wear test results published by several investigators lead to a different conclusion [36.76, 78–83]. In general, they find high stiction in both dry and moist environments. However, stiction is substantially suppressed at intermediate humidity levels (approximately 15–40 % relative humidity at room temperature). This is equivalent to a moisture level of 3,000–10,000 ppm. Note that the limits cited by different authors vary considerably due to the use of different materials and test conditions. Stiction at low humidity is probably caused by high surface energy on inorganic oxides that do not have adsorbed surface films. The low stiction and wear rates observed at intermediate humidity are attributed to the passivating and lubricating effects of adsorbed water. Capillary forces caused by the high surface tension of water become dominant at high humidity.

36.3.2 System Interface

MEMS products are mounted in systems that have much greater mass than the device itself. For this reason, it is difficult to transmit a mechanical shock wave to a MEMS sensor that is sufficient to cause failure after it is installed in a module. However, failures are possible when the devices are handled during module assembly. Improper functional response is also possible if the module mount system has a mechanical resonance that amplifies or reduces mechanical transmission under different conditions. This is a particular concern in automotive safety applications because every automobile platform has a different mechanical signature. Proper module mount design is required to ensure that the impact signal from an automobile crash is properly transmitted to air bag modules.

Suppliers of large IC die closely examine board-level stresses in order to minimize solder joint fatigue. Such stresses are more serious in MEMS products because the device, as well as the package are susceptible to mounting conditions. Furthermore, customer processes impose these stresses, so a supplier may have little information with respect to their origin or existence. One solution is to define and qualify a product-specific mount system, as Texas Instruments has done with the DMD. The ideal, however, is to devise products that are compatible with standard board footprint and assembly processes. This challenge became more difficult with the shift from through-hole to surface mount technology, because the mechanical isolation provided by the package pins is no longer present. Small package size, in-package, and within-chip isolation techniques, as well as thorough package-board stress analysis are used to address this concern.

References

36.1 J. B. Angell, S. C. Terry, P. W. Barth: Silicon micromechanical devices, Sci. Am. **248** (1983) 44–50

36.2 T. A. Core, W. K. Tsang, S. J. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. **36**(10) (1993) 39–47

36.3 W. Kuehnel, S. Sherman: A surface micromachined silicon accelerometer with on-chip detection circuitry, Sens. Actuators A **45** (1994) 7–16

36.4 K. H.-L. Chau, R. E. Sulouff Jr.: Technology for the high-volume manufacturing of integrated surface-micromachined accelerometer products, Microelectron. J. **29** (1998) 579–586

36.5 B. Sulouff: Integrated surface micromachined technology. In: *Sensors for Automotive Technology*, Sensors Applications, Vol. 4, ed. by J. Marek, H.-P. Trah, Y. Suzuki, I. Yokomori (Wiley-VCH, Weinheim 2003) Chap. 5.2

36.6 M. A. Mignardi, R. O. Gale, D. J. Dawson, J. C. Smith: The digital micromirror device – A micro-optical electromechanical device of display applications. In: *MEMS and MOEMS Technology and Applications*, ed. by P. Rai-Choudhury (SPIE, Bellingham 2000) Chap. 4

36.7 K. Nunan, G. Ready, J. Sledziewski: LPCVD & PECVD operations designed for iMEMS sensor devices, Vac. Technol. Coat. **2**(1) (2001) 26–37

36.8 M. Williams, J. Smith, J. Mark, G. Matamis, B. Gogoi: Development of low stress, silicon-rich nitride film for micromachined sensor applications, *Micromachining and Microfabrication Process Technology VI*, Proc. SPIE **4174**, ed. by J. Karam, J. Yasaitis, Proc. SPIE **4174** (2000) 436–442

36.9 Z. Zhang, K. Eskes: Elimination of wafer edge die yield loss for accelerometers, *Micromachining and Microfabrication Process Technology VI*, ed. by J. Karam, J. Yasaitis, Proc. SPIE **4174** (2000) 477–484

36.10 G. Bitko, A. C. McNeil, D. J. Monk: Effect of inorganic thin film material processing and properties on stress in silicon piezoresistive pressure sensors, Proc. Mat. Res. Soc. Symp. **444** (1997) 221–226

36.11 A. Hein, S. Finkbeiner, J. Marek, E. Obermeier: Material related effects on wet chemical micromachining of smart MEMS devices, *Micromachined Devices and Components V*, ed. by P. French, E. Peeters, Proc. SPIE **3876** (1999) 29–36

36.12 B. Sulouff: Commercialization of MEMS automotive accelerometers, 7th International Conference on the Commercialization of Micro and Nano Systems (COMS), Ypsilanti 2002 (MANCEF, Albuquerque 2002) 267–270

36.13 H. Weinberg: MEMS sensors are driving the automotive industry, Sensors **19**(2) (2002) 36–41

36.14 J. Marek, M. Illing: Microsystems for the automotive industry, Electron Devices Meeting, IEDM Technical Digest International, San Francisco 2000 (IEEE, New York 2000) 3–8

36.15 D. S. Eddy, D. R. Sparks: Application of MEMS technology in automotive sensors and actuators, Proc. IEEE **86**(8) (1998) 1747–1755

36.16 R. Verma, I. Baskett, B. Loggins: Micromachined electromechanical sensors for automotive applications, SAE Special Pub. **1312** (1998) 55–59

36.17 R. Maboudian, R. T. Howe: Critical review: Adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B **15**(1) (1997) 1–20

36.18 U. Srinivasan, M. R. Houston, R. T. Howe, R. Maboudian: Alkyltrichlorosilane-based self-assembled monolayer films for stiction reduction in silicon micromachines, J. Microelectromech. Syst. **7**(2) (1998) 252–260

36.19 B. H. Kim, T. D. Chung, C. H. Oh, K. Chun: A new organic modifier for anti-stiction, J. Microelectromech. Syst. **10**(1) (2001) 33–40

36.20 S. Pamidighantam, W. Laureyn, A. Salah, A. Verbist, H. Tilmans: A novel process for fabricating slender and compliant suspended poly-Si micromechanical structures with sub-micron gap spacing, 15th IEEE 2002 Micro Electro Mechanical Systems (MEMS) Conf. (IEEE, New York 2002) 661–664

36.21 B. C. Bunker, R. W. Carpick, R. A. Assink, M. L. Thomas, M. G. Hankins, J. A. Voigt, D. Sipola, M. P. de Boer, G. L. Gulley: Impact of solution agglomeration on the deposition of self-assembled monolayers, Langmuir **16** (2000) 7742–7751

36.22 Y. Jun, V. Boiadjiev, R. Major, X.-Y. Zhu: Novel chemistry for surface engineering in MEMS, *Materials and Devices Characterization in Micromachining III*, ed. by Y. Vladimirsky, P. Coane, Proc. SPIE **4175** (2000) 113–120

36.23 R. Maboudian, W. R. Ashurst, C. Carraro: Self-assembled monolayers as anti-stiction coatings for MEMS: Characteristics and recent developments, Sens. Actuators A **82**(1) (2000) 219–223

36.24 W. R. Ashurst, C. Yau, C. Carraro, R. Maboudian, M. T. Dugger: Dichlorodimethylsilane as an anti-stiction monolayer for MEMS: A comparison to the octadecyltrichlorosilane self-assembled monolayer, J. Microelectromech. Syst. **10** (2001) 41–49

36.25 W. R. Ashurst, C. Yau, C. Carraro, C. Lee, G. J. Kluth, R. T. Howe, R. Maboudian: Alkene based monolayer films as anti-stiction coatings for polysilicon MEMS, Sens. Actuators A **91**(3) (2001) 239–248

36.26 S. J. Jacobs, S. A. Miller, J. J. Malone, W. C. McDonald, V. C. Lopes, L. K. Magel: Hermeticity and stiction in MEMS packaging, 40th Annual IEEE In-

ternational Reliability Physics Symp., Dallas 2002 (IEEE, New York 2002) 136–139
36.27 Nexus Task Force Report: *Market Analysis for Microsystems 1996–2002* (NEXUS, Grenoble 1998) www.nexus-mems.com
36.28 T. Maudie, J. Wertz: Pressure sensor performance and reliability, IEEE Industry Appl. Mag. **3**(3) (1997) 37–43
36.29 Lj. Ristic, R. Gutteridge, B. Dunn, D. Mietus, P. Bennett: Surface micromachined polysilicon accelerometer, IEEE 1992 Solid State Sensor and Actuator Workshop, IEEE 5th Technical Digest, Hilton Head 1992 (IEEE, New York 1992) 118–121
36.30 M. Furtsch, M. Offenberg, H. Munzel, J. R. Morante: Influence of anneals in oxygen ambient on stress of thick polysilicon layers, Sens. Actuators **76** (1999) 335–342
36.31 P. Lange, M. Kirsten, W. Riethmuller, B. Wenk, G. Zwicker, J. R. Morante, F. Ericson, J. A. Schweitz: Thick polycristalline silicon for surface micromechanical applications: Deposition, structuring and mechanical characterization, Proc. 8th International Conf. on Solid State Sensors and Actuators and Eurosensors IX, Transducers '95, Vol. 1 (IEEE, New York 1995) 202–205
36.32 F. Laermer, A. Schilp: Method of anisotropically etching silicon, (1996) U.S. Patent 5,501,893
36.33 M. Offenberg, H. Munzel, D. Schubert, O. Schatz, F. Laermer, E. Muller, B. Maihofer, J. Marek: Acceleration sensor in surface micromachining for airbag applications with high signal/noise ratio, SAE Special Pub. **1133** (1996) 35–41
36.34 F. Laermer, A. Schilp, K. Funk, M. Offenberg: Bosch deep silicon etching: Improving uniformity and etch rate for advanced MEMS applications, Proc 12th International Conf. on Micro Electro Mechanical Systems MEMS (IEEE, New York 1999) 211–216
36.35 J. R. Martin, C. M. Roberts Jr.: Package for sealing an integrated circuit die, (2001) U.S. Patent 6,323,550
36.36 J. A. Geen, S. J. Sherman, J. F. Chang, S. R. Lewis: Single chip surface micromachined integrated gyroscope with 50 deg/hour allan deviation, IEEE J. Solid-State Circuits **37**(12) (2002) 1860–1866
36.37 K. Funk, H. Emmerich, A. Schilp, M. Offenberg, R. Neul, F. Larmer: A surface micromachined silicon gyroscope using a thick polysilicon layer, Proc. 12th International Conf. on Micro Electro Mechanical Systems MEMS (IEEE, New York 1999) 57–60
36.38 M. Lutz, W. Golderer, J. Gerstenmeier, J. Marek, B. Maihofer, S. Mahler, H. Munzel, U. Bischof: A precision yaw rate sensor in silicon micromachining, 11th International Conf. on Solid State Sensors and Actuators, Transducers '97 (IEEE, New York 1997) 847–850
36.39 T. A. Core, R. T. Howe: Method for fabricating microstructures, (1994) U.S. Patent 5,314,572
36.40 M. Offenberg, F. Laermer, B. Elsner, H. Munzel, W. Reithmuller: Novel process for a monolithic integrated accelerometer, Proc. 8th International Conf. on Solid State Sensors and Actuators and Eurosensors IX, Transducers '95 (IEEE, New York 1995) 589–592
36.41 J. Anguita, F. Briones: HF/H_2O vapor etching of SiO_2 sacrificial layer for large-area surface-micromachined membranes, Sens. Actuators A **64** (1998) 247–251
36.42 M. A. Schmidt: Wafer-to-wafer bonding for microstructure formation, Proc. IEEE **86**(8) (1998) 1575–1585
36.43 A. R. Mirza: Wafer-level bonding technology for MEMS, Proc. 7th Intersociety Conf. on Thermal and Thermomechanical Phenomena in Electronic Systems, ITHERM 2000, Vol. 1 (IEEE, New York 2000) 113–119
36.44 A. R. Mirza: One micron precision, wafer-level aligned bonding for interconnect, MEMS and packaging applications, Proc. 50th Electronic Components & Technology Conf. (IEEE, New York 2000) 676–680
36.45 S. Mack, H. Baumann, U. Goesele: Gas tightness of cavities sealed by silicon wafer bonding, Proc. 10th Annual International Workshop on Micro Electro Mechanical Systems, MEMS; IEEE Micro Electro Mechanical Systems (MEMS) (IEEE, New York 1997) 488–493
36.46 G. Wallis: Field assisted glass sealing, SAE Automotive Engineering Congress, Detroit 1971 (Society of Automotive Engineers, New York 1971) Paper 71023
36.47 S. A. Audet, K. M. Edenfeld: Integrated sensor wafer-level packaging, Proc. International Conf. on Solid State Sensors and Actuators, Transducers '97, Vol. 1 (IEEE, New York 1997) 287–289
36.48 C. Gormley, A. Boyle, V. Srigengan, S. Blackstone: HARM processing techniques for MEMS and MOEMS devices using bonded SOI substrates and DRIE, *Micromachining and Microfabrication Process Technology VI*, ed. by J. Karam, J. Yasaitis, Proc. SPIE **4174** (2000) 98–110
36.49 K. Somasundram, D. Cole, C. McNamara, A. Boyle, P. McCann, C. Devine, A. Nevin: Fusion-bonded multilayered SOI for MEMS applications, Smart Sensors, Actuators and MEMS, Proc. SPIE **5116**, ed. by J.-C. Chiao, V. Varadan, C. Cané (2003) 12–19
36.50 C. M. Roberts Jr., L. H. Long, P. A. Ruggerio: Method for separating circuit dies from a wafer, (1994) U.S. Patent 5,362,681
36.51 T. Maudie, T. Miller, R. Nielsen, D. Wallace, T. Ruehs, D. Zehrbach: Challenges of MEMS device characterization in engineering development and final manufacturing, Proc. 1998 IEEE AUTOTESTCON (IEEE, New York 1998) 164–170

36.52 S. Petrovic, A. Ramirez, T. Maudie, D. Stanerson, J. Wertz, G. Bitko, J. Matkin, D. J. Monk: Reliability test methods for media-compatible pressure sensors, IEEE Trans. Industrial Electron. **45**(6) (1998) 877–885

36.53 S. Petrovic: Progress in media compatible pressure sensors, Proc. of InterPACK'01, the Pacific Rim/International Intersociety Electronic Packaging Technical/Business Conf. & Exhibition (ASME, New York 2001) IPACK2001-15517

36.54 H.-J. Kress, J. Marek, M. Mast, O. Schatz, J. Muchow: Integrated pressure sensors with electronic trimming, Automotive Eng. **103**(4) (1995) 65–68

36.55 C. Fung, R. Harris, T. Zhu: Multifunction polysilicon pressure sensors for process control, Sensors **16**(10) (1999) 75–79

36.56 K. H.-L. Chau, C. D. Fung, P. R. Harris, J. G. Panagou: High-stress and overrange behavior of sealed cavity polysilicon pressure sensors, IEEE 4th Technical Digest, Solid State Sensor and Actuator Workshop (IEEE, New York 1990) 181–183

36.57 A. B. Sontheimer: Digital micromirror device (DMD) hinge memory lifetime reliability modeling, Proc. 40th Annual IEEE International Reliability Physics Symp. (IEEE, New York 2002) 118–121

36.58 J. P. O'Connor: Packaging design considerations and guidelines for the digital micromirror device™, Proc. of InterPACK'01, the Pacific Rim/International Intersociety Electronic Packaging Technical/Business Conf. & Exhibition (ASME, New York 2001) IPACK2001-15526

36.59 C. Bang, V. Bright, M. A. Mignardi, D. J. Monk: Assembly and test for MEMS and optical MEMS. In: *MEMS and MOEMS Technology and Applications*, ed. by P. Rai-Choudhury (SPIE, Bellingham 2000) Chap. 7

36.60 F. Poradish, J. T. McKinley: Package for a semiconductor device, (1994) U.S. Patent 5,293,511

36.61 T. W. Migl: Interfacing to the digital micromirror device for home entertainment applications, Proc. of InterPACK'01, the Pacific Rim/International Intersociety Electronic Packaging Technical/Business Conf. & Exhibition (ASME, New York 2001) IPACK2001-15712

36.62 S. J. Jacobs, J. J. Malone, S. A. Miller, A. Gonzalez, R. Robbins, V. C. Lopes, D. Doane: Challenges in DMD™ assembly and test, Proc. Mater. Res. Soc. **657** (2001) EE6.1.1–EE6.1.12

36.63 Automotive Industry Action Group: *Quality Systems Requirements QS-9000*, 3rd edn. (Automotive Industry Action Group of the American Society for Quality, Milwaukee 1998)

36.64 J. A. Walraven, B. A. Waterson, I. De Wolf: Failure analysis of micromechanical systems (MEMS). In: *Microelectronic Failure Analysis*, 4th edn. (ASM, Materials Park 2002) 2002 Suppl.

36.65 M. R. Douglass: Lifetime estimates and unique failure mechanisms of the digital micromirror device (DMD), Proc. 1998 36th Annual IEEE International Reliability Physics Symp. (IEEE, New York 1998) 9–16

36.66 J. R. Martin: Surface characteristics of integrated MEMS in high volume production. In: *Nanotribology: Critical Assessment and Research Needs*, ed. by S. M. Hsu, Z. C. Ying (Kluwer, Dordrecht 2002) Chap. 14

36.67 H. R. Shea, A. Gasparyan, C. D. White, R. B. Comizzoli, D. Abusch-Magder, S. Arney: Anodic oxidation and reliability of MEMS poly-silicon electrodes at high relative humidity and high voltages, *MEMS Reliability for Critical Applications*; Proc. SPIE **4180** (2000) 117–122

36.68 J. Martin: Stiction suppression in high volume MEMS products, Proc. 2003 STLE/ASME Joint International Tribology Conf. (ASME, New York 2003) 2003TRIB-266

36.69 G. X. Li, F. A. Shemansky Jr.: Drop Test and analysis on micro-machined structures, Sens. Actuators **85** (2000) 280–286

36.70 R. T. Howe, H. J. Barber, M. Judy: Apparatus to minimize stiction in micromachined structures, (1996) U.S. patent 5,542,295

36.71 L. J. Hornbeck, W. E. Nelson: Spatial light modulator and method, (1992) U.S. Patent 5,096,279

36.72 L. J. Hornbeck: Low reset voltage process for DMD, (1994) U.S. Patent 5,331,454

36.73 E. C. Fisher, R. Jascott, R. O. Gale: Method of passivating a micromechanical device within a hermetic package, (1999) U.S. Patent 5,936,758

36.74 B. A. Miksic: Use of vapor phase inhibitors for corrosion protection of metal products, Proc. 1983 NACE Annual Conf., Corrosion 83 (National Assoc. Corrosion Engineers, Houston 1983) Paper 308

36.75 D. Vanderpool, S. Akin, P. Hassett: Corrosion inhibitors in the electronics industry: Organic copper corrosion inhibitors, Proc. 1986 NACE Annual Conf., Corrosion 86 (National Assoc. Corrosion Engineers, Houston 1986) Paper 1

36.76 J. R. Martin, Y. Zhao: Micromachined device packaged to reduce stiction, (1997) U.S. Patent 5,694,740

36.77 J. R. Martin: Process for wafer level treatment to reduce stiction and passivate micromachined surfaces and compounds used therefor, (2001) International Patent Application WO 01/57920 A1

36.78 S. T. Patton, J. S. Zabinski: Failure mechanisms of a MEMS actuator in very high vacuum, Tribol. Int. **35**(6) (2002) 373–379

36.79 S. T. Patton, K. C. Eapen, J. S. Zabinski: Effects of adsorbed water and sample aging in air on the μN level adhesion force between Si(100) and silicon nitride, Tribol. Int. **34**(7) (2001) 481–491

36.80 S. T. Patton, W. D. Cowan, K. C. Eapen, J. S. Zabinski: Effect of surface chemistry on the tribological performance of a MEMS electrostatic lateral output motor, Tribol. Lett. **9** (2000) 199–209

36.81 S. T. Patton, W. D. Cowan, J. S. Zabinski: Performance and reliability of a new MEMS electrostatic lateral output motor, Proc. 1999 37th Annual IEEE International Reliability Physics Symp. (IEEE, New York 1999) 179–188

36.82 D. M. Tanner, J. A. Walraven, L. W. Irwin, M. T. Dugger, N. F. Smith, W. P. Eaton, W. M. Miller, S. L. Miller: The effect of humidity on the reliability of a surface micromachined microengine, Proc. 1999 37th Annual IEEE International Reliability Physics Symp. (IEEE, New York 1999) 189–197

36.83 M. P. de Boer, P. J. Clews, B. K. Smith, T. A. Michalske: Adhesion of polysilicon microbeams in controlled humidity ambients, Proc. Mater. Res. Soc. Symp. **518** (1998) 131–136

37. MEMS Packaging and Thermal Issues in Reliability

The potential of MEMS/NEMS technologies has been viewed as a comparable or even bigger revolution than that of microelectronics. These scientific and engineering advancements in MEMS/NEMS could bring applications to reality previously unthinkable, from space systems, environmental instruments, to daily life appliances. As presented in previous chapters, the development of core MEMS/NEMS processes has already demonstrated a lot of commercial applications as well as future potentials with elaborated functionalities. However, a low cost and reliable package for the protection of these MEMS/NEMS products is still a very difficult task. Without addressing the packaging and reliability issues, no commercial products can be sold on the market. Packaging design and modeling, packaging material selection, packaging process integration, and packaging cost are main issues to be considered when developing a new MEMS packaging process. In this chapter, we will present the fundamentals of MEMS/NEMS packaging technology, including packaging processes, hermetic and vacuum encapsulations, thermal issues, packaging reliability, and future packaging trends. The future development of MEMS packaging will rely on the success of the implementation of several unique echniques, such as packaging design kits for system and circuit designer, low cost and high yield wafer level, chip-scale packaging techniques, effective testing techniques at the wafer-level to reduce overall testing costs; and reliable fabrication of an interposer [37.1] with vertical through-interconnects for device integrations.

37.1 **MEMS Packaging** 1111
37.1.1 MEMS Packaging Fundamentals 1112
37.1.2 Contemporary MEMS Packaging Approaches 1113

37.2 **Hermetic and Vacuum Packaging and Applications** 1116
37.2.1 Integrated Micromachining Processes 1117
37.2.2 Post-Packaging Processes 1118
37.2.3 Localized Heating and Bonding 1119

37.3 **Thermal Issues and Packaging Reliability** 1122
37.3.1 Thermal Issues in Packaging 1122
37.3.2 Packaging Reliability 1124
37.3.3 Long-Term and Accelerated MEMS Packaging Tests 1125

37.4 **Future Trends and Summary** 1128

References 1129

37.1 MEMS Packaging

MEMS are miniaturized systems in a size of micrometer to millimeter that may have mechanical, chemical, or biomedical features integrated with IC circuitry for sensor or actuator applications [37.2]. For example, pressure [37.3], temperature, flow [37.4], accelerometers [37.5], gyroscopes [37.6], and chemical sensors [37.7] can be fabricated by MEMS technologies for sensing applications. Fluidic valves [37.8], pumps [37.9], and inkjet printer heads are examples of actuation devices for medical, environmental, office, and industrial applications. Silicon is typically used as the primary substrate material for MEMS fabrication because it can provide unique electrical, thermal, and mechanical properties however also can be easily micromachined in a form of batch-processing and be incorporated with a microelectronic circuits by using most of the conventional semiconductor manufacturing processes and tools. As a result, smaller size, lighter weight, lesser power consumption and cheaper fabrication cost become the advantages of MEMS devices as compared with the existing macroscale systems with similar functionalities. With the advances of MEMS fabrication technology in the past decades, the MEMS market at the component-level is currently in excess of

$ 5 billion and is driving the end-product markets of large than $ 100 billion [37.10].

Nevertheless, the road to the commercialization of MEMS doesn't look as promising as expected. Many industrial companies took the advantages of MEMS technology of high production volumes and high added value created by product integrations. Therefore, cost-efficiency becomes the major factor driving MEMS toward commercialization. Several MEMS devices have been developed for and applied in automotive industry and information technology, and they dominate the MEMS market due to high production volumes. Most of custom-designed MEMS products are still very diverse, aiming for different applications, and their initial costs in small-to-medium-scale production are still much higher than market-acceptable levels. Consequently, high packaging and testing costs have slowed the commercialization of MEMS. Furthermore, based on the past experience of IC industry, the cost of packaging processes is about 30%, and sometimes can be more than 70%, of the total production expenses. MEMS packaging process is expected to be even more costly because of the challenging and stringent packaging issues with regard to the additional MEMS components, in addition to the microelectronic circuitry, in a typical MEMS product [37.11].

37.1.1 MEMS Packaging Fundamentals

The functions of conventional IC packaging are to protect, power, and cool the microelectronic chips or components and provide electrical and mechanical connection between the microelectronic part and the outside world. With the increase of the needs of high performance and multifunctional consumer electronic products, IC packaging processes have incorporated more complex designs and advanced fabrication technologies, such as Cu interconnects [37.12], flip-chip bonding [37.13], ball grid array [37.14], wafer-level chip scale packaging [37.15], and 3-D packaging [37.16], etc., in order to satisfy the need for high I/O density, large die area, and high clock frequency. In addition to the requirements of electrical interconnects, MEMS components may need to interface outside environment (for example, fluidic interconnectors [37.17]), some other components may need to be hermetically sealed in vacuum (for example, accelerometers [37.5]). Therefore, MEMS packaging processes have to provide more functionalities including better mechanical protection, thermal management, hermetic sealing, and complex electricity and signal distribution.

It has been suggested that MEMS packaging should be incorporated in the device fabrication stage as part of the micromachining process. Although this approach solves the need for some specific devices, it does not solve the packaging need for general microsystems. Many MEMS devices are now fabricated by various foundry services [37.18, 19], and there is a tremendous need for a uniform packaging process. Figure 37.1 shows a typical MEMS device being encapsulated by a packaging cap. The most fragile part of this device is the suspended mechanical sensor, which is a freestanding mechanical, mass-spring microstructure. This mechanical part must be protected during the packaging and handling process. Moreover, vacuum encapsulation may be required for these microstructures in applications such as resonant accelerometers [37.5] or gyroscopes [37.6, 20]. A "packaging cap" with properly designed micro cavity is to be fabricated to encapsulate and protect the fragile MEMS structure as the first-level MEMS post-packaging process. The wafer can be diced afterward, and the well-established packaging technology in IC industry can follow and finish the final packaging step. Unlike the packaging requirements for ICs, however, the common MEMS packaging requirement is hermetic seal and sometimes vacuum encapsulations. Hermetic seal is important to ensure that no moisture or contaminant can enter the package

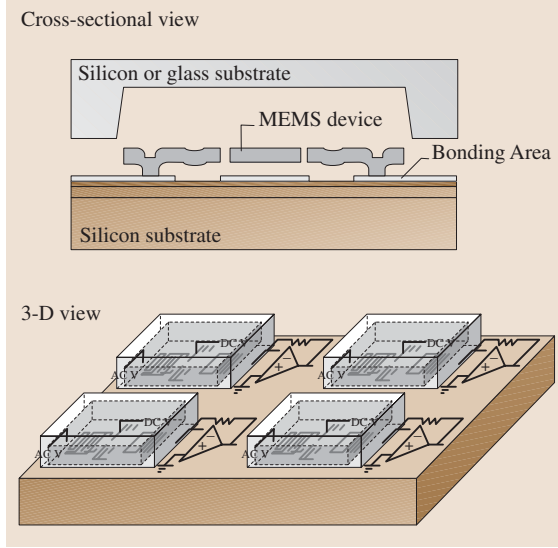


Fig. 37.1 A typical MEMS packaging illustration in which a MEMS structure is encapsulated and protected by the packaging cap

and affect the functionality of microstructures. This increases the difficulty of common IC packaging processes tremendously. Although most single function MEMS chips can employ typical IC packaging techniques, such as die-attached processes, wiring interconnects, molded plastic, ceramic, and metal for packaging [37.21], with the increase of complexity MEMS devices, more advanced packaging techniques, especially wafer-level packaging, are required for the integration of multi-chips for multifunctional applications.

Before the state-of-the-art MEMS packaging processes are discussed, several primary micro-fabrication processes for packaging applications are described. These processes may include Flip-Chip (FC) technique, Ball Grid Arrays (BGA), through-wafer etching, and plating. Other silicon-based processes, such as thin-film deposition, wet and dry chemical etching, lithography, lift-off, and wiring bonding processes can be found in many textbooks [37.22].

Flip-Chip Technique (FC)

This technique is commonly used in the assembly process between a chip with microelectronics and a package substrate [37.13]. The microelectronic chip is "flipped joined" with the packaging substrate, and metal solder bumps are used as both the bonding agents and electrical paths between bond pads on the microelectronic chip and metal pads on the package substrate. Because the vertical bonding space can be very small as controlled by the heights of the solder bumps, and the lateral distributions of bond pads can be on the whole chip instead of being only on the edge, this technique can provide high density Input/Output (I/O) connections. In the FC technique, solder bumps are generally fabricated by means of electroplating. Before the bumping process, multiple metal layers, such as TiW-Cu, Cr-Cu, Cr-Ni, TaN/Ta/Ni, have to be deposited as a seed layer for electroplating and as a barrier to prevent the diffusion of solder into its underneath electrical interconnect.

Ball Grid Arrays (BGA)

This technology is very similar to the FC technique. An area array of solder balls on a single chip module or multi-chip module are used in the packaging process as electrical, thermal, and mechanical connects to join the module with the next level package, usually a printed circuit board [37.14]. The major difference between a typical BGA and FC is the size of solder bump. In BGA chip, the bumps are in the order of 750 μm in diameter which is ten times larger than those commonly used in FC.

Through-Wafer Etching

This is a chemical etching process to make through-wafer channels on a silicon substrate for the fabrication of vertical through-wafer interconnects. The chemical etching process can be either wet or dry. Anisotropic or isotropic etching solutions can be used in the wet etching process. The dry etch process is based on plasma and ion-assisted chemical reactions that can be either isotropic or anisotropic. In order to create high density and high aspect ratio through-wafer vias, Deep Reactive Ion Etching (DRIE) is typically used. Two popular DRIE approaches, Bosch and Cyro, are well described in literature [37.23].

Electroplating

Electroplating is another common micro-fabrication process. It can be conducted for the deposition of an adherent metallic layer onto a conductive or nonconductive substrate. The process on a conductive substrate is called electrolytic plating, which utilizes a seed layer as the anode to transfer metal ions onto the cathode surface when a DC current is passed through the plating solution. The plating process without applying electrical current is called electroless plating, which can happen in both conducive and nonconductive surfaces. Electroless plating process requires a layer of noble metal such as Pd, Pt, or Ru on the substrate as the catalysis to trigger the self-decomposition reaction in the plating solution. The electroplating processes are very important for electrical interconnect and solder bump fabrications for packaging applications because of low process temperature and cost.

These processes are generally developed for providing electrical and thermal paths for various IC/MEMS packaging approaches.

37.1.2 Contemporary MEMS Packaging Approaches

Several MEMS packaging issues and approaches before 1985 were discussed in [37.24], and researchers have been working on MEMS packaging approaches continuously. For example, *Senturia* and *Smith* [37.25] discussed the packaging and partitioning issues for microsystems. *Smith* and *Collins* [37.26] used epoxy to bond glass and silicon for chemical sensors. Several MCM methods have been proposed. *Butler* et al. [37.27] proposed adapting multi-chip module foundries using chip-on-flex (COF) process. *Schuenemann* et al. [37.28] introduced a 3-D stackable packaging concept for the Top-Bottom Ball Grid Array (TB-BGA) that includes

electric, fluidic, optic, and communication interfaces. *Lee* et al. [37.29] and *Ok* et al. [37.10] presented a direct-chip-attach MEMS packaging using through-wafer electrical interconnects. *Laskar* and *Blythe* [37.30] developed a multi-chip modules (MCM) type packaging process by using epoxy. *Reichl* [37.31] discussed different materials for bonding and interconnection. *Grisel* et al. [37.32] designed a special process to package micro-chemical sensors. Special processes have also been developed for MEMS packaging, such as packaging for microelectrodes [37.33], packaging for biomedical systems [37.34], and packaging for space systems [37.35]. These specially designed, device-oriented packaging methods are meant for individual systems. There is no reliable method yet that would qualify as a versatile post-packaging process for MEMS with the rigorous process requirements of low temperature, hermetic sealing and long-term stability.

An integrated process by using surface-micromachined microshells has previously been developed [37.36]. This process applies the concepts of sacrificial layer and LPCVD sealing to achieve wafer-level post-packaging. Similar processes have been demonstrated. For example, *Guckel* et al. [37.37] and *Sniegowski* et al. [37.38] developed a reactive sealing method to seal vibratory micromachined beams. *Ikeda* et al. [37.39] adopted epitaxial silicon to seal microstructures. *Mastrangelo* et al. [37.40] used silicon nitride to seal mechanical beams as light sources. *Smith* et al. [37.41] accomplished a new fabrication technology by embedding microstructures and CMOS circuitry. All of these methods have integrated the MEMS process together with the post-packaging such that no extra bonding is required. These schemes are highly process dependent, however, and not suitable for prefabricated circuitry.

Recently several new efforts for MEMS post-packaging processes have been reported. *Butler* et al. [37.27] demonstrated an advanced MCM packaging scheme. It adopts the high density interconnect (HDI) process consisting of embedding bare die into pre-milled substrates. Because the MEMS structures have to be released after the packaging process, it is undesirable for general microsystems. *Van der Groen* et al. [37.42] reported a transfer technique for CMOS circuits based on epoxy bonding. This process overcomes the surface roughness problem, but epoxy is not a good material for hermetic sealing. In 1996, *Cohn* et al. [37.43] demonstrated a wafer-to-wafer to-wafer vacuum packaging process by using Silicon–Gold eutectic bonding with a 2-μm-thick polysilicon microcap. But experimental results show substantial leakage after a period of 50 days. *Cheng* et al. [37.44] developed a vacuum packaging technology using localized aluminum/silicon-to-glass bonding. In 2002, *Chiao* and *Lin* [37.45] demonstrated vacuum packaging of microresonators by rapid thermal processing. These recent and ongoing research efforts indicate the strong need for a versatile MEMS post-packaging process.

Bonding Processes for MEMS Packaging Applications

Previously, silicon-bonding technologies have been used in many MEMS fabrication and packaging applications, including epoxy bonding, eutectic bonding, anodic bonding, fusion bonding, and solder bonding. For example, such devices as pressure sensors, micro-pumps, biomedical sensors or chemical sensors require mechanical interconnectors to be bonded on the substrate (see for example [37.7, 17, 46]). Glass has been commonly used as the bonding material by anodic bonding at a temperature of about 300 ∼ 450 °C (see for example [37.47, 48]). *Klaassen* et al. [37.49] and *Hsu* et al. [37.50] have demonstrated different types of silicon fusion bonding and Si–SiO_2 bonding processes at very high temperatures of over 1,000 °C. *Ko* et al. [37.24], *Tiensuu* et al. [37.51], *Lee* et al. [37.52] and *Cohn* et al. [37.43] have used eutectic bonding for different applications. All of these bonding techniques have different mechanisms that determine the individual bonding characteristics and process parameters. This section discusses the details of these processes.

Fusion Bonding for MEMS Packaging

Silicon fusion bonding is an important fabrication technique of the SOI (Silicon On Insulator). The bonding is based on Si−O, Si−N, or Si−Si strong covalent bond. But very high bonding temperature (higher than 1,000 °C) and flat bonding surfaces (less than 6 nm) are the two basic requirements for strong, uniform, and hermetic bonding. Although hydrophilic surface treatment can lower the bonding temperature, an annealing step higher than 800 °C is still needed to remove possible bubble formation at the bonding interface. *Bower* et al. [37.53] proposed that low temperature Si_3N_4 fusion bonding could be achieved at the temperature under 300 °C. *Takagi* et al. [37.54] proposed that silicon fusion bonding could be made at room temperature by using Ar^+ beam treatment on the wafer surface, and the bond strength is comparable to the conventional fusion bonding. In summary, fusion bonding has been

a popular fabrication technique in MEMS fabrication and packaging.

Anodic Bonding for MEMS Packaging Applications

The invention of anodic bonding dates back to 1969 when *Wallis* and *Pomerantz* [37.55] found that glass and metal can be bonded together at about 200 ∼ 400 °C below the melting point of glass with the aid of a high electrical field. This technology has been widely used for protecting on-board electronics in biosensors (see for example [37.56–58]) and sealing cavities in pressure sensors (see for example [37.59]). Many reports have also discussed the possibility of lowering the bonding temperature by different mechanisms [37.60, 61]. Anodic bonding forms Si−O or Si−Si covalent bond and is one of the strongest chemical bonds available for silicon-based systems. The bonding process can be accomplished on a hot plate with a temperature between 180 °C and 500 °C in atmosphere or vacuum environment. When a static electrical field is built within the Pyrex glass (7740 from Dow Corning) and silicon, the electrostatic force can pull two surfaces close for a strong bond. In order to create high electrical field, a flat bonding surface with less than 50 nm roughness is required. In addition, the electrical field required for bonding is larger than 3×10^6 V/cm [37.24]. Such a high electrical field is generated by a power supply of 200 to 1,000 V, which may cause damages on integrated circuits. *Hanneborg* et al. [37.62] have successfully bonded silicon with other thin solid films, such as silicon dioxide, nitride, and polysilicon, together with an intermediate glass layer using anodic bonding technique. In practice, electrostatic bonding has become widely accepted in MEMS fabrications and packaging applications. Unfortunately, the possible contamination due to excessive alkali metal in the glass; possible damage to microelectronics due to the high electrical field; and the requirement of flat surface for bonding limit the application of anodic bonding to MEMS post-packaging applications [37.63].

Epoxy Bonding (Adhesive Bonding)

Four major components comprise epoxy : epoxy resin, filler – like silver slake, solvent or reactive epoxy diluent, and additives like hardener and catalyst [37.64, 65]. The bonding mechanism of epoxy is very complicated and depends on the type of epoxy. In general, the main source of bonding strength is the van der Waals force. Because epoxy is a soft polymer material and its curing temperature for bonding is only around 150 °C, low residual stress and process temperature are the major advantages of epoxy bonding. But the properties of epoxy can be easily changed with environmental humidity and temperature so the bonding strength decays over time. In addition, epoxy bonding has low moisture resistance and, due to its additives, is a dirty process. These disadvantages have made epoxy unfavorable for the special MEMS packaging need for hermetic or vacuum sealing.

Eutectic Bonding

In many binary systems, a eutectic point corresponds to the alloy composition with the lowest melting temperature. If the environmental temperature is kept higher than the eutectic point, two contacted surfaces containing two elements with the eutectic composition can form liquid phase alloy. The solidification of the eutectic alloy forms "eutectic bonding" at a temperature lower than the melting temperature of either element in the alloy. Eutectic bonding can be a strong bond. For example, in the case of Au/Si alloy system, eutectic temperature is only 363 °C when the composition is at the atomic ratio of 81.4% Au to 18.6% Si and bonding strength is higher than 5.5 GPa [37.66]. Because other alloy systems may have lower eutectic temperatures than Al/Si system, they present great potential for MEMS packaging applications. In addition to Au/Si, Al/Ge/Si, Au/SnSi, and Au/Ge/Si systems have been applied for MEMS packaging.

Solder Bonding

Solder bonding has been widely applied in microelectronic packaging [37.67]. Both low bonding temperature and high bonding strength are good characteristics for packaging. Furthermore, there are a variety of choices of solder material for specific applications. *Singh* et al. [37.68] have successfully applied solder bump bonding in the integration of electronic components and mechanical devices for MEMS fabrication [37.69]. In this case, indium metal was used for bonding two separated silicon surfaces together by applying 350 MPa pressure and the bonding strength was as strong as 10 MPa. Glass frits can also be treated as a solder material and have been extensively used for vacuum encapsulation in MEMS industry. Glass frits are ceramic materials that can provide strong bonding strength with silicon with good hermeticity. Its bonding temperature is lower than 400 °C and is suitable for electronic components. But a bonding area more than 200 μm wide is required to achieve good results, which may become a drawback because area is the measure of manufacturing cost in IC industry. Nevertheless, glass frit is the

Table 37.1 Summary of bonding mechanisms

Bonding methods	Temperature	Roughness	Hermeticity	Post-packaging	Reliability
Fusion bonding	very high	highly sensitive	yes	yes by LH	good
Anodic bonding	medium	highly sensitive	yes	difficult	good
Epoxy bonding	low	low	no	yes	???
Integrated process	high	medium	yes	no	good
Low Temp. bonding	low	highly sensitive	???	no	???
Eutectic bonding	medium	low	yes	yes by LH	???
Brazing	very high	low	yes	yes by LH	good
???: no conlusive data		LH: Localized Heating			

most popular bonding process used in the current MEMS products.

Localized Heating and Bonding

Low bonding temperature and short process time are desirable process parameters in MEMS packaging fabrication to provide less thermal budget and high throughput. Most chemical bonding reactions, however, require a minimum and sufficient thermal energy to overcome the reaction energy barrier, or activation energy, to start the reaction and to form a strong bond. As a result, high bonding temperature generally results in shorter processing time to reach the same bonding quality at a lower bonding temperature [37.70]. The common limitations for the above techniques are their individual bonding characteristics and temperature requirements. In general, MEMS packaging requires a good bonding for hermetic sealing while the processing temperature must be kept low at the wafer-level to have less thermal effects on the existing devices. For example, a MEMS device may have prefabricated circuitry, biomaterial, or other temperature-sensitive materials such as organic polymer, magnetic metal alloy, or pizeo-ceramic. Since the packaging step comes after the MEMS device fabrication processes, bonding temperature should be kept low to avoid high temperature effects on the system. Possible temperature effects include residual stress due to the mismatch of thermal expansion coefficient of bonding materials and substrates, electrical contact failure due to atom inter-diffusion at the interface, and contamination due to the outgas or evaporation of materials. In addition to the control of bonding temperature, Other factors to consider are the magnitude of applied force to create intimate contact for bonding and atmospherics environment control. Based on the heat transfer simulation study [37.71], it is possible to confine high temperature area in a small region by localized heating without heating the whole substrate. Assembly steps can therefore always be processed after device fabrication without having detrimental effects. As such, localized heating and bonding technique is introduced and implemented for the fabrication of MEMS packaging for post-processing approaches [37.70, 72].

Table 37.1 summarizes these MEMS packaging technologies and their limitations, including the localized heating and bonding approach. The localized heating approach introduces several new opportunities. First, better and faster temperature control can be achieved. Second, higher temperature can be applied to improve the bonding quality. Third, new bonding mechanisms that require high temperature such as brazing [37.73] may now be explored in MEMS applications. As such, it has potential applications for a widerange of MEMS devices and is expected to advance the field of MEMS packaging.

37.2 Hermetic and Vacuum Packaging and Applications

Hermetic packaging is desirable for MEMS packaging requirements in most cases because it provides a moisture free environment to avoid charge separation in capacitive devices, corrosion in metallization, or electrolytic conduction in order to prolong the lifetime of the electronic circuitry. In several device applications, vacuum encapsulation is necessary but can be costly. Many surface micromachined resonant devices need

vacuum to improve the performance including comb-shape μ-resonators and ring-type μ-gyroscopes that have very large surface-to-volume ratios and vibrate in a very tight space [37.20,39]. Two major approaches of MEMS hermetic and vacuum packaging have been demonstrated: (1) the integrated encapsulation approach and (2) the post-process packaging approach. Both are discussed in this section. Vacuum encapsulation by means of localized heating and bonding is discussed separately as another example for issues related to hermetic and vacuum packaging.

37.2.1 Integrated Micromachining Processes

Several MEMS hermetic and vacuum packaging processes have been demonstrated before based on the integrated micromachining processes. An integrated vacuum sealing process by LPCVD is presented here as the illustration example. This integrated process can encapsulate comb-shape microresonators [37.74] in vacuum at the wafer level. Figure 37.2 illustrates the cross-sectional view of the manufacturing process. First, standard surface-micromachining process [37.75] is conducted by using four masks to define the first polysilicon layer, anchors to the substrate, dimples and, the second polysilicon layer as shown in Fig. 37.2a. The process so far is similar to the MCNC MUMPs process [37.19], and comb-shape microstructures are fabricated at the end of these steps. In the standard surface micromachining process, the sacrificial layer (oxide) is etched away to release the microstructures. In the MEMS post-packaging process, a thick PSG (phosphorus-doped glass) of 7 μm is deposited to cover the microstructure and patterned by using 5:1 BHF (buffered HF) to define the microshell area as shown in Fig. 37.2b. A thin PSG layer of 1 μm is then deposited and defined to form etch channels as illustrated in Fig. 37.2c. The microshell material, low-stress silicon nitride, is now deposited with a thickness of 1 μm. Etch holes are defined and opened on the silicon nitride layer by using a plasma etcher. Silicon dioxide inside the packaging shell is now etched away by concentrated HF, and the wafer is dried using the supercritical CO_2 drying process [37.76]. After these steps, Fig. 37.2d applies. A 2-μm-thick LPCVD low-stress nitride is now deposited at a deposition pressure of 300 mtorr to seal the shell in the vacuum condition. Finally, the contact pads are opened as shown in Fig. 37.2e. Figure 37.3 is the SEM (Scanning Electron Microscope) microphoto of a finished device with protected microshell on top. The total packaging area

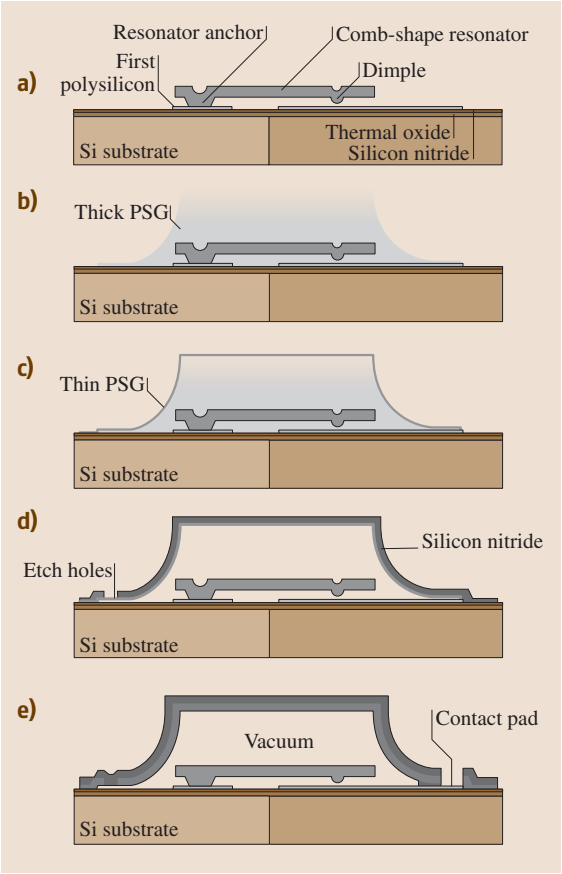


Fig. 37.2 An integrated vacuum encapsulation process using LPCVD nitride sealing to package micro-mechanical resonators [37.71]

(microshell) is about $400 \times 400\,\mu m^2$. A contact pad is shown with the covering nitride layer removed. The shape of the microresonator, with beams of 150-μm-long and 2-μm-wide is reflected on the surface of the microshell due to the integrated packaging process. The total height of the nitride shell is 12 μm, as seen standing above the substrate. Spectrum measurement of the comb resonator inside the packaging reveals that a vacuum level of about 200 mtorr has been accomplished [37.77].

Although the above vacuum sealing process successfully achieves MEMS hermetic and vacuum packaging, it has several drawbacks. First, several high temperature steps were used after the standard surface-micromachining process. As such, no circuitry (such as those shown in Fig. 37.1) or temperature-sensitive materials

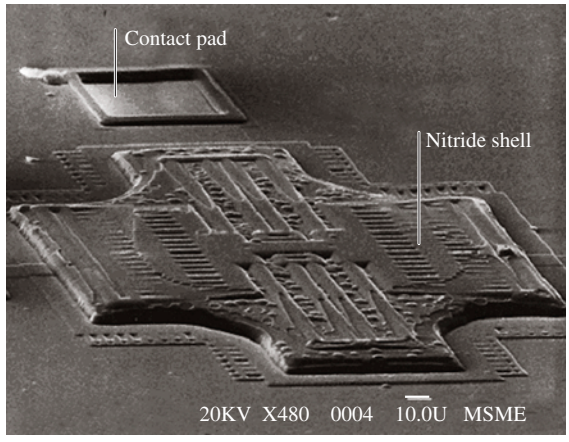


Fig. 37.3 SEM microphoto showing a vacuum packaged MEMS mechanical comb-shape resonator by an the integrated LPCVD sealing process as depicted in Fig. 37.2 [37.71]

will survive due to the global heating effect. Second, this post-packaging process is very specific and process dependent. MEMS companies or researchers have to adopt this post-packaging process with their own device manufacturing process. It can't be conducted in the multiuser MEMS process such as the MCNC MUMPs. Third, the thickness of the microshell is limited by the thin film deposition step that generally is in the range of a few micrometers. Wether the thin microshell can survive the high-pressure plastic molding process afterward in some device packaging design is a big concern. Finally, although integrated encapsulation can achieve low pressure by wafer-level fabrication and provide lower manufacturing cost, it does not provide the controllability of the cavity pressure.

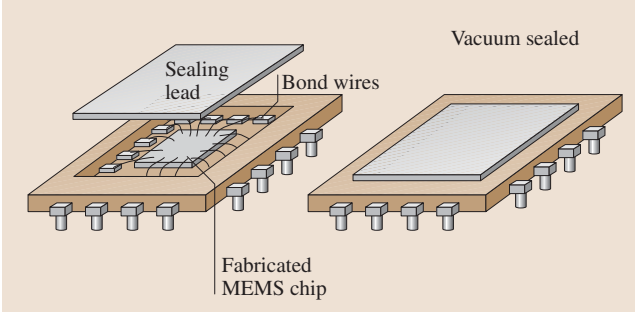


Fig. 37.4 The schematic diagram of industrial post-packaging (DIPS) using a ceramic holder to be covered by a sealing lid

37.2.2 Post-Packaging Processes

The second approach is defined as post-packaging process. The packaging process starts when the device fabrication processes are completed, so this approach has high flexibility for various microsystems. For example, Fig. 37.4 shows a common industrial post-hermetic packaging called Dual-in-line packaging (DIP) [37.78, 79]. A die is placed inside a ceramic holder covered by a sealing lid. Solder or ceramic joining is generally used for assembling lid and holder under a pressure-controlled environment. High cost is the major drawback of this method because of expensive ceramic holder and low fabrication throughput. Another example of post-packaging method is based on wafer bonding techniques combined with a microshell encapsulation. Devices are sealed by stacking another micromachined silicon or glass substrate as illustrated in Fig. 37.1. Integrated microsystems and protection shells are fabricated on different wafers, either silicon or glass, at the same time. After the two substrates are assembled together using silicon fusion, anodic, or low temperature solder bonding to achieve the final encapsulation, these micro-shells will provide mechanical support, thermal path, or electrical contact for the MEMS devices. Low packaging cost can be expected due to wafer-level processing.

The related packaging issues are discussed by using a specific example of hermetic and vacuum post-packaging by rapid thermal processing (RTP) that provides the advantages of wafer-level processing and low thermal budget. Figure 37.5 shows the

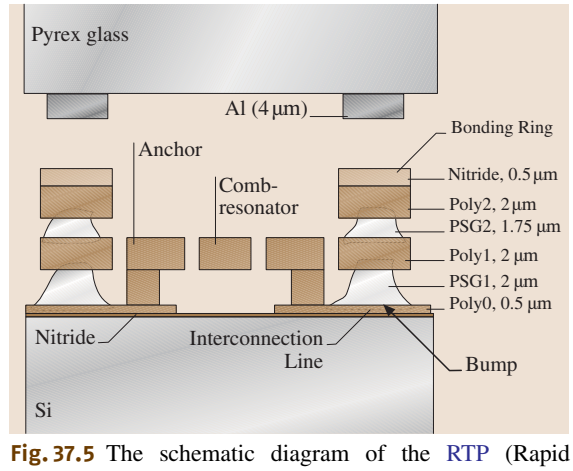


Fig. 37.5 The schematic diagram of the RTP (Rapid Thermal Processing) aluminum-to-nitride bonding set-up [37.45]

schematic illustration of the RTP MEMS vacuum packaging process. Comb-shape microresonators are chosen as the vacuum packaging examples in this chapter, and a standard surface-micromachining process is used to fabricate these microresonators [37.19]. One major addition to a regular microstructure is that an integrated sealing ring using silicon nitride as the topmost layer is incorporated in the manufacturing process for the purpose of aluminum-to-nitride bonding [37.45]. A glass cap wafer is deposited and patterned with 4-μm-thick aluminum sealing-rings with width ranging from 100 to 250 μm and bonding area ranging from 450×450 to 1,000×1,000 μm^2. Before the vacuum packaging process, both the device and cap wafers are baked in vacuum at 300 °C for over four hours to dry water and gas species adhere at the surface. Afterward, the device and cap wafers are flip-chip assembled immediately and loaded into a quartz chamber and put into a RTP chamber. The base vacuum estimated at 10 mtorr inside the quartz chamber is achieved by using a mechanical pump. After heating for 10 seconds at 750 °C, the aluminum-to-nitride bond is formed. Figure 37.6 shows the measured spectrum of a vacuum-packaged, double-folded beam comb-drive resonator by using a micro-stroboscope [37.80]. The central resonant frequency is about 18,625 Hz and the quality factor is extracted as 1,800 ± 200 corresponding to a pressure level about 200 mtorr inside the package [37.81].

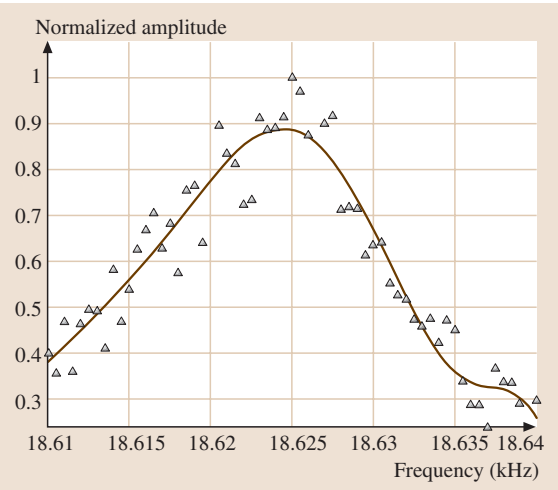


Fig. 37.6 Spectrum measurement results of a vacuum encapsulated comb-shape resonator by using the RTP aluminum-to-nitride bonding method [37.45]

This type of post-packaging process at wafer level has become the favorite approach to fabricate a hermetic encapsulation because it can provide lower cost and more process flexibility. But the packaging process relies on "good" bonding techniques. A strong and reliable bonding between two substrates should be provided, and this procedure should be compatible with the other microsystem fabrication processes.

37.2.3 Localized Heating and Bonding

The approach of MEMS post-packaging by localized heating and bonding is proposed to address the problems of global heating effects. In this chapter, resistive micro-heaters are used as the examples to provide localized heating, although several other means of localized heating have been demonstrated recently, including laser welding [37.82], inductive heating [37.83], and ultrasonic bonding [37.84]. The principle of localized heating is to achieve high temperature for bonding while maintaining low temperature globally at the wafer level. Resistive heating by using micro-heaters on top of device substrate is applied to form strong bond with silicon or glass cap. According to the results of a 2-D heat conduction finite element analysis as shown in Fig. 37.7, the steady-state heating region of a 5-μm-wide polysilicon micro-heater capped with a Pyrex glass substrate

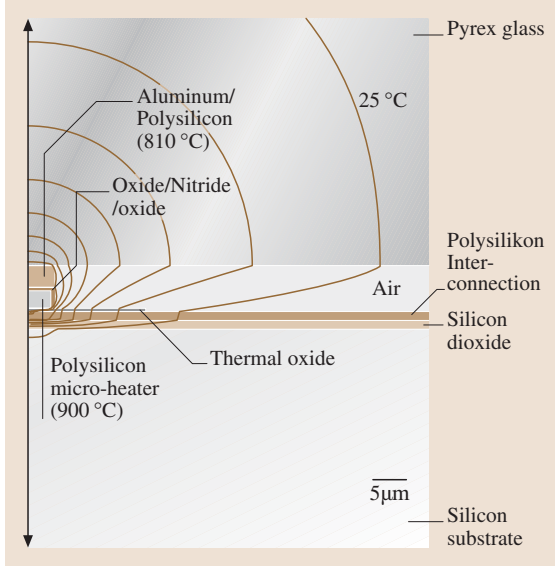


Fig. 37.7 The schematic diagram of 2-D heat transfer model, geometry and boundary conditions

can be confined locally as long as the bottom of the silicon substrate is constrained to the ambient temperature. The physics of localized heating behind this design can be understood by solving the governing heat conduction equations of the device structure without a cap [37.81]. As long as the width of micro-heater and the thickness of silicon substrate are much smaller than the die size and a good heat sink is placed underneath the silicon substrate, the heating can be confined locally. The temperature of silicon substrate can be kept low or close to room temperature. Several localized resistive heating and bonding techniques have been successfully developed for packaging applications including localized silicon-to-glass fusion bonding, gold-to-silicon eutectic bonding, and localized solder bonding. Several solder materials have been successfully tested, including PSG, indium, and aluminum alloy [37.85].

The vacuum packaging example presented in this chapter is based on localized aluminum/silicon-to-glass solder bonding technique. Built-in folded-beam comb drive μ-resonators are used to monitor the pressure of the package. Figure 37.8 shows the fabrication process of the package and resonators. Thermal oxide (2 μm) and LPCVD Si_3N_4 (3,000 Å) are first deposited on a silicon substrate for electrical insulation followed by the deposition of 3,000 Å LPCVD polysilicon. This polysilicon is used as both the ground plane and the electrical interconnect to the μ-resonators as shown in Fig. 37.8a. Figure 37.8b shows a 2 μm LPCVD SiO_2 layer that is deposited and patterned as a sacrificial layer for the fabrication of polysilicon μ-resonators using a standard surface micromachining process. A 2-μm-thick phosphorus doped polysilicon is used for both the structural layer of microresonators and the on-chip micro-heaters. This

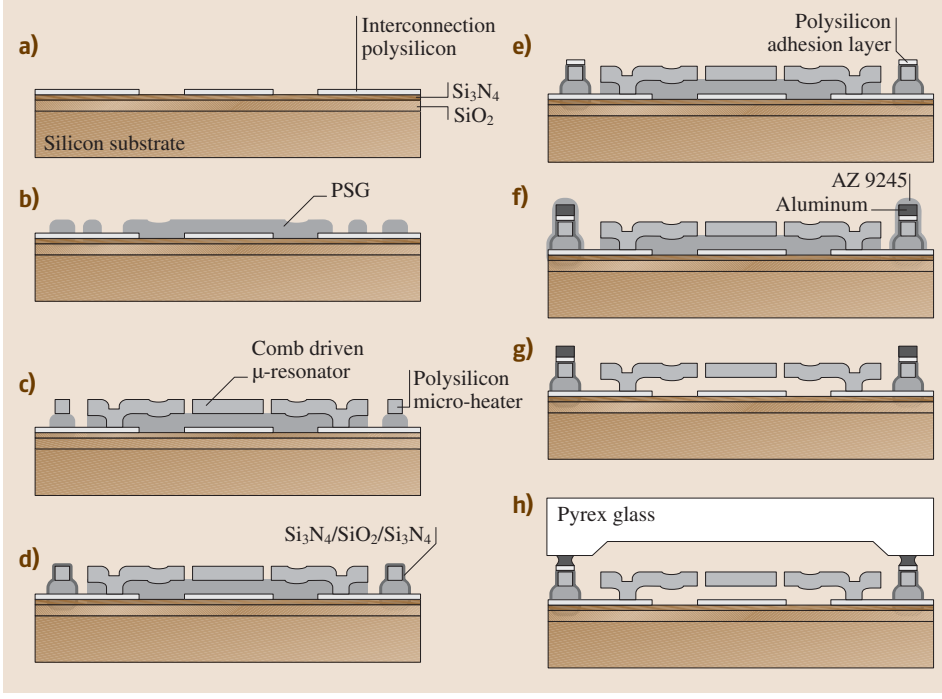


Fig. 37.8a–h The schematic process flow of vacuum encapsulation using localized aluminum/silicon-to-glass bonding, **(a)** electrical ground plane and device interconnects fabrication **(b)** a 2 μm LPCVD SiO_2 sacrificial layer deposition for the fabrication of polysilicon μ-resonator **(c)** a 2 μm-thick phosphorus doped polysilicon deposition and patterning for both the structural layer of micro resonators and the on-chip microheaters **(d)** a LPCVD Si_3N_4 (750 Å)/SiO_2 (1,000 Å)/Si_3N_4 (750 Å) sandwich layer deposition and patterning on top of the microheater for the electrical isolation **(e)** an aluminum (2.5 μm) and polysilicon (5,000 Å) bonding materials are deposition and patterning. **(f)** a thick AZ 9245 photoresist deposition and patterning for the protection of aluminum/silicon-to-glass bonding layers **(g)** 20 minutes sacrificial release in concentrated HF **(h)** vacuum encapsulation with 0.2 MPa contact pressure and 3.4 W input power under a 25 mtorr vacuum environment

layer is formed over the sacrificial oxide in two steps to achieve a uniform doping profile. Lower input power and better process compatibility are two major advantages in using the on-chip micro-heater in the glass package. The resonators are separated from the heater by a short distance, 30 μm, to effectively prevent their exposure to the high heater temperature as shown in Fig. 37.8c. This concludes the fabrication of μ-resonators.

In order to prevent the current supplied to the micro-heater from leaking into the aluminum solder during bonding, a LPCVD Si_3N_4 (750 Å)/SiO_2 (1,000 Å)/Si_3N_4 (750 Å) sandwich layer is grown and patterned on top of the micro-heater as shown in Fig. 37.8d. Figure 37.8e and f show that aluminum (2.5 μm) and polysilicon (5,000 Å) bonding materials are deposited and patterned. The sacrificial release is the final step to form freestanding μ-resonators. Figure 37.8f shows a thick AZ 9245 photoresist is applied to cover aluminum/silicon-to-glass bonding system to ensure that the system withstands the attack from concentrated hydrofluoric acid. After 20 minutes sacrificial release in concentrated HF, the system as shown in Fig. 37.8g is ready for vacuum packaging. A Pyrex glass cap with a 10-μm deep recess is then placed on top with an applied pressure of ∼ 0.2 MPa under a 25 mtorr vacuum, and the heater is warmed using 3.4 watt input power (exact amount depends on the design of the micro-heaters) for 10 minutes to complete the vacuum packaging process as shown in Fig. 37.8h.

To evaluate the integrity of the resonators packaged using localized aluminum/silicon-to-glass solder bonding, the glass cap is forcefully broken and removed from the substrate. No damage is found on the μ-resonator and a part of the micro-heater is stripped away as shown in Fig. 37.9, demonstrating that a strong and uniform bond can be achieved without detrimental effects on the encapsulated device. Figure 37.10 shows a vacuum encapsulated unannealed μ-resonator (∼ 57 kHz) after

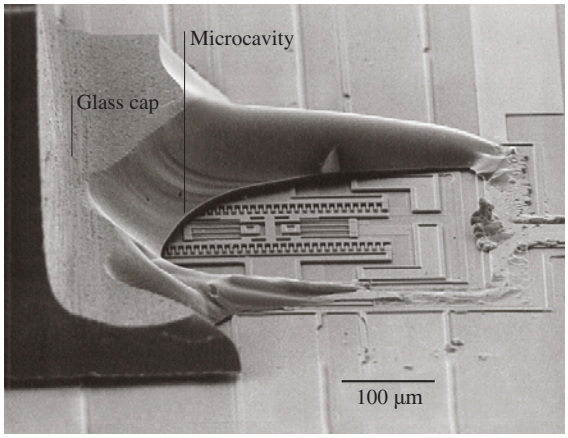


Fig. 37.9 The SEM microphoto of encapsulated micro-resonators after the glass cap is forcefully broken

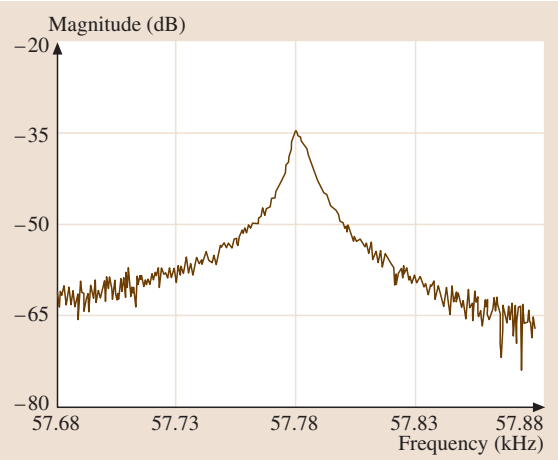


Fig. 37.10 The transmission spectrum of a glass-encapsulated μ-resonator after 120 minute pump down time in vacuum environment ($Q = 9{,}600$)

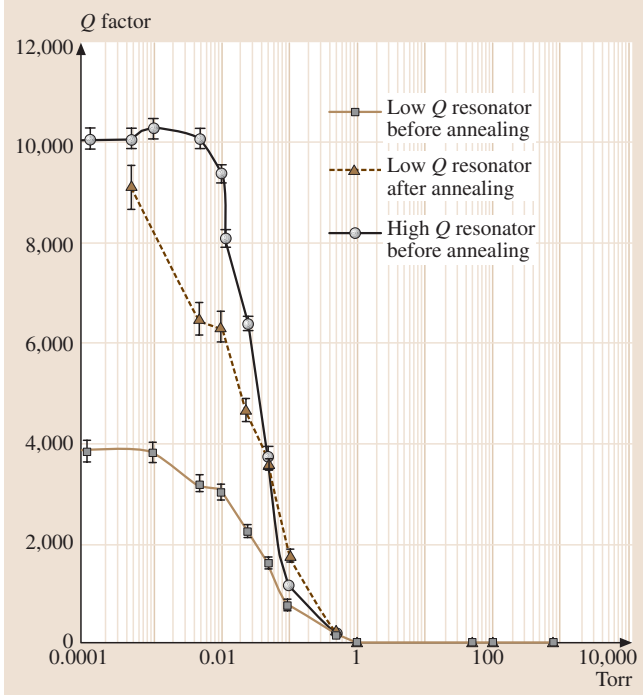


Fig. 37.11 Measured Q factor vs. pressure of unpackaged μ-resonators

120 minutes of wait time. The measured Q-factor after packaging is 9,600. Based on the measurement of Q vs. pressure of a high Q unpackaged μ-resonator as shown in Fig. 37.11, it is demonstrated that the pressure inside the packaging is comparable to the vacuum level of packaging chamber.

The post-process packaging using localized heating and bonding technique includes four basic components: (1) an electrical and thermal insulation layer such as silicon dioxide or silicon nitride should be used for localized heating, (2) resistive micro-heaters will be fabricated to provide the heating source for localized bonding, (3) materials, including metal and polysilicon that can provide good bonding and hermeticity with silicon or glass substrates are considered as the bonding materials, and (4) a good heat sink under the device substrate for localized heating is provided during the bonding experiments. MEMS devices will be fabricated on the device chip and hermetically sealed in the cavity formed by the device chip, resistive micro-heaters, and protection cap. The process can be either die level or wafer level. The schematic design of wafer-level packaging process is shown in Fig. 37.12. The resistive micro-heaters are parallel to each other and connected in order to ensure that identical current density is applied for individual packages at the same time. These heaters can be fabricated either on the chip or protection cap and can be built in a larger wafer for current inputs. The interconnections for these packaging cavities can be built in the dicing area, so no extra space is required for the packaging process.

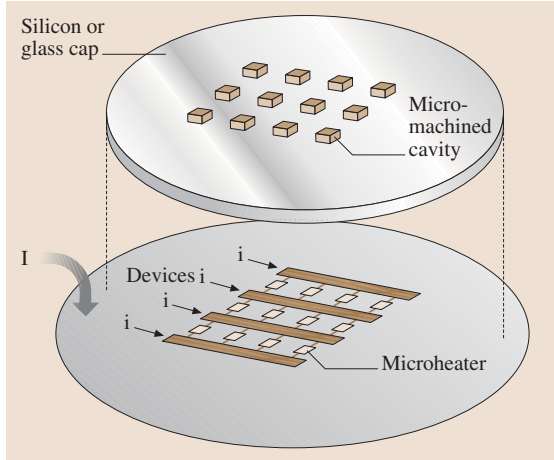


Fig. 37.12 Illustration of wafer-level vacuum packaging in the wafer level using localized heating and bonding technique

37.3 Thermal Issues and Packaging Reliability

37.3.1 Thermal Issues in Packaging

The two key thermal issues related to MEMS packaging are: 1) heat dissipation from actuators and integrated circuitry components and 2) thermal stress generated during the packaging process. These two topics are discussed separately.

Heat Dissipation Issues

In IC, heat dissipation is a serious problem when the size of transistor continues to shrink and the density of transistor on a chip continue to increase with the advance of IC fabrication technology. The trend of power packing into smaller packages has created increasing thermal management challenges [37.86]. Since the electrical characteristics of transistors change with working temperature, inefficient power dissipation could raise the working temperature and affect device performance. Present MEMS devices don't need high power, high performance microprocessors so power dissipation is not a problem. Nevertheless, some functional components in packaged MEMS are very sensitive to temperature variation, such as biomaterials or laser diodes. Several MEMS chemical sensors and other applications such as micro PCR chamber for DNA replication actually require elevated temperature for operation and micro-thermal platforms are built for these devices. Thermal management to maintain working temperature on these chips for stable operation is still an essential issue for packaging considerations. The geometrical complexity of MEMS resulting from packing various functional components in the tight space increases the difficulty of the thermal management. As the MEMS packaging integration process becomes more complex, the fabrication constraint in the packaging process will have great impact on the heterogeneous integration process in the front-end MEMS and IC processes. For example, low temperature requirement at the packaging process generally limits the possible choices of materials in the back-end process. In general, conventional IC packaging employs a heat sink attached to the chip to remove heat. The heat sink is generally made of a copper or

stainless steel bar with an array of fin structures on one side for better natural or forced heat convection to dissipate heat into the environment. In addition to heat sinks, thermal vias, heat pipe cooling, immersion cooling, and thermoelectric cooling can also be used for effective heat removal. Because most MEMS packages still follow the typical IC packaging architecture, one promising thermal management method, heat pipe, is discussed for possible MEMS packaging applications.

A heat pipe is a sealed slender tube containing a wick structure and a working fluid, typically water in electronics cooling. It is composed of three sections, the evaporator section at one end, condenser section at the other end, and adiabatic section in the middle. In the evaporator section, heat is absorbed by the working fluid via phase transformation from liquid to vapor. In the section of condenser, heat dissipates into the outside environment. Thus the fluid goes back to liquid phase. Vapor phase is at a high pressure and temperature state that forces the vapor to flow into the condenser section at a lower temperature. Once the vapor condenses and gives up its latent heat, the condensed fluid is pumped back to the evaporator section by the capillary force developed in the wick structure. Therefore the middle adiabatic section contains two phases, the vapor phase in the core region and the liquid phase in the wick, flowing opposite to each other with no significant heat transfer between the fluid and the surrounding medium. Silicon has good thermal conductivity (1.41 W/cm/°C) and is easily micro-machined to fabricate the heat pipe. The implementation of the silicon micro heat pipe in IC and MEM packaging has therefore great potential and several approaches have been proposed on this topic [37.87–89].

Packaging Induced Thermal Stresses

Thermal based bonding processes have been used in MEMS packaging applications for many years, as introduced previously in this chapter. Thermal management is extremely important during the bonding process in order to avoid fracture in the substrate or MEMS device itself. Extremely high temperatures or rapid cooling conditions can cause damage and should be carefully evaluated both analytically and experimentally. There are many ways to provide heating energy including electrical resistive heating, oven heating or induction heating [37.83]. These bonding processes may be put into two categories. The first is localized bonding in which the heat is directly applied only to the adhesive material used to bond the package to the MEMS device. The second category is global heating where the entire system (MEMS device, adhesive, and packaging material) is heated to produce bonding of the materials. The latter is the common approach for all MEMS packaging processes. This section focuses therefore on the thermal stress effects during the heating and cooling procedures in MEMS packaging. The RTP aluminum-to-glass bonding process is used as the specific example for the discussion of thermal stresses [37.90]. The bonding process heats the packaging system to 750 °C for 10 seconds, and then cools it back to room temperature. To simulate this process, an ANSYS program [37.91] is established to examine the shear stress due to CTE (Coefficient of Thermal Expansion) variations in the bonding system as the result of temperature changes. The shear stress was recorded from the ANSYS analysis on the aluminum/Pyrex glass interface and the aluminum/silicon interface.

Two different models were analyzed. The first was the quartz-aluminum-silicon bonding system, and the second was the Pyrex glass-aluminum-silicon bonding system [37.92]. The results of the ANSYS analysis were then analyzed and compared with experimental observations. Figure 37.13 shows the ANSYS results for a Pyrex glass bonding system: the width of the aluminum solder is 100 μm and the maximum residual stress is be 60 MPa in glass, slightly lower than the fracture strength of Pyrex glass at 70 MPa. It was discovered that increasing the aluminum width leads to lower residual stresses. This likely occurred because the length of the Pyrex glass, quartz, and silicon remains constant. As a result, it will always want to contract the same amount with a constant temperature change independent of the aluminum width. When increasing the aluminum width, however, the stress is not in such a concentrated area and therefore decreases. For example, the maximum residual stress an-

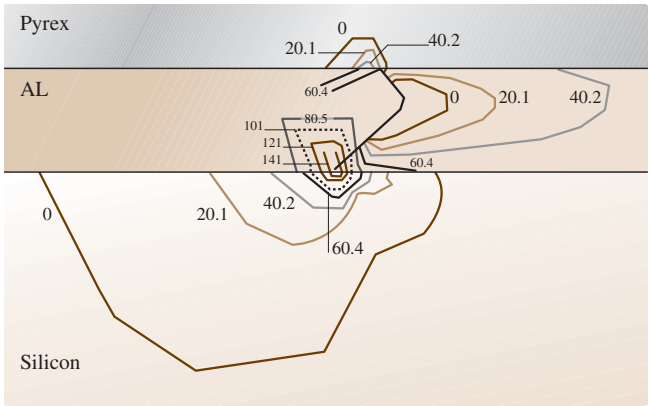


Fig. 37.13 Residual Stress (GPa) for an aluminum solder width of 100 μm under the RTP silicon-aluminum-glass bonding system

alyzed from ANSYS in the Pyrex glass bonding systems is 74.5 Gpa, 58 Gpa, and 60 GPa for aluminum widths of 30, 50, and 100 μm, respectively [37.92]. Pyrex glass has a documented strength of around 69 GPa [37.93]. Fracture should always occur with an aluminum width of 30 μm or less, according to the ANSYS analysis. Fracture may occur sporadically at widths of 50 or 100 μm, depending on the amount and magnitude of the flaws in the Pyrex glass. Experiments were done on the Pyrex glass bonding systems with a width of 100 μm. The samples were heated to 750 °C and then cooled by taking them out of the oven. In all four experimental cases, small cracks were observed in Pyrex glass as shown in Fig. 37.14. These cracks may have occurred consistently for several reasons. First, it may be a result of handling of the Pyrex glass before bonding. The Pyrex glass samples were kept in containers with each other, which may have resulted in abrasive contact and possibly caused flaws in the materials. These flaws can cause a reduced strength from the 69 GPa and therefore the predictions from the ANSYS analysis could be correct. Second, the observation is made that the cracks are small, only occurring tens of microns away from the aluminum and not propagating completely through the Pyrex glass. These cracks could be caused by the high stress applied, but the crack has not reached a critical size and therefore has not propagated completely through the Pyrex glass. Therefore, the strength remains at the theoretical strength of 69 GPa, and the Pyrex glass is only partially cracked. Experimental analysis done by *Chaio* and *Lin* [37.90] shows that fracture was not observed when using aluminum widths greater than 150 μm. This is consistent with the results of the ANSYS analysis which show that as the width of the aluminum is increased, the residual stress decreases.

The quartz bonding system ANSYS stress predictions are conducted, and the maximum stress predicted in the analysis is 207 GPa, 117 GPa, and 100 GPa, for aluminum widths of 30 μm, 50 μm, and 100 μm, respectively. All three of these stresses are much larger than the theoretical strength of Quartz at 48 GPa, and therefore fracture should always occur. Quartz has a much larger coefficient of thermal expansion (CTE) than silicon, which is why this was predicted. Experimentally, a quartz substrate was tested for the silicon-aluminum-quartz bonding test and the result is shown in Fig. 37.15.Cracks happen all over the place and cause serious damages on the quartz wafer. These cracks could be the failure mechanism of the hermetic package. Pyrex glass is therefore identified as a better bonding substrate than quartz.

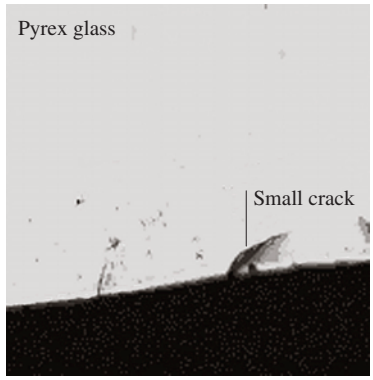


Fig. 37.14 Microphoto of the experimental result on Pyrex Glass-Aluminum-Silicon system. Small cracks can be observed

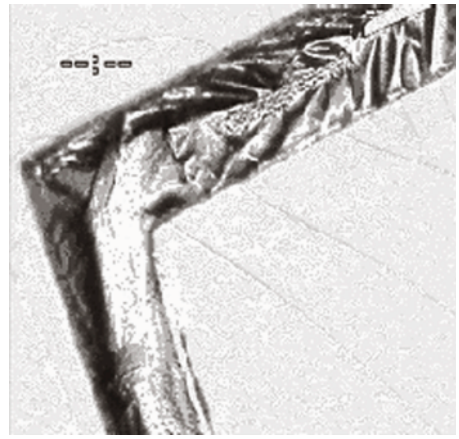


Fig. 37.15 Bonding result of Quartz-Aluminum-Silicon system, fracture can be observed

The thermal stresses generated in the packaging process in quartz are much larger than the stresses in Pyrex glass because of the difference between the CTE in the two systems. Quartz has a low CTE ($0.54\times10^{-6}\,K^{-1}$) compared to the aluminum CTE ($23\times10^{-6}\,K^{-1}$) and silicon CTE ($3.5\times10^{-6}\,K^{-1}$). On the other hand Pyrex glass has a much closer CTE ($3.2\times10^{-6}\,K^{-1}$) to silicon and aluminum, resulting in smaller stresses. The practical implications of the ANSYS results and the preceding information are that materials must be chosen carefully when carrying out bonding. Material should not be used with a much higher or lower CTE than silicon to ensure fracture will not occur. This finding is a valuable component of this particular packaging system and should benefit the other packaging processes involving bonding MEMS packaging because it supports the prediction that Pyrex glass is an excellent material to use when bonding with silicon in MEMS packaging, as long as sufficiently wide adhesive material is used.

37.3.2 Packaging Reliability

The thermal stress induced by the CTE mismatch is one of the main factors that affect the packaging reliability. As a matter of fact, the formation of the stress can happen not only during packaging process but also during the operation of devices. In particular, during device operation, the package will go through various temperature cycles because of environment change. Such temperature variation causes the expansion of packaging materials when they are constrained by the packaged assembly. As a result of such thermal mismatch, significant stresses are induced in the package and finally cause the device to fail. In addition to thermal mismatch, corrosion creep, fracture fatigue, crack initiation and propagation, and delamination of thin films are all possible factors to cause failure of packaged devices [37.94]. These failure mechanisms could be prevented or deferred under proper packaging designs. For instance, the thermally induced strain inside the packaging material is generally below the tolerance of the material and can't cause immediate catastrophic damage. But cyclic loading can generate and accumulate stresses, eventually causing failure. Several common designs in IC packaging have been used to prolong the lifetime of devices. For example, the strain in solder interconnects of BGA or flip-chip packaging can be effectively reduced by introducing a polymer underfill material between the chip and the substrate to distribute thermal stress induced by CTE mismatch [37.95]. The strain can further be reduced if excellent thermal paths are built around interconnects to diminish the thermal stress originated from the temperature gradient between ambient and operation temperatures. Delamination phenomenon is another source to cause reliability problems, and it happens in the interface of adjacent material layers. In MEMS, components made of dissimilar materials are commonly bonded together to provide specific functions. Delamination can result in electrical or mechanical failures of devices in the package such as mechanically cracking through the electrical via wall to make an electrical open because of the propagation of the delamination of metal line from the dielectric layer or over heating of the die because of the delamination of die from its underneath to make an opening on the heat dissipation path. Because the stress and thermal loading, the geometry, and the material properties are complex in MEMS, the development of the packaging designs to increase the reliability is very important and requires more extensive investigations.

Reliability testing is required before a new device can be delivered to the market. The test results can provide the information for the following improvement of packaging design and fabrication. How to analyze the failure data, called the "reliability metrology", is thus very important in the packaging industry. The analysis method is to use the mathematical tools of probability and statistical distributions to evaluate data to understand the patterns and identify the sources of failure. For example, a failure density function is defined as the time derivative of the cumulative failure function

$$f(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t}\,, \tag{37.1}$$

$$F(t) = \int_0^t f(s)\,\mathrm{d}s\,. \tag{37.2}$$

The cumulative failure function, $F(t)$, is the fraction of a group of original devices that has failed at time t. The Weibull distribution function is one of the analytic mathematic models commonly used in the packaging reliability evaluation to represent the failure density function [37.21].

$$f(t) = \frac{\beta}{\lambda}\left(\frac{t}{\lambda}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right]\,, \tag{37.3}$$

where β and λ are the Weibull parameters. The parameters β, called a shape factor, measures how the failure frequency is distributed around the average lifetime. The parameter λ is called the lifetime parameter and indicates the time at which 63.2% of the devices faile. By integrating both sides of the equation (37.3). $F(t)$ becomes:

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right]\,. \tag{37.4}$$

Using the Weibull distribution function with the two parameters extrapolated by experimental data, one can estimate the number of failures at any time during the test. Moreover, by knowing the meaning and values of the parameters, one can compare two sets of test data. For example, the λ with greater number means this set of samples has longer lifetime. Because all of the mathematical models are statistical approximations based on the real experimental data, more testing samples can provide more accurate estimation.

37.3.3 Long-Term and Accelerated MEMS Packaging Tests

The reliability of MEMS packages is best characterized by means of long-term tests with statistical data analyses. But it is very difficult and time-consuming to obtain

these experimental data so accelerated tests that put samples in extreme environment for accelerated failure are commonly used to predict the lifetime of the devices. Unfortunately, few research publications deal with these two issues. In this section, two MEMS packaging examples that aim to address these two tests we discuss.

Figure 37.16 shows long-term measurements of the Q factor of a vacuum packaged μ-resonators using localized aluminum/silicon-to-glass bonding [37.81] vacuum encapsulation process is described in details as shown in Fig. 37.8. It is found that the vacuum package by means of localized heating and bonding provides stable vacuum environments for the μ-resonator and a quality factor of 9,600 has been achieved with no degradation for at least one year. Since the performance of high Q μ-resonator is very sensitive to environmental pressure as shown in Fig. 37.11, any leakage can be easily detected. The fact that this high Q value can hold for one year indicates the packaging process is well performed and both aluminum and Pyrex glass are suitable materials for vacuum packaging applications. According to a previous study of hermeticity in different materials, metal has lower permeability to moisture than other materials such as glass, epoxy, and silicon. With a width of 1 μm, metal can effectively block moisture for more than ten years [37.21]. In this vacuum package system, the bonding width is 30 μm so it can sufficiently block the diffusion process of moisture. On the other hand, the diffusion effects of air molecules into these tiny cavities have not been studied extensively, and the design guidelines for vacuum encapsulations are not clearly defined. Further investigations will be needed in this area, and the example presented here serves as a good starting point.

On the other hand, accelerated testing puts a large amount of samples in harsh environments, such as elevated temperature, elevated pressure, and 100% humidity, to accelerate the corrosion process. The statistical failure data are gathered and analyzed to predict the lifetime of packages under normal usage environment. As a result, the long-term reliability of the package can be predicted without going through the true long-term tests. Unfortunately, accelerated tests have not been addressed in MEMS research papers. Although the MEMS industry must have done some extensive reliability tests, they do not publish their results, probably due to liability concerns. Among the very limited publications, this section uses as the illustration example a specific MEMS packaging system that has gone through accelerated tests [37.45].

The MEMS package is accomplished by means of RTP (Rapid Thermal Processing) bonding as described previously in this chapter. The goal of the accelerated test is to examine the failure rate at the bonding interface. The accelerated tests start by putting the packaged samples into the autoclave chamber filled with high temperature and pressurized steam at 130 °C, 2.7 atm, and 100% relative humidity for accelerated testing. The pressurized steam can penetrate small cracks if any defect exists at the bonding interface [37.96]. Elevated temperature and humid environment speed up the corrosion process. A package is considered a failure if water is condensed or diffused into the package. Statistical failure data are gathered every 24 h under optical examination for a period of 864 h, during which time new failure is seldom observed. Because of the robustness of the samples, it is difficult to conduct the tests to the point at which all the packages may fail. The cumulative failure function is recorded, and it is found that most failures occur in the first 96 h and such high early failure rate reflects the yielding issue of the bonding process. Weibull and Lognormal models are compared to predict the lifetime of the packages, and it is found that Lognormal model better describes the statistical data. Figure 37.17 shows the inverse standard normal distribution function vs. ln(time) and the Maximum Likelihood Estimator (MLE) is then used to predict the mean, standard deviation, and the MTTF (mean time to failure). Table 37.2 shows the MLE calculation results of MTTF. The wide interval of confidence level comes from the fact that only a small number of samples failed at the end of the test. It is also observed that packages with larger bonding width and smaller bonding areas have larger MTTF values. The lower bound of the MTTF provides the worst-case scenario. For example, only 4 our of 31 samples failed when tests stopped in the case of ring width of 200 μm and sealing area of $450 \times 450\ \mu m^2$. The MTTF predicts, in the worst-case scenario, that there

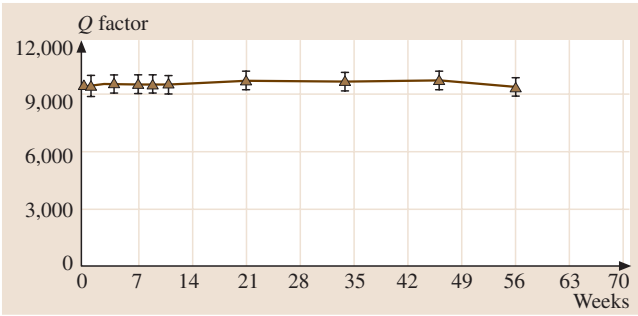


Fig. 37.16 Long-term measurement of encapsulated μ-resonators. No degradation of Q factors is found after 56 weeks

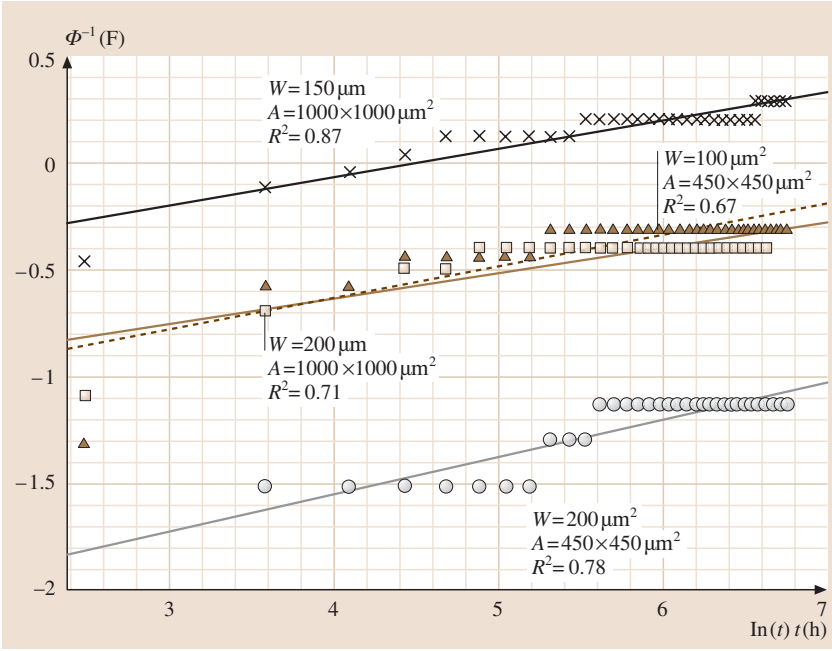


Fig. 37.17 Life data fitted by Lognormal distribution. R^2 is the coefficient of determination

Table 37.2 The Maximum Likelihood Estimation for Mean Time To Failure (MTTF)

MLE calculation resuls of MMTF[a]				
Bonding width, W (μm)	Area, A (μm²)	MTTF		Worst cases in jungle condition (years)
		UB (years)	LB (years)	
200	450× 450	1.8×10^7	0.57	1,700
100	450× 450	5.3	0.10	300
200	1,000×1,000	6.5×10^3	0.09	270
150	1,000×1,000	0.50	0.017	50

[a] UB is the upper bound, and LB is the lower bound of the 90% confidence interval, respectively. The MTTF LB times AF is the worst case MTTF used in jungle condition.

is 90% chance a package will fail in 0.57 years in the autoclave environment.

It is widely accepted that the acceleration factor (AF) for autoclave tests follows the Arrhenius equation [37.21] and can be modeled as:

$$\mathrm{AF} = \frac{(\mathrm{RH}^{-n}\,\mathrm{e}^{\Delta E_a/kT})_{\mathrm{normal}}}{(\mathrm{RH}^{-n}\,\mathrm{e}^{\Delta E_a/kT})_{\mathrm{accelerated}}}\,, \tag{37.5}$$

where RH is relative humidity (85%, RH = 85), k the Boltzmann constant and T the absolute temperature. The recommended value for n, an empirical constant, is 3.0 [37.97], and ΔE_a, the activation energy, is 0.9 eV for plastic dip package and 0.997 eV for anodically bonded glass-to-silicon package [37.58]. If $\Delta E_a = 0.9$ is used to estimate the AF for the accelerated testing condition as compared with the jungle condition (35 °C, 1 atm and 95% RH), AF is about 3,000, and the worst-case lifetime values in jungle condition are also listed in Table 37.2. The high values of estimated MTTF in jungle condition could be a result of overestimation of AF because plastic dip package may have smaller AFs than those of glass packages. Nevertheless, these data and analyses provide important guidelines in the area of accelerated tests for MEMS packages.

37.4 Future Trends and Summary

In the past, the development of MEMS packaging is mainly originated from IC packaging advancement because existing packaging techniques can significantly reduce the development cost of MEMS. It is expected that the situation will be changed very soon, however, in a way the MEMS packaging approaches will assist the IC packaging development. Recent advances in IC packaging aim to provide high I/O density and more chip integration capability for the needs of high speed and high data communication rate. In order to satisfy those requirements, several packaging concepts and techniques are developed, including 3-D packaging, wafer-level packaging, BGA, and flip-chip technique. Although all of the concepts and methods can provide a package with more I/O density, flexibility in chip integration, and lower manufacturing cost for IC fabrication, they are still insufficient for providing solutions for the future applications because of the increase of complexity and requirements of MEMS packaging. On the contrary, with the progress of MEMS fabrication technologies, several key processes such as DRIE (Deep Reactive Ion Etching), wafer bonding, and thick photoresist processes [37.98] have been utilized for IC packaging fabrication. Technologies developed in MEMS fabrication can therefore also assist the development of new IC packaging approaches.

In order to address the future needs of process integration, adaptive multi-chip module (MCM) [37.28] or 3-D packaging combined with vertical through substrate interconnects [37.10, 99] are promising approaches for the development of future MEMS packaging processes. Based on low temperature flip-chip solder bonding technique, these packaging methods can provide more flexibility in device fabrication and packaging. Devices can be fabricated before they are integrated to form microsystems to dramatically reduce the packaging cost. Vertical through substrate interconnects can have higher I/O density, smaller resistance, parasitic capacitance, and mutual inductance. Although this approach provides many possible advantages, technical challenges exist. For instance, metal is commonly used as the fill material inside the vertical vias as electrical interconnects, and it may introduce large thermal mismatch with respect to silicon substrate and generate huge thermal stress to cause packaging reliability problems. Moreover, it will be an interesting engineering challenge to fill the materials into those high aspect ratio vias.

The future development of MEMS packaging depends on the successful implementation of unique techniques as described in the following:

1. Development of mechanical, thermal, and electrical models for packaging designs and fabrication processes;
2. Wafer-level, chip-scale packaging with low packaging cost and high yield;
3. Effective testing techniques at the wafer level to reduce the testing costs;
4. Device integrations by vertical through-interconnects as an interposer [37.1] to avoid thermal mismatch problems.

In addition to these approaches and challenges, many other possibilities have not been listed but also require dedicated investigations. For example, several key nanotechnologies have been introduced in the previous chapters, but the packaging solutions for the NEMS devices have not been addressed. Because it is feasible to use MEMS as the platform for NEMS fabrication, all of the packaging issues discussed in the chapter can be directly applied to NEMS devices. On the other hand, nanotechnology may introduce new opportunities for MEMS/NEMS packaging applications by providing superior electrical, mechanical, and thermal properties [37.100–103]. For example, carbon nanotubes have very high thermal conductivity [37.102] and may be suitable to increase the thermal cooling effects for better IC/MEMS/NEMS packaging applications.

In summary, MEMS packaging issues have been introduced in the areas of fabrication, application, reliability, and future development. Packaging design and modeling, packaging material selection, packaging process integration, and packaging cost are the main issues to be considered when developing a new MEMS packaging process.

References

37.1 M. Matsuo, N. Hayasaka, K. Okumura, E. Hosomi, C. Takubo: Silicon interposer technology for high-density package, IEEE ECTC (2000) 1455–1459

37.2 K. E. Peterson: Silicon as a mechanical material, Proc. IEEE **70** (1982) 420–457

37.3 L. Lin, H.-C. Chu, Y.-W. Lu: A simulation program for the sensitivity and linearity of piezoresistive pressure sensors, IEEE/ASME J. Microelectromech. Syst. **8** (1999) 514–522

37.4 Y. C. Tai, R. S. Muller: Lightly-doped polysilicon bridge as a flow meter, Sens. Actuators **15** (1988) 63–75

37.5 D. Hicks, S.-C. Chang, M. W. Putty, D. S. Eddy: Piezoelectrically activated resonant bridge micro-acceleromenter, 1994 Solid-State Sensors and Actuators Workshop (1994) 225–228

37.6 J. Berstein, S. Cho, A. King, A. Kourepenis, P. Maciel, M. Weinberg: A micromachined comb-drive tuning fork rate gyroscope, 6th IEEE Int. Conf. on MEMS (1993) 143–148

37.7 M. Madou: Compatibility and incompatibility of chemical sensors and analytical equipment with micromachining, Solid-State Sensor and Actuator Workshop, Hilton Head (1994) 164–171

37.8 M. J. Zdeblick, J. B. Angell: A microminiature electric-to-fluidic valve, Proceedings of Transducers'87, the 4th International Conference on Solid-State Transducers and Actuators (1987) 827–829

37.9 J. H. Tsai, L. Lin: A thermal bubble actuated micro nozzle-diffuser pump, IEEE/ASME J. Microelectromech. Syst. **11** (2002)

37.10 S. J. Ok, D. Baldwin: High density, aspect ratio through-wafer electrical interconnect vias for low cost, generic modular MEMS packaging, 8th IEEE Int. Symp. On Adv. Pack. Mater. (2002) 8–11

37.11 H. Reichl, V. Grosser: Overview and development trends in the field of MEMS packaging, 14th IEEE Int. Conf. on MEMS (2001) 1–5

37.12 P. Kapur, P. M. McVittie, K. Saraswat: Technology and reliability constrained future copper interconnects=Part II: Performance implication, Trans. Electron. Dev. **49** (2002) 598–604

37.13 J. Lau: *Flip Chip Technologies* (McGraw-Hill, New York 1996)

37.14 R. Prasad: *Surface Mount Technology: Principles and Practice*, 2nd edn. (Chapman, New York 1989)

37.15 M. Töpper, J. Auersperg, V. Glaw, K. Kaskoun, E. Prack, B. Keser, P. Coskina, D. Jäger, D. Petter, O. Ehrmann, K. Samulewicz, C. Meinherz, S. Fehlberg, C. Karduck, H. Reichl: Fab integrated packaging (FIP) a new concept for high reliability wafer-level chip size packaging, IEEE, ECTC (2000) 74–81

37.16 S. Savastiouk, O. Siniaguine, Ed. Korczynski: 3D wafer level packaging, IEEE, ECTC. (2000) 26–31

37.17 J. H. Tsai, L. Lin: Micro-to-macro fluidic interconnectors with an integrated polymer sealant, J. Micromech. Microeng. **11** (2001) 577–581

37.18 T. A. Core, W. K. Tsang, S. Sherman: Fabrication technology for an integrated surface-micromachined sensor, Solid State Technol. (October 1993) 39–47

37.19 K. Koester, R. Majedevan, A. Shishkoff, K. Marcus: *Multi-Sser MEMS Processes (MUMPS) Introduction and Design Rules* (MCNC MEMS Technology Applications Center, Research Triangle Park, NC 27709 1996)

37.20 R. Lengtenberg, H. A. C. Tilmans: Electrically driven vacuum-encapsulated polysilicon resonantor, Part I: Design and fabrication, Sens. Actuators A **45** (1994) 57–66

37.21 R. R. Tummala, E. J. Rymaszewski, A. G. Klopfenstein: *Microelectronics Packaging Handbook, Semiconductor Packaging* (Chapman, New York 1997)

37.22 S. Wolf: *Silicon Processing for the VLSI era, Vol. I: Process Technology* (Lattice Press, Sunset Beach 1995)

37.23 J. K. Bhardway, H. Ashraf: Advanced silicon etching using high density plasmas, SPIE Micromach. Fabr. Technol. **2639** (1995) 224–233

37.24 W. H. Ko, J. T. Suminto, G. J. Yeh: *Bonding Techniques for Microsensors, Micromachining and Micropackaging for Transducers* (Elsevier, Amsterdam 1985)

37.25 S. D. Senturia, R. L. Smith: Microsensor packaging and system partitioning, Sens. Actuators **15** (1988) 221–234

37.26 R. L. Smith, S. D. Collins: Micromachined packaging for chemical microsensors, IEEE Trans. Electron Dev. **35** (1988) 787–792

37.27 J. T. Butler, V. M. Bright, P. B. Chu, R. J. Saia: Adapting multichip module foundries for MEMS packaging, Int. Conf. on Multi. Mod. High Den. Pack. (1998) 106–111

37.28 M. Schuenemann, A. J. Kourosh, V. Grosser, R. Leutenbauer, G. Bauer, W. Schaefer, H. Reichl: MEM modular packaging and interfaces, IEEE ECTC (2000) 681–688

37.29 D. W. Lee, T. Ono, T. Abe, M. Esashi: Fabrication of microprobe array with sub-100nm nano-heater for nanometric thermal imaging and data storage, Proc. IEEE Micro Electro Mechanical Systems Conf., IEEE Int. Conf. on MEMS (January 2001) 204–207

37.30 A. S. Laskar, B. Blythe: Epoxy multichip modules, a solution to the problem of packaging and interconnection of sensors and signal-processing chips, Sens. Actuators A **36** (1993) 1–27

37.31 R. Reichl: Packaging and interconnection of sensors, Sens. Actuators A **25–27** (1991) 63–71
37.32 A. Grisel, C. Francis, E. Verney, G. Mondin: Packaging technologies for integrated electrochemical sensors, Sens. Actuators **17** (1989) 285–295
37.33 J. L. Lund, K. D. Wise: Chip-level encapsulation of implantable CMOS microelectrode arrays, 1994 Solid-State Sensor and Actuator Workshop, Hilton Head (1994) 29–32
37.34 T. Akin, B. Siaie, K. Najafi: Modular micromachined high-density connector for implantable biomedical systems, IEEE Micro Electro Mechanical Systems Workshop (1996) 497–502
37.35 L. Muller, M. H. Hecht, L. M. Miller et al.: Packaging qualification for MEMS-based space systems, IEEE Micro Electro Mechanical Systems Workshop 503–508
37.36 L. Lin, K. McNair, R. T. Howe, A. P. Pisano: Vacuum encapsulated lateral microresoantors, 7th Int. Conference on Solid State Sensors and Actuators (June 1993) 270–273
37.37 H. Guckel: Surface micromachined pressure transducers, Sens. Actuators A **28** (1991) 133–146
37.38 J. J. Sniegowski, H. Guckel, R. T. Christenson: Performance characteristics of second generation polysilicon resonating beam force transducers, IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island (1990) 9–12
37.39 K. Ikeda, H. Kuwayama, T. Kobayashi, T. Watanabe, T. Nishikawa, T. Oshida, K. Harada: Three dimensional micromachining of silicon pressure sensor integrating resonant strain gauge on diaphragm, Sens. Actuators A **21–23** (1990) 1001–1010
37.40 C. H. Mastrangelo, R. S. Muller: Vacuum-sealed silicon micromachined incandescent light source, IEEE IEDM (1989) 503–506
37.41 J. Smith, S. Montague, J. Sniegowski, R. Manginell, P. McWhorter, R. Huber: Characterization of the embedded micromechanical device approach to the monolithic integration of MEMS with CMOS, SPIE 2879
37.42 S. Van der Groen, M. Rosmeulen, P. Jansen, K. Baert, L. Deferm: CMOS compatible wafer scale adhesive bonding for circuit transfer, 1997 International Conference on Solid-State Sensors and Actuators, Transducers'97 (1997) 629–632
37.43 M. B. Cohn, Y. Liang, R. Howe, A. P. Pisano: Wafer to wafer transfer of microstructures for vacuum package, Solid State Sensor and Actuator Workshop, Hilton Head (1996) 32–35
37.44 Y. T. Cheng, Y. T. Hsu, L. Lin, C. T. Nguyen, K. Najafi: Vacuum packaging using localized aluminum/silicon-to-glass bonding, Int. Conf. on MEMS (2001) 18–21
37.45 M. Chiao, L. Lin: Vacuum packaging of microresonators by rapid thermal processing, Proc. of SPIE on Smart Electronics, MEMS, and Nanotechnology (2002) 17–21
37.46 M. Esashi, S. Shoji, A. Nakano: Normally closed microvalve and micropump fabricated on a silicon wafer, Sens. Actuators **20** (1989) 163–169
37.47 M. E. Poplawski, R. W. Hower, R. B. Brown: A simple packaging process for chemical sensors, Solid-State Sensor and Actuator Workshop, Hilton Head (1994) 25–28
37.48 S. F. Trautweiler, O. Paul, J. Stahl, H. Baltes: Anodically bonded silicon membranes for sealed and flush mounted microsensors, Micro Electro Mechanical Systems Workshop (1996) 61–66
37.49 E. H. Klaassen, K. Petersen, J. M. Noworolski, J. Logan, N. I. Malfu, J. Brown, C. Storment, W. McCulley, G. T. A. Kovac: Silicon fusion bonding and deep reactive ion etching: A new technology for microstructures, Sens. Actuators A **52** (1996) 132–139
37.50 C. H. Hsu, M. A. Schmidt: Micromachined structures fabricated using a wafer-bonded sealed cavity process, Solid State Sensor and Actuator Workshop, Hilton Head (1994) 151–155
37.51 A. L. Tiensuu, J. A. Schweitz, S. Johansson: In situ investigation of precise high strength micro assembly using Au-Si eutectic bonding, Int. Conf. Solid-State Sens. Actuators, Eurosens. IX (1995) 236–239
37.52 A. P. Lee, D. R. Ciarlo, P. A. Krulevitch, S. Lehew, J. Trevino, M. A. Northrup: Practical microgripper by fine alignment, eutectic bonding and SMA actuation, Int. Conf. Solid-State Sens. Actuators, Eurosens. IX (1995) 368–371
37.53 R. W. Bower, M. S. Ismail, B. E. Roberds: Low temperature Si_3N_4 direct bonding, Appl. Phys. Lett. **62** (1993) 3485–3487
37.54 H. Takagi, R. Maeda, T. R. Chung, T. Suga: Low temperature direct bonding of silicon and silicon dioxide by the surface activation method, Int. Conf. on Solid-State Sensors and Actuators, Transducer 97 (1997) 657–660
37.55 G. Wallis, D. Pomerantz: Filed assisted glass-metal sealing, J. Appl. Phys. **40** (1969) 3946–3949
37.56 L. Bowman, J. Meindl: The packaging of implantable integrated sensors, IEEE Trans. Biomed. Engin. **33** (1986) 248–255
37.57 M. Esashi: Encapsulated micro mechanical sensors, Microsyst. Technol. **1** (1994) 2–9
37.58 B. Ziaie, J. Von Arx, M. Dokmeci, K. Najafi: A hermetic glass-silicon micropackages with high-density on-Chip feedthroughs for sensors and actuators, J. Microelectromech. Syst. **5** (1996) 166–179
37.59 Y. Lee, K. Wise: A batch-fabricated silicon capacitive pressure transducer with low temperature sensitivity, IEEE Trans. Electron Dev. **29** (1982) 42–48
37.60 S. Shoji, H. Kicuchi, H. Torigoe: Anodic bonding below 180 degree C for packaging and assembling of MEMS using lithium aluminosilicate-beta-quartz glass-ceramic, Proc. 10th Annual Int. Workshop Micro Electro Mechanical Systems (1997) 482–487

37.61 M. Esashi, N. Akira, S. Shoji, H. Hebiguchi: Low-temperature silicon-to-silicon anodic bonding with intermediate low melting point glass, Sens. Actuators A **23** (1990) 931–934

37.62 A. Hanneborg, M. Nese, H. Jakobsen, R. Holm: Silicon-to-thin film anodic bonding, J. Micromech. Microeng. **2** (1992) 117–121

37.63 S. A. Audet, K. M. Edenfeld: Integrated sensor wafer-level packaging, International Conference on Solid-State Sensors and Actuators, Transducers'97 (1997) 287–289

37.64 R. C. Benson, N. deHaas, P. Goodwin, T. E. Phillips: Epoxy adhesives in microelectronic hybrid applications, Johns Hopkins APL Tech. Dig. **13** (1992) 400–406

37.65 M. Shimbo, J. Yoshikawa: New silicon bonding method, J. Electrochem. Soc. **143** (1996) 2371–2377

37.66 P. M. Zavracky, B. Vu: Patterned eutectic bonding with Al/Ge thin film for MEMS, SPIE **2639** (1995) 46–52

37.67 G. Humpston, D. M. Jacobson: Principles of soldering and brazing, ASM Int. (1993) 241–244

37.68 A. Singh, D. Horsely, M. B. Cohn, R. Howe: Batch transfer of microstructures using flip-chip solder bump bonding, International Conference on Solid State Sensors and Actuators, Transducer 97 (1997) 265–268

37.69 M. M. Maharbiz, M. B. Cohn, R. T. Howe, R. Horowitz, A. P. Pisano: Batch micropackaging by compression-bonded wafer-wafer transfer, 12th Int. Conf. MEMS (1999) 482–489

37.70 Y. T. Cheng, L. Lin, K. Najafi: Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging, IEEE/ASME J. Microelectromech. Syst. **9** (2000) 3–8

37.71 L. Lin: Selective encapsulations of MEMS: micro channels, needles, resonators, and electromechanical filters. Ph.D. Thesis (University of California, Berkeley 1993)

37.72 Y. C. Su, L. Lin: Localized plastic bonding for micro assembly, packaging and liquid encapsulation, Proceedings of IEEE Micro Electro Mechanical Systems Conference, Interlaken (2001) 50–53

37.73 M. Schwartz: *Brazing* (Chapman, London 1995)

37.74 L. Lin, R. T. Howe, A. P. Pisano: Microelectromechanical Filters for Signal Processing, IEEE/ASME J. Microelectromech. Syst. **7** (1998) 286–294

37.75 W. C. Tang, C. T.-C. Nguyen, R. T. Howe: Laterally driven polysilicon resonant microstructures, Sens. Actuators A **20** (1989) 25–32

37.76 G. T. Mulhern, D. S. Soane, R. T. Howe: Supercritical carbon dioxide drying of microstructures, 7th International Conference on Solid State Sensors and Actuators, Yokohama (1993) 296–299

37.77 M. Judy: Micromechanisms using sidewall beams. Ph.D. Thesis (University of California, Berkeley 1994)

37.78 A. M. Leung, J. Jones, E. Czyzewska, J. Chen, B. Woods: Micromachined accelerometer based on convection heat transfer, International Conference on MEMS (1998) 627–630

37.79 D. R. Spark, L. Jordan, J. H. Frazee: Flexible vacuum-packaging method for resonating micromachines, Sens. Actuators A **55** (1996) 179–183

37.80 SensArray Corporation, 47451 Fremont Blvd. Fremont, CA 94538

37.81 Y. T. Cheng, W. T. Hsu, K. Najafi, C. T. Nguyen, L. Lin: Vacuum packaging technology using localized aluminum/silicon-to-glass bonding, IEEE/ASME J. Microelectromech. Syst. **11** (2002) 556–565

37.82 C. Luo, L. Lin: The application of nanosecond-pulsed laser welding technology in MEMS packaging with a shadow mask, Sens. Actuators A **97–98** (2002) 398–404

37.83 A. Cao, M. Chiao, Liwei. Lin: Selective and localized wafer bonding using induction heating, Technical Digest of Solid-State Sensors and Actuators Workshop (2002) 153–156

37.84 J. B. Kim, M. Chiao, L. Lin: Ultrasonic Bonding of In/Au and Al/Al for Hermetic Sealing of MEMS Packaging, Proceedings of IEEE Micro Electro Mechanical Systems Conference (2002) 415–418

37.85 Y. T. Cheng, L. Lin, K. Najafi: Localized bonding with PSG or indium solder as intermediate layer, 12th International Conference on MEMS (1999) 285–289

37.86 G. Thyrum, E. Cruse: Heat pipe simulation, a simplified technique for modeling heat pipe assisted heat sinks, Adv. Packaging (2001) 23–27

37.87 G. P. Peterson, A. B. Duncan, M. H. Weichold: Experimental investigation of micro heat pipes fabricated in silicon wafers, J. Heat Trans. (1993) 751–756

37.88 L. Jiang, M. Wong, Y. Zohar: Forced convection boiling in a microchannel heat sink, J. Microelectromech. Syst. **10** (2001) 80–87

37.89 F. Arias, S. R. J. Oliver, B. Xu, E. Holmlin, G. M. Whitesides: Fabrication of metallic exchangers using sacrificial polymer mandrils, IEEE/ASME J. Microelectromech. Syst. **10** (2001) 107–112

37.90 M. Chiao, L. Lin: Hermetic wafer bonding based on rapid thermal processing, Sens. Actuators A **91** (2001) 404–408

37.91 *ANSYS Modeling and Meshing Guide*, 3rd edn. (SAS IP, Inc., Canousburg 2002)

37.92 D. Bystrom, L. Lin: Residual stress analysis of silicon-aluminum-glass bonding processes, ASME International Mechanical Engineering Congress and Exposition, MEMS Symposium (November 2002)

37.93 H. Scholze: *Glass: Nature, Structure, and Properties*, 1st edn. (Springer, Berlin, Heidelberg 1991)

37.94 R. R. Tummala: *Fundamentals of Microsystems packaging* (McGraw-Hill, New York 2001)

37.95 S. J. Adamson: BGA, CSP, and flip chip, Adv. Packaging (2002) 21–24

37.96 D. Steoehle: On the penetration of water vapor into packages with cavities and on maximum allow-

able leak rate, 15th Annual Proceeding Reliability Physics Symposium (1977) 101–106
37.97 W. D. Brown: *Advanced Electronic Packaging* (IEEE Press, 1999)
37.98 F. Niklaus, P. Znoksson, E. Käluesten, G. Stemme: Void free full wafer adhesion bonding, Proceedings of IEEE Micro Electro Mechanical Systems Conf. 2000) 241–252
37.99 C. H. Cheng, A. S. Ergun, B. T. Khuri-Yakub: Electrical through-wafer interconnects with subpicofarad parasitic capacitance, IEEE ECTC (2002) 18–21
37.100 D. Routkevitch, A. A. Tager, J. Haruyama, D. Almawlawi, M. Moskovits, J. M. Xu: Nonlithographic nano-array arrays: fabrication, physics, and device applications, IEEE Trans. Electron. Dev. **43** (1996) 1646–1658
37.101 J. Gou, M. Lundstrom, S. Datta: Performance projections for ballistic carbon nanotube fieldeffect transistors, Appl. Phys. Lett. **80** (2002) 3192–3194
37.102 S. U. S. Choi, Z. G. Zhang, W. Yu, F. E. Lockwood, E. A. Grulke: Anomalous thermal conductivity enhancement in nanotube suspensions, Appl. Phys. Lett. **79** (2001) 2252–2254
37.103 K. Velikov, A. Moroz, A. Blaaderen: Photonic crystals of core-shell colloidal particles, Appl. Phys. Lett. **80** (2002) 49–51

Part F Social and

Part F Social and Ethical Implication

38 Social and Ethical Implications of Nanotechnology
William S. Bainbridge, Arlington, USA

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

38. Social and Ethical Implications of Nanotechnology

Nanotechnology will have very broad applications across all fields of engineering, so it will be an amplifier of the social effects of other technologies. There is an especially great potential for it to combine with three other powerful trends – biotechnology, information technology, and cognitive science – based on the material unity of nature at the nanoscale and on technology integration from that scale. Technological convergence highlights such existing issues as the treatment of the disabled, communication breakdowns, economic stagnation, and threats to national security. Nanotechnology itself may possibly raise distinctive ethical and social issues in the future, but much of the public discussion to this point has been misdirected and misinformed, lacking a firm social scientific basis. Thus it will be important to integrate social and ethical studies into nanotechnology developments from their very beginning.

Social scientific and economic research can help manufacturers and governments make the right decisions when deploying a new technology, maximizing its benefit for human beings. In addition, technically competent research on the societal implications of nanotechnology will help give policymakers and the general public a realistic picture free of unreasonable hopes or fears. The costs of premature or excessive regulation would be extremely high, harming the very people it was intended to protect, and failure to develop beneficial nanotechnology applications would be unethical.

The significance of nanotechnology depends largely on how its development relates to wider trends going on in the world such as the impending population declines of most advanced industrial nations, the apparent diminishing returns to increased medical research and health care investment, and the threatened deceleration of progress in microelectronics. Well established social-scientific explanations for unethical behavior – such as learning, strain, control, and subculture theories – could help us understand possible future cases in nanotechnology industries. Ethics and social implications are largely matters of social perception, and the public conception of nanotechnology is still in the early stages of developing. Social science can now begin to examine its unfolding impacts in all sectors of the economy, in most spheres of life, and both short-term and long-term time scales.

38.1 **Applications and Societal Impacts** 1136
38.2 **Technological Convergence** 1139
38.3 **Major Socio-technical Trends** 1141
38.4 **Sources of Ethical Behavior** 1143
38.5 **Public Opinion** 1145
38.6 **A Research Agenda** 1148
References 1149

The views expressed in this essay do not necessarily represent the views of the National Science Foundation or the United States.

Science and engineering have only begun to explore the potential for discovery and creativity at the nanoscale, but already some intemperate voices call for government regulation or outright banning of nanotechnology [38.1–4]. Such action would be extremely premature, because we have just started on a very long road of research that we must traverse before we will know what is technically practical, what real-world applications nanotechnology might actually have, and what the appropriate societal responses would be to any hazards that might be directly or indirectly related to these applications. Some writers in the popular press seem to treat nanotechnology as

Springer Handbook of Nanotechnology
B. Bhushan • © Springer 2004

a new kind of Frankenstein's Monster, but it is worth remembering that biological science is still not able to create such a monster, nearly two centuries after Mary Shelly imagined one.

A good starting point for consideration of social and ethical issues is a formal definition of the topic, but unfortunately a somewhat different definition has lodged in the public mind from that employed by professionals working in the field. For example, *Merriam-Webster's Collegiate Dictionary* defines *nanotechnology* as: "the art of manipulating materials on an atomic or molecular scale especially to build microscopic devices (as robots)." Similarly, the online reference About.com defines it thus: "The development and use of devices that have a size of only a few nanometres." Later, we will examine public conceptions more closely, but it is worth noting here that these definitions focus on robot-like devices constructed at the nanoscale and ignore the topics of most chapters in this volume.

A more authoritative if long-winded definition of *nanotechnology* has been provided by the U.S. Government's National Nanotechnology Initiative: "Research and technology development at the atomic, molecular or macromolecular levels, in the length scale of approximately 1–100 nm range, to provide a fundamental understanding of phenomena and materials at the nanoscale and to create and use structures, devices and systems that have novel properties and functions because of their small and/or intermediate size. The novel and differentiating properties and functions are developed at a critical length scale of matter typically under 100 nm. Nanotechnology research and development includes manipulation under control of the nanoscale structures and their integration into larger material components, systems and architectures."

While this definition includes the word "devices," it also covers materials, structures, and systems. Significantly, it envisions that many nanotechnology advances will be integrated into "larger material components, systems and architectures," which means that we would need to understand the social and ethical implications of the larger-scale systems of which nanoscale structures are only components.

There is good reason to believe that engineering at the nanoscale will have very broad applications across all fields of technology and most fields of science [38.5–7]. Therefore, it will be an enabler or amplifier of the effects of other technologies. This means that it is foolish to ask, "Is nanotechnology harmful?" Nanotechnology is not one thing, either ethical or unethical. Rather, it is a myriad of different things – many nanotechnologies – each of which has a range of potential uses and misuses in conjunction with other technologies and applied to different goals.

This line of reasoning could lead some people to conclude that we did not need to worry at all about the social and ethical implications of nanotechnology, because they will be subsumed by the implications of larger systems. Others might conclude that it is far too early in the history of nanoscale engineering to consider societal implications, because we cannot study the impact of a technology that does not yet exist.

However, the very fact that some writers are urging regulation or banning of nanotechnology suggests that the debate about its value to humanity has begun, and we need to bring the best possible thinking and information to that debate [38.8–10]. In addition, a few relatively simple but significant applications have begun to appear, so in those cases at least an impact can be observed. Social scientists, natural scientists, philosophers, and engineers have indeed begun considering the future of nanotechnology, so we can draw upon the very first serious intellectual efforts in this important field.

38.1 Applications and Societal Impacts

In the year 2000, the National Science and Technology Council sponsored a major workshop at the National Science Foundation, which led to a published report, *Societal Implications of Nanoscience and Nanotechnology* [38.11]. Sixty-four representatives of academic science, government laboratories, and corporate research listed a large number of potential nanotechnology applications, a few of which had already entered production, but many of which could be expected only after decades of research. Following are diverse examples of possible applications suggested by workshop participants concerning a variety of sectors of the economy and spheres of life:

- More efficient components for the semiconductor industry such as integrated circuits containing transistors constructed from carbon nanotubes.

- Nanostructured catalysts for the chemical industry and for more effective converters to handle pollution from automobile exhausts and other combustion.
- Lighter and stronger nanomaterials in bulk quantities to enable safer and more efficient transportation vehicles, including automobiles, aircraft, and train systems.
- Improved pharmaceuticals with features such as programmed delivery to desired targets like tumors and the ability to employ substances that are not soluble in water as medications.
- Cost-effective and reliable filters for water decontamination, desalination, containment of industrial pollutants, and air purification.
- More efficient solar energy conversion, thereby reducing current reliance on oil and offering an alternative to nuclear power for future electricity needs.
- Efficient fuel cells and hydrogen storage systems, leading to nonpolluting automobiles, trucks, and buses.
- More durable composite materials, such as nanoparticle-reinforced polymers, designed for optimal performance in specific uses with reduced waste and greatly increased lifetime.
- Molecularly engineered biodegradable fertilizers and insecticides designed for precision agriculture with efficient delivery to where they are needed and prevention of unwanted side effects.
- Revolutionary launch vehicles with high mass-ratios, nonpolluting but powerful engines, and low electric power needs in order to finally realize the promise of space exploration.
- Nanoscale components in sensor systems that can quickly detect and identify pollutants and disease organisms, as well as chemical or biological warfare agents to allow quick and appropriate medical treatment or security responses.
- Coatings that would give ordinary materials extraordinary properties like self-cleaning window glass and heat-reflecting exterior architectural surfaces.

Knowledgeable scientists and engineers reported that these and many other similar applications were plausible benefits of nanotechnology research, but they recognized that progress has costs. For example, the creation of a new industry often results in the decline or complete destruction of an older industry, as in the proverbial case of buggy-whip manufacturers who were rendered obsolete by the automobile. The potential benefit from a new technology must be calculated in a sophisticated manner, taking into account the full life-cycle costs [38.12]. For example, an electric car may reduce some forms of pollution while increasing others such as the release of lead or other dangerous substances into the environment during the manufacture of the electric battery and the disposal of it when the car is eventually junked.

Many ethical issues of fairness will arise within nanotechnology industries themselves. For example, it will be important to invest in appropriate safeguards for workers engaged in hazardous production processes. In general, scientific discoveries cannot be patented, whereas new products and production methods can, so there will be serious issues of intellectual property protection in nanoscience, a nontraditional field where the legal line between discovery and invention has not yet been clearly drawn. Often, it will be necessary to bring together information and expertise from several different sources in order to create an economically profitable application, drawing upon government-funded university research projects, as well as several specialized industrial corporations, which will require improved models of collaboration across organizations.

As the *Societal Implications* report observed, the general public has a significant stake in the National Nanotechnology Initiative (NNI) that must be managed both directly through public participation and indirectly through the involvement of social scientists:

> It is important to include a wide range of interests, values, and perspectives in the overall decision process that charts the future development of nanotechnology. Involvement of members of the public or their representatives has the added benefit of respecting their interests and enlisting their support. ... The inclusion of social scientists and humanistic scholars, such as philosophers of ethics, in the social process of setting visions for nanotechnology is an important step for the NNI. As scientists or dedicated scholars in their own right, they can respect the professional integrity of nanoscientists and nanotechnologists, while contributing a fresh perspective. Given appropriate support, they could inform themselves deeply enough about a particular nanotechnology to have a well-grounded evaluation. At the same time, they are professionally trained representatives of the public interest and capable of functioning as communicators between nanotechnologists and the public or government officials. Their input may help maximize the societal benefits of the technology while reducing the possibility of debilitating public controversies [38.11].

Social scientific and economic research can help manufacturers and the government make the right decisions when deploying a new technology, maximizing its benefit for human beings. In addition, technically competent research on the societal implications of nanotechnology will help give policymakers and the general public a realistic picture free from unreasonable hopes or fears. As the report notes, the National Nanotechnology Initiative is just now commencing, so "there is a rare opportunity to integrate the societal studies and dialogues from the very beginning and to include societal studies as a core part of the NNI investment strategy" [38.11].

The workshop developed a number of recommendations for government, academia, and industry suggesting how the social and ethical implications of nanotechnology could best be addressed [38.11]. Here are some of the most important principles in education, social science, and ethics:

- Nanoscale concepts should be introduced into science and engineering education at all levels, thereby giving the widest possible range of students a fundamental understanding of the field while intellectually linking nanotechnology to many other fields.
- The training of nanotechnologists should include societal implications and ethical sensitivity, so their future work will be guided by principles that will maximize human benefit.
- A sufficient number and variety of social and economic scientists should receive effective multidisciplinary training, so they will be well prepared to work in the nanotechnology area.
- Formal measurement methods, such as social and economic indicators, should be developed and consistently employed to chart the actual widespread changes in industry, education, and public welfare as they occur.
- Government agencies, private foundations, and industry should support a wide range of high-quality, theory-based social and economic research studies on nanotechnology, examining both the decision processes that shape the emerging technology and itsspecific societal impacts once it is deployed.
- A knowledge base and institutional infrastructure must be created to evaluate the probable future intellectual and societal impacts of nanoscience and nanotechnology over the short, medium, and long term.
- Government and the private sector should establish effective channels for informing the public promptly about new concepts, projects, potential applications, and ethical issues as they emerge.
- There should be formal organizations and mechanisms to ensure participation by diverse societal institutions and the general public in setting priorities for research and development and in providing timely input to decision makers.
- Government and industry should develop management plans and policies that can effectively incorporate all relevant information and legitimate stakeholder interests to ensure that we can respond flexibly to social, ethical, legal, and economic implications as they appear.

It will take time and effort to implement these recommendations, but some progress has already been made. The National Science Foundation has recently funded Rosalyn Berne at the University of Virginia to study the developing ethics of nanotechnology through a narrative analysis of discourse from engineers and scientists in the field, and a team headed by Michael Gorman at the same institution was funded to explore the potential for multidisciplinary research on social and ethical dimensions of nanotechnology. The Environmental Protection Agency, which has supported several research projects on how nanotechnology might remedy conventional forms of pollution, has funded Darrell Velegol and Kristen Fichthorn at Pennsylvania State University to study "Green Engineering of Dispersed Nanoparticles," exploring how valuable nanoparticles might be produced without the use of polluting additives. In its review of the National Nanotechnology Initiative, the National Research Council has recommended the development of "a new funding strategy to ensure that the societal implications of nanoscale science and technology become an integral and vital component of the NNI" [38.13].

38.2 Technological Convergence

In the year following the *Societal Implications* workshop, NSF and the U.S. Department of Commerce staged an even larger meeting to examine the potential for *nanotechnology* to combine with three other powerful scientific and engineering trends: *biotechnology*, *information technology*, and *cognitive science* – four "NBIC" fields. The resulting report, *Converging Technologies for Improving Human Performance* [38.14], examined benefits of technological convergence that would specifically enhance the ability of humans to satisfy their needs and achieve their legitimate goals.

Convergence of diverse NBIC technologies will be based on the material unity of nature at the nanoscale and on technology integration from that scale. This is the scale at which much future biotechnology and the hardware for information technology will operate, and it is the scale from which fundamental research on the human nervous system will take place. Key transforming tools, including research instruments and production methods, will arise in the conjunction of previously separated fields of science. Drawing upon advances in mathematics and computation, it will be possible to model phenomena from the nanoscale to the cosmic scale as complex, coupled, hierarchical systems.

The *Converging Technologies* scientists, engineers, and policymakers sketched six general areas in which the human benefits and social issues are likely to result:

- Overall potential of converging technologies
- Expanding human cognition and communication
- Improving human health and physical capabilities
- Enhancing group and societal outcomes
- National security
- Unifying science and education

Numerous valuable applications were sketched in greater or lesser detail that could reasonably be expected to result from a decade or two of vigorous research and development, including the following ideas that depend on partnerships between nanotechnology and one or more of the other NBIC fields:

- Comfortable, wearable sensors and computers will enhance every person's awareness of his or her health condition, the environment, chemical pollutants, potential hazards, and information of interest about local businesses, natural resources, and the like.
- The human body will be more durable, healthier, more energetic, easier to repair, and more resistant to many kinds of stress, biological threats, and aging processes.
- Machines and structures of all kinds, from homes to aircraft, will be constructed of materials that have exactly the desired properties, including the ability to adapt to changing situations, high energy efficiency, and environmental friendliness.
- A combination of technologies and treatments will compensate for many physical and mental disabilities and will eradicate altogether some handicaps that have plagued the lives of millions of people.
- National security will be greatly strengthened by lightweight, information-rich war fighting systems, capable uninhabited combat vehicles, adaptable smart materials, invulnerable data networks, superior intelligence-gathering systems, and effective measures against biological, chemical, radiological, and nuclear attacks.
- The ability to control the genetics of humans, animals, and agricultural plants will greatly benefit human welfare; widespread consensus about ethical, legal, and moral issues will be built in the process.
- Agriculture and the food industry will greatly increase yields and reduce spoilage through networks of cheap, smart sensors that constantly monitor the condition and needs of plants, animals, and farm products.
- Transportation will be safe, cheap, and fast due to ubiquitous real-time information systems, extremely high-efficiency vehicle designs, and the use of synthetic materials and machines fabricated from the nanoscale for optimum performance.
- The work of scientists will be revolutionized by importing approaches pioneered in other sciences, for example, genetic research employing principles from natural language processing and cultural research employing principles from genetics.
- Formal education will be transformed by a unified but diverse curriculum based on a comprehensive, hierarchical intellectual paradigm for understanding the architecture of the physical world from the nanoscale through the cosmic scale.

Several of these application areas depend upon the development of nanoscale components for sensor, computation, and communication networks. Others involve a synergy between inorganic nanoscale engineering and organic chemistry or biology. Others extrapolate that it will be possible to build large-scale structures that

are engineered at the nanoscale to enhance the qualities of architecture, transportation vehicles, or even the human body. Fundamental to all of the application areas is a conceptual revolution that unifies science and provides humanity with a comprehensive view of nature and technology.

Contributors to the report showed how nanoscale science and engineering were essential for progress in biology, information science, and cognitive science. The result of NBIC convergence will be the strengthening of science and the empowerment of humanity, as we will find solutions to many of humanity's problems in health, natural resources, environment, security, and even communication with fellow human beings. Often, the technologies based in the sister sciences will raise ethical issues. In such cases, nanotechnology may become implicated in controversies without being central to them. The correct policy response would then be to ignore the superficial nanotechnology aspects and focus directly on the core problem that lies in a different field. *Converging Technologies* contributors examined how technological convergence highlighted such existing issues as the treatment of the disabled, communication breakdowns, economic stagnation, and threats to national security.

For greater clarity, we can consider here the very specific example of the proliferation of nuclear weapons. Nuclear physics and atomic technology were specifically not included in the constellation of NBIC converging technologies considered by the workshop, so the following discussion supplements its findings. As scientists already understood in the early 1940s, there are fundamentally two ways that a rogue nation can build up its own supply of weapons-grade nuclear materials. One is breeding plutonium in a nuclear reactor fueled by uranium, transmuting the common uranium isotope of atomic weight 238. But this is hard to conceal from other nations, especially if reactors are open to international inspection. The other way is separation of the relatively rare uranium isotope ^{235}U, which, like plutonium, is suitable for use in weapons. The original separation method was gas diffusion, but then the centrifuge method was developed, and other approaches are possible such as laser or electromagnetic separation.

In the gas diffusion method, a chemical containing uranium (uranium hexaflouride) is passed through a porous barrier such as a ceramic wall that has tiny holes in it. Molecules containing ^{235}U travel through the barrier slightly faster than molecules containing ^{238}U. This is a very inefficient process, because the isotopes are very similar to each other, so many passes through a barrier are required to enrich the proportion of ^{235}U sufficiently for use in a nuclear weapon. The other approach uses centrifuges, rapidly spinning rotors containing the uranium hexaflouride to separate the two isotopes on the basis of their slight difference in mass. Both approaches are very difficult and costly, however. Conceivably, nanotechnology could slightly reduce the difficulty, by producing porous barriers with exactly the optimum sized holes and by providing lighter and stronger centrifuge rotors that could operate at higher speeds.

Therefore, without a more careful analysis someone might leap to the conclusion that nanotechnology is a danger because it facilitates proliferation of nuclear weapons. This ignores several crucial facts. First, the world already faces a very severe crisis of nuclear proliferation without nanotechnology. Second, the fundamental requirement for production of nuclear weapons is the availability of uranium, so the most logical preventive would be international monitoring and control over its mining and distribution. Third, there are many other technical requirements for production and delivery of nuclear weapons, and it is the whole collection of them that presents a threat, not any one alone. Fourth, nanotechnology is inherently nonnuclear, because nuclear reactions occur at a smaller scale than the nanoscale, so its connections to proliferation of nuclear weapons will always be highly indirect if they exist at all. Finally, the fundamental problem is not nuclear proliferation, but weapons of mass destruction in general, and nanotechnology has a major role to play in the defense against chemical and biological warfare agents in sensor systems that can detect them before they have caused harm.

Perhaps a very few nano products will need to be "controlled substances," analogous to narcotics, insecticides, asbestos, and antibiotics [38.15]. Conceivably, porous barriers with exactly the right pore size for uranium isotope separation would be controlled in some way. Some products could be produced and used under strict regulations, whereas a few others might be banned altogether. Most would not require restrictions of any kind.

Such decisions would need to be based on the careful analysis of facts related to the specific issue at hand. For example, it may be that no improvements in gas diffusion separation are possible through nanotechnology, or they are so slight as to be negligible. Perhaps technological developments in laser or electromagnetic separation, completely unrelated to nanotechnology, may render

both the gas diffusion and centrifuge methods obsolete. Therefore, the mere fact that we can imagine a way in which nanotechnology might possibly encourage the proliferation of nuclear weapons is far too weak a justification for international regulation. We would need to go beyond vague fears to precise analysis based on rigorous scientific research to distinguish real dangers from imaginary ones and design appropriate safeguards in those rare cases when they are necessary.

The costs of excessive regulation would be extremely high, harming the very human beings it was intended to protect. Porous barriers precisely designed at the nanoscale are among the high benefit applications we can anticipate early in the history of nanotechnology. They would have a wide range of applications, from water purification and desalination to separation of valuable substances in biotechnology production processes. Similarly, new high-performance materials with great strength but low weight will have a myriad of applications throughout the civilian sectors of the economy. Among the products that would benefit millions of ordinary people are light but indestructible eyeglass frames, safety helmets, and hand tools.

Nanotechnology has the potential to increase the effectiveness of biotechnology, information technology, and technologies based on cognitive science through NBIC convergence. Therefore, it could become entangled in ethical issues that already involve those other technologies such as the conflict over genetically modified foods, the dispute over information privacy, and the debate over treatment of mental illness. These are not fundamentally nanotechnology issues, however, and it will be important to keep a proper perspective on their real nature.

38.3 Major Socio-technical Trends

The significance of nanotechnology depends not only on its own accomplishments and on convergence with other technologies, but also on how these purely technical developments relate to wider trends going on in the world [38.16, 17]. At present, many of these trends are confused and are the subject of great controversy in the social science and public policy arenas. It is always hazardous to extrapolate trends from any particular moment in history, but the present time appears so chaotic that the events of a single day might send humanity down a very different road.

A few years ago, there was a broad consensus that the rapid growth in the number of people on the planet, the so-called *population explosion*, was the dominant quantitative trend that needed to be factored into any projections [38.18]. But by the mid-1980s, it became obvious that population growth had been replaced by the threat of population decline in most advanced industrial nations, with the possible exception of the United States [38.19, 20]. Even before collapsing fertility begins to reduce the number of people in the affected nations, their average age and pension costs increase, social dynamism and creativity decrease, and governments become gridlocked trying to deal with ultimately insoluble problems of funding demanded services. This is one of the explicit reasons why the United States has begun to discount the influence of its European allies and Japan [38.21, 22].

Although nanotechnology has little obvious direct connection with fertility, there are two potentially counterbalancing, indirect connections. Rapid progress in nanotechnology could revitalize economic expansion and technical creativity, thereby to some extent offsetting the negative consequences of fertility collapse. However, the aversion to change that marks societies having a high proportion of elderly citizens may dampen the intellectual inquisitiveness and investment risk-taking required to develop nanotechnology and technological convergence.

A related trend is the constant improvement in health resulting from progress in medical science and the economic growth needed to fund increasingly costly health care. Continued progress in health is by no means assured, however. During the nineteenth century, American health may actually have declined significantly for a number of decades, despite economic growth [38.23]. We tend to attribute the undeniable improvement in health over the twentieth century to medical progress, but health education and public sanitation may have been more important [38.24]. The U.S. Centers for Disease Control argue that substantial improvements in health and longevity could be achieved by lifestyle changes, notably, exercising more, reducing the fats in our diet, and avoiding smoking [38.25,26]. The introduction of antibiotics helped increase longevity, and modern cardiology saves the lives of many people who might otherwise die prematurely of heart attacks, but the progress against

cancer, AIDS, and most aging-related illnesses has been agonizingly slow. On balance, economists find that the increasing investments in heath care are paying off, but not necessarily in all areas, or with very great benefit [38.27–29].

There are several reasons to be pessimistic about the future benefits of medicine, unless there are very major scientific breakthroughs. We have yet to find a cure for any chronic viral infections, of which AIDS is merely the most publicized example. Bacteria are rapidly evolving resistance to antibiotics, yet the rate of development of new prescription medicines is declining.

The average white American male born in 1900 could expect to live 48 years, but in 2000 this life expectancy had increased to 74 years. For white females, the average life expectancy increased from 51 to 80. Projecting these figures forward at the same rate of increase suggests that life expectancy in 2100 might be 114 for males and 125 for females. However, using more realistic assumptions, the U.S. Census Bureau has projected that life expectancy for Americans born in the year 2100 might be only around 88 for males and 92 for females [38.30]. This does represent progress, but at a steeply declining rate, and it depends upon continued economic growth.

Perhaps the most economically significant technical trend of recent years has been *Moore's Law*, Gordon Moore's observation that the density of transistors on the most advanced microchip doubles about every 18 months. The rapid improvement of computer and communications hardware has been fundamental to the implementation of new software applications, networking, and information technology in general. These, in turn, have been responsible for much economic growth. However, the semiconductor industry is approaching physical limits in the traditional methods of making integrated circuits, including computer chips, so Moore's Law will stall without breakthroughs in nanoscale technologies such as molecular logic gates and carbon nanotube transistors [38.31–33].

Arguably, several of the great thrusts of twentieth-century technology have already stalled. Except for the improved capacity of scientific instruments to collect and process astronomical data, the space program has hardly advanced since the mid-1970s. Civilian aviation and ground transportation have improved only in relatively minor design details since the 1950s. Despite great efforts, little progress has been made in controlling nuclear fusion for power generation. Nuclear fission power generation has largely been halted by public controversies, and no technological breakthroughs in safe operation of reactors or disposal of radioactive wastes have assuaged the public's concerns. The social sciences seem no better able to define good public policies than they were decades ago. Outside of nanotechnology itself, the two main areas where rapid progress obviously continues are information technology and genetic engineering. Convergence with nanotechnology is essential for further progress in both of these areas.

What would happen to the world economy if technological progress slowed or even halted? First, the advanced industrial nations would lose the advantage they currently hold over developing countries. For example, the American semiconductor industry would probably collapse if it lost its technological superiority, because nations with lower wage rates could produce the same things more cheaply. Technological products, from computer chips to pharmaceuticals, would become commodities, manufactured wherever it was cheapest to do so. With their heavy pension burdens and costly social services, governments of advanced nations would face fiscal crises of unprecedented magnitude. Indeed, the economies of these nations could fall rapidly down to the average of all nations, unleashing social discontent that could lead to unpredictable violence.

At the same time, some developing nations would advance economically, which would mean increased net world industrial production with concomitant increase in environmental pollution and resource depletion. Without technological progress to solve these problems, the average human welfare could decline, rather than remaining stable. In order to improve human welfare, technological progress must continue. Perhaps the best way to achieve that is through nanoscience and nanotechnology in convergence with the other NBIC fields, gaining understanding and control at the physical scale that is the basis for most science and technology.

It is impossible to predict how rapidly the conditions in the richest nations would deteriorate, or whether humanity could navigate through these crises to achieve a stable world. But clearly a halt in technological progress would be a shock to the world economy, global political institutions, and human welfare – which it would be wise to avoid.

38.4 Sources of Ethical Behavior

Ethical questions related to nanotechnology are not limited to the ways people might use it to harm others intentionally, but also include obligations to avoid potentially harmful unintended consequences [38.34]. An example of the former might be a weapon and of the latter, a kind of environmental pollution. In either case, the harm might be morally justified, as when a nation employs a weapon to defend against attack, or when limited pollution of the environment is offset by substantial benefits to humanity. Other ethical questions concern harm to the owners of nanotechnology, for example, if a competing company violates a nanotech patent, or if a government bans a nano-related product without careful examination of scientific evidence about its value. News reporters or popular writers who spread false information about nanotechnology may also be acting unethically, whether they are willfully lying or merely failing to be professionally diligent in checking their facts.

It is important to recognize that there are sins of omission as well as commission. To fail to develop a beneficial nanotechnology application could also be unethical. Imagine, for instance, that a laboratory had discovered how nanotechnology could enable a much more effective form of chemotherapy to cure cancer reliably and cheaply, but the company owning the patents prevented it from being developed, because its less effective traditional chemotherapy products were more profitable. This could be considered just as unethical as doing aggressive harm.

For over a century, social scientists have been studying the origins of ethical behavior – or its opposite, unethical behavior – in fields like the social psychology of groups, the sociology of deviance, and criminology. Although it is impossible to predict when a particular person will commit a specific ethical or unethical act, we can identify with good confidence the general factors that steer human behavior in one direction or another. An individual's behavior is determined by a complex of factors, few of which can be measured accurately, and chance seems to play a significant role in human affairs. Thus, the findings of social science tend to be statistical or conditional in nature. This does not render them useless, because one can develop rational laws, government policies, and investment strategies based on probabilities rather than certainties.

The theoretical and empirical scientific literature on the topic is vast, including literally thousands of publications dating back to the 1840s, expressing the views of countless schools of thought, often employing idiosyncratic terminology. For present purposes, we must distill that huge and incoherent intellectual heritage down to its most meaningful essentials. A good start is the following set of four themes, each of which can be expressed as a theory about why people might violate an ethical rule:

1. Learning – A person violates a rule because the person has learned it is rewarding to do so.
2. Strain – A person violates a rule because the person is frustrated in the attempt to fulfill one of the expectations of society.
3. Control – A person violates a rule because the person lacks stable social bonds that might prevent it.
4. Subculture – A person violates a rule because the person belongs to a group or network of other people who encourage violation.

When each of these ideas first appeared in the social-scientific literature, the theory's proponents tended to present it as the complete explanation for deviant behavior, but now we understand that each of these four has merit and the correct explanation is usually a mixture of two or more. We will illustrate each of the theories with hypothetical illustrations, because actual examples may not exist yet.

Learning Theory is sometimes interpreted in economic or "rational choice" terms, as meaning that a person will do whatever he or she perceives it is profitable to do. But human life is not spent entirely in a marketplace where one can compare prices and shop for a good deal. Rather, there are many quite different contexts in which people make decisions. Regardless of whether options can be quantified in dollars, people tend to rely upon their own past experiences in selecting between familiar courses of action [38.35–37].

If the managers of an industrial corporation have learned they can achieve higher profits by ignoring regulations against polluting the environment, and they have not suffered heavy fines from the government for doing so, they will be more likely to pollute in the future. Therefore, a corporation guilty of polluting with conventional technologies is apt to continue with this history of pollution when it adopts new nanotechnologies. This suggests that one way to reduce the likelihood that industrial companies will pollute when they adopt new nanotechnologies is to enact and enforce strong regulations against pollution in general, to make them learn to avoid polluting.

In Strain Theory, a person violates a norm because he or she is frustrated in the attempt to achieve one of the values endorsed by society [38.38, 39]. This idea has been used to explain why many members of poor communities or groups that have been subject to discrimination turn to crime to gain wealth and status, but it can also explain why a major corporation might begin to engage in dubious accounting practices when its growth in real business stalls. The central idea is that an individual or group becomes committed to a particular kind of success that society encourages – whether it is wealth, political power, widespread fame, or honor within a particular sector of society – then experiences frustration in attaining success, which motivates ethical violations.

Conceivably, a nanotechnology corporation could be affected by strain if it has committed itself to a new and untested line of products that turns out to be less effective, more harmful, or more costly than originally estimated. The corporation has committed not only its money and effort, but also its public prestige to success, but success may elude it. It may then unethically advertise its products as more effective than they really are, conceal knowledge of their danger to human health or the environment, and cook its financial records to deceive investors. Of course, many ordinary companies have done this in the past, and, ironically, they sometimes solve their problems before the public becomes aware of them, thereby becoming ethical and successful again (e.g., [38.40]).

When Merton proposed the most influential version of Strain Theory in 1938, he identified three modes of adaptation to three increasing levels of strain. If a person was unable to achieve society's values by following conventional norms, frustration might drive him or her to violate the norms in pursuit of the value, what Merton called *innovation*. If this did not work, he or she might substitute both new values and new norms for those of society, in what he called *rebellion*. But if this failed as well, the person might abandon values, as well as norms and fall into a demoralized state of inaction called *retreatism*. Retreatism is inconsequential, and successful rebellion tends to radicalize society, but innovation (in Merton's sense of the term) can promote scientific and technical progress. In other work, Merton [38.41, 42] was one of the founders of the sociology of science and technology, but he never followed through on the implications of Strain Theory for science, and his distinctive concept of innovation was chiefly used to explain criminal behavior by poor people who violated the law in order to the get the values (i. e., money) endorsed by capitalist society.

There are many examples in the history of science in which Strain Theory can be used to explain scientific and technological innovations that turned out to be beneficial, rather than unethical, and others where the result was a mixture of benefit and harm that presumably could have been managed to emphasize the benefit. In the seventeenth century, England led the world in scientific progress, perhaps in a nationalistic reaction to its inferiority in population and other resources relative to France and Spain [38.43]. In the late nineteenth and early twentieth centuries, Germany lacked access to as abundant a supply of natural resources as had the great colonial powers, Britain and France. This inferiority may have stimulated many advances in German chemistry such as Adolf Bayer's synthesis of indigo dye and the nitrogen fixation method developed by Fritz Haber and Karl Bosch, which allowed Germany to continue producing explosives for World War I after it had lost access to foreign sources of nitrates. When the Treaty of Versailles concluding World War I limited German long-range artillery, the Germans developed the V-2 rocket largely to get around this restriction, thereby ushering in the age of space flight [38.44, 45].

In Control Theory, a person violates norms because he or she lacks stable social bonds that might prevent it [38.46–48]. This is a very successful scientific explanation of many common crimes, such as ordinary theft, and many kinds of criminals tend to lack stable friendships or family relationships that might restrain their deviant behavior. Control Theory might seem irrelevant to nanotechnology ethics because no isolated individual is in a position to make major decisions about new applications. The theory perfectly fits the Hollywood image of Dr. Frankenstein living like a hermit in a remote castle, where he is free to conduct any ghastly experiment he wishes, but no real scientist lives like this.

However, a variant of Control Theory explains that a community or organization can suffer from a partial but pervasive form of isolation, called *social disorganization*, in which communication channels are poor and many people deviate to a moderate degree because they are somewhat detached from each other. In a socially disorganized research laboratory studying the properties of a new nanostructured material, the quality of the science could suffer, pieces of information could fail to get to scientists who need them, and management would get incorrect impressions about the material's properties. This could lead to a chain reaction of unethical behav-

ior in which no one individual was primarily to blame. For instance, the result could be a corporate decision to put into production a new kind of aircraft component that was subject to unpredictable catastrophic failure, causing crashes and loss of life. In a case such as this, the unethical behavior is scattered in many small pieces, committed by many individuals, any time one of them presents a dubious observation as a scientific fact. The best way to reassert the truth-oriented professional norms of science would be to rebuild good channels of communication and cooperation, reattaching the researchers to each other and to the scientific community.

Subculture Theory is similar to Control Theory, but operates on the level of groups rather than individuals. A person violates a rule because he or she belongs to a subculture that rejects the rule and even encourages violation of it [38.49, 50]. A subculture is a group or network having a set of norms, roles, and values that differ from those of the surrounding culture. At the extreme, it can have a radical ideology or dogma, but practically every social group is a subculture to at least a very mild extent. Social psychologist *Janis* [38.51] has shown that even very powerful groups can become deviant subcultures, if they isolate themselves from the wider community, as the Nixon White House did during the Watergate episode.

Even if a research laboratory is functioning properly, the organization of which it is a part can make harmful decisions if the top management is a subculture practicing what Janis calls "groupthink," the refusal to accept information that conflicts with the group's beliefs and commitments. Sometimes this may result in harm to others, for instance, if a new nanotechnology-enabled treatment for cancer were put into production before safe methods for administering it to patients in hospital settings were worked out. Or the harm could fall on the organization itself, for example, if a semiconductor company decided to invest its future in molecular logic gates without fully assessing the potential of carbon nanotube transistors instead. In that case, we might say the company deserves what it gets, but there is the ethical issue of its responsibility to stockholders and employees, including to the scientists in the research laboratory who were ready to do the studies needed for a correct decision.

As the example of Strain Theory illustrates, sometimes strain has positive consequences, rather than negative ones. Similarly, Learning Theory explains why an individual or organization might develop habits of ethical innovation, and both Control Theory and Subculture Theory could show that a person or group may innovate best when somewhat detached from conventional thinking. Thus, the issue is not one of avoiding these four factors, but of finding the right levels of them to encourage healthy innovation without unleashing serious unethical behavior. This is the fundamental trade-off in any free society: How to liberate human creativity while restraining greed, immorality, and inhumanity.

Strain Theory has been used to explain the emergence of collective behavior and social movements [38.52], including reform movements such as the environmental movement or crusades to establish professional codes of ethics. Concern about the social and ethical implications of nanotechnology can, therefore, energize the development of appropriate institutions to manage it wisely. Learning Theory suggests that a rewarding market for beneficial nanotechnology products could teach scientists, engineers, and managers to invest their energy in projects that would serve the needs of humanity. Both Control Theory and Subculture Theory stress the importance of strong networks of communication in restraining unethical behavior. Convergence of nanotechnology with the other NBIC fields would support healthy communication by unifying science, providing a universal technical language for interaction across fields, and creating a network of collaborative relationships across sciences, institutions, and sectors of society.

38.5 Public Opinion

Ethics and social implications are largely matters of social perception, and the public conception of nanotechnology is still in the early stages of developing. Nanoscale science and engineering are evolving from conventional work in materials science, physics, chemistry, and related fields, as the gradual emergence of new methodologies permits many kinds of observation and experimentation with inorganic structures having at least one dimension less than about 100 nm. This progress in *real nanotechnology* is solidly based in existing scientific knowledge, even as it achieves new discoveries and inventions.

At the same time, something quite different has been emerging that also calls itself nanotechnology: a social movement based on metaphors, approximations, and hopes. *Tolles* [38.53] calls it an "irrational vision." In social science, popular enthusiasm for a loosely defined set of unreasonable hopes is often called a *mania* [38.54], so I will call this movement *nanomania*.

The key moment in the emergence of nanomania, but not of nanoscience and nanotechnology themselves, was the 1986 publication of *Engines of Creation* by Drexler. This is a popular book, inspirational rather than technical, filled with verbal metaphors about what might possibly be accomplished if nanoscale assemblers could be created, rather like industrial robot arms and assembly lines. The plausibility of Drexler's argument, such as it is, comes from his constant reminder that biological systems include "machines" that assemble themselves and manufacture or manipulate other nanoscale structures: proteins, RNA, and DNA.

In a rambling discourse that reminds us that Leonardo da Vinci could not have predicted the details of late twentieth century technology, Drexler asserts that some time in the future it will be possible to build nanoscale assemblers that are able to make *anything that can be designed*. At this point he has outstripped his biological analogy, because the "nanomachines" of biology cannot do this; they can make only those specific things that evolution has programmed them to make [38.55]. Drexler says that suitably programmed assemblers will be able to make copies of themselves, as living organisms do.

The specific image he offers is of tiny robot arms grabbing atoms from the surrounding raw material and putting them together in the correct configurations. These imaginary assemblers can make anything, he suggests, so they can quickly replace all the factories and construction companies on Earth, giving humankind perfect abundance of anything we might want. The book then scans quickly through other dreams of science fiction, including artificial intelligence, cheap space flight, near immortality, and the danger to human life of uncontrolled self-replicating intelligent machines.

In some ways the book is laudable, inspiring nontechnical readers with many conceivable applications of nanotechnology, some of which might actually turn out to be practical in the long run, discussing their social implications, and offering potential solutions to problems that might arise. The problem is that many people have come to the unfounded conclusion that Drexler's self-replicating universal assemblers are not only theoretically possible, but likely to be created in the near future. This is the fundamental error of nanomania, and it has called forth a reaction in the form of a *nanophobia* – the fear that all kinds of horrible evils will soon emerge from nanotechnology.

At the 2000 *Societal Implications* workshop, Drexler's assemblers were debunked by several participants, including Nobel prize-winner Smalley, discoverer of fullerenes and a pioneer of carbon nanotubes. *Smalley* observes that technical details of chemistry and a correct understanding of the properties of atoms and molecules are largely absent from nanomania [38.56]. In particular, the universal nanoscale assemblers envisioned by Drexler are physically impossible: "Because the fingers of a manipulator arm must themselves be made out of atoms, they have a certain irreducible size. There just isn't enough room in the nanometer-size reaction region to accommodate all the fingers of all the manipulators necessary to have complete control of the chemistry. . . . Manipulator fingers on the hypothetical self-replicating nanobot are not only too fat; they are also too sticky: the atoms of the manipulator hands will adhere to the atom that is being moved. So it will often be impossible to release this miniscule building block in precisely the right spot" [38.57].

The difficulty of developing reasonable public policies about nanotechnology when popular thinking is dominated by fanciful notions is illustrated by Chen's paper, "The Ethics of Nanotechnology", distributed online both by the Markkula Center for Applied Ethics at Santa Clara University and by BioScience Productions, Inc., which describes itself as "a non-partisan organization whose goal is the promotion of public literacy in the biosciences". Following Drexler's approach, *Chen* takes assumes that nanotechnology means the creation of molecular machines, "which could be used as molecular assemblers and disassemblers to build, repair, or tear down any physical or biological objects" [38.58]. Recognizing that such "nanites" may not be practical in the near future, Chen nonetheless suggests that we consider taking three possible actions soon:

1. Nanotechnology R&D should be banned
2. A nongovernmental regulatory or advisory commission should be set up
3. Adopt design guidelines
   - Nanomachines should only be specialized, not general purpose
   - Nanomachines should not be self-replicating

- Nanomachines should not be made to use an abundant natural compound as fuel
- Nanomachines should be tagged so they can be tracked

Chen notes that the first choice would undesirably prevent beneficial applications of molecular machines, but might not prevent "rogue researchers" from developing nanites anyway. However, readers of his article who are unfamiliar with real nanotechnology may not realize that most research in the field has nothing to do with these hypothetical nanites. It would take considerable wisdom to figure out how a regulatory commission might be set up, empowered, and managed effectively in such a way that it could ensure nanotechnology will be used for human benefit without stifling worthwhile innovations. The design guidelines flow from Chen's assumptions about what nanotechnology is, but any formal regulations embodying them could have serious negative unintended consequences, given that they start with a flawed definition of nanomachines.

Chen's rule that nanomachines be "specialized, not general purpose" may mean little, because a fully "general purpose" machine is as mythical as a "universal solvent" or "perpetual motion machine." However, a regulatory commission could do great harm by adopting poorly conceived definitions of *nanomachine* and *general purpose*. Suppose a very useful new abrasive were developed consisting of nanoparticles that changed shape, perhaps in response to temperature or force fluctuations, in some complex way that helped them do their intended job. *Webster's II New Riverside University Dictionary* defines *machine* thus: "a system, usually of rigid bodies, constructed and connected to change, transmit, and direct applied forces in a predetermined way to accomplish a particular objective, as performance." By this definition, the nanoscale abrasive particles would be machines. If they could be used to shape a wide range of industrial materials, they could be called general purpose. Yet this abrasive could be no more dangerous to humanity than ordinary sandpaper.

Strictly speaking, no entities on Earth are "self-replicating." Living organisms cannot replicate themselves without the help afforded them by the particular environment in which they dwell. All life is part of an ecology. Any regulation prohibiting self-replication would have to define very clearly what *self-replication* meant in the context of particular environments. Today, factory robots build parts of robots, but nobody worries about the ethical implications, perhaps because anybody can see that the manufacture process is entirely under the control of human beings. Because nanites are imaginary, and only vaguely defined, there are no facts to give nontechnical people comparable confidence.

Chen's rule against nanomachines that use "an abundant natural compound as fuel" presumably would prevent a nanite from rampaging uncontrolled in the natural environment. But this could outlaw the development of non-replicating nanoscale devices that can function only in very restricted laboratory or industrial systems, because most chemicals used in motors or batteries are abundant natural compounds.

Requiring that all nanoscale machines be "tagged so that they can be tracked" is reminiscent of the British Red Flag Act of 1865, which required that each horseless carriage be preceded by someone on foot waving a red flag. In the context of Chen's recommendations, *tracked* would seem to mean located, not merely identified once found. This could add an extremely heavy technical burden such as radioactive tagging, addition of a distinctive adulterating chemical, or even complex design features above the nanoscale to allow the nanomachine to be detected by microwave radio locators. If the machines are incapable of escaping into the natural environment, this burden would be not only ill-advised but ludicrous.

Has nanomania distorted the reception real nanotechnology is receiving? A survey done in 1999 found that the American business community was rather unaware of nanotechnology, only 2% saying they knew what it was and a further 2% reporting they had heard of it, but did not know what it was [38.59]. Two years later, a Web-based survey sponsored by the National Geographic Society found that science-attentive members of the general public are very enthusiastic about nanotechnology, and a rather large number of ideas about its benefits have already entered popular culture [38.60]. Statistical analysis of responses from 3,909 people revealed that they associate nanotechnology with other science-based technology, but do not connect it with pseudoscience. Significantly, when 598 people wrote paragraphs about nanotechnology, not a single respondent expressed concern about hypothetical dangers from self-reproducing nanites. Future surveys can examine the deepening popular awareness of nanotechnology, as the field itself progresses, and assess any public concerns that develop.

38.6 A Research Agenda

Sociologist *Etzkowitz* [38.61] has noted that the social sciences can play three different but mutually supportive roles in the development of nanotechnology:

1. Analyzing and contributing to the improvement of the processes of scientific discoveries that increasingly involve organizational issues where the social sciences have a long-term research and knowledge base.
2. Analyzing the effects of nanotechnology, whether positive or negative, expected or unintended, hypothetically and proactively and as they occur in real-time.
3. Evaluation of public and private programs to promote nanoscience and nanotechnology.

There are several ways social scientists can study the nanotechnology innovation process as it occurs [38.11, 62]. Cultural anthropologists and participant observation sociologists can enter nanoscience research teams, ethnographically documenting the behavior of the researchers over time as they frame their scientific problems, seek solutions, and labor collectively to understand the meaning of research results [38.63]. Fruitful sites for research on the innovation process include academe, industry, government laboratories, federal agencies, and professional societies.

Historical methods for collecting and analyzing written documents are also useful, even in cases where the documents are only days or weeks old, rather than the decades-old materials employed by most professional historians. Formal interviews, content analysis of communications such as e-mail messages, and questionnaires can provide more systematic kinds of data.

Application of a scientific idea to a technical problem, a technology transfer, and the introduction of products into the marketplace can be tracked through economic statistics on research and development investments, patent applications, and advertisements for new products and services. Some geographic areas, strata of society, and kinds of individuals and institutions experience technological change earlier than others, notably, the first adopters who try out an innovation before influencing other people to try it as well [38.64, 65]. Start-up companies are in a sense early adopters, even though they are also innovators, and thus they may be an especially fruitful if logistically difficult field for research.

There is a long tradition of systematic research on scientific publication, tracking the introduction of new concepts, charting the changing patterns of citation in scientific literature, and measuring the bulk of publications in different areas [38.66]. The progress of nanotechnology in modern culture could also be studied by tracking the introduction of new topics and courses in school and university curricula, mentions of nanoscale phenomena in the popular press, and the proliferation of both commercial and noncommercial Websites dedicated to nanotechnology. Professional societies and entrepreneurs are already creating series of new forums, symposia, journals, and job fairs oriented toward nanoscience and nanotechnology.

Nanotechnology may have significant impacts in all sectors of the economy, in most spheres of life, and on both short-term and long-term time scales. The intended consequences of any particular innovation can be determined simply by interviewing its investors or promoters and by inspecting their publications about it. Unintended consequences are much harder to study, because they require not only extensive research on the impacts on diverse areas of society, but also good scientific intuition about where to look for them. Consequences have consequences, in a continuing cascade, so the second-order consequences will also have to be studied [38.67]. Ultimately, this would require an extensive, vigorous interdisciplinary scientific community that is dedicated to research on social and technological change in general.

Technology is an elaborate system, embedded in the even more elaborate system that is global society. Through feedback among complex social subsystems, major phenomena can produce chaotic and unpredictable effects. Nanotechnology will have substantial impacts on many aspects of society in often different ways, so predictions of its influence will be difficult to make and evaluate scientifically. This means that researchers will have to invest considerable sophistication and effort in their work.

The social and ethical implications of nanotechnology are the results not merely of what the technology itself does, but also of how people react to it. Therefore, it will be important to understand the social acceptance, resistance, or rejection of nanotechnology at different times and in different places and human contexts. Interviews, focus groups, and questionnaire surveys can measure emotional, cognitive, and psychosocial parameters [38.68]. The combined methodologies of political science, sociology, and socio-legal studies will be required to chart the process of regulatory review and approval, court decisions that actively permit or prohibit

the use of the technology, mobilization of political support and opposition, and the activities of relevant social movements.

Ultimately, the test of the various nanotechnologies will be their benefit for human beings, as measured by economic growth, improved health and longevity, environmental protection, strengthened security, social vitality, and enhanced human capabilities. Convergence with other NBIC technologies will complicate the challenge for researchers and policymakers who wish to understand the social and ethical implications of nanotechnology, but it will also magnify the potential benefits to humanity many times over. Success will be most complete and most probable if scientific research on these implications is an integral part of the effort from the very beginning.

References

38.1 B. Joy: Why the future doesn't need us, Wired **8**(4) (April 2000) 238–262

38.2 B. J. Feder: Opposition to nanotechnology, New York Times (August 19 2002) (http://www.nytimes.com/)

38.3 G. H. Reynolds: *Forward to the Future: Nanotechnology and Regulatory Policy* (Pacific Research Institute, San Francisco 2002)

38.4 J. S. Brown, P. Duguid: Don't count society out: A response to Bill Joy. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 37–45

38.5 R. W. Siegel, E. Hu, M. C. Roco (Eds.): *Nanostructure Science and Technology* (Kluwer, Dordrecht 1999)

38.6 M. C. Roco, R. S. Williams, P. Alivisatos (Eds.): *Nanotechnology Research Directions* (Kluwer, Dordrecht 2000)

38.7 M. Roco, R. Tomellini: *Nanotechnology: Revolutionary Opportunities and Societal Implications* (Office for Official Publications of the European Communities, Luxembourg 2002)

38.8 M. S. Meyer: Socio-economic research on nanoscale science and technology: A european overview and illustration. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 278–311

38.9 R. H. Smith: Social, ethical, and legal implications of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 257–271

38.10 V. Weil: Ethical issues in nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 244–251

38.11 M. C. Roco, W. S. Bainbridge (Eds.): *Societal Implications of Nanoscience and Nanotechnology* (Kluwer, Dordrecht 2001)

38.12 L. B. Lave: Lifecycle/sustainability implications of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 205–212

38.13 National Research Council: *Small Wonders, Endless Frontiers: A Review of the National Nanotechnology Initiative* (National Academy Press, Washington D.C. 2002)

38.14 M. C. Roco, W. S. Bainbridge (Eds.): *Converging Technologies to Improve Human Performance* (Kluwer, Dordrecht 2003)

38.15 M. C. Suchman: Envisioning life on the nanofrontier. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 271–276

38.16 N. Gingrich: The age of transition. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 29–35

38.17 G. Yonas, S. T. Picraux: National needs drivers for nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 46–55

38.18 P. R. Ehrlich: *The Population Bomb* (Ballantine, New York 1968)

38.19 K. Davis, M. S. Bernstam, R. Ricardo-Campbell (Eds.): *Below-Replacement Fertility in Industrial Societies: Causes, Consequences, Policies (A supplement to Population and Development Review)* 1986)

38.20 B. J. Wattenberg: *The Birth Dearth* (Ballantine, New York 1987)

38.21 Central Intelligence Agency: *Long-Term Global Demographic Trends: Reshaping the Geopolitical Landscape* (Central Intelligence Agency, Washington D.C. 2001)

38.22 Demography and the west, Economist **364**(8287) (August 2002) 20–22

38.23 D. L. Costa, R. H. Steckel: Long-term trends in health, welfare, and economic growth in the United States. In: *Health and Welfare During Industrialization*, ed. by R. H. Steckel, R. Floud (Univ. Chicago Press, Chicago 1997) pp. 47–89

38.24 S. H. Preston: *American Longevity: Past, Present, and Future* (Center for Policy Research, Syracuse University, Syracuse 1996)
38.25 HHS (U.S. Department of Health and Human Services): *Chronic Diseases and Their Risk Factors: The Nation's Leading Causes of Death* (Centers for Disease Control, Atlanta 1999)
38.26 HHS (U.S. Department of Health and Human Services): *Reducing Tobacco Use: A Report of the Surgeon General* (Centers for Disease Control, Atlanta 2000)
38.27 D. M. Cutler, M. McClellan, J. P. Newhouse, D. Remler: Are medical prices declining, Quarterly J. Economics **113** (1998) 991–1024
38.28 D. M. Cutler, E. Richardson: Your money and your life: The value of health care and what affects it. In: *Frontiers in Health Policy Research*, ed. by A. M. Garber (MIT, Cambridge 1999) pp. 99–132
38.29 V. R. Fuchs: The future of health economics, J. Health Economics **19** (2000) 141–157
38.30 F. W. Hollmann, T. J. Mulder, J. E. Kallan: *Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100* (United States Census Bureau, Washington D.C. 2000)
38.31 R. Doering: Social implications of scaling to nanoelectronics. In: *Social Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 84–93
38.32 R. S. Williams, P. J. Kuekes: W've only just began. In: *Social Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 103–108
38.33 Central Intelligence Agency: *Experts See Risks to US IT Superiority Beyond Moore's Law*, report OTI IA 2002-151 (Central Intelligence Agency, Directorate of Intelligence, Washington D.C. 2002) pp. 2002–151
38.34 E. Tenner: Nanotechnology and unintended consequences. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 311–318
38.35 G. C. Homans: *Social Behavior: Its Elementary Forms* (Harcourt Brace Jovanovich, New York 1974)
38.36 R. L. Akers: *Deviant Behavior: A Social Learning Approach* (Wadsworth, Belmont 1985)
38.37 R. Stark, W. S. Bainbridge: *A Theory of Religion* (Rutgers University Press, New Brunswick 1987)
38.38 R. K. Merton: *Social structure and anomie* (Free Press, New York 1968) pp. 185–214, first published 1938
38.39 R. A. Cloward, L. E. Ohlin: *Delinquency and Opportunity* (Free Press, New York 1960)
38.40 A. P. Sloan: *My Years with General Motors* (Doubleday, Garden City 1964)
38.41 R. K. Merton: *Science, Technology and Society in Seventeenth Century England* (H. Fertig, New York 1970)
38.42 R. K. Merton: *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, Chicago 1973)
38.43 L. Greenfeld: *Nationalism: Five Roads to Modernity* (Harvard Univ. Press, Cambridge 1992)
38.44 W. S. Bainbridge: *The Spaceflight Revolution* (Wiley, New York 1976)
38.45 W. S. Bainbridge: *Goals in Space: American Values and the Future of Technology* (State Univ. New York Press, Albany 1991)
38.46 T. Hirschi: *Causes of Delinquency* (University of California Press, Berkeley 1969)
38.47 C. Shaw, H. D. McKay: *Delinquency Areas* (University of Chicago Press, Chicago 1929)
38.48 M. R. Gottfredson, T. Hirschi: *A General Theory of Crime* (Stanford Univ. Press, Stanford 1990)
38.49 E. H. Sutherland: *Principles of Criminology* (Lippincott, Philadelphia 1947)
38.50 G. Sykes, D. Matza: Techniques of neutralization: A theory of delinquency, Am. Sociol. Rev. **22** (1957) 664–670
38.51 I. L. Janis: *Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (Houghton Mifflin, Boston 1982)
38.52 N. J. Smelser: *Theory of Collective Behavior* (Free Press, New York 1962)
38.53 W. M. Tolles: National security aspects of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 218–237
38.54 W. S. Bainbridge: Collective behavior and social movements. In: *Sociology*, ed. by R. Stark (Wadsworth, Belmont 1985) pp. 493–523
38.55 V. Vogel: Societal impacts of nanotechnology in education and medicine. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 180–186
38.56 R. F. Service: Is nanotechnology dangerous?, Science **290** (2000) 1526–1527
38.57 R. E. Smalley: Of chemistry, love and nanobots, Sci. Am. **285** (2001) 76–77
38.58 A. Chen: *The Ethics of Nanotechnology* (BioScience Productions and Markkula Center for Applied Ethics, Santa Clara University, 2002) www.actionbioscience.org/newfrontiers/chen.html and www.scu.edu/ethics/publications/submitted/chen/nanotechnology.html
38.59 J. Canton: The strategic impact of nanotechnology on the future of business and economics. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 114–121
38.60 W. S. Bainbridge: Public attitudes toward nanotechnology, J. Nanopart. Res. **4** (2002) 561–570

38.61 H. Etzkowitz: Nano-science and society: Finding a social basis for science policy. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 121–128
38.62 J. S. Carroll: Social science research methods for assessing societal implications of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 238–244
38.63 B. A. Nardi: A cultural ecology of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 318–324
38.64 E. Katz, P. F. Lazarsfeld: *Personal Influence* (Free Press, Glencoe 1955)
38.65 E. M. Rogers: *Social Change in Rural Society* (Appleton-Century-Crofts, New York 1960)
38.66 D. Crane: *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (University of Chicago Press, Chicago 1972)
38.67 R. A. Bauer: *Second-Order Consequences: A Methodological Essay on the Impact of Technology* (MIT, Cambridge 1969)
38.68 P. B. Thompson: Social acceptance of nanotechnology. In: *Societal Implications of Nanoscience and Nanotechnology*, ed. by M. C. Roco, W. S. Bainbridge (Kluwer, Dordrecht 2001) pp. 251–256

Acknowledgements

4 Nanowires
by Mildred S. Dresselhaus, Yu-Ming Lin, Oded Rabin, Marcie R. Black, Gene Dresselhaus

The authors gratefully acknowledge the stimulating discussions with Professors Charles Lieber, Gang Chen, S. T. Lee, Arun Majumdar, Peidong Yang, and Jean-Paul Issi, Dr. Joseph Heremans, and Ted Harman. The authors are grateful for support for this work by the ONR Grant #000140-21-0865, the MURI program subcontract PO #0205-G-7A114-01 through UCLA, and DARPA contract #N66001-00-1-8603.

6 Stamping Techniques for Micro and Nanofabrication: Methods and Applications
by John A. Rogers

The author extends his deepest thanks to all of the collaborators who contributed the work described here.

8 MEMS/NEMS Devices and Applications
by Darrin J. Young, Christian A. Zorman, Mehran Mehregany

The authors wish to thank Wen H. Ko for the helpful discussions and suggestions and Michael Suster and Joseph Seeger for preparing the figures.

10 Therapeutic Nanodevices
by Stephen C. Lee, Mark Ruegsegger, Philip D. Barnes, Bryan R. Smith, Mauro Ferrari

The authors gratefully acknowledge many friends and colleagues whose input helped shape this article. Particularly, we acknowledge Beth S. Lee for many, many helpful discussions and for unflagging support. We also acknowledge Phil Streeter for many of the same services, performed in the office of friend rather than spouse. We acknowledge the support services of Anita Bratcher in manuscript preparation and the artistic stylings of Vladimir Marukhlenko for the figures incorporated in the manuscript. We also acknowledge Carol Bozarth for her kind and spontaneous support of the physical process of manuscript editing. This work is dedicated to the memories of Mildred A. Lee, Antonio Ferrari and Marialuisa Ferrari and the multitude of others whose lives have been tragically shortened by cancer, with the determination that therapeutic nanotechnology be used to help abolish the terrible power of the disease over human life.

13 Noncontact Atomic Force Microscopy and Its Related Topics
by Seizo Morita, Franz J. Giessibl, Yasuhiro Sugawara, Hirotaka Hosoi, Koichi Mukasa, Akira Sasahara, Hiroshi Onishi

Thanks to Tom Albrecht, Alexis Baratoff, Hartmut Bielefeldt, Gerd Binnig, Dominik Brändlin, Peter van Dongen, Urs Dürig, Christoph Gerber, Stefan Hembacher, Markus Herz, Lukas Howald, Christian Laschinger, Ulrich Mair, Jochen Mannhart, Thomas Ottenthal, Calvin Quate, Marco Tortonese, and the BMBF for funding under project no. 13N6918.

18 Surface Forces and Nanorheology of Molecularly Thin Films
by Marina Ruths, Alan D. Berman, Jacob N. Israelachvili

This work was supported by ONR grant N00014-00-1–0214. M. R. Luths thanks the Academy of Finland for financial support.

19 Scanning Probe Studies of Nano-scale Adhesion Between Solids in the Presence of Liquids and Monolayer Films
by Robert W. Carpick, James D. Batteas

We gratefully acknowledge the help of Ms. Erin Flater, who provided valuable assistance and insights into the literature on capillary formation. RWC acknowledges support of a career award from the National Science Foundation, grant #CMS-0134571.

24 Mechanics of Biological Nanotechnology
by Rob Phillips, Prashant K. Purohit, Jané Kondev

We happily acknowledge useful discussions with Kai Zinn, Jon Widom, Bill Gelbart, Andy Spakowitz, Zhen-Gang Wang, Ken Dill, Carlos Bustamante, Tom Powers, Larry Friedman, Jack Johnson, Pamela Bjorkman, Paul Wiggins, Steve Williams, Wayne Falk, Adrian Parsegian, Alasdair Steven and Steve Quake. RP and PP acknowledge support of the NSF through grant number 9971922, the NSF supported CIMMS center, and the support of the Keck Foundation. JK is supported

by the NSF under grant number DMR-9984471 and is a Cottrell Scholar of Research Corporation.

30 Nanotechnology for Data Storage Applications

by Dror Sarid, Brendan McCarthy, Ghassan E. Jabbour

The authors would like to thank Digital Instruments (Veeco) for contributing their Multi-Mode Nanoscope III, the Department of Energy (DE-FG-3-02ER46013/A001) for a generous grant, EMC for their generous gift, and the Vice President for Research, University of Arizona, for equipment support.

31 The "Millipede" – A Nanotechnology-Based AFM Data-Storage System

by Gerd K. Binnig, G. Cherubini, M. Despont, Urs T. Duerig, Evangelos Eleftheriou, H. Pozidis, Peter Vettiger

It is our pleasure to acknowledge our colleagues T. Albrecht, T. Antonakopoulos, P. Bächtold, A. Dholakia, U. Drechsler, B. Gotsmann, W. Häberle, D. Jubin, M. A. Lantz, T. Loeliger, H. E. Rothuizen, R. Stutz, and D. Wiesmann for their invaluable contributions to the millipede project.

In addition, thanks and appreciation go to H. Rohrer for his contribution to the initial millipede vision and concept and to our former collaborators, J. Brugger, now at the Swiss Federal Institute of Technology, Lausanne (Switzerland), M. I. Lutwyche, now at Seagate, Pittsburg, IL, and W. P. King, now at Georgia Tech, Atlanta, GA, as well as to K. Goodson, T. W. Kenny, and C. F. Quate of Stanford University, CA.

We are also pleased to acknowledge stimulating discussions with and encouraging support from our colleagues W. Bux and P. F. Seidler of the IBM Zurich Research Laboratory, J. Mamin, D. Rugar, and B. D. Terris of the IBM Almaden Research Center, San Jose, CA, and G. Hefferon of IBM, East Fishkill, NY.

Special thanks go to J. Frommer, C. Hawker, J. Mamin, and R. Miller of the IBM Almaden Research Center for their enthusiastic support in identifying and synthesizing alternative polymer media materials, and to H. Dang, A. Sharma, and S. Sri-Jayantha of the IBM T. J. Watson Research Center, Yorktown Heights, NY, for their contributions to the work on servo control.

35 Thermo- and Electromechanics of Thin-Film Microstructures

by Martin L. Dunn, Shawn J. Cunningham

We are most grateful for the assistance of Ms. Yanhang Zhang, a Ph.D. student at the University of Colorado, for her help with the preparation of figures. Many of the calculations and measurements also draw from her work. MLD acknowledges support from DARPA, Sandia National Laboratories, and the AFOSR for support of aspects of his research that appear in this work.

About the Authors

Chong H. Ahn

University of Cincinnati
Department of Electrical
and Computer Engineering
and Computer Science
Cincinnati, OH, USA
chong.ahn@uc.edu
http://www.BioMEMS.uc.edu/

Dr. Chong Ahn is a Professor of Electrical and Computer Engineering at the University of Cincinnati. He obtained his Ph.D. degree in Electrical Engineering from the Georgia Institute of Technology in 1993 and then worked as a postdoctoral fellow at IBM T.J. Watson Research Center. His research interests include all aspects of design, fabrication, and characterization of magnetic MEMS devices, microfluidic devices, protein chips, lab-on-a-chips, nano biosensors, point-of-care testing and BioMEMS systems. He is an associate editor of the IEEE Sensors Journal.

Boris Anczykowski

nanoAnalytics GmbH
Münster, Germany
anczykowski@nanoanalytics.com
http://www.nanoanalytics.com

Dr. Boris Anczykowski is a physicist with an extensive research background in the field of dynamic Scanning Force Microscopy. He co-invented the Q-Control technique and received the Innovation Award Münsterland for Science and Economy in 2001 for this achievement. He is a managing director and co-founder of nanoAnalytics GmbH, a company specialized in the characterization of surfaces and interfaces on the micro- and nanometer scale.

Massood Z. Atashbar

Western Michigan University
Department of Electrical
and Computer Engineering
Kalamazoo, MI, USA
massood.atashbar@wmich.edu
http://homepages.wmich.edu/~zandim

Professor Atashbar's main area of research is nanotechnology and wireless sensors. He has been working in the field of micro/nanosensors for chemical and physical sensing specifically integrated smart wireless surface acoustic wave sensors and systems. He is a senior member of IEEE and an associate editor of the International Journal of Modelling and Simulation

Wolfgang Bacsa

Université Paul Sabatier
Laboratoire de Physique des Solides (LPST)
Toulouse, France
bacsa@lpst.ups-tlse.fr
http://www.lpst.ups-tlse.fr/users/
wolfgang/

Wolfgang Bacsa is an expert in the emerging field of Nano-Optics and Carbon Nanotubes. He has a Ph.D. from the Swiss Federal Institute of Technology (ETH) Zurich in Physics and has extensive experience in condensed matter physics, optics, microscopy, synthesis of ultra-thin films and nanostructured carbon. Professor Bacsa worked at the ETH Zürich, PennState Universiy and EPFL Lausanne.

William Sims Bainbridge

National Science Foundation
Division of Information
and Intelligent Systems
Arlington, VA, USA
wbainbri@nsf.gov
http://mysite.verizon.net/william.
bainbridge

William Sims Bainbridge earned his Ph.D. in Sociology from Harvard University in 1975 and is the author of 10 books, 4 textbook-software packages, and about 150 shorter publications in information science, social science of technology, and the sociology of culture. He has represented the social and behavioral sciences on five US government advanced technology initiatives: High Performance Computing and Communications, Knowledge and Distributed Intelligence, Digital Libraries, Information Technology Research, and Nanotechnology.

Antonio Baldi

Institut de Microelectronica
de Barcelona (IMB)
Centro National Microelectrónica
(CNM-CSIC)
Barcelona, Spain
Antoni.baldi@cnm.es

Antonio Baldi received his PhD from the Universitat Autonoma de Barcelona, Spain, in 2001. He is a postdoctoral fellow at the University of Minnesota. His current research is focused on the fabrication and testing of bioMEMS, on the development of new sensors and microfluidic devices incorporating stimuli-sensitive hydrogels and on the development of inductively coupled wireless MEMS.

Philip D. Barnes

Ohio State University
Biomedical Engineering Center
Columbus, OH, USA
d_skill@yahoo.com

Phillip D. Barnes is a graduate student at the Biomedical Engineering Center at the Ohio State University. His background is in electrical engineering and his current interests are in sensors and imaging devices.

James D. Batteas

National Institute
of Standards and Technology
Surface and Microanalysis Science
Division
Gaithersburg, MD, USA
james.batteas@nist.gov
http://www.cstl.nist.gov/div837/837.03/

Dr. James D. Batteas is a research chemist in the Surface and Interface Research Group of the National Institute of Standards and Technology. The group's current research involves developing and charcterizing the organization and properties of nanoscale materials and devices on surfaces, this includes nanoparticles for use in optoelectric and optomagnetic gating, molecular electronics, and sensors. His group is also investigating friction, adhesion and wear mechanisms on the nanoscale.

Roland Bennewitz

McGill University
Physics Department
Montreal, QC, Canada
roland@physics.mcgill.ca

Roland Bennewitz studied physics in Freiburg and Berlin where he received his PhD for work on defects at surfaces of insulators. He is now assistant at the University of Basel where his research activities focus on high-resolution force microscopy as a tool in nanotribology and surface science.

Alan D. Berman

Monitor Venture Enterprises
Los Angeles, CA, USA
alan.berman.2001@anderson.ucla.edu

Alan Berman received his Ph.D. from the Chemical Engineering Department at the University of California, Santa Barbara. He is involved in technology commercialisation, assisting large technology companies to maximize return on R&D investments by finding and developing applications for their intellectual property outside of their core business. Value is realized through technology licensing, joint ventures and venture capital funded new entities. Technical interests are in fields of tribology, MEMS, and nanotechnology.

Bharat Bhushan

The Ohio State University
Nanotribology Laboratory
for Information Storage
and MEMS/NEMS
Columbus, OH, USA
bhushan.2@osu.edu

Dr. Bharat Bhushan is an Ohio Eminent Scholar and The Howard D. Winbigler Professor in the Department of Mechanical Engineering, a Graduate Research Faculty Advisor in the Department of Materials Science & Engineering, and the Director of the Nanotribology Laboratory for Information Storage & MEMS/NEMS (NLIM) at the Ohio State University, Columbus, Ohio. He holds two M.S., a Ph.D. in mechanical engineering/mechanics, an MBA, and three semi-honorary and honorary doctorates. His research interests are in micro/nanotribology and its applications to magnetic storage devices and MEMS/NEMS (Nanotechnology). He has authored 5 technical books, 45 handbook chapters, more than 450 technical papers in referred journals, and more than 60 technical reports, edited more than 25 books, and holds 14 U.S. patents.

Gerd K. Binnig

IBM Zurich Research Laboratory
Micro-/Nanomechanics
Rüschlikon, Switzerland
gbi@zurich.ibm.com

Gerd Binnig obtained his Ph.D. from the Johann Wolfgang Goethe University, Frankfurt, Germany, and joined IBM Research in 1978. He was corecipient of the 1986 Nobel Prize in Physics for the invention of the scanning tunnelling microscope, and he also invented the atomic force microscope. His current research interests are micro- and nanosystem techniques and "Fractal Darwinism", a theory he developed to describe complex systems.

Marcie R. Black

Massachusetts Institute
of Technology
Department of Electrical Engineering and
Computer Science
Cambridge, MA, USA
marcie@alum.mit.edu
http://web.mit.edu/mrb/www/prof.html

Marcie Black recently received her Ph.D. from Prof. Dresselhaus's research group at MIT studying the optical properties of nanowires. In particular, she identified the dominant optical absorption mechanism in the IR of bismuth nanowires as an indirect interband transition that is enhanced over bulk bismuth. Currently she is studying organic opto-electronics with an emphasis on photovoltaics.

Jean-Marc Broto

University Toulouse III
Laboratoire National des Champs
Magnétiques Pulsés (LNCMP)
Toulouse, France
broto@insa-tlse.fr
http://www.insa-tlse.fr/

Jean-Marc Broto is Professor at the Universite Toulouse III, France. He is a specialist in electronic transport and magnetization properties under high magnetic fields and contributed to the discovery of the Giant Magnetoresistance in 1988.

Robert W. Carpick

University of Wisconsin-Madison
Department of Engineering Physics
Madison, WI, USA
carpick@engr.wisc.edu
http://www.engr.wisc.edu/ep/faculty/
carpick_robert.html

Robert Carpick has a Ph.D. in Physics (1997) from the University of California at Berkeley. He has been an Assistant Professor at the University of Wisconsin-Madison since 2000. He carries out research and publishes in the areas of nanotribology, nanomechanics, nanostructured materials, and scanning probe development. He serves on the editorial board of Review of Scientific Instruments. In 2002 he received a Faculty Early Career Development award from the U.S. National Science Foundation.

Tsung-Lin Chen

National Chiao Tung University
Department
of Mechanical Engineering
Shin Chu, Taiwan
tsunglin@mail.nctu.edu.tw

Tsung-Lin Chen received his B.S. and M.A. degrees in mechanical engineering from the National Tsing-Hua University, Taiwan in 1990 and 1992, respectively. He received Ph.D. degree in mechanical engineering from University of California at Berkeley, USA in 2001. He is currently an assistant professor in National Chiao Tung University in Taiwan. His research interests include MEMS devices design and MEMS fabrication process development.

Yu-Ting Cheng

National Chiao Tung University
Department
of Electronics Engineering
& Institute of Electronics
HsinChu, Taiwan
ytcheng@faculty.nctu.edu.tw

Yu-Ting Cheng received his Ph.D. degree in Electrical Engineering at the University of Michigan, Ann Arbor, in 2000. After his graduation, he worked for IBM Watson Research Center, Yorktown Heights, as a research staff member. He is a member of IEEE, IOP, and Phi Tau Phi. Currently he is Assistant Professor at the National Chiao Tung University, Taiwan. His research interests include the development of novel materials and fabrication technologies for MEMS/NEMS applications and microsystems integration.

Giovanni Cherubini

IBM Zurich Research Laboratory
Storage Technologies
Rüschlikon, Switzerland
cbi@zurich.ibm.com

Dr. Giovanni Cherubini received a Ph.D. degree in electrical engineering from the University of California, San Diego, in 1986, and joined IBM Research in 1987. His interests include high-speed data transmission and data storage systems. He is Editor for CDMA systems, for IEEE Transactions on Communications, and served as Guest Editor for the IEEE Journal on Selected Areas in Communications issues on access technologies (1995) and on multiuser detection techniques (2001–2002). He is co-author of the book Algorithms for Communications Systems and their Applications.

Jin-Woo Choi

Louisiana State University
Department of Electrical
and Computer Engineering
Baton Rouge, LA, USA
choi@ece.lsu.edu

Jin-Woo Choi received his B.S. and M.S. degree in Electrical Engineering from Seoul National University in Korea in 1994 and 1996, respectively. He received his Ph.D. degree in Electrical Engineering from the University of Cincinnati in 2000. Now he is an Assistant Professor at Louisiana State University, Baton Rouge, Louisiana. His current research activities include magnetic particle separators, microfluidic systems for biochemical detection, micro total analysis systems (μ-TAS), bioelectronics, and BioMEMS components and systems.

Shawn J. Cunningham

WiSpry, Inc.
Colorado Springs Design Center
Colorado Springs, CO, USA
shawn.cunningham@wispry.com

Shawn Cunningham is working on the development of RF MEMS switch and associated processes with wiSpry, Inc. His interests include materials characterization, reliability, and Design for MEMS Manufacturability. Prior to joining wiSpry, Shawn pursued MEMS research and product development at Coventor, Ford Microelectronics, and the University of Utah's Center for Engineering Design and in collaboration with the Univerisity of Colorado.

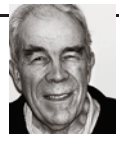
Michel Despont

IBM Zurich Research Laboratory
Micro-/Nanomechanics
Rüschlikon, Switzerland
dpt@zurich.ibm.com

Michel Despont received his Ph.D. in physics from the University of Neuchâtel, Switzerland, in 1996. After a postdoctoral fellowship at IBM's Zurich Research Laboratory, he was visiting scientist at the Seiko Instrument Research Laboratory in Japan in 1997. His current research at IBM focuses on the development of micro- and nanomechanical devices and of processes to fabricate so-called system-on-chip.

Gene Dresselhaus

Massachusetts Institute
of Technology
Francis Bitter Magnet Laboratory
Cambridge, MA, USA
gene@mgm.mit.edu
http://web.mit.edu/fbml/cmr/



Gene Dresselhaus received his Ph.D. in physics from the University of California in 1955. He was a faculty member at the University of Chicago, and assistant professor at Cornell before joining MIT Lincoln Laboratory in 1960 as a staff member. In 1976 he assumed his current position at the MIT Francis Bitter Magnet Laboratory. His area of interest is the electronic structure of nanomaterials and he has co-authored with M.S. Dresselhaus several books on fullerenes, nanowires, and nanotubes.

Mildred S. Dresselhaus

Massachusetts Institute
of Technology
Department
of Electrical Engineering
and Computer Science
and Department of Physics
Cambridge, MA, USA
millie@mgm.mit.edu
http://www.eecs.mit.edu/faculty/index.
html

Mildred Dresselhaus received her Ph.D. in physics from the University of Chicago in 1958. She joined the MIT faculty in 1967. She has been active in research across broad areas of solid state physics, especially in carbon science. Her present research activities focus on carbon nanotubes, bismuth nanowires, low dimensional thermoelectricity, and novel forms of carbon. She is the recipient of the National Medal of Science and 17 honorary degrees.

Martin L. Dunn

University of Colorado at Boulder
Department of Mechanical Engineering
Boulder, CO, USA
martin.dunn@colorado.edu

Martin L. Dunn received the Ph.D. in Mechanical Engineering from the University of Washington and was a postdoctoral appointee at Sandia National Laboratories. His research focuses on the micromechanical behavior of materials and structures. He has published over 75 articles in archival journals, and his research has been sponsored by NSF, DOE, NIST, DARPA, AFOSR, and Sandia National Laboratories.

Urs T. Dürig

IBM Zurich Research Laboratory
Micro-/Nanomechanics
Rüschlikon, Switzerland
drg@zurich.ibm.com

Urs Dürig received a Ph.D. degree from the Swiss Federal Institute of Technology, Zurich, in 1984. He joined IBM as a post-doc working on near-field optical microscopy. He is Research Staff Member since 1986: He worked in the field of scanning tunnelling and dynamic force microscopy. In 1997, he joined the Micro/Nanomechanics group focusing on polymer material issues and thermal modelling.

Evangelos Eleftheriou

IBM Zurich Research Laboratory
Storage Technologies
Rüschlikon, Switzerland
ele@zurich.ibm.com

He received a Ph.D. in Electrical Engineering from Carleton University, Ottawa, Canada, in 1985, and joined IBM Research in 1986. His research focuses on signal processing and coding for recording and transmission systems. His current research interests include nanotechnology, in particular probe-storage techniques.
Dr. Eleftheriou was elected IEEE Fellow in 2002.

Mauro Ferrari

Ohio State University
Biomedical Engineering Center
Columbus, OH, USA
Ferrari.5@osu.edu
http://bmew.bme.ohio- state.edu/bmeweb3/bme_faculty.htm

Mauro Ferrari is the Edgar Hendrickson Professor of Biomedical Engineering of the Ohio State University. A mechanical engineer by training, he also holds appointments in Internal Medicine, Mechanical Engineering and Materials Science. Professor Ferrari is a globally recognized figure in biomedical nanotechnology and micro and nanofabrication. His current interests are application of novel fabrication technologies to innovative drug delivery devices for treatment of cardiovascular disease, palliative care and oncology.

Emmanuel Flahaut

Université Paul Sabatier
CIRIMAT (Centre Interuniversitaire de Recherche et d'Ingénierie des Matériaux)
Toulouse, France
flahaut@chimie.ups-tlse.fr
http://eflahaut.nano.free.fr

Emmanuel Flahaut obtained his Ph.D in Materials Science in Toulouse working on CCVD synthesis of carbon nanotubes (CNTs) and dense ceramic-based composites including CNTs. He spent then more than one year as a post-doctoral researcher in Malcolm Green's Group in Oxford to work mainly on the filling of CNTs. He is now a permanent CNRS researcher at the University of Toulouse.

Lásló Forró

Swiss Federal Institute of Technology (EPFL)
Institute of Physics of Complex Matter
Lausanne, Switzerland
laszlo.forro@epfl.ch
http://nanotubes.epfl.ch

Professor László Forró is working on the synthesis, physical properties and manipulation of carbon nanostuctures and nanostuctured arrays, as well as, mechanical properties of carbon nanotubes, carbon onions, biological tubular systems. Transport and electron spin resonance studies of molecular materials, quasi-one-dimensional organic metals, organic superconductors, cuprates, manganates and fullerenes – up to high pressures. Tunneling spectroscopy in cuprate and fullerene superconductors. Optical properties of strongly correlated systems and biomaterials.

Jane Frommer

IBM Almaden Research Center
Department of Science and Technology
San Jose, CA, USA
frommer@Almaden.ibm.com

Following a Ph.D. in Organometallic Chemistry from Caltech, Dr. Jane Frommer has been involved in a diverse set of research areas, including electronically conducting polymers and scanning probe microscopy. Her present AFM lab at IBM collaborates with a wide variety of laboratories involved in materials research, including lithography, chromatography, and storage. Common to all these studies is Frommer's interest in the properties and structure of molecules in confined geometries.

Harald Fuchs

Universität Münster
Physikalisches Institut
Münster, Germany
fuchsh@uni-muenster.de
http://www.uni-muenster.de/Physik/PI/Fuchs

1984 Ph.D. Universität des Saarlandes with Prof. H. Prof Gleiter (nano crystalline Systems), 1984–1985, Post doc with IBM Research Lab. Zurich in the group of G. Binnig and H. Rohrer, 1985–1993 Project manager 'Ultrathin Organic Films' with BASF AG, Ludwigshafen, Germany. Since 1993 he is full Professor and Director at the Physical Institute of the University of Münster, 2000: Cofounder and Scientific Director of the Center for Nanotechnology (CeNTech).

Franz J. Giessibl

Universität Augsburg
Lehrstuhl für Experimentalphysik VI
Augsburg, Germany
franz.giessibl@physik.uni-augsburg.de
http://www.physik.uni-augsburg.de/exp6/

Dr. Giessibl is working in atomic force microscopy and scanning tunnelling microscopy in ultrahigh vacuum at room temperature and low temperatures. He is a Steering Committee Member of the International Conference on Noncontact Atomic Force Microscopy. He received the R&D 100 Award Chicago 1994, the German Nanoscience Award 2000, and the Rudolf Kaiser Price in 2001.

Enrico Gnecco

University of Basel
Department of Physics
Basel, Switzerland
Enrico.Gnecco@unibas.ch
http://monet.physik.unibas.ch/~gnecco/

Enrico Gnecco studied physics in Genoa, where he received his PhD for work on the mechanism of growth of nanostructured carbon films. He is now assistant at the University of Basel, where his research activities focus on friction force microscopy and molecular machinery.

Gérard Gremaud

Swiss Federal Institute
of Technology (EPFL)
Institute of Physics
of Complex Matter
Lausanne, Switzerland
gremaud@epfl.ch

Dr. Gérard Gremaud is a physicist and senior lecturer at the Ecole Polytechnique Fédérale de Lausanne (EPFL). He is active in the research fields of dislocation dynamic, acoustic and atomic force microscopy and granular physics. He is also responsible for the teaching of metrology and practical works to the physics students.

Jason H. Hafner

Rice University
Department of Physics
& Astronomy
Houston, TX, USA
hafner@rice.edu
http://www.ruf.rice.edu/~hafner

Jason Hafner earned his Ph.D. in physics from Rice University in 1998. He then held an NIH postdoctoral fellowship at Harvard University working on nanotube probes for high resolution biological atomic force microscopy. In 2001 he returned to Rice as an assistant professor where his group is pursuing various biophysical applications of scanned probe microscopy and nanomaterials. Dr. Hafner received a Beckman Young Investigator award in 2002.

Stefan Hengsberger

University of Applied Science
of Fribourg
Fribourg, Switzerland
stefan.hengsberger@eif.ch
http://www.anwalt-homburg.de/Stefan-index.htm

Stefan Hengsberger received his diploma in physics from the University of Saarbrücken in 1997. He started as a research scientist at Fraunhofer/Miami (Florida) where he worked until mai 1998. In July 1998 he joined the swiss federal institute of Technology (EPFL) where he earned his PhD in biomechanics in spring 2002. He stayed for another year at EPFL as a postdoc in biomechanics and physics until summer 2003. Since October 2003 he is Professor of Physics at the University of Applied Science in Fribourg, Switzerland.

Peter Hinterdorfer

Johannes Kepler University of Linz
Institute for Biophysics
Linz, Austria
peter.hinterdorfer@jku.at
http://at22.bphys.uni-linz.ac.at/bioph/staf/hipe/hipe.htm

Peter Hinterdorfer earned a Dr. tech. from the University of Linz, Austria, Institute for Biophysics in 1992. He was a postdoctoral fellow at the University of Virginia, Department of Molecular Physiology and Biological Physics (1992/1993). Since then he is at the University of Linz, Institute for Biophysics, where he holds a position as Associate Professor. His current and ongoing research includes single molecule force spectroscopy and high resolution topography and recognition imaging of biological samples.

Roberto Horowitz

University of California at Berkeley
Department
of Mechanical Engineering
Berkeley, CA, USA
horowitz@me.berkeley.edu
http://www.me.berkeley.edu/~horowitz

Roberto Horowitz joined the Department of Mechanical Engineering at the University of California at Berkeley in 1982, where he is currently a Professor. Dr. Horowitz teaches and conducts research in the areas of adaptive, learning, nonlinear and optimal control, with applications to Micro-Electromechanical Systems (MEMS), computer disk file systems, robotics, mechatronics and intelligent vehicle and highway systems (IVHS).

Hirotaka Hosoi

Japan Science and Technology
Corporation
Sapporo, Japan
hosoi@sapporo.jst-plaza.jp

Hirotaka Hosoi received the D.E. degree in electronic enginnering from Hokkaido University in 1999. Since 2002 he is at Innovation Plaza Hokkaido, Japan Science and Technology Corporation (JST). His main research focus is in high-resolution magnetic imaging of magnetic materials surfaces using a scanning force microscope. His current research interests includes magnetism on metal-oxide surfaces.

Jacob N. Israelachvili

University of California
Department of Chemical Engineering and
Materials Department
Santa Barbara, CA, USA
Jacob@engineering.ucsb.edu
http://www.chemengr.ucsb.edu/people/faculty/israelachvili

Jacob Israelachvili earned his Ph.D. 1971 at the Cavendish Laboratory, University of Cambridge, UK. He held various positions at the Department of Applied Mathematics at the Australian National University (1974–1986), including those of Professional Fellow and Head of Department. In 1986 he joined the faculty at University of California, Santa Barbara as Professor in the Department of Chemical Engineering and Materials Department. In 1988 he was elected a Fellow of the Royal Society of London, and in 1991 he was awarded the Alpha Chi Sigma Award for Chemical Engineering Research by the AIChE. He was elected as a foreign associate of the US National Academy of Engineering in 1996.

Ghassan E. Jabbour

University of Arizona
Optical Sciences Center
Tucson, AZ, USA
gej@optics.arizona.edu

Ghassan E. Jabbour is the head of the research group Organic Optoelectronic Materials and Devices at the Optical Sciences Center working on organic and hybrid materials and their applications to light emitting dvices, solar cells, memory storage, solid state lighting, and other areas. Professor Jabbour is a SPIE Fellow, Track Chair of the Nanotechnology Program for SPIE, and Associate Editor of the Journal of the Society for Information Display (JSID). He has over 200 publications, invited talks, and conference proceedings.

Harold Kahn

Case Western Reserve University
Department of Materials Science
and Engineering
Cleveland, OH, USA
kahn@cwru.edu

Harold Kahn is Researcher Associate Professor of Materials Science and Engineering at Case Western Reserve University, Cleveland, Ohio. His research is focused on MEMS device processing and testing, particularly wafer-level mechanical testing and shape-memory actuated microfluidics. He received a B.S. in metallurgical engineering from Lafayette College and a Ph.D. in electronic materials from the Massachusetts Institute of Technology.

András Kis

Swiss Federal Institute
of Technology (EPFL)
Institute of Physics
of Complex Matter
Lausanne, Switzerland
andras@igahpse.epfl.ch
http://www.andras.kis.name

András Kis is a Ph.D. student of Prof. László Forró at the Swiss Federal Institute of Technology (EPFL) where he works together with Andrzej Kulik. His main subject is the measurement of the mechanical properties of nanoscale objects, including carbon nanotubes and microtubules with the aim of understanding the interplay between their nanoscale structure and mechanical properties.

Jané Kondev

Brandeis University
Physics Department
Waltham, MA, USA
kondev@brandeis.edu
http://mattter.cc.brandeis.edu

Professor Kondev graduated with a PhD from Cornell University and did postdoctoral work at Brown University and Princeton University before joining the faculty at Brandeis University. He is a recipient of the CAREER award from the National Science Foundation and is a Cotrell Scholar of the Research Corporation.

Andrzej J. Kulik

Swiss Federal Institute
of Technology (EPFL)
Institute of Physics
of Complex Matter
Lausanne, Switzerland
andrzej.kulik@epfl.ch

Andrzej Kulik is the Head of the Biostructures and Nanomechanics Laboratory of the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. His research concentrates on quantitative nanoscale materials properties, nanotribology. scanning probe microscopy, contact mechanics, nanoindentation, near-field ultrasonics, optical tweezers, nanolithography, and nanomanipulation.

Christophe Laurent

Université Paul Sabatier
CIRIMAT (Centre Interuniversitaire
de Recherche et d'Ingénierie
des Matériaux)
Toulouse, France
laurent@chimie.ups-tlse.fr
http://ncn.f2g.net/

Dr. Ch. Laurent is Professor of Materials Chemistry at University Paul Sabatier and is the head of the Nanocomposites and Carbon Nanotubes group of CIRIMAT. His research include in the synthesis, characterization and mechanical properties of ceramic-matrix nanocomposites, and since 1994, carbon nanotubes (synthesis of single- and double-walled CNTs, formation mechanisms, characterization, localized growth, hydrogen storage, ceramic-matrix composites).

Stephen C. Lee

Ohio State University
Biomedical Engineering Center
Columbus, OH, USA
Lee@bme.ohio-state.edu
http://medicine.osu.edu/mcbiochem/

Stephen C. Lee is a pioneer in the field of semi-biological nanodevices (nanobiological devices), having published the first monograph devoted to the topic in 1998. His interests are in enabling technologies for the incorporation of functional proteins and nucleic acids into nanodevices, particularly for application in oncology and cardiovascular disease. He is currently Associate Professor of Cellular and Molecular Biochemistry, Chemical Engineering and Biomedical Engineering at the Ohio State University.

Yunfeng Li

University of California at Berkeley
Department
of Mechanical Engineering
Berkeley, CA, USA
yunfeng@me.Berkeley.edu

Yunfeng Li received the B.S. and M.S. degrees from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1992 and 1995, respectively. He is currently working towards the Ph.D. degree in the Department of Mechanical Engineering, University of California, Berkeley, CA, USA. His research interests include motion control, vibration control, control of MEMS devices, and mechatronics.

Liwei Lin

University of California at Berkeley
Mechanical Engineering Department
Berkeley, CA, USA
lwlin@me.berkeley.edu
http:// me.berkeley.edu/faculty/lin/index.html

Liwei Lin received his Ph.D. from the University of California, Berkeley, in 1993. He was an Associate Professor in the Institute of Applied Mechanics, National Taiwan University, Taiwan (1994–1996) and an Assistant Professor in Mechanical Engineering Department, University of Michigan (1996–1999). He joined UC-Berkeley in 1999 and is now an Associate Professor at Mechanical Engineering Department and Co-Director at Berkeley Sensor and Actuator Center.

Yu-Ming Lin

Massachusetts Institute
of Technology
Department of Electrical Engineering and
Computer Science
Cambridge, MA, USA
yming@mgm.mit.edu

Lin Yu-Ming performed experimental and theoretical studies on Bi-based nanowires for next-generation thermoelectric materials. The electronic transport properties of these systems are studied to investigate quantum size effects. He received the Masterworks Award of the Department of EECS, MIT (2000) and the Gold Medal, MRS Graduate Student Award (2002)

Huiwen Liu

Ohio State University
Nanotribology Laboratory
for Information Storage
and MEMS/NEMS
Columbus, OH, USA
liu.403@osu.edu

Dr. Huiwen Liu is the associate director of the Nanotribology Laboratory for Information Storage and MEMS/NEMS at The Ohio State University. His research interests are study of mechanical, tribological, and physical properties of advanced materials and MEMS/NEMS devices on micro/nano scale. He obtained an Alexander von Humboldt fellowship and the Japanese Science and Technology Agency Fellowship in 1997 and 1998, respectively

Adrian B. Mann

Rutgers University
Department of Ceramics
and Materials Engineering
Piscataway, NJ, USA
abmann@rci.rutgers.edu

Dr. Mann's research focuses on the nanomechanics of materials and the fabrication of nanostructured materials. His research is predominantly on biomedical materials, but also includes ceramics, polymers and metals. He is currently an Assistant Professor at Rutgers University, New Jersey. Prior to this he was a lecturer at the University of Manchester, England and a Fulbright scholar at The Johns Hopkins University, Maryland.

Othmar Marti

University of Ulm
Department of Experimental Physics
Ulm, Germany
Othmar.Marti@physik.uni-ulm.de
http://servex.physik.uni-ulm.de/marti

Profesor Othmar Marti is Head of the Department of Experimental Physics at the University of Ulm. His main research topics are polymers, scanning force microscopy, friction, near-field optics, and the optics of nanoparticles. Studies: diploma from ETH Zürich, Dr. sc. nat. ETH Zürich, Habilitation University Konstanz. He worked at IBM Research Zurich, Switzerland, University of California, USA, ETH Zurich, Switzerland and University of Konstanz, Germany.

Jack Martin

Analog Devices, Inc.
Micromachined Products Division
Cambridge, MA, USA
jack.martin@analog.com

Jack Martin has been a technologist and manager in the design, development and manufacture of industrial MEMS products for 25 years. His accomplishments include development of wafer fab and packaging processes that are used by Analog Devices to produce iMEMS integrated accelerometers and gyroscopes. He has a Ph.D. in Materials Science, a BS and MS in Chemical Engineering, and is a Licensed Professional Engineer.

Brendan McCarthy

University of Arizona
Optical Sciences Center
Tucson, AZ, USA
bmccarthy@optics.arizona.edu

Brendan Mc Carthy is a postdoctoral researcher in the Optical Sciences Center, University of Arizona. He received his PhD in 2001, from Trinity College, University of Dublin, Ireland. His thesis topic was the characterisation of composite materials based on carbon nanotubes and conjugated polymers, using optical and vibrational spectroscopy, scanning tunnelling microscopy, and transmission electron microscopy.

Mehran Mehregany

Case Western Reserve University
Department of Electrical Engineering and Computer Science
Cleveland, OH, USA
mxm31@cwru.edu
http://mems.cwru.edu/

Mehran Mehregany is currently the Chairman of the Electrical Engineering and Computer Science Department at Case and the Goodrich Professor of Engineering Innovation. His research interests include: silicon and silicon carbide micro/nano systems technology (including MEMS/NEMS); micromachining and microfabrication technologies; materials and modeling issues related to MEMS/NEMS and (in some cases) integrated circuits technologies; and MEMS packaging.

Ernst Meyer

University of Basel
Institute of Physics
Basel, Switzerland
Ernst.Meyer@unibas.ch
http://monet.unibas.ch/cgi-bin/people/meyer/

Ernst Meyer is professor of physics at the University of Basel. He is interested in friction force and dynamic force microscopy with true atomic resolution. He is also active in the field of sensors based upon micromechanics and magnetic spin resonance detection with force microscopy. Awarded from the Swiss Physical Society, he is member of the Swiss and American Physical Society, of the Editorial Board of Tribology Letters, and co-editor of books on atomic force microscopy.

Marc Monthioux

UPR A-8011 CNRS
Centre d'Elaboration des Matériaux et d'Etudes Structurales (CEMES)
Toulouse, France
monthiou@cemes.fr

Marc Monthioux has been working on carbon materials for more than 20 years. He is involved in research on carbon nanotubes since 1998, discovered the ability of single-wall nanotubes in being filled by foreign molecules the same year, associated with B. Smith and Prof. D.E. Luzzi from University of Pennsylvania. He is currently Director of Research at the French National Center for Scientific Research and European Associate Editor of CARBON Journal.

Markus Morgenstern

University of Hamburg
Institute of Applied Physics
Hamburg, Germany
mmorgens@physnet.uni-hamburg.de
http://www.nanoscience.de/group_r/stm-sts/

Markus Morgenstern earned his Ph.D. from the Institute of Interface Research and Vacuum Physics of the Forschungszentrum Jülich, Germany in 1996. After one year of research at the University of Paris VII he joined the Group of Prof. Dr. R. Wiesendanger in 1997 as a Senior Scientist. In 2002 he completed his Habilitation at the University of Hamburg with the subject Scanning tunnelling spectroscopy on semiconductor systems and nanostructures

Seizo Morita

Osaka University
Department of Electronic Engineering
Suita-Citiy, Osaka, Japan
smorita@ele.eng.osaka-u.ac.jp
http://www-e2.ele.eng.osaka-u.ac.jp/

Prof. Seizo Morita works in the atomic force microscopy (AFM). He has discovered two-dimensional friction with a lattice periodicity, and two-dimensional solid phase of densely contact-electrified electrons on SiO_2 thin films under ambient conditions. He has achieved mapping, discrimination and control of atomic force and atom with atomic resolution, and also atom manipulation based on a mechanical method using the noncontact AFM apparatus.

Koichi Mukasa

Hokkaido University
Nanoelectronics Laboratory
Sapporo, Japan
mukasa@nano.eng.hokudai.ac.jp
http://www.nano.eng.hokudai.ac.jp/nano/

K. Mukasa is a Professor of electronics at Hokkaido University, Sapporo, Japan. In 1980 he joined Alps Electric Co.Ltd, where he worked on the magnetic thin film heads and materials. In 1987 he moved to the university. His research interests include spin-polarized STM, exchange force microscopy, magnetic force microscopy, Mott spin detectors, nanostructure concerning electron spin and molecular/biological materials and devices.

Martin H. Müser

University of Western Ontario
Department of Applied Mathematics
London, Ontario, Canada
mmuser@uwo.ca
http://publish.uwo.ca/~mmuser/

Prof. Martin Müser received his Ph.D. (1995) and habilitation (2002) from the University of Mainz, Germany. He was a post doctoral researcher at Columbia University (1996/97) as Feodor Lynen fellow and at Johns Hopkins University (1998). He joined the faculty of the University of Western Ontario (U.W.O.) in 2002. His research interests are statistical physics with emphasis on computational materials science, tribology, and quantum solids.

Kenn Oldham

University of California at Berkeley
Department
of Mechanical Engineering
Berkeley, CA, USA
oldham@newton.berkeley.edu

Kenn Oldham is a graduate student in Mechanical Engineering at the University of California at Berkeley, pursuing M.S. and Ph.D. through National Science Foundation Graduate Fellowship. He received his B.S. from Carnegie Mellon University, Pittsburgh, PA. He is currently researching microdevices for hard disk drives. Research interests include MEMS design for control and reliability, optimal and robust control, and materials for microdevices.

Hiroshi Onishi

Kanagawa Academy
of Science and Technology
Surface Chemistry Laboratory
Kanagawa, Japan
oni@net.ksp.or.jp
http://home.ksp.or.jp/onishipro/home.eng/

Dr. Onishi Hiroshi is an experimental chemist at the Kanagawa Academy of Science and Technology interested in molecule-scale reaction kinetics at interfaces. He likes to observe molecules moving and reacting over metal oxide surfaces by time-lapse imaging with scanning probe microscopes. Domestic societies encouraged him with awards to further develop his research towards nano-scale chemistry.

René M. Overney

University of Washington
Department of Chemical Engineering
Seattle, WA, USA
roverney@u.Washington.edu
http://depts.washington.edu/nanolab/

René Overney received his Ph.D. in Physics at the University of Basel in 1992. After his postdoctoral years in Japan and at Exxon, CR, Annandale (NJ) he joined the University of Washington in 1996. His research interests are in mesoscale sciences involving nanorheological interpretations of processes that are as diverse as lubrication, membrane transport, or quantum yield in optoelectronic devices.

Alain Peigney

Université Paul Sabatier
CIRIMAT (Centre Inter-universitaire de
Recherches et d'Ingénierie
des Matériaux) – UMR CNRS 5085
Toulouse, France
peigney@chimie.ups-tlse.fr
http://ncn.f2g.net/

Ceramic Engineer and Doctor in Physical-Chemistry – Associate Professor of Materials Chemistry at the Paul Sabatier University – Researches in the synthesis, sintering and microstructural characterization of ceramics and ceramic matrix nanocomposites, and from 1994 on the synthesis on single- and double-walled carbon nanotubes and preparation of nanocomposites containing carbon nanotubes.

Oliver Pfeiffer

University of Basel
Institute of Physics
Basel, Switzerland
Oliver.Pfeiffer@stud.unibas.ch

Oliver Pfeiffer is PhD student at the University of Basel. The main topic of his work is the energy dissipation of oscillating cantilevers in non-contact AFM. Related to this research field is the examination of damping of torsional oscillations of cantilevers when approaching the sample.

Rob Phillips

California Institute of Technology
Mechanical Engineering
and Applied Physics
Pasadena, CA, USA
phillips@aero.caltech.edu
http://www.rpgroup.caltech.edu/people/phillips.html

Rob Phillips is Professor of Mechanical Engineering and Applied Physics at CALTECH, Pasadena. Phillips' work aims to examine nanomechanics of both crystalline solids and biological molecules and their assemblies. Recent efforts have been aimed at investigating mechanics of DNA packing in viruses, the tension-induced gating of ion channels and the mechanical response of bio-functionalized cantilevers.

Haralampos Pozidis

IBM Zurich Research Laboratory
Storage Technologies
Rüschlikon, Switzerland
hap@zurich.ibm.com

He received the Ph.D. in electrical engineering from Drexel University, Philadelphia, PA, in 1998. After working with Philips Research, Eindhoven, the Netherlands, on signal processing and coding for optical storage technologies, with focus on DVD and Blue-Ray-Disc, Dr. Pozidis joined IBM Research in 2001. His research focuses on receiver design for alternative storage technologies, particularly scanning probe microscopy-based techniques.

Prashant K. Purohit

California Institute of Technology
Mechanical Engineering
Pasadena, CA, USA
prashant@caltech.edu
http://www.rpgroup.caltech.edu/~prashant/

Purohit Prashant is currently a Postdoctoral Scholar in Applied Mechanics in the Mechanical Engineering department at Caltech. The overall theme of his research is to develop systematic methods to understand the physics operative at the nanometer scales. His research falls broadly into two categories. The first concerns the subject of coarse-graining methods for crystalline solids and the second mechanics problems in biology. His group hopes to extend the coarse-graining methods in solids to the arena of biological macromolecules.

Oded Rabin

Massachusetts Institute
of Technology
Department of Chemistry
Cambridge, MA, USA
oded@mgm.mit.edu

Oded Rabin is a Ph.D. candidate in Prof. Mildred S. Dresselhaus' group working on the thermoelectric properties of bismuth-antimony nanowire systems, and on electrochemistry-based nanowire synthesis methods. He earned a B.A. degree in Chemistry from the Technion – Israel Institute of Technology, Haifa, Israel and an M.A. degree in Chemistry from the Weizmann Institute of Science, Rehovot, Israel.

Françisco M. Raymo

University of Miami
Department of Chemistry
Coral Gables, FL, USA
fraymo@miami.edu
http://www.as.miami.edu/chemistry/FMRaymo.html

Françisco M. Raymo is Assistant Professor in the Department of Chemistry at the University of Miami. His research interests lie at the interface of chemistry and materials science. In particular, he is exploring innovative strategies to process optical signals with molecular switches, design fluorescent probes for chemical sensing and assemble nanostructured films from electroactive building blocks.

Manitra Razafinimanana

Université Paul Sabatier
Centre de Physique des Plasmas
et leurs Applications (CPPAT)
Toulouse, France
razafinimanana@cpat.ups-tlse.fr
http://cpat.ups-tlse.fr

Manitra Razafinimanana was born in Analalava, Madagascar, on June 28, 1951. He received the 3rd cycle degree, and Doctorat d'Etat ès Sciences Physiques degree from the Université Paul Sabatier, Toulouse, France, in 1982 and 1986 respectively. Since 2000, he was held the position of Professor. He has worked on plasma diagnostics, arc-electrode interaction, transport coefficients and thermodynamical properties calculation.

Mark O. Robbins

Johns Hopkins University
Department of Physics
and Astronomy
Baltimore, MD, USA
mr@jhu.edu
http://www.pha.jhu.edu/~mr

Mark Robbins is a Professor in the Department of Physics and Astronomy of the Johns Hopkins University. His research focuses on non-equilibrium processes like friction, adhesion, spreading, fluid invasion, and shear-induced phase transitions. A common goal is to understand the atomic origins of macroscopic behaviour. He is a fellow of the American Physical Society and was a Sloan Fellow and Presidential Young Investigator.

John A. Rogers

University of Illinois
Department of Materials Science and
Engineering
Urbana, IL, USA
jrogers@uiuc.edu
http://www.mse.uiuc.edu/faculty/
Rogers.html

John A. Rogers earned the Ph.D. degree in physical chemistry from M.I. T. in 1995. Until 2002, he was at Bell Labs serving first as a Member of Technical Staff and subsequently as research Director. John is currently Founder Professor of Engineering at University of Illinois at Urbana/ Champaign. His research interests include methods for micro/ nanofabrication, plastic and molecular electronics, photonics

Mark Ruegsegger

Ohio State University
Biomedical Engineering Center
Columbus, OH, USA
mark@bme.ohio-state.edu

Mark Ruegsegger received his PhD in biomedical engineering at Case Western Reserve Unviersity and is currently an Assistant Professor of Biomedical Engineering at The Ohio State University. His research focus is the development of superparamagnetic particles that can target specific cell or tissue types, in particular atherosclerotic plaque. Other projects include biomimetic surface coatings on biomaterials and characterization of biomolecular flow in nanochannels.

Marina Ruths

Åbo Akademi University
Department of Physical Chemistry
Åbo, Finland
mruths@abo.fi

Marina Ruths received her Ph.D. from the University of California, Santa Barbara in 1996 followed by postdoctoral research at the University of Illinois at Urbana-Champaign, at the Max-Planck-Institute for Polymer Research, an at Åbo Akademi University. Her current research includes adhesion, friction and nanorheology of surfactant, polymer, and liquid crystal systems. She received an ASLA-Fulbright grant in 1991 and an Alexander von Humboldt fellowship in 1998.

Dror Sarid

University of Arizona
Optical Sciences Center
Tucson, AZ, USA
sarid@optics.arizona.edu
http://www.optics.arizona.edu/spm

Dror Sarid is Professor of Optical Sciences and Director of the Optical Data Storage Center. He is conducting research in scanning tunneling microcopy, atomic force microscopy and related systems and in particular probe storage, nano-optics and nano-technology. He published three books and more than 150 papers, mostly on optics and nanotechnology related topics.

Akira Sasahara

Kanagawa Academy
of Science and Technology
Surface Chemistry Laboratory
Kanagawa, Japan
ryo@net.ksp.or.jp
http://home.ksp.or.jp/onishipro/home.
eng/

Dr. Sasahara Akira is interested in local structures formed on solid surfaces. His current research focuses on elucidation of chemical and physical properties of nano-scale structures on metal oxide surfaces and their effect on chemical reaction.

André Schirmeisen

University of Münster
Institute of Physics
Münster, Germany
schira@uni-muenster.de
http://www.andre-schirmeisen.de

Dr. André Schirmeisen is currently working with his research group on nanoscale mechanical phenomena at Münster University. First to combine field ion microscopy with force microscopy to investigate atomically defined nanocontacts. He spent several years in Canada at McGill University to earn his PhD degree in physics. Before he worked as a strategic business consultant and is always interested in connecting nanoscience research with business applications.

Alexander Schwarz

University of Hamburg
Institute of Applied Physics
Hamburg, Germany
aschwarz@physnet.uni-hamburg.de
http://www.nanoscience.de/group_r/

Dr. Alexander Schwarz belongs to the scientific staff of the Center of Microstructure Research at the Institute of Applied Physics at the University of Hamburg, Germany. The group has 10 years (since 1993) experience in scientific research in the field of force microscopy and spectroscopy at cryogenic temperatures in ultrahigh vacuum and high magnetic fields. He is a Senior Scientist and works on Magnetic Force Microscopy (MFM) at low temperatures

Udo D. Schwarz

Yale University
Department
of Mechanical Engineering
New Haven, CT, USA
udo.schwarz@yale.edu
http://www.eng.yale.edu/
nanomechanics/

Udo D. Schwarz received his Ph.D. from the University of Basel in 1993 already using scanning force microscopy. Subsequently he moved to the University of Hamburg where he specialised on low-temperature scanning force microscopy and nanotribology. After spending a year at the Lawrence Berkeley National Laboratory he accepted a position as Associate Professor of Mechanical Engineering at Yale University in 2002.

Philippe Serp

Ecole Nationale Supèrieure d'Ingénieurs
en Arts Chimiques
et Technologiques
Laboratoire de Catalyse,
Chimie Fine et Polymères
Toulouse, France
Philippe.Serp@ensiacet.fr
http://www.ensiacet.fr/

Philippe Serp is associate professor in the LCCFP at ENSIACET. After receiving is PhD from Paul Sabatier University, Toulouse, in 1994 where he worked on the preparation of supported catalyst, he moved to Universidade do Porto to carry out post-doctoral research on catalytic CVD to prepare carbon fibers. His current research interests include CVD preparation of nanostructured materials and catalysis.

Bryan R. Smith

Ohio State University
Biomedical Engineering Center
Columbus, OH, USA
bryan@bme.ohio-state.edu

Bryan Smith is a Ph.D. candidate in biomedical engineering at The Ohio State University, specializing in the microfabrication and bioconjugation of micro- and nanoparticles for proteomics as well as imaging and drug delivery in breast cancer and atherosclerosis.

Anisoara Socoliuc

University of Basel
Institute of Physics
Basel, Switzerland
A.Socoliuc@unibas.ch
http://monet.physik.unibas.ch/gue/
uhvafm/

Anisoara Socoliuc studied physics in Iasi, Romania. She is now PhD student at the University of Basel. The main topic of her work is the study of wear processes on the nanometer scale by friction force microscopy.

Yasuhiro Sugawara

Osaka University
Department of Applied Physics
Suita, Japan
sugawara@ap.eng.osaka-u.ac.jp

Yasuhiro Sugawara received his Ph.D. in 1988 from Tohoku University and is Professor in the Department of Applied Physics of the Graduate School of Engineering at Osaka University since 2002. His research focuses on the further development of scanning probe microscopes and their applications, especially the noncontact atomic force microscope for the observation of solid surfaces at the atomic and molecular level. His aim is also to develop new nanomaterials and nanodevices by manipulation of single atoms and molecules using the atomic force microscope.

George W. Tyndall

IBM Almaden Research Center
Science and Technology
San Jose, CA, USA
tyndallgw@netscape.net

Dr. George Tyndall's current research focuses on the friction and adhesion in boundary lubricants especially as it pertains to the tribology of the magnetic recording head-disk interface in hard-disk drive applications.

Peter Vettiger

IBM Zurich Research Laboratory
Manager Micro-/Nanomechanics
Rüschlikon, Switzerland
pv@zurich.ibm.com

In 1963 Peter Vettiger joined IBM Research (Rüschlikon). He established and headed micro/nanoscale fabrication activities for superconducting, electronic and opto-electronic devices. Together with G.K. Binnig, he initiated Millipede probe-storage activities in 1995. His research interests are micro/nanomechanical devices and systems for probe-storage and biological applications. He is an IEEE Fellow and received a Doctor honoris causa from the University of Basel, Switzerland.

Darrin J. Young

Case Western Reserve University
Electrical Engineering
and Computer Science
Cleveland, OH, USA
djy@po.cwru.edu
http://home.cwru.edu/~djy

Darrin J. Young received his BS, MS, and PhD degrees from the EECS Department at University of California at Berkeley in 1991, 1993, and 1999, respectively. He pioneered RF MEMS high-Q tunable passive devices for wireless communications. He joined the EECS Department at Case Western Reserve University as an assistant professor in 1999. His main research interests include MEMS device design and fabrication.

Babak Ziaie

University of Minnesota
Department of Electrical
and Computer Engineering
Minneapolis, MN, USA
ziaie@ece.umn.edu

Dr. Babak Ziaie is an assistant professor in the Electrical and Computer Engineering Department of the University of Minnesota. His research interests are located at the boundaries between engineering and biology. He collaborates with physicians and biologists of all kinds and his lab is actively involved in the design and fabrication of biomedical micro and nanoelectromechanical systems.

Christian A. Zorman

Case Western Reserve University
Department of Electrical Engineering
and Computer Science
Cleveland, OH, USA
caz@po.cwru.edu
http://mems.cwru.edu/

Christian A. Zorman received a B.S. cum laude in physics and a B.A. cum laude in economics from The Ohio State University in 1988. He received M.S. and Ph.D. degrees in physics from Case Western Reserve University (CWRU) in 1991 and 1994, respectively. Dr. Zorman served in several research capacities at CWRU between 1994 and 2002, before being appointed an Associate Professor in 2002. His research interests involve materials and processes for MEMS.

Philippe K. Zysset

Technische Universität Wien
Institut für Leichtbau
und Flugzeugbau (ILFB)
Wien, Austria
philippe.zysset@epfl.ch
http://ilfb.tuwien.ac.at/

Philippe Zysset graduated in physics in 1987 and obtained a Ph.D. in biomechanics in 1994 at Ecole Polytechnique Fédérale de Lausanne (EPFL) in Lausanne, Switzerland. He completed his postdoctoral training at the Orthopaedic Research Laboratories of the University of Michigan and returned to EPFL as an assistant professor in 1997. His current research interests are in bone biomechanics.

Detailed Contents

List of Tables XXIX
List of Abbreviations XXXIII

1 **Introduction to Nanotechnology** 1
1.1 Background and Definition of Nanotechnology 1
1.2 Why Nano? 2
1.3 Lessons from Nature 2
1.4 Applications in Different Fields 3
1.5 Reliability Issues of MEMS/NEMS 4
1.6 Organization of the Handbook 5
References 5

Part A Nanostructures, Micro/Nanofabrication, and Micro/Nanodevices

2 **Nanomaterials Synthesis and Applications: Molecule-Based Devices** 9
2.1 Chemical Approaches to Nanostructured Materials 10
2.1.1 From Molecular Building Blocks to Nanostructures 10
2.1.2 Nanoscaled Biomolecules: Nucleic Acids and Proteins 10
2.1.3 Chemical Synthesis of Artificial Nanostructures 12
2.1.4 From Structural Control to Designed Properties and Functions 12
2.2 Molecular Switches and Logic Gates 14
2.2.1 From Macroscopic to Molecular Switches 15
2.2.2 Digital Processing and Molecular Logic Gates 15
2.2.3 Molecular AND, NOT, and OR Gates 16
2.2.4 Combinational Logic at the Molecular Level 17
2.2.5 Intermolecular Communication 18
2.3 Solid State Devices 22
2.3.1 From Functional Solutions to Electroactive and Photoactive Solids 22
2.3.2 Langmuir–Blodgett Films 23
2.3.3 Self-Assembled Monolayers 27
2.3.4 Nanogaps and Nanowires 31
2.4 Conclusions and Outlook 35
References 36

3 **Introduction to Carbon Nanotubes** 39
3.1 Structure of Carbon Nanotubes 40
3.1.1 Single-Wall Nanotubes 40
3.1.2 Multiwall Nanotubes 43

3.2 Synthesis of Carbon Nanotubes 45
3.2.1 Solid Carbon Source-Based Production Techniques for Carbon Nanotubes 45
3.2.2 Gaseous Carbon Source-Based Production Techniques for Carbon Nanotubes 52
3.2.3 Miscellaneous Techniques 57
3.2.4 Synthesis of Aligned Carbon Nanotubes 58
3.3 Growth Mechanisms of Carbon Nanotubes 59
3.3.1 Catalyst-Free Growth 59
3.3.2 Catalytically Activated Growth 60
3.4 Properties of Carbon Nanotubes 63
3.4.1 Variability of Carbon Nanotube Properties 63
3.4.2 General Properties 63
3.4.3 SWNT Adsorption Properties 63
3.4.4 Transport Properties 65
3.4.5 Mechanical Properties 67
3.4.6 Reactivity 67
3.5 Carbon Nanotube-Based Nano-Objects 68
3.5.1 Hetero-Nanotubes 68
3.5.2 Hybrid Carbon Nanotubes 68
3.5.3 Functionalized Nanotubes 71
3.6 Applications of Carbon Nanotubes 73
3.6.1 Current Applications 73
3.6.2 Expected Applications Related to Adsorption 76
References 86

4 **Nanowires** 99
4.1 Synthesis 100
4.1.1 Template-Assisted Synthesis 100
4.1.2 VLS Method for Nanowire Synthesis 105
4.1.3 Other Synthesis Methods 107
4.1.4 Hierarchical Arrangement and Superstructures of Nanowires 108
4.2 Characterization and Physical Properties of Nanowires 110
4.2.1 Structural Characterization 110
4.2.2 Transport Properties 115
4.2.3 Optical Properties 126
4.3 Applications 131
4.3.1 Electrical Applications 131
4.3.2 Thermoelectric Applications 133
4.3.3 Optical Applications 134
4.3.4 Chemical and Biochemical Sensing Devices 137
4.3.5 Magnetic Applications 137
4.4 Concluding Remarks 138
References 138

5 **Introduction to Micro/Nanofabrication** 147
5.1 Basic Microfabrication Techniques 148
5.1.1 Lithography 148
5.1.2 Thin Film Deposition and Doping 149
5.1.3 Etching and Substrate Removal 153
5.1.4 Substrate Bonding 157
5.2 MEMS Fabrication Techniques 159
5.2.1 Bulk Micromachining 159
5.2.2 Surface Micromachining 163
5.2.3 High-Aspect-Ratio Micromachining 166
5.3 Nanofabrication Techniques 170
5.3.1 E-Beam and Nano-Imprint Fabrication 171
5.3.2 Epitaxy and Strain Engineering 172
5.3.3 Scanned Probe Techniques 173
5.3.4 Self-Assembly and Template Manufacturing 176
References 180

6 **Stamping Techniques for Micro and Nanofabrication: Methods and Applications** 185
6.1 High Resolution Stamps 186
6.2 Microcontact Printing 187
6.3 Nanotransfer Printing 190
6.4 Applications 193
6.4.1 Unconventional Electronic Systems 193
6.4.2 Lasers and Waveguide Structures 198
6.5 Conclusions 200
References 200

7 **Materials Aspects of Micro- and Nanoelectromechanical Systems** 203
7.1 Silicon 203
7.1.1 Single Crystal Silicon 204
7.1.2 Polysilicon 205
7.1.3 Porous Silicon 208
7.1.4 Silicon Dioxide 208
7.1.5 Silicon Nitride 209
7.2 Germanium-Based Materials 210
7.2.1 Polycrystalline Ge 210
7.2.2 Polycrystalline SiGe 210
7.3 Metals 211
7.4 Harsh Environment Semiconductors 212
7.4.1 Silicon Carbide 212
7.4.2 Diamond 215
7.5 GaAs, InP, and Related III-V Materials 217
7.6 Ferroelectric Materials 218
7.7 Polymer Materials 219
7.7.1 Polyimide 219
7.7.2 SU-8 220

7.7.3 Parylene 220
7.8 Future Trends 220
References 221

8 **MEMS/NEMS Devices and Applications** 225
8.1 MEMS Devices and Applications 227
8.1.1 Pressure Sensor 227
8.1.2 Inertial Sensor 229
8.1.3 Optical MEMS 233
8.1.4 RF MEMS 239
8.2 NEMS Devices and Applications 246
8.3 Current Challenges and Future Trends 249
References 250

9 **Microfluidics and Their Applications to Lab-on-a-Chip** 253
9.1 Materials for Microfluidic Devices and Micro/Nano Fabrication Techniques 254
9.1.1 Silicon 254
9.1.2 Glass 254
9.1.3 Polymer 255
9.2 Active Microfluidic Devices 257
9.2.1 Microvalves 258
9.2.2 Micropumps 260
9.3 Smart Passive Microfluidic Devices 262
9.3.1 Passive Microvalves 262
9.3.2 Passive Micromixers 265
9.3.3 Passive Microdispensers 266
9.3.4 Microfluidic Multiplexer Integrated with Passive Microdispenser 267
9.3.5 Passive Micropumps 269
9.3.6 Advantages and Disadvantages of the Passive Microfluidic Approach 269
9.4 Lab-on-a-Chip for Biochemical Analysis 270
9.4.1 Magnetic Micro/Nano Bead-Based Biochemical Detection System 270
9.4.2 Disposable Smart Lab-on-a-Chip for Blood Analysis 273
References 276

10 **Therapeutic Nanodevices** 279
10.1 Definitions and Scope of Discussion 280
10.1.1 Design Issues 281
10.1.2 Utility and Scope of Therapeutic Nanodevices 285
10.2 Synthetic Approaches: "top-down" versus "bottom-up" Approaches for Nanotherapeutic Device Components 285
10.2.1 Production of Nanoporous Membranes by Microfabrication Methods: A top-down Approach 285

10.2.2 Synthesis of Poly(amido) Amine (PAMAM) Dendrimers: A bottom-up Approach ... 286
10.2.3 The Limits of top-down and bottom-up Distinctions with Respect to Nanomaterials and Nanodevices ... 287
10.3 Technological and Biological Opportunities ... 288
10.3.1 Assembly Approaches ... 288
10.3.2 Targeting: Delimiting Nanotherapeutic Action in Three-Dimensional Space ... 296
10.3.3 Triggering: Delimiting Nanotherapeutic Action in Space and Time ... 298
10.3.4 Sensing Modalities ... 302
10.3.5 Imaging Using Nanotherapeutic Contrast Agents ... 304
10.4 Applications for Nanotherapeutic Devices ... 307
10.4.1 Nanotherapeutic Devices in Oncology ... 307
10.4.2 Cardiovascular Applications of Nanotherapeutics ... 310
10.4.3 Nanotherapeutics and Specific Host Immune Responses ... 311
10.5 Concluding Remarks: Barriers to Practice and Prospects ... 315
10.5.1 Complexity in Biology ... 315
10.5.2 Dissemination of Biological Information ... 315
10.5.3 Cultural Differences Between Technologists and Biologists ... 316
References ... 317

Part B Scanning Probe Microscopy

11 **Scanning Probe Microscopy – Principle of Operation, Instrumentation, and Probes** ... 325
11.1 Scanning Tunneling Microscope ... 327
11.1.1 Binnig et al.'s Design ... 327
11.1.2 Commercial STMs ... 328
11.1.3 STM Probe Construction ... 330
11.2 Atomic Force Microscope ... 331
11.2.1 Binnig et al.'s Design ... 333
11.2.2 Commercial AFM ... 333
11.2.3 AFM Probe Construction ... 338
11.2.4 Friction Measurement Methods ... 342
11.2.5 Normal Force and Friction Force Calibrations of Cantilever Beams ... 346
11.3 AFM Instrumentation and Analyses ... 347
11.3.1 The Mechanics of Cantilevers ... 347
11.3.2 Instrumentation and Analyses of Detection Systems for Cantilever Deflections ... 350
11.3.3 Combinations for 3-D-Force Measurements ... 358
11.3.4 Scanning and Control Systems ... 359
References ... 364

12 **Probes in Scanning Microscopies** 371
12.1 Atomic Force Microscopy 372
12.1.1 Principles of Operation 372
12.1.2 Standard Probe Tips 373
12.1.3 Probe Tip Performance 374
12.1.4 Oxide-Sharpened Tips 375
12.1.5 FIB tips 376
12.1.6 EBD tips 376
12.1.7 Carbon Nanotube Tips 376
12.2 Scanning Tunneling Microscopy 382
12.2.1 Mechanically Cut STM Tips 382
12.2.2 Electrochemically Etched STM Tips 383
References 383

13 **Noncontact Atomic Force Microscopy and Its Related Topics** 385
13.1 Principles of Noncontact Atomic Force Microscope (NC-AFM) 386
13.1.1 Imaging Signal in AFM 386
13.1.2 Experimental Measurement and Noise 387
13.1.3 Static AFM Operating Mode 387
13.1.4 Dynamic AFM Operating Mode 388
13.1.5 The Four Additional Challenges Faced by AFM 388
13.1.6 Frequency-Modulation AFM (FM-AFM) 389
13.1.7 Relation Between Frequency Shift and Forces 390
13.1.8 Noise in Frequency-Modulation AFM – Generic Calculation . 391
13.1.9 Conclusion 391
13.2 Applications to Semiconductors 391
13.2.1 Si(111)7×7 Surface 392
13.2.2 Si(100)2×1 and Si(100)2×1:H Monohydride Surfaces 393
13.2.3 Metal-Deposited Si Surface 395
13.3 Applications to Insulators 397
13.3.1 Alkali Halides, Fluorides, and Metal Oxides 397
13.3.2 Atomically Resolved Imaging of a NiO(001) Surface 402
13.3.3 Atomically Resolved Imaging Using Noncoated and Fe-Coated Si Tips 402
13.4 Applications to Molecules 404
13.4.1 Why Molecules and What Molecules? 404
13.4.2 Mechanism of Molecular Imaging 404
13.4.3 Perspectives 407
References 407

14 **Low Temperature Scanning Probe Microscopy** 413
14.1 Microscope Operation at Low Temperatures 414
14.1.1 Drift 414
14.1.2 Noise 415
14.1.3 Stability 415
14.1.4 Piezo Relaxation and Hysteresis 415
14.2 Instrumentation 415

14.2.1 A Simple Design for a Variable Temperature STM 416
14.2.2 A Low Temperature SFM Based on a Bath Cryostat 417
14.3 Scanning Tunneling Microscopy and Spectroscopy 419
14.3.1 Atomic Manipulation 419
14.3.2 Imaging Atomic Motion 420
14.3.3 Detecting Light from Single Atoms and Molecules 421
14.3.4 High Resolution Spectroscopy 422
14.3.5 Imaging Electronic Wave Functions 427
14.3.6 Imaging Spin Polarization: Nanomagnetism 431
14.4 Scanning Force Microscopy and Spectroscopy 433
14.4.1 Atomic-Scale Imaging 434
14.4.2 Force Spectroscopy 436
14.4.3 Electrostatic Force Microscopy 438
14.4.4 Magnetic Force Microscopy 439
References 442

15 **Dynamic Force Microscopy** 449
15.1 Motivation: Measurement of a Single Atomic Bond 450
15.2 Harmonic Oscillator: A Model System for Dynamic AFM 454
15.3 Dynamic AFM Operational Modes 455
15.3.1 Amplitude-Modulation/ Tapping-Mode AFMs 456
15.3.2 Self-Excitation Modes 461
15.4 *Q*-Control 464
15.5 Dissipation Processes Measured with Dynamic AFM 468
15.6 Conclusion 471
References 471

16 **Molecular Recognition Force Microscopy** 475
16.1 Ligand Tip Chemistry 476
16.2 Fixation of Receptors to Probe Surfaces 478
16.3 Single-Molecule Recognition Force Detection 479
16.4 Principles of Molecular Recognition Force Spectroscopy 482
16.5 Recognition Force Spectroscopy: From Isolated Molecules to Biological Membranes 484
16.5.1 Forces, Energies, and Kinetic Rates 484
16.5.2 Complex Bonds and Energy Landscapes 486
16.5.3 Live Cells and Membranes 489
16.6 Recognition Imaging 489
16.7 Concluding Remarks 491
References 492

Part C Nanotribology and Nanomechanics

17 **Micro/Nanotribology and Materials Characterization Studies Using Scanning Probe Microscopy** 497
17.1 Description of AFM/FFM and Various Measurement Techniques 499

17.1.1 Surface Roughness and Friction Force Measurements 500
17.1.2 Adhesion Measurements 502
17.1.3 Scratching, Wear and Fabrication/Machining 503
17.1.4 Surface Potential Measurements 503
17.1.5 In Situ Characterization of Local Deformation Studies 504
17.1.6 Nanoindentation Measurements 504
17.1.7 Localized Surface Elasticity and Viscoelasticity Mapping 505
17.1.8 Boundary Lubrication Measurements 507
17.2 Friction and Adhesion 507
17.2.1 Atomic-Scale Friction 507
17.2.2 Microscale Friction 507
17.2.3 Directionality Effect on Microfriction 511
17.2.4 Velocity Dependence on Microfriction 513
17.2.5 Effect of Tip Radii and Humidity on Adhesion and Friction .. 515
17.2.6 Scale Dependence on Friction 518
17.3 Scratching, Wear, Local Deformation, and Fabrication/Machining ... 518
17.3.1 Nanoscale Wear 518
17.3.2 Microscale Scratching 519
17.3.3 Microscale Wear 520
17.3.4 In Situ Characterization of Local Deformation 524
17.3.5 Nanofabrication/Nanomachining 526
17.4 Indentation 526
17.4.1 Picoindentation 526
17.4.2 Nanoscale Indentation 527
17.4.3 Localized Surface Elasticity and Viscoelasticity Mapping 528
17.5 Boundary Lubrication 530
17.5.1 Perfluoropolyether Lubricants 530
17.5.2 Self-Assembled Monolayers 536
17.5.3 Liquid Film Thickness Measurements 537
17.6 Closure 538
References 539

18 **Surface Forces and Nanorheology of Molecularly Thin Films** 543
18.1 Introduction: Types of Surface Forces 544
18.2 Methods Used to Study Surface Forces 546
18.2.1 Force Laws 546
18.2.2 Adhesion Forces 547
18.2.3 The SFA and AFM 547
18.2.4 Some Other Force-Measuring Techniques 549
18.3 Normal Forces Between Dry (Unlubricated) Surfaces 550
18.3.1 Van der Waals Forces in Vacuum and Inert Vapors 550
18.3.2 Charge Exchange Interactions 552
18.3.3 Sintering and Cold Welding 553
18.4 Normal Forces Between Surfaces in Liquids 554
18.4.1 Van der Waals Forces in Liquids 554
18.4.2 Electrostatic and Ion Correlation Forces 554
18.4.3 Solvation and Structural Forces 557

18.4.4 Hydration and Hydrophobic Forces 559
18.4.5 Polymer-Mediated Forces 561
18.4.6 Thermal Fluctuation Forces 563
18.5 Adhesion and Capillary Forces 564
18.5.1 Capillary Forces 564
18.5.2 Adhesion Mechanics 566
18.5.3 Effects of Surface Structure, Roughness, and Lattice Mismatch 566
18.5.4 Nonequilibrium and Rate-Dependent Interactions: Adhesion Hysteresis 567
18.6 Introduction: Different Modes of Friction and the Limits of Continuum Models 569
18.7 Relationship Between Adhesion and Friction Between Dry (Unlubricated and Solid Boundary Lubricated) Surfaces 571
18.7.1 Amontons' Law and Deviations from It Due to Adhesion: The Cobblestone Model 571
18.7.2 Adhesion Force and Load Contribution to Interfacial Friction 572
18.7.3 Examples of Experimentally Observed Friction of Dry Surfaces 576
18.7.4 Transition from Interfacial to Normal Friction with Wear 579
18.8 Liquid Lubricated Surfaces 580
18.8.1 Viscous Forces and Friction of Thick Films: Continuum Regime 580
18.8.2 Friction of Intermediate Thickness Films 582
18.8.3 Boundary Lubrication of Molecularly Thin Films: Nanorheology 584
18.9 Role of Molecular Shape and Surface Structure in Friction 591
References 594

19 **Scanning Probe Studies of Nanoscale Adhesion Between Solids in the Presence of Liquids and Monolayer Films** 605
19.1 The Importance of Adhesion at the Nanoscale 605
19.2 Techniques for Measuring Adhesion 606
19.3 Calibration of Forces, Displacements, and Tips 610
19.3.1 Force Calibration 610
19.3.2 Probe Tip Characterization 611
19.3.3 Displacement Calibration 612
19.4 The Effect of Liquid Capillaries on Adhesion 612
19.4.1 Theoretical Background 612
19.4.2 Experimental and Theoretical Studies of Capillary Formation with Scanning Probes 614
19.4.3 Future Directions 618
19.5 Self-Assembled Monolayers 618
19.5.1 Adhesion at SAM Interfaces 618
19.5.2 Chemical Force Microscopy: General Methodology 619
19.5.3 Adhesion at SAM-Modified Surfaces in Liquids 620
19.5.4 Impact of Intra- and Inter-Chain Interactions on Adhesion 621

19.5.5 Adhesion at the Single-Bond Level 622
19.5.6 Future Directions 623
19.6 Concluding Remarks 624
References 624

20 **Friction and Wear on the Atomic Scale** 631
20.1 Friction Force Microscopy in Ultra-High Vacuum 632
20.1.1 Friction Force Microscopy 632
20.1.2 Force Calibration 632
20.1.3 The Ultra-high Vacuum Environment 635
20.1.4 A Typical Microscope in UHV 635
20.2 The Tomlinson Model 636
20.2.1 One-dimensional Tomlinson Model 636
20.2.2 Two-dimensional Tomlinson Model 637
20.2.3 Friction Between Atomically Flat Surfaces 637
20.3 Friction Experiments on Atomic Scale 638
20.3.1 Anisotropy of Friction 642
20.4 Thermal Effects on Atomic Friction 642
20.4.1 The Tomlinson Model at Finite Temperature 642
20.4.2 Velocity Dependence of Friction 644
20.4.3 Temperature Dependence of Friction 645
20.5 Geometry Effects in Nanocontacts 646
20.5.1 Continuum Mechanics of Single Asperities 646
20.5.2 Load Dependence of Friction 647
20.5.3 Estimation of the Contact Area 647
20.6 Wear on the Atomic Scale 649
20.6.1 Abrasive Wear on the Atomic Scale 649
20.6.2 Wear Contribution to Friction 650
20.7 Molecular Dynamics Simulations of Atomic Friction and Wear 651
20.7.1 Molecular Dynamics Simulation of Friction Processes 651
20.7.2 Molecular Dynamics Simulations of Abrasive Wear 652
20.8 Energy Dissipation in Noncontact Atomic Force Microscopy 654
20.9 Conclusion 656
References 657

21 **Nanoscale Mechanical Properties – Measuring Techniques and Applications** 661
21.1 Local Mechanical Spectroscopy by Contact AFM 662
21.1.1 The Variable-Temperature SLAM (T-SLAM) 663
21.1.2 Example One: Local Mechanical Spectroscopy of Polymers 664
21.1.3 Example Two: Local Mechanical Spectroscopy of NiTi 665
21.2 Static Methods – Mesoscopic Samples 667
21.2.1 Carbon Nanotubes – Introduction to Basic Morphologies and Production Methods 667
21.2.2 Measurements of the Mechanical Properties of Carbon Nanotubes by SPM 668
21.2.3 Microtubules and Their Elastic Properties 673

21.3 Scanning Nanoindentation: An Application to Bone Tissue 674
21.3.1 Scanning Nanoindentation 674
21.3.2 Application of Scanning Nanoindentation 674
21.3.3 Example: Study of Mechanical Properties of Bone Lamellae Using SN 675
21.3.4 Conclusion 681
21.4 Conclusions and Perspectives 682
References 682

22 **Nanomechanical Properties of Solid Surfaces and Thin Films** 687
22.1 Instrumentation 688
22.1.1 AFM and Scanning Probe Microscopy 688
22.1.2 Nanoindentation 689
22.1.3 Adaptations of Nanoindentation 690
22.1.4 Complimentary Techniques 691
22.1.5 Bulge Tests 691
22.1.6 Acoustic Methods 692
22.1.7 Imaging Methods 693
22.2 Data Analysis 694
22.2.1 Elastic Contacts 694
22.2.2 Indentation of Ideal Plastic Materials 694
22.2.3 Adhesive Contacts 695
22.2.4 Indenter Geometry 696
22.2.5 Analyzing Load/Displacement Curves 696
22.2.6 Modifications to the Analysis 699
22.2.7 Alternative Methods of Analysis 700
22.2.8 Measuring Contact Stiffness 701
22.2.9 Measuring Viscoelasticity 702
22.3 Modes of Deformation 702
22.3.1 Defect Nucleation 702
22.3.2 Variations with Depth 704
22.3.3 Anisotropic Materials 704
22.3.4 Fracture and Delamination 704
22.3.5 Phase Transformations 705
22.4 Thin Films and Multilayers 707
22.4.1 Thin Films 707
22.4.2 Multilayers 709
22.5 Developing Areas 711
References 712

23 **Atomistic Computer Simulations of Nanotribology** 717
23.1 Molecular Dynamics 718
23.1.1 Model Potentials 719
23.1.2 Maintaining Constant Temperature 720
23.1.3 Imposing Load and Shear 721
23.1.4 The Time-Scale and Length-Scale Gaps 721
23.1.5 A Summary of Possible Traps 722

23.2 Friction Mechanisms at the Atomic Scale 723
23.2.1 Geometric Interlocking 723
23.2.2 Elastic Instabilities 724
23.2.3 Role of Dimensionality and Disorder 727
23.2.4 Elastic Instabilities vs. Wear in Atomistic Models 727
23.2.5 Hydrodynamic Lubrication and Its Confinement-Induced Breakdown 729
23.2.6 Submonolayer Films 731
23.3 Stick-Slip Dynamics 732
23.4 Conclusions 734
References 735

24 **Mechanics of Biological Nanotechnology** 739
24.1 Science at the Biology–Nanotechnology Interface 740
24.1.1 Biological Nanotechnology 740
24.1.2 Self-Assembly as Biological Nanotechnology 740
24.1.3 Molecular Motors as Biological Nanotechnology 740
24.1.4 Molecular Channels and Pumps as Biological Nanotechnology 741
24.1.5 Biologically Inspired Nanotechnology 742
24.1.6 Nanotechnology and Single Molecule Assays in Biology 743
24.1.7 The Challenge of Modeling the Bio-Nano Interface 744
24.2 Scales at the Bio-Nano Interface 746
24.2.1 Spatial Scales and Structures 747
24.2.2 Temporal Scales and Processes 749
24.2.3 Force and Energy Scales: The Interplay of Deterministic and Thermal Forces 750
24.3 Modeling at the Nano-Bio Interface 752
24.3.1 Tension Between Universality and Specificity 752
24.3.2 Atomic-Level Analysis of Biological Systems 753
24.3.3 Continuum Analysis of Biological Systems 753
24.4 Nature's Nanotechnology Revealed: Viruses as a Case Study 755
24.5 Concluding Remarks 760
References 761

25 **Mechanical Properties of Nanostructures** 763
25.1 Experimental Techniques for Measurement of Mechanical Properties of Nanostructures 765
25.1.1 Indentation and Scratch Tests Using Micro/Nanoindenters 765
25.1.2 Bending Tests of Nanostructures Using an AFM 765
25.1.3 Bending Tests Using a Nanoindenter 769
25.2 Experimental Results and Discussion 770
25.2.1 Indentation and Scratch Tests of Various Materials Using Micro/Nanoindenters 770
25.2.2 Bending Tests of Nanobeams Using an AFM 773
25.2.3 Bending Tests of Microbeams Using a Nanoindenter 777

25.3 Finite Element Analysis of Nanostructures with Roughness and Scratches ... 778
25.3.1 Stress Distribution in a Smooth Nanobeam ... 779
25.3.2 Effect of Roughness in the Longitudinal Direction ... 781
25.3.3 Effect of Roughness in the Transverse Direction and Scratches ... 781
25.3.4 Effect on Stresses and Displacements for Materials That Are Elastic, Elastic–Plastic, or Elastic–Perfectly Plastic ... 784
25.4 Closure ... 785
References ... 786

Part D Molecularly Thick Films for Lubrication

26 **Nanotribology of Ultrathin and Hard Amorphous Carbon Films** ... 791
26.1 Description of Commonly Used Deposition Techniques ... 795
26.1.1 Filtered Cathodic Arc Deposition Technique ... 798
26.1.2 Ion Beam Deposition Technique ... 798
26.1.3 Electron Cyclotron Resonance Chemical Vapor Deposition Technique ... 799
26.1.4 Sputtering Deposition Technique ... 799
26.1.5 Plasma-Enhanced Chemical Vapor Deposition Technique ... 799
26.2 Chemical Characterization and Effect of Deposition Conditions on Chemical Characteristics and Physical Properties ... 800
26.2.1 EELS and Raman Spectroscopy ... 800
26.2.2 Hydrogen Concentrations ... 804
26.2.3 Physical Properties ... 804
26.2.4 Summary ... 805
26.3 Micromechanical and Tribological Characterizations of Coatings Deposited by Various Techniques ... 805
26.3.1 Micromechanical Characterization ... 805
26.3.2 Microscratch and Microwear Studies ... 813
26.3.3 Macroscale Tribological Characterization ... 822
26.3.4 Coating Continuity Analysis ... 826
References ... 827

27 **Self-Assembled Monolayers for Controlling Adhesion, Friction and Wear** ... 831
27.1 A Primer to Organic Chemistry ... 834
27.1.1 Electronegativity/Polarity ... 834
27.1.2 Classification and Structure of Organic Compounds ... 835
27.1.3 Polar and Nonpolar Groups ... 838
27.2 Self-Assembled Monolayers: Substrates, Head Groups, Spacer Chains, and End Groups ... 839
27.3 Tribological Properties of SAMs ... 841
27.3.1 Surface Roughness and Friction Images of SAMs Films ... 844
27.3.2 Adhesion, Friction, and Work of Adhesion ... 844

27.3.3 Stiffness, Molecular Spring Model, and Micropatterned SAMs 848
27.3.4 Influence of Humidity, Temperature, and Velocity on Adhesion and Friction ... 850
27.3.5 Wear and Scratch Resistance of SAMs ... 853
27.4 Closure ... 856
References ... 857

28 **Nanoscale Boundary Lubrication Studies** ... 861
28.1 Lubricants Details ... 862
28.2 Nanodeformation, Molecular Conformation, and Lubricant Spreading ... 864
28.3 Boundary Lubrication Studies ... 866
28.3.1 Friction and Adhesion ... 866
28.3.2 Rest Time Effect ... 869
28.3.3 Velocity Effect ... 871
28.3.4 Relative Humidity and Temperature Effect ... 873
28.3.5 Tip Radius Effect ... 876
28.3.6 Wear Study ... 879
28.4 Closure ... 880
References ... 881

29 **Kinetics and Energetics in Nanolubrication** ... 883
29.1 Background: From Bulk to Molecular Lubrication ... 885
29.1.1 Hydrodynamic Lubrication and Relaxation ... 885
29.1.2 Boundary Lubrication ... 885
29.1.3 Stick Slip and Collective Phenomena ... 885
29.2 Thermal Activation Model of Lubricated Friction ... 887
29.3 Functional Behavior of Lubricated Friction ... 888
29.4 Thermodynamical Models Based on Small and Nonconforming Contacts ... 890
29.5 Limitation of the Gaussian Statistics – The Fractal Space ... 891
29.6 Fractal Mobility in Reactive Lubrication ... 892
29.7 Metastable Lubricant Systems in Large Conforming Contacts ... 894
29.8 Conclusion ... 895
References ... 895

Part E Industrial Applications and Microdevice Reliability

30 **Nanotechnology for Data Storage Applications** ... 899
30.1 Current Status of Commercial Data Storage Devices ... 901
30.1.1 Non-Volatile Random Access Memory ... 904
30.2 Opportunities Offered by Nanotechnology for Data Storage ... 907
30.2.1 Motors ... 907
30.2.2 Sensors ... 909
30.2.3 Media and Experimental Results ... 913
30.3 Conclusion ... 918

References 919

31 **The "Millipede" – A Nanotechnology-Based AFM Data-Storage System** 921
31.1 The Millipede Concept 923
31.2 Thermomechanical AFM Data Storage 924
31.3 Array Design, Technology, and Fabrication 926
31.4 Array Characterization 927
31.5 *x*/*y*/*z* Medium Microscanner 929
31.6 First Write/Read Results with the 32×32 Array Chip 931
31.7 Polymer Medium 932
31.7.1 Writing Mechanism 932
31.7.2 Erasing Mechanism 935
31.7.3 Overwriting Mechanism 937
31.8 Read Channel Model 939
31.9 System Aspects 943
31.9.1 [peserror]PES Generation for the Servo Loop 943
31.9.2 Timing Recovery 945
31.9.3 Considerations on Capacity and Data Rate 946
31.10 Conclusions 948
References 948

32 **Microactuators for Dual-Stage Servo Systems in Magnetic Disk Files** 951
32.1 Design of the Electrostatic Microactuator 952
32.1.1 Disk Drive Structural Requirements 952
32.1.2 Dual-Stage Servo Configurations 953
32.1.3 Electrostatic Microactuators: Comb-Drives vs. Parallel-Plates 954
32.1.4 Position Sensing 956
32.1.5 Electrostatic Microactuator Designs for Disk Drives 958
32.2 Fabrication 962
32.2.1 Basic Requirements 962
32.2.2 Electrostatic Microactuator Fabrication Example 962
32.2.3 Electrostatic Microactuator Example Two 963
32.2.4 Other Fabrication Processes 966
32.2.5 Suspension-Level Fabrication Processes 967
32.2.6 Actuated Head Fabrication 968
32.3 Servo Control Design of MEMS Microactuator Dual-Stage Servo Systems 968
32.3.1 Introduction to Disk Drive Servo Control 969
32.3.2 Overview of Dual-Stage Servo Control Design Methodologies 969
32.3.3 Track-Following Controller Design for a MEMS Microactuator Dual-Stage Servo System 971
32.3.4 Dual-Stage Seek Control Design 976
32.4 Conclusions and Outlook 978
References 979

33 **Micro/Nanotribology of MEMS/NEMS Materials and Devices** 983
33.1 Introduction to MEMS 985
33.2 Introduction to NEMS 988
33.3 Tribological Issues in MEMS/NEMS 989
33.3.1 MEMS 989
33.3.2 NEMS 994
33.3.3 Tribological Needs 995
33.4 Tribological Studies of Silicon and Related Materials 995
33.4.1 Tribological Properties of Silicon and the Effect of Ion Implantation 996
33.4.2 Effect of Oxide Films on Tribological Properties of Silicon 998
33.4.3 Tribological Properties of Polysilicon Films and SiC Film 1000
33.5 Lubrication Studies for MEMS/NEMS 1003
33.5.1 Perfluoropolyether Lubricants 1003
33.5.2 Self-Assembled Monolayers (SAMs) 1004
33.5.3 Hard Diamond-like Carbon (DLC) Coatings 1008
33.6 Component-Level Studies 1009
33.6.1 Surface Roughness Studies of Micromotor Components 1009
33.6.2 Adhesion Measurements 1011
33.6.3 Static Friction Force (Stiction) Measurements in MEMS 1014
33.6.4 Mechanisms Associated with Observed Stiction Phenomena in Micromotors 1016
References 1017

34 **Mechanical Properties of Micromachined Structures** 1023
34.1 Measuring Mechanical Properties of Films on Substrates 1023
34.1.1 Residual Stress Measurements 1023
34.1.2 Mechanical Measurements Using Nanoindentation 1024
34.2 Micromachined Structures for Measuring Mechanical Properties 1024
34.2.1 Passive Structures 1025
34.2.2 Active Structures 1028
34.3 Measurements of Mechanical Properties 1034
34.3.1 Mechanical Properties of Polysilicon 1034
34.3.2 Mechanical Properties of Other Materials 1036
References 1037

35 **Thermo- and Electromechanics of Thin-Film Microstructures** 1039
35.1 Thermomechanics of Multilayer Thin-Film Microstructures 1041
35.1.1 Basic Phenomena 1041
35.1.2 A General Framework for the Thermomechanics of Multilayer Films 1046
35.1.3 Nonlinear Geometry 1054
35.1.4 Nonlinear Material Behavior 1058
35.1.5 Other Issues 1061
35.2 Electromechanics of Thin-Film Microstructures 1061
35.2.1 Applications of Electromechanics 1061
35.2.2 Electromechanics Analysis 1063

35.2.3 Electromechanics – Parallel-Plate Capacitor 1064
35.2.4 Electromechanics of Beams and Plates 1066
35.2.5 Electromechanics of Torsional Plates 1068
35.2.6 Leveraged Bending 1069
35.2.7 Electromechanics of Zipper Actuators 1070
35.2.8 Electromechanics for Test Structures 1072
35.2.9 Electromechanical Dynamics: Switching Time 1073
35.2.10 Electromechanics Issues: Dielectric Charging 1074
35.2.11 Electromechanics Issues: Gas Discharge 1075
35.3 Summary and Mention of Topics not Covered 1078
References 1078

36 **High Volume Manufacturing and Field Stability of MEMS Products** 1083
36.1 Manufacturing Strategy 1086
36.1.1 Volume 1086
36.1.2 Standardization 1086
36.1.3 Production Facilities 1086
36.1.4 Quality 1087
36.1.5 Environmental Shield 1087
36.2 Robust Manufacturing 1087
36.2.1 Design for Manufacturability 1087
36.2.2 Process Flow and Its Interaction with Product Architecture . 1088
36.2.3 Microstructure Release 1095
36.2.4 Wafer Bonding 1095
36.2.5 Wafer Singulation 1097
36.2.6 Particles 1098
36.2.7 Electrostatic Discharge and Static Charges 1098
36.2.8 Package and Test 1099
36.2.9 Quality Systems 1101
36.3 Stable Field Performance 1102
36.3.1 Surface Passivation 1102
36.3.2 System Interface 1105
References 1106

37 **MEMS Packaging and Thermal Issues in Reliability** 1111
37.1 MEMS Packaging 1111
37.1.1 MEMS Packaging Fundamentals 1112
37.1.2 Contemporary MEMS Packaging Approaches 1113
37.2 Hermetic and Vacuum Packaging and Applications 1116
37.2.1 Integrated Micromachining Processes 1117
37.2.2 Post-Packaging Processes 1118
37.2.3 Localized Heating and Bonding 1119
37.3 Thermal Issues and Packaging Reliability 1122
37.3.1 Thermal Issues in Packaging 1122
37.3.2 Packaging Reliability 1124
37.3.3 Long-Term and Accelerated MEMS Packaging Tests 1125

37.4 Future Trends and Summary 1128
References 1129

Part F Social and Ethical Implication

38 **Social and Ethical Implications of Nanotechnology** 1135
38.1 Applications and Societal Impacts 1136
38.2 Technological Convergence 1139
38.3 Major Socio-technical Trends 1141
38.4 Sources of Ethical Behavior 1143
38.5 Public Opinion 1145
38.6 A Research Agenda 1148
References 1149

Acknowledgements 1153
About the Authors 1155
Detailed Contents 1171
Subject Index 1189

Subject Index

1, 1′-biphenyl-4-thiol (BPT) 843
4, 4′-dihydroxybiphenyl (DHBp) 843
μCP (microcontact printing) 187, 197, 834
μTAS (micro-total analysis systems) 257
γ-modified geometry 700
1,4-phenylenediamine (pDA) 916
1/f noise 387, 957
16-mercaptohexadecanoic acid thiol (MHA) 843
1-D localization
– effects 120
– theory 120
1-D to 3-D transition in magnetic field 122
1-ethyl-3-(3-diamethylaminopropyl) carbodiimide (EDC) 80, 296
2-D FKT model 638
2-DEG
– two-dimensional electron gas 438
2-D-histogram technique 647
2-amino-4,5-imidazoledicarbonitrile (AIDCN) 918
2-mercaptoethylamine HCl 478
2-pyridyldithiopropionyl (PDP) 478
3,3′-dimethyl bipyridinium 14
3-D bulk state 428
3-D coil inductor 242
3-D switching architecture 238
3-D-force measurements 358
3-nitrobenzal malonitrile (NBMN) 916

A

abrasive wear 649, 652
Abrikosov lattice 440
ABS (air-bearing surface) 968
ac electrochemical deposition 104
ac electrodeposition 103
accelerated friction 822
acceleration energy 798
accelerometer 230
– lateral 231
– packaging 1100
– test 1101
– three-axis 233
– vertical 232
– wafer fabrication 1091
acetonitrile 14, 19
acetylene (C_2H_2) 799
acid-treated SWNT 64
acoustic emission 690
actin-myosin motor 751
activation energy barrier 482
active linearization 361
active matrix electronic paper 195
active microfluidic devices 257
active microvalves 258
active structures 1028
actuated head 954
– fabrication 968
actuated slider 954, 968
– configuration 954
actuated suspension 953
actuation
– pneumatic 258
actuator
– miniaturized 986
actuator failure modes 1074
adatoms 392
addressing nanoscale bits of data 901
adenine 10
adenosine triphosphate (ATP) 741
adhesion 325, 498, 507, 515, 605, 799, 844, 930, 983
– at SAM interfaces 618
– control 832
– energy 619
– force 452, 546, 547, 552, 564, 566, 567, 619, 621
– force, quantized 558, 584
– hysteresis 567, 568, 573
– hysteresis, relation to friction 573, 576–579, 585
– influence of humidity on 850
– measurement 1011
– measurement techniques 606
– mechanics 564–566
– performance 866
– primary minimum 544, 555, 559
– promoter 188
– quantization 623
– rate-dependent 567, 568
adhesion force
– total 616
adhesion-controlled friction 571, 572, 574
adhesive
– coating 479
– interaction 732
– material 610
– surfaces 731
– wear 728
adhesive force 335, 646, 662, 867, 1004
– increase 876
– intrinsic 832
– mapping 537, 869
– measurement 503
adhesive tape film 907
adiabatic limit 482
adsorbate 404, 695
adsorbed
– insulator substrate 400
– water 875
– water film 878
adsorption 64, 79, 476
– capacity 80
– sites in MWNT 65
AES (Auger electron spectroscopy) measurement 826
AFAM (atomic force acoustic microscopy) 326
affinity maturation 316
affinity selection 297
affinity-based targeting 301
AFM (atomic force microscope) 173–175, 743, 765
– adhesion measurement 620
– *Binnig* design 333
– calculated sensitivity 355
– cantilever 332, 350, 610
– cantilever array 923
– commercial 333
– construction 288
– contact mode 331
– control electronics 362
– design optimization 348
– designs 376
– feedback loop 362
– for UHV application 635
– image 392
– imaging signal 386
– instrumentation 347
– manufacturers 333
– microscratch technique 853
– probe construction 338
– probes 378
– pull-off force 619

– resolution 375
– set-up 450, 453
– spectroscopy 458
– surface height map 817
– test 818
– thermomechanical recording 924
– tip 288, 379, 386, 450, 452, 477, 480, 1013
– tip containing antibodies 489
– tip radius 518
– tip sensor design 476
– tip size 622
AFM image
– tapping mode 914
Ag 395, 397
– on Nb(110) 429
– trimer 395
air bag sensor 1084, 1091
air damping 470
air gap capacitance 1075
air induced oscillations 348
air/water interface 479
air-bearing surface (ABS) 968
Al_2O_3-TiC head 825
Al_2O_3 54, 462, 633
– grains 508
– ultrafiltration membrane 669
Al_2O_3-TiC composite 508
Al_2O_3-TiC slider 998
$Al_xGa_{1-x}As$ 248
AlAs 248
alcohol 837
aldehydes 837
aligned carbon nanotubes 58
alignment accuracy 195
alignment of nanowires 109
alkali halides 397
alkaline phosphatase (AP) 271
alkane 835, 888, 889
alkanethiolate 187
– film 841
– on gold 187
– on palladium 187
– on silver 187
alkanethiols 404
alkylsilane film 619
all-fiber interferometer 418
all-MEMS configuration 907
all-metal cantilever 912
all-optical logic gates 21
all-optical switching network 237
alloy 662
all-silicon cantilever 926
alumina template 127
aluminum
– gates 34
– mirror array 1088
aluminum oxide 54, 462, 633
aluminum-based surface micromachining 241
AM (amplitude modulation)
– AFM 388
– mode 455
amide 838
amine 838
Amontons law 454, 517, 571, 574, 577, 580, 646, 718
amorphous
– oxide layer 106
– surfaces 544
amorphous carbon 49, 793
– chemical structure 800
– coatings 800, 806
– phase 48
amperometric time-based detection method 272
amphiphilic molecular building blocks 23
amplitude feedback 465
amplitude modulation (AM) 388, 456
– mode 331
– SFM (scanning force microscope) 433
amplitude of vibration 911
anchor group 840
AND operator 15
angle of helicity 41
angle of twist formula 355
anion-terminated tip 399
anisotropy 704
– of friction 642
annealing effect 1036
anodic alumina (Al_2O_3) 100
– oxide films 102
– templates 103
anodic bonding 1096, 1115
anodization 100
ANSYS
– analysis 1123
– stress prediction 1124
anthracene 20
– channel 21
antibiotic-resistant pathogens 293
antibody 296
– directed enzyme-prodrug therapy (ADEPTS) 301
– production 312
antibody–antigen 475
– binding 283
– complex 484
– recognition 491
antibody-coated beads 271
antiferromagnetic spin ordering 403
antigen presenting cell (APC) 312
antigen recognition 282
antimicrobial chemotherapy 293
anti-stiction
– agent (ASA) 1104
– material 1104
AP (alkaline phosphatase) 271
APC (antigen presenting cell) 312
APCVD (atmospheric pressure chemical vapor deposition) 770
apparent viscosity 888
applications
– electrical 131
arc discharge 798
areal density 900, 901, 907, 925, 932, 947, 952
Arg-Gly-Asp (RGD) 489
armchair-type SWNT 42
array
– cantilever 927
– characterization 927
– chip 923
– of aligned nanowires 115
– of nanotubes 53
Arrhenius equation 1127
artifacts 415
artificial double helix 12
as-deposited film 1035
aspect ratio 127
asperity-asperity interaction 618
as-released curvature 1059
assembled nanotube probes 378
assembly approaches 288
assembly of nanostructures 294
association process 476
atmospheric pressure chemical vapor deposition (APCVD) 770
atomic
– interaction force 452
– manipulation 419
– motion imaging 420
atomic force acoustic microscopy (AFAM) 326
atomic force microscope (AFM) 114, 288, 289, 303, 325, 331, 371, 385, 404, 476, 498, 546, 547, 549, 585, 588, 688, 717, 743, 841, 861, 862, 899, 922
atomic resolution 330, 382, 388, 389, 427, 476, 480
– imaging 391, 394–397, 462
atomic-scale
– dissipation 655
– force measurement 347

– friction 507
– hysteresis 468
– image 327, 342, 434
– roughness 721
atomistic computer simulation 398, 724
ATP hydrolysis 742
ATP(adenosintriphosphate) synthase 741
attraction
– long-range 550, 562
attractive force–distance profile 480
attractive interaction 864
Au film 771
Au microbeam 777, 778, 785
Au tip 619
Au(111) 423, 848, 850
– surface 619
Aubry transition 728
Au-coated AFM tip 619
austenite 665
average
– distance 749
– lifetime 482
– stress 1049
axial strength 672
azopyridine 19

B

backbone chain 839
backflow 261
bacmid 283
bacteriophage 744, 755
bacteriorhodopsin 742
BaF_2 398
ball and spring model 724
ball grid arrays (BGA) 1113
ballistic phonon transport 126
ballistic transport 115, 121
balloon angioplasty 189
bamboo texture 52, 61
band gap tunability 128
band structure
– electronic 128
barcode reading 233
barcode tags 136
barrier-hopping 887
– fluctuations 890
basic switching operations 15
batch fabrication 249
– techniques 373
batch nanotube tip fabrication 380
B-cell epitopes 311
bc-MWNT (bamboo-concentric) 44
BDCS (biphenyldimethylchlorosilane) 849
beam
– two-layer 1045
beam bending energy 745
beam deflection 455
beam failure 775
beam-deflection FFM 633
beam-steering mirror 239
behavior
– unethical 1143
Bell's formula 483
bending
– moment 767, 783
– stiffness 349
– strength 767, 774, 775, 785
– stress 767, 781
– test 765, 766, 769
Berkovich
– indenter 765
– indenter tip 677
– pyramid 696
– tip 815
BGA (ball grid arrays) 1113
bh-MWNT (bamboo-herringbone) 44
$Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCCO) 431
$Bi_{1-x}Sb_x$ nanowires 123
bias voltage 328
biaxial strain 1042
biaxial stress 1053
BiCMOS 229
bidirectional micropump 261
bifurcation 1056
bilayer
– beam 1054
– cantilever 1046
– film 1059
– model-lipid-lubricant 889
– plate 1050
– system 1042
bimetallic catalyst 51, 62
binary compound 397
bio force probe (BFP) 480
biochemical fluids 254
biochemical reaction 262, 270
biocompatibility 263, 311
bioconjugate chemistry 295, 296
bioconjugation method 288
biofilter 271
– surface-mounted 272
biofluidic chip 273, 987
biofunctionalized cantilever 742, 743, 745
biogenic amorphous silicon shells 314
biological
– activity 316
– affinity 297
– affinity reagents 303
– device 742
– device components 304
– evolution 751
– macromolecules 299
– nanotechnology 739, 760
– self-assembly 293
biomaterials 711
biomechanics 675
biomedical device 189
BioMEMS (Biological or Biomedical Microelectromechanical Systems) 253, 262, 764, 987
– actuation 269
biomimetic valves 262
biomolecules bound to carbon nanotubes 303
biomolecules in therapeutic nanodevices 293
bio–nano interface 744
BioNEMS (Biological or Biomedical Nanoelectromechanical Systems) 764, 989
bio-sampling
– magnetic bead-based 271
biosensors 80, 302
bio-surfaces 489
biotechnology 1139
– limits of 281
biotin 80, 476
biotin-avidin 622
– spectrum 487
biotin-directed IgG 476
biotinylated AFM tip 489
biphenyl 419
bipyridine
– building block 12, 31
– centered LUMOs 26
birefringent crystal 352
bismuth nanowire 118, 134
bistability 459, 916
bit pitch (BP) 943
bit strings 938
bits per inch (BPI) 952
blister test 691
Bloch states 428
Bloch wave 427
block co-polymer 291
block-like debris 823
blood cell analysis 310
blood flow measurement 304
blood-pool agents 306
BN nanotubes 68

Bode plot 976
Bohr radius of excitons 135
Boltzmann ansatz 482
Boltzmann distribution 887
bond
– angle 67
– breakage 483
– lifetime 482
– rupture 487
– scission 623
– strength 550
bonded lubricant film 864
bonded PFPE 1012
bonding
– energy 386
– kinetics 893
– techniques 1114
bone
– lamellae 675
– lamellation 679
– tissue 675
boron
– diffusion 227
– doped Si 205
– ion implantation 330
bottom-up
– approach 99, 107, 285
– chemical strategies 10
– materials 281
– scheme 133
boundary
– element (BE) 1063
– film formation 861
– lubricant 729, 861
– lubricant film 832
– lubrication 530, 569–571, 580, 583, 585, 852, 866, 885
– lubrication measurement 507
– lubrication regime 884
– scattering effects 125
bovine serum albumin (BSA) 476
bpsi (bits per square inch) 901
BPT (1, 1′-biphenyl-4-thiol) 854, 1007
BPTC (cross-linked BPT) 1007
branched hydrocarbon lubricants 889
breakdown
– electrical 1076
bridging of polymer chains 562, 563
broken beams 777
broken coating chip 817
brush *see* polymer brush
brush-like structures 109
buckling 808, 812, 1026
– force 377
– stress 813
bulge test 691
bulk
– addressing 22
– atoms 387
– conduction band 424
– diamond 793, 801, 803
– etched silicon wafer 263
– fluid transport 261
– graphitic carbon 824
– micromachining 229
– Si wafers 205
– state 428
– viscosity 731
– xenon 436

C

C_{70} 792
C_2 moieties 51
C_{60} 32, 41, 49, 404, 647, 792
– film 330
– fullerenes 289
– island 641, 642
– multilayered film 404
– SWNT (single wall nanotube) 70
C_{60}-terminated film 843
Ca ion 399
cadherin-mediated adhesion 488
CaF_2 398
CaF_2 tip 652
CaF_2(111) 399
CaF_2(111) surface 399, 652
calculated tapping behavior 911
calculus of variations 745
calibration 610, 698, 699
calorimetric experiments 482
cantilever 387, 390, 457, 481
– all-metal 912
– all-silicon 926
– array 926, 927
– axis 418
– base 469
– beam 1015, 1061
– beam array (CBA) 1011
– beam model 1067
– biofunctionalized 742, 743, 745
– biosensors 303
– cell 926
– deflection 332, 372, 387, 389, 418, 465, 480, 632, 745, 848
– deflection calculation 344
– diode array 928
– driven 455
– effective mass 633
– eigenfrequency 415
– elasticity 636
– fabrication 909
– flexible 332
– foil 373
– heater 925
– material 339, 387
– motion 331, 913
– mount 336
– oscillation 457, 469
– Q-factor 433
– resonance 451
– resonance behavior 350
– resonance frequency 481, 611, 890, 1029
– sensor 925
– spring 454
– spring constant 485
– stainless steel 813
– stiffness 338, 505, 610, 926
– temperature 940
– thickness 632, 633
– tip 303, 454
– triangular 339, 349, 354
– untwisted 335
cantilever-based probes 433
capacitance detection 356
capacitive
– accelerometers 230
– detection 361
– detector 331
– displacement sensor 1033
– forces 465
– position sensing 956
– sensor 227, 230
– transducer 245
capacitor-like device 913
capacity growth 903
capillarity-driven stop valve 265
capillary
– effect on adhesion 607
– electrophoresis (CE) 255
– force 544, 559, 564, 619, 662
– force curve 617
– formation 614
– pressure 265
– stiffness 614
capsid wall 760
carbon 792
– crystalline 792
– film 865
– magnetron sputtered 799
– shells 52
– source 57
– spacer chain 853
– superactivated 77

carbon coating
– unhydrogenated 804
carbon nanofiber
– vapor grown 56
carbon nanotube (CNT) 55, 105, 287, 425, 667
– adsorption properties 76
– application 73
– catalyst-free growth 59
– catalyst-supporting materials 75
– chemical reactivity 67
– diameters 63
– FETS 303
– field emission 74
– formation 46
– growth mechanisms 59
– heterogeneity 63
– in situ filling 69
– maximum current density 74
– mechanical properties 376
– molten state filling 69
– oxidation 72
– production 46
– properties 63
– quantum wires 66
– sublimation filling 70
– synthesis conditions 61
– tip 342, 375, 376, 501
– wet chemistry filling 69
carbon–carbon distance 436
carboxyl acid 837
carboxylates ($RCOO^-$) 404
cardiovascular
– sensors 310
– system 310
– tissue engineering 310
carrier
– density 118, 122
– gas 804
– mean free path 116, 118, 121
– mobility 118
Casimir force 545
casting 186
catalysis 397
catalysis-enhanced disproportionation 54
catalysis-enhanced thermal cracking 54
catalyst 49, 106, 668
– nanoparticles 58
– preparation 54
catalyst-based SWNT 61
catalyst-free 57
catalytic chemical vapor deposition (CCVD) 52, 55, 58, 60, 61
catalytic decomposition 54
catalytically grown MWNT 673
cathode deposit 49, 51
cathodic arc carbon 806
CCVD *see* catalytic chemical vapor deposition
CCVD method 61
CDS (correlated double sampling) technique 961
CdSe nanorods 136
cell adhesion 489
cellular immune responses 311
cellular phone 904
$CeO_2(111)$ 400
ceramic
– matrix composites 81
– slider 824
– tip 74
cerpacs 1101
CFM (chemical force microscopy) 73, 489, 619
chain adsorption 894
chalcogenide 906, 910, 915
change of meniscus 1005
changing of surface conditions 117
channel etching process 255
channel-liquid pair 261
chaotic mixer 266
characteristic distance 588
characteristic slip time 894
characteristic velocity 889
charge density wave (CDW) 428
charge exchange interactions 545, 552, 553
charge fluctuation forces *see* ion correlation forces
charge separation 1116
charge transfer interactions *see* charge exchange interactions
charge trapping 913
charge-controlled pull-in equation 1064
check valve design 260
chelate metal cations 12
chemical
– bond 386
– bonding 388, 863
– bonding force 331
– characterization 800
– composition 115
– detection devices 137
– force 386
– force microscopy (CFM) 73, 489, 619
– heterogeneity 568
– input 16
– interaction force 452
– sensors 74
– signal 19, 20
– synthesis 12
– vapor deposition (CVD) 104, 793
– vapor infiltration 57
chemisorption 861
chemistry route 57
chemoselective conjugation 295, 296
chemotherapy 307
chip level integration 1093
chip-on-flex (COF) 1113
clamped-clamped mechanical beams 247
classical finite size 116
classical size effect 120
clock field 946
clogging 56
closed loop sensitivity 973
clustered acetate 406
CMOS electronics 230
CMOS static random access memory (SRAM) 235
CMOS sustaining electronics 245
c-MWNT (concentric multiwall nanotube) 43, 48, 60, 67
Co clusters 425
CO disproportionation 54
CO on Cu(110) 419
coated particles 306
coating
– continuity 826
– damage 823
– failure 820
– hardness 818
– hydrophobic 1014
– mass density 801
– microstructure 800
– thickness 810, 812
coating–substrate interface 819
coefficient
– effective 647
coefficient of friction 501, 510, 516, 535, 579, 634, 771, 773, 809, 813, 823, 824, 844, 868, 872, 996, 1005
– Si(100) 531
– Z-15 531
– Z-DOL (BW) 531
coefficient of friction *see* friction coefficient
coefficient of friction relationship 343
coefficient of thermal expansion (CTE) 1124
co-energy 1064
coercivity 137

COF (chip-on-flex) 1113
cognate receptors 298
cognitive ligands 476
cognitive science 1139
coherence length 423
cohesive
– potential 759
– surface model 759
coil inductor 242
cold welding 553, 579
collagen fibers 679
collective phenomenon 885, 888
colloidal
– forces 545
– probe 548
colossal magneto resistive effect 439
comb storage device 907
comb-drive 955, 1030
– actuator 955
– resonator 245
combined AFM nanoindenter device 676
commensurate Cu(111) 728
commensurate system 723
commercial data-storage device 907
commercial MEMS devices 991
communicating
– between compatible molecular components 19
– force 748
– molecular switches 19
compact bone
– lamellae 678
– mean hardness 680
complex
– bonds 486
– dielectric function 127
– fluidic systems 886
– logic functions with molecular switches 19
component
– failure 764
– supplier 902
– surface 1009
composite 662
composite material 81
compositionally modulated nanowires 106
compression energy 887
compressive stress 805, 812, 1035, 1044
computer simulation 718
– forces 553, 557, 561, 564, 565, 585
– friction 553, 573, 574, 579, 580, 583, 585, 586, 590, 591, 594
concentration
– critical 650
concentric
– graphenes 43
– texture 61
– type (c-MWNT) 43
conditional release of therapeutics 288
conductance 117
conductance quantization in metallic nanowires 115
conducting polymers 103, 304
conductive polymeric materials 304
conductivity 136
– electrical 133
confined
– complex liquids 891
– nanostructure 68
– simple liquids 888
confinement 544, 545, 570, 584, 593
conformation 864
conformational defect 849
conforming contacts 894
connector
– miniaturized 986
constant
– amplitude (CA) 464
– amplitude FM mode 468
– current mode 328
– excitation mode (CE) 464
– force mode 362
– height mode 328, 399
constrained elastic media 1061
consumer electronics 904
consumer nanotherapeutic device 288
contact
– AFM (atomic force microscope) 455
– AFM dynamic method 662
– angle 616, 833
– angle goniometer 847
– angle of SAM 846
– conductance 649, 913
– elastic 703
– mechanics *see* adhesion mechanics
– meniscus 619
– mode 373, 909
– mode photolithography 195
– printing 186, 200, 833
– radius 678
– resistance 690
– stiffness 677, 689, 698, 701, 810, 894
– stress 872
– value theorem 575
contact area 548, 650, 651, 677, 699, 709
– apparent macroscopic 571, 576, 580
– true molecular 547, 549, 566, 571, 574–576, 594
contamination 332
continuous micropump 269
continuous stiffness measurement (CSM) 810
continuum
– analysis 753
– model 638
– theory 554, 556, 558, 559, 569, 580
contour plot 1057
contour-mode disk resonator 246
contraction 1044
contrast 435
– agent design 306
– agents 304
– enhancing agents for medical imaging 304
– enhancing nanoparticle 306
– formation 397
control
– of nanotherapeutic action 302
– of position 109
– over nanowire diameter distribution 106
– system 359, 362
– theory 1144
controlled desorption 419
controlled evaporation 269
controlled geometry (CG) 330
controlled nanoscale architecture 294
controlled substances 1140
controlled triggering of therapeutic action 298
controller design 969
CoO 402
Co-O film on PET 524
Cooper pairs 429
coordinated molecular process 886
copper adhesion 963
copper hexadecafluorophthalocyanine 196
copper tip 652, 727
corannulene 61
cores of carbon nanotubes 105
core-sheath nanowire 114

core-shell structure 132
Coriolis acceleration 232, 1095
correlated-double-sampling (CDS) 960
corrosion 1116, 1125
cost of goods (COGs) 285, 307
co-transporter 489
Couette flow 569, 593, 885
Coulomb
– force 396
– interaction 66
– law of friction 644
coupled electromechanics 1063
covalent
– bonds 10
– scaffolds 12
Cr binding layer 619
Cr coating 432
Cr(001) 427
crack spacing 524
cracks 807
creativity at the nanoscale 1135
creep 415, 691, 1125
– effect 528
– relaxation 1060
creep model
– linear 890
critical
– concentration 650
– curvature 1056
– degree of bending 1031
– load 809
– magnetic field strength 122
– normal load 853
– position 636
– shear stress 570–572, 577, 585, 586
– temperature 423, 434, 1059
– time 894
– velocity 585, 587, 590, 591
crosshair alignment 195
crosslinked BPT (BPTC) 843
crosstalk 363, 927, 1066
cross-track distance 945
cryostat 415
crystal
– growth direction 103
– orientation 206
– structure 333
– surfaces 375
crystal growth orientation, preferred 104
crystalline
– carbon 792
– silicon 254
– surfaces 544
CSM (continuous stiffness measurement) 810
CTE (coefficient of thermal expansion) 1124
CTL (cytotoxic T-cell) responses 311
Cu(001) 397
Cu(100) 640
Cu(100) substrate 728
Cu(100) tip 728
Cu(111) 423, 639, 640, 655
Cu(111) surface states 427
Cu(111) tip 653
cube corner 696, 704
culture of engineering 315
cuprates 431
current density effect 1036
current rectification 116
current-voltage characteristics 116, 916
curvature 1044
– critical 1056
– effects 723
– measured 1053
– variation 1046
curved electrode
– actuator 1070
– shape 1071
custom-designed MEMS product 1112
cut-off distance 552, 573
CVD (chemical vapor deposition) 104, 793
cyclic
– fatigue 810, 1036
– olefin copolymers (COC) 257
– thermomechanical load 1061
– voltammograms 29
cytoplasmic surface 478
cytosine 10
cytoskeletal filament 747
cytotoxic therapeutics 307
cytotoxic T-lymphocytes (CTLs) 311
cytotoxin deposition 309

D

(*d*, *k*) codes 947
D_2O 425
damage 553, 569–588
damage mechanism 818
damped harmonic oscillator 454
damping 664
– effective 470
– effects 451
– pneumatic 418
data
– lifetime 905
– rate limitation 922
– rates 948
– storage 900
– storage application 912
– storage device 922
– storage system 994
dc electrochemical deposition 103, 104
de Broglie wavelength 116
Deborah number 583
debris 823
Debye exponential relaxation 885, 892
Debye length 555, 557
decoupled design 970
decoupled track-following control 972
deep reactive ion etching (DRIE) 156, 162, 254, 963
defect motion 415
defect nucleation 702
defect production 651
defects in channel 255
deflection 457, 1054
– curve 1071
– maximum 1071
– measurement 387
– noise 433
deformation 547, 549, 1041, 1047
– behavior 1058
– elastic 450, 566, 581, 647
– of microtubule 674
– plastic 553, 580
– regime 1046
deformed beam 778
degradability 311
degrees of freedom 753
delamination 704, 804, 808, 812, 1125
delivery of therapeutics 296
Demnum-type PFPE 537
– lubricant film 869
dendrimer-based therapeutic 284
dendritic
– cells (DCs) 312
– polymers 310
density
– functional theory (DFT) 78
– modulation 730
– of defects 193, 195
density of states (DOS) 66
– electronic 133
dental enamel 711

depletion
– attraction 545, 562
– interaction 562
– stabilization 545, 562
deployment of functional biomolecules 288
deposition
– conditions 206
– rate 798
– techniques 795, 796
Derjaguin approximation 546, 557, 559
Derjaguin–Landau–Verwey–Overbeek (DLVO) theory 555
Derjaguin–Muller–Toporov (DMT)
– model 647
– theory 566
design rule 348
designed molecules 12
design-for-manufacturability 1086
design-for-volume 1086
detection 939
– circuit 939
– systems 332, 350
device
– architectures 288
– components communication 304
– design 280
– electronic 133
– fluidic 253
– molecule-based 24, 28
– scaling 989
dewetting 868, 1003, 1013
DFM (dynamic force microscopy) 433, 471, 490, 501
DFS (dynamic force spectroscopy) 484
DHBp (4, 4′-dihydroxybiphenyl) 843, 854, 1007
diameter-dependent 116
diamond 67, 373, 379, 673, 698, 792
– coated cantilever 914
– coating 804, 833
– film 793
– like amorphous carbon coating 800
– like carbon (DLC) 524, 1003
– powder 51
– tip 504, 651, 662, 676, 814, 819, 841, 866, 999
diatoms 314
diblock copolymers 102, 103, 138
dicationic BIPY 17
dielectric
– breakdown 691
– capacitance 1075
– charging 1074
– function 127, 129
differential scanning calorimetry (DSC) 664
differential-drive configuration 955
diffuser micropumps 261
diffusion 1061
– based extractor 266
– coefficient 265, 865
– flame synthesis 58
– limited reaction 893
– parameters 420
– thermally activated 415
diffusive
– communication 750
– relaxation 482
– transport 115, 116
digital
– feedback 362
– light processing (DLP) 987
– micromirror device (DMD) 235, 987, 992
– signal processor (DSP) 328
– transmission between molecules 20
dilation 573, 584, 586, 593
dimension 387
dimensions of spatial structures 747
dimer structure 394
dimer-adatom-stacking (DAS) 392
Diophantine equation 973
DIP (dual in-line packaging) 1118
dip coating technique 863
diphtheria toxin (DT) 300
dipole molecule 834
dip-pen nanolithography (DPN) 175, 288
diprotic acid (11-thioundecyl-1-phosphonic acid) 621
direct overwriting 938
direct parallel design 970
direct write electron beam/focused ion beam lithography 186
directly growing nanotubes 379
discontinuous sliding motion 887
disjoining pressure 871
disk drive
– manufacturer 902
– servo control 968
dislocation
– line tension 704
– motion 570
– nucleation 696
dispersion
– force 406
– mixing 266
displacement 1047, 1057
– amplitude 930
– calibration 612
– controlled scanning probe 610
– field 1055
– maximum 1070
– resolution 681
– vertical 779
display
– paper-like 193
disposable biosensors 274
disposable plastic lab-on-a-chip 993
dissemination of biological information 315
dissipated power 939
dissipation 468
– force 461
– measurement 631
dissipative tip-sample interaction 468
dissociation 476
distortion 360
distributed
– Bragg reflector (DBR) 199
– feedback (DFB) 199
– laser resonators 199
dithio-bis(succinimidylundecanoate) 476
dithio-phospholipids 479
D-L amino acids (aas) 291
DLC 1012
– coating 524, 794, 796, 809, 819, 826, 1008
– coating microstructure 812
DLP technology 1088
DLVO interactions 549, 555–558
DMD (digital mirror device)
– fabrication 1088
– manufacturing 1095
– packaging 1100
DMT (Derjaguin–Muller–Toporov)
– model 695
– theory 724
DNA 404, 464, 467, 476, 747
– analysis 272
– analysis system 265
– nanostructure 294
– nanowire 33
– packing 756
domain
– pattern 439
– re-orientation 913
donor impurity concentration 127

doped
– anode 49
– polysilicon film 1001
– silicon 996
– silicon wafer 31
DOS structure 423
double barrier resonant tunneling 124
double-layer
– force 621
– interaction 545, 555–557
double-wall nanotubes (DWNTs) 51
doubly clamped beam 1028
doxorubicin 308
DRAM 923
Drexler 1146
DRIE (deep reactive ion etching) 156, 162, 166, 168–170, 254, 954, 963
dried tissue properties 676
drive amplitude 470
drive current
– maximum 976
driven cantilever 455
driving frequency 351
driving/sensing circuit integration 960
drug
– delivery 285
– resistant organism 293
– targeting 297
dry
– etching 205
– nitrogen environment 999
– (plasma) etching 254
– thermal oxidation 286
dry surfaces 571, 587
– forces 550, 552
– friction 571, 576, 588, 589
dual-axis gyroscope 233
dual-in-line packaging (DIP) 1118
dual-stage
– configuration 952
– sensitivity 973
– servo system 953
– short span seek 977
Dupré equation 569
DWNT (double-wall nanotubes) 51
dynamic
– interactions 544–546, 570, 581, 584
– mode 331
– operation mode 388
– oscillation force 456
– spectral analysis 891
– states of friction 894
– viscosity 264
dynamic AFM 387, 453–455, 461, 468, 469
dynamic force microscopy (DFM) 457, 471, 490, 501
dynamic force spectroscopy (DFS) 484
dynamic friction *see* kinetic friction

E

E. coli
– bacterium 748
– cells 756
early failure 1126
early tumor detection 310
EBD (electron beam deposited) tips 376
ECR (electron cyclotron resonance) CVD 795, 815, 1009
– coating 822
EDP (ethylene diamine pyrocatechol) 254
EELS (electron resonance loss spectroscopy) 800
EFC (electrostatic force constant) 678
effect of surface roughness 779, 781
effective
– coefficient of friction 647
– damping constant 470
– force gradient 463
– ligand concentration 485
– mass 433
– medium 127
– medium theory 127
– shear stress 648
– spring constant 637, 645, 721
– structural stiffness 1053
– tether length 485
– viscosity 583, 593, 885
effects of doping and annealing 120
EHD
– induction type 262
– injection type 262
eigenstrain 1041
elastic
– contact 703
– energy 732
– force 727, 757
– Frenkel–Kontorova chain 727
– Hamiltonian 755
– instabilities 727
– mismatch 1051, 1056
– modulus 664, 698, 767, 770, 774, 805, 806, 809
– properties 673
– properties of ssDNA 746
– stiffness coefficient 1048
elastic deformation 450, 647
– long-range 722
elasticity 433, 663
– of DNA 757
– two-dimensional 754
elastic-plastic deformation 785
elastohydrodynamic lubrication 569, 570, 581, 583
elastomer precursor 256
elastomeric stamp 190
electric charging 913
electric field gradient microscopy (EFM) 114
electric force gradient 338
electrical
– applications 131
– breakdown 1076
– conductivity 133
– isolation 956
– surface passivation 1103
electric-arc
– method 61
– reactor 48
electroactive
– fragments 14
– layer 24
– solids 22
electrochemical
– AFM 338
– cell 24
– degradation 304
– deposition 103, 675
– detection 270
– etch 383
– etching 373
– immunosensor 272
– STM 330
electrochemically etched tips 383
electrode
– arrays 26
– geometry 1076
– pair 1077
electrodeposition 103, 153, 180
electrodes with molecular layers 27
electrohydrodynamic (EHD) pumping 261
electroless deposition 188
electro-luminescence (EL) 132, 135
electrolytic
– conduction 1116

– method 57
electromagnetic
– actuators 259
– forces 550
– microvalves 259
electromechanical
– coupling 245
– load 1040
– mirror 1069
electromechanics
– analysis 1063
– application 1061
– test structure 1072
electromigration 419
electron
– beam deposition (EBD) 376
– cyclotron resonance chemical vapor deposition (ECR-CVD) 795, 799
– energy loss spectrometer (EELS) 113
– energy loss spectroscopy (EELS) 798
– Fermi wavelength 116
– field emission 133
– interactions 428
– mean free path 115
– scattering 120
– tunneling 327
electron-beam lithography 674
electronegativity 834
electron-electron interaction 423
electronic
– band structure 128
– density of states 133
– device 133
– ink 195
– newspaper 196
– noise 333
electron-phonon interaction 428
electro-optical modulation 136
electro-osmosis (EO) 261
electro-osmotic
– flow (EOF) 255
– pumping 260
electroplated microactuator 966
electroplating 153, 1113
electro-rheological fluid 260
electrostatic
– actuation 236, 244, 248
– binding 478
– coupling 1063
– discharge (ESD) 1098
– interaction 393, 402, 464
– microactuator 952
– micromotor 1061
– potential 398, 438
– short-range interaction 398
– spring constant 956
electrostatic force 386, 437, 544–558, 1062, 1103
– constant (EFC) 677
– interaction 393
– microscopy 438
electrostatically actuated devices 258
electrostatic-induced errors 1098
embedded atom method (EAM) 720
embedded system 721
embossing 833
– technique 191
end caps 47
end deflection 1026
end group 864
endofullerene 71
endothelial cell surfaces 488
energy
– barrier 482
– conservation 469
– dissipation 468, 567–569, 572–575, 578, 587, 654, 849
– elastic 732
– electrical 1063
– mechanical 1063
– resolution 422
– scales 750
– separation between subband 115
– total 1063
energy dispersive X-ray spectrometer (EDS) 113
engineered nanostructures 13
engineering materials 758
enhanced permeability and retention effect (EPR) 307
entangled states 570
entanglement 622, 886
– strength 886
entropic force 564
entropically
– confined systems 891
– cooled layer 888
entropy
– structural 886, 887
environmental shield 1087
enzymatic cleavage 287
enzymatic prodrug activator 301
enzymes 482
epitaxial silicon seal 1114
epitaxy 152, 170, 172
epitope mapping 491
epoxy bonding 1115
EPR tumor targeting 310
equibiaxial
– deformation state 1050
– stress distribution 1051
equilibrium
– equipartition 720
– interactions *see* static interactions
– true (full) or restricted 562
equivalent mechanical circuit 1073
erasing mechanism 936
ESD protective circuitry 1098
ester 837
etching 186, 1041
ethanolamine 477
ether 837
ethical rule violation 1143
ethics of nanotechnology 1138
ethylene diamine 287
ethylene diamine pyrocatechol (EDP) 254
Euler
– buckling criterion 1026
– equation 349
Euler–Bernoulli beam equation 1063
eutectic bonding 1115
evolution of curvature 1059
examples of NEMS 988
exchange
– force 402
– force interaction 402
– interaction energy 402
exchangeable carrier plate 636
excitation
– external 456
excitation amplitude 461
exciton 127, 136
– binding energy 135
– confinement 135
– laser action 135
extended exponential Kohlrausch relaxation function 892
external
– excitation 456
– excitation frequency 466
– noise 415
– normal load 867
– triggering 284
– vibrations 329
extraction of mechanical properties 1073
extravasation 297, 307
extrinsic stress 1041
Eyring
– model 644, 887, 890
– theory 887

F

Fab molecule 476
fabrication
– of membrane 314
– of microfluidic systems 254
– of mold masters 255
– techniques 254
– technology 225
face-centered cubic (fcc) 792
faceted 109, 111, 135
failure
– density function 1125
– mechanism 825, 1125
– mode 1101
– mode effect analysis (FMEA) 1101
– of MEMS/NEMS device 1012
– rate 1126
Fano resonance 425
fatigue 810
– behavior 769
– crack 776, 823, 1125
– damage 810
– failure 1036
– life 809
– measurement 1032
– properties 768, 776
– resistance 1028
– strength 776
– test 776, 813
fatty acid monolayer 841
FC (flip chip) technique 1113
FCA (filtered cathodic arc) 815
– coating 806, 811
$Fe(NO_3)_3$ 381
feasible simulation
– maximum 719
Fe-coated tip 403
feedback
– architecture 228
– circuit 333, 464, 465
– loop 363, 373, 450, 456, 461, 465, 485, 634, 957
– network 327
– signal 388
Fe-N/Ti-N multilayer 675
FeO 402
Feridex 306
Fermi
– energy 122
– level 33, 422
– points 428
ferrocene 59
– hydrophobic 24
ferroelectric film 913
ferromagnetic
– particles 305
– probe 439
– tip 441
FE-SEM image 136
FFM (friction force microscopy) 632
– dynamic mode 652
– on atomic scale 644
– tip 507, 636, 639
FIB (focused ion beam) tips 376
fiber
– optic telecommunication 233
– optical interferometer 352
– optics 299
FIB-milled probe 342
FID
– free induction decay 305
field
– emission 910
– emission tip 378, 910
– emission-based display 58
– emission-based screen 74
– ion microscope (FIM) 452
field-effect transistor (FET) 85, 116, 132, 137, 289, 290, 302
figure of merit 133
filled nanotube 69
filling efficiency 71
film/substrate system 1045
films on substrate 1023
film–substrate interface 1035
film–substrate system 1060
filtered cathodic arc (FCA) deposition 796–811
filtering 246
FIM (field ion microscope) 452
– technique 464
finite
– difference (FD) 1063
– element (FE) 1063
– element analysis 778
– element method (FEM) 764, 1049
– element modeling (FEM) 699
– size effect 131
first generation NEMS devices 247
first principle calculation 396
first principle simulation 434
FKT (Frenkel–Kontorova–Tomlinson)
– model 637
flagellar motor rotation rate 750
flap type valves 261
flash memory 899, 904–907
flat punch 697
flavin adenine dinucleotide (FAD) 303
flavoenyzme glucose oxidase (Gox) 303
flexible cantilever 332
flexible electronic devices 193
flip chip technique (FC) 1113
floating catalyst method 55
flocculation 555
Flory temperature 561
flow
– activation energy 865
– laminar 264
– rate 269
– resistance 266
fluid mosaic membranes 293
fluidic
– devices 253
– motion 270
– sampling 253
fluorescence
– quantum yield 17
– spectroscopy 127
fluoride 397, 398
fluoride (111) surface 400
fluorinated DLC 833
fluorinated silane monolayer 190
FM (frequency modulation)
– AFM 464
– AFM images 435
– mode 461
FMEA (failure mode effect analysis) 1101
focused ion beam (FIB) 330, 342, 376, 693, 709
foil cantilever 373
Fokker–Planck 891
folded protein structures 293
force
– advection 266
– and displacement calibration 618
– between macroscopic bodies 551, 556
– between surfaces in liquids 554
– calibration 610, 632
– calibration mode 526
– calibration plot (FCP) 867, 1003
– cantilever-based 433
– constant 618
– curve 382
– detection 348
– distribution 1043
– effective gradient 463
– elastic 727, 757
– extension curve 744
– induced unbinding 487

– long-range 389, 398, 437
– mapping 490
– measurement 387
– measuring techniques 546, 547, 549
– modulation mode (FMM) 663
– modulation technique 505
– packing curves 758
– repulsive 480
– resolution 433, 481
– scales 750
– sensing tip 358
– sensitivity 481
– sensor 332, 389, 635
– spectroscopy (FS) 362, 434, 436, 476
– titration 621
– undulation 545, 564
– velocity curve 752
force field spectroscopy
– three-dimensional 437
force-displacement
– curve 608, 677, 681
– plot 607
force-distance
– calibration 848
– curve 338, 451, 456, 621, 662, 668
– cycle 480
– diagram 469
form factor 900, 907
formate
– covered surface 405
– ($HCOO^-$) 404
formation of SAM 840
four helical bundle proteins 316
four-quadrants photodetectors 632
Fowler–Nordheim
– equation 1076
– plot 1077
– tunneling 905
fractal
– mobility 892
– space 891
– time 892
– time dependence 893
fracture 704, 1124, 1125
– strength 1036
– stress 767
– surfaces 777
– toughness 765, 767, 771, 775, 806, 807, 809, 810, 818
– toughness measurement 1027, 1030
Frank-Read source 703
free carrier density 127
free induction decay (FID) 305
free surface energy 874
freezing-melting transitions 588, 590, 886
Frenkel–Kontorova–Tomlinson (FKT) 637
– elastic chain 727
– model 725
frequency
– measurement precision 358
– modulation (FM) 388
– modulation (FM) mode 331
– modulation AFM (FM-AFM) 389
– modulation SFM (FM-SFM) 415
– shift calculation 390
– shift curve 463, 464
friction 325, 498, 507, 515, 606, 618, 631, 717, 793, 844, 983
– anisotropy 642
– characteristics 1009
– coefficient 569–572, 576, 577, 579, 580, 583, 590, 592, 593, 886
load dependence 518
– control 832
– directionality effect 512
– effect of humidity on 515
– effect of tip on 515
– experiments on atomic scale 638
– global internal 667
– image 640
– kinetic 571–573, 583, 585–587, 589, 590, 593, 718, 728, 729, 884, 885
– lateral 453
– loop 637, 638, 649
– macroscale 842, 998
– macroscopic 732
– map 583, 640
– measurement 1033
– measurement methods 342
– measuring techniques 547, 548, 576, 583, 585
– mechanism 510, 723, 843, 848, 871, 1013
– molecular dynamics simulation 651
– performance 866
– rate experiments 888
– scale dependent 518
– torque 1015
– torque model 1014
– total 726
friction force 331, 335, 501, 516, 642, 813, 843, 850, 866, 1004, 1005
– calibration 335, 346
– curve 363
– decrease of 875
– image 845
– kinetic 722
– magnitude 345
– map 511, 648
– mapping 869
– microscope (FFM) 325, 498, 499, 841
– microscopy (FFM) 537, 631, 632
– Z-15 531
– Z-DOL (BW) 531
friction models
– cobblestone model 572, 574
– Coulomb model 572
– creep model 589
– distance-dependent model 589
– interlocking asperity model 572
– phase transitions model 587, 589, 590
– rate-and-state dependent 590
– rough surfaces model 588
– surface topology model 588
– velocity-dependent model 587, 589
fringe detector 1028
fringing field 1068
frustules 314
fullerene 41, 792
fullerene-like structure 49
fullerite 42
functionalized nanotube 71, 72
fundamental resonant frequency 350
fused silica 699, 700, 704
– hardness 678
fusion bonding 1114

G

GaAs 248
– stamp 191
– wafers 192
GaAs/AlGaAs
– heterostructures 438
– quantum well heterostructures 248
gage marker 1028
gain 465
gain control circuit 389
GaN nanowires 135
gap stability 327
GaP(110) 435
gas separation 79
gate potential
– floating 905

gauche defect 622
Gaussian
– fluctuation–dissipation relation 889, 891
– statistics 884, 891, 892
Gd on Nb(110) 429
Ge membranes 210
gear
– miniaturized 986
generalized actuator 1064
genetic engineering 1142
genomic libraries 756
genuine graphite 40
geometric
– interlocking 724
– nonlinearity 1054
– parameters 228
– restrictions 249
geometry effects in nanocontacts 646
germanium 705
g-factor 424
giant magnetoresistance (GMR) 900
gimballed microactuator design 959
gimbal-mounted mirror 1097
GIO (grazing impact oscillator) 911
glass 619
– bulk etching 255
– frit bonding 1096
– frit sealing 1093
– transition 894
– transition temperature 256, 570, 579, 665
glassiness 593
glass-like behavior 891
glass-to-glass direct bonding 271
glass-to-silicon package 1127
global internal friction 667
glucose oxidase (GOD) 274
GMR
– giant magnetoresistance 900
gold 330
– coated tip 476
– electrodes 27
– film on polysilicon substrate 1051
– nanowires 136
– to-silicon eutectic bonding 1120
Grahame equation 555
grain
– boundary 553, 1035
– boundary scattering 120, 121
– size 189, 793
graphene 40, 63, 667
– defect 70
graphite 415, 417, 507, 638, 651, 792
– cathode 798
– pellet 47
– surface 463
graphite(0001) 427
gratings 200
grazing impact oscillator (GIO) 911
Griffith fracture theory 778
grinding 382
growth
– by pressure injection 102
– of nanotubes 668
guanine 10
gyro
– packaging 1100
– test 1101
gyroscope 230, 232, 1094

H

H_2O 425
Hagen–Poiseuille equation 264
half-generation precursor 287
Hall resistance 438
Hall–Petch behavior 710
Hamaker constant 386, 464, 551–577, 621, 871
Hamilton–Jacobi method 390
hand-held biochip analyzer 274
handling damage 1105
haptenization 313
haptenized nanostructure 313
hard amorphous carbon coatings 795
hard disk drive (HDD) 899–907, 952
– head 901
– performance 903
hard disk drive technology
– limits for 904
hard-core or steric repulsion 545
hardness 504, 770, 800, 802, 805, 809, 1024
harmonic oscillator 454
harpooning interaction 545, 553
HARPSS 166, 169, 170
HDT (hexadecanethiol) 188, 843, 854, 1006, 1012
head
– displacement 953
– gimbal assembly (HGA) 971
– group 839
– position 969
health care 1141
heat
– curable prepolymer 186
– dissipation 1122
– pipe cooling 1123
– transfer model 1119
heater
– cantilever 925
– temperature 934
heavy ion bombardment 441
heavy metal toxicity 306
helicity 41
helper T-cell epitopes 311
hermetic
– packaging 1116
– seal 1112
herringbone
– MWNT 44, 60, 76
– texture 61
Hershey–Chase experiment 755
Hertz model 694
Hertzian contact model 695
Hertz-plus-offset relation 647
heterodyne interferometer 351
heterogeneous CCVD 53
hetero-nanotube 68
hexadecanethiol (HDT) 188, 843, 854, 1006, 1012
HexSil 166, 168, 169, 954
– micro-molding process 962
hierarchical arrangement 108
hierarchy
– of structures 747
– of temporal processes 749
high modulus elastomer 188
high quality manufacturing 1087
high reflectivity 136
high speed scanning 467
high temperature conditions 61
high temperature superconductivity (HTCS) 419, 431, 440
high volume production 1093
high-aspect-ratio
– MEMS (HARMEMS) 986
– tips 376, 377
high-definition television 235
higher orbital tip states 422
highest resolution images 382
highly ordered monolayers 187
highly oriented pyrolytic graphite (HOPG) 337, 507
high-performance material 1141
high-resolution
– electrodes 196
– FM-AFM 436
– image 305
– imaging 407

– MR images 305
– patterning 193
– printing 186
– spectroscopy 422
– stamps 186
– tips 379
high-temperature operation STM 417
Hill coefficient 488
hinged cantilever 1034
HiPCo process 56
HiPCo technique 48
h-MWNT (herringbone) 44, 60, 76
homodyne interferometer 351
honeycomb
– chained trimer (HCT) 395
– lattice 68
– pattern 399
Hooke's law 433, 450, 454, 469, 480
horizontal coupling 329
host immune responses 311
hot embossing 257
HtBDC (hexa-tert-butyl-decacyclene) on Cu(110) 420
HTCS (high temperature superconductivity) 419, 431, 440
Huber-Mises 695
human bone tissue 675
humanity 1136
humidity 645
– effect 1105
hybrid
– carbon nanotube 68
– manufacturing 1093
– nanotube tip fabrication 380
hybridization 41, 397
hydration forces 545–575
hydration regulation 560
hydrocarbon 53, 57, 835
– chains 731
– precusors 799
– unsaturated 836
hydrodynamic
– forces 545, 549
– lubrication 729, 884
– radius 561
hydrofluoric (HF) 832
hydrogen 394
– bonding 10, 561
– bonding force 622
– concentrations 804
– content 804
– end group 841
– flow rate 805
– in carbon nanotube 77
– sensor 137
– storage 77
– storage capacity 78
– termination 394
hydrogenated carbon 800
hydrogenated coating 800
hydrophilic
– control 262
– surfaces 544, 560
– tip 616
hydrophobic
– coating 1014
– control 262
– ferrocene 24
– force 559, 560
– surfaces 544, 560
– tip 616
hydroxyl polymethacrylamide (HPMA) 308
hydroxylated surfaces 29
hydroxylation 840
hydroxyl-terminated perfluoropolyether 892
Hysitron 676
hysteresis 360, 382, 388, 415, 439, 459, 469, 681
– loop 333, 363

I

IB (ion beam)
– coating 822
IBD (ion beam deposition) 796, 798, 815
IC (integrated circuit) 226, 988
– industry 1112
– package technology 1101
– packaging 1112
ideal nanobiological devices 282
IFM (interfacial force microscope) 610
IgG
– repetitive 484
image
– effects 710
– processing software 364
– topography 331, 406
image projection 1100
– system 1084
imaging
– bandwidth 339
– electronic wave functions 427
– signal noise 389, 391
– tools 111
immunoassay
– magnetic bead-based 271
immunogenic synthetic nanomaterials 313
immunoisolation 286, 313
immunosensor 271
immunosuppression 313
immunosurveillance 313
impregnation method 54
imprinted polymers 200
in situ
– environmental TEM chamber 112
– screening 137
– sharpening of the tips 328
in vivo properties 676
InAs 417
InAs(110) 423, 435
incipient wetness impregnation 76
incommensurate surface 725, 731
indentation 498, 526, 923, 932, 935, 936
– creep process 528
– depth 527, 678, 806, 809, 998
– fatigue damage 812
– hardness 325, 527
– induced compression 812
– modulus 679
– size 527, 924
– size effect (ISE) 704
– technique 765
independent molecular operators 19
indirect sensor-device coupling 304
indirect transition in bismuth 129
indium tin oxide (ITO) 193
induction type EHD 262
inductive heating 1119
inelastic tunneling 419
inertial sensor 229
infected cell 756
influence of humidity
– on adhesion 850
information technology 1139, 1142
initial contact 703
injection type EHD 262
ink
– electronic 195
ink jet printing 193
inorganic-organic solar cells 136
InP(110) 392
in-phase signal 944
in-plane
– actuation 929
– mechanical strain 188
instability position 461
instability, jump-to-position 388, 610

instrumented suspension 967
insulator 397
insuline pump 269
integrated
– circuit (IC) 226, 988
– MEMS 259, 1084, 1092
– MEMS accelerometer 1091
– micromachining process 1117
– silicon suspension 967
– suspension 967
– tip 341, 373
– vacuum sealing 1117
Intel 915
intellectual property protection 1137
interaction
– energy 757
– force 434, 757
– potential 718
interatomic
– attractive force 344
– bonding 393
– force 332
– force constants 387
– interaction 1035
– spring constant 332
interbulk interactions 728
intercalated MoS_2 833
intercellular adhesion molecule-1 (ICAM-1) 489
intercellular concentration 748
intercept length 855
interchain hydrogen bonding 621
interchain interactions 484
interconnect 900
interconnect technology 1094
interconnected operators 18
interdiffusion 568
interdigitation 568
interface sensor components 302
interfacial
– chemistry 607
– defects 813
– energy (tension) *see* surface energy (tension)
– force microscope (IFM) 610
– friction *see* also boundary lubrication, 571–574, 579, 580, 585, 862
– liquid structuring 888
– potential 726
– stress 819
– structuring 888
– wear 606
interferometeric detection sensitivity 353
interlocking macrocycles 26
intermediate or mixed lubrication 569, 570, 582, 583
intermittent contact mode 373, 460
intermolecular communication 18
intermolecular force 289, 481, 484, 606
internal friction 935
internal stress 1025, 1041
intersymbol interference 947
intertube distance 43
intraband transitions 423
intramolecular
– forces 484
– self-assembly 293
intrinsic
– adhesive force 832
– damping 469
– mean stress 1045
– stress 804
iodine 419
iodobenzene 419
ion
– correlation forces 554, 557
– displacement 398
– implantation 205, 528, 925, 996
– implanted silicon 997
– plating techniques 795
– source 798
ion beam
– deposition (IBD) 796, 798
– etching 248
– sputtered carbon 798
ion correlation forces 545
ionic bond 545
ionic strength 619
Ising spin model 892
isolated nanotubes 381
isotropic etching 189
isotropic layer 1050
itinerant nanotube levels 425
ITO (indium-tin-oxide) 195
I–V characterization 116, 124

J

JKR (Johnson–Kendall–Roberts)
– model 619, 621, 695
– relation 646
– theory 566
Joule
– dissipation 655
– heating 439
jump length distribution 889
jump-to-contact 388, 436, 452
junction area 132, 135

K

KBr 397
KBr(100) 640, 650
$KCl_{0.6}Br_{0.4}$(001) surface 398
Kelvin
– equation 564, 612
– radius 613, 614
Kelvin probe force microscopy (KPFM) 407, 503
ketones 837
key scale at bio-nano interface 746
kidney exclusion limit 296
kinesin 741, 745
kinetic
– friction 571–573, 583, 585–587, 589, 590, 593, 718, 728, 729, 884, 885
– friction force 722
– friction ultralow 591
– measurements 892
– off-rate 487, 488
– processes 703
– rate constant 482, 486
kinetics 883
– of capillary formation 614
Kirchoff hypothesis 1047
knife-edge blocking 361
Koanda effect 266
KOH 330
Kohlrausch
– exponent 892
– relaxation 884
Kondo
– effect 414, 425
– temperature 425
K-shell EELS spectra 802

L

L amino acid 291
lab-on-a-chip 270, 988
– application 270
– concept 253
– disposable plastic 993
– systems 265
– technology 255
lac repressor 754
Lagrangian approach 1064
Lamb waves 693
laminar
– flow 264
– flow characteristics 266
laminated n-channel transistor 196
Landau
– levels 122, 415, 438

– quantization 424
Langevin equation 891, 892
Langmuir–Blodgett (LB) 861
– deposition 536, 833
– film (ethyl-2,3-dihydroxyoctadecanoate) 467
– methodology 35
Laplace
– force 844, 868
– pressure 544, 564, 613–618, 832
– pressure force 616
large angular deflection 1069
large array of nanowires 109
larger-scale systems 1136
large-scale stick-slip motion 733
Larmor frequency 305
laser
– ablation method 61
– beam deflection method 453
– deflection sensing 453
– deflection technique 333
– devices 46
– interference lithography 186
– vaporization method 61
– welding 1119
laser-to-fiber coupler 236
lateral
– accelerometer 231
– contact stiffness 631
– cracks 693
– deflection 671
– deflection signal 1016
– displacement 669
– force 632, 643, 669, 718, 1015
– force calculation 354
– force microscope (LFM) 325, 499, 632
– force microscopy (LFM) 187, 404, 548
– friction 453
– resolution 328, 331, 477, 500, 606, 663
– resonator 1037
– scanning range 929
– spring constant 338, 636
– stiffness 362, 376, 649
lattice
– constant 725
– imaging 332
– vibration 726
layer of resist 186
layered structure 1045
LCC (Leadless Chip Carrier) 1101
leakage 1126
– rate 263
learning theory 1143
length scale 722, 747
Lennard-Jones potential 436, 457, 719, 730
Lennard-Jonesium 727
leukocyte function-associated antigen-1 (LFA-1) 489
leveraged bending 1070
levers 932
lever-sample displacement 608
levitated copper inductor 242
Lévy flight model 891
LiF 397
LiF(100) surface 652
life-cycle costs 1137
lifetime broadening 419, 423
lifetime-force relation 485
Lifshitz theory 551, 552, 554
lift mode 338, 503
LIGA 166–168, 986
– fabrication of MEMS 987
– technique 765
ligand concentration
– effective 485
ligand-receptor affinity 288, 316
ligand-receptor interaction 476
light beam deflection galvanometer 353
light emission 135
light-emitting diode (LED) 132, 135
limits for hard disk drive technology 904
limits of biotechnology 281
linear
– creep model 890
– fractional transformations (LFT) 975
– friction–velocity dependence 889
– quadratic Gaussian (LQG) 971
– recording density 947
– stress gradient 1045
– variable differential transformers (LVDT) 361
lipid 747
lipid film 644
liposome 742
liquid
– capillaries 612
– capillary condensation 844
– film meniscus 612
– film thickness 537, 885
– helium 415
– helium operation STM 417
– lubricant 530, 832
– lubricated surfaces 572, 576, 578, 587
– mediated adhesion 832
– nitrogen (LN) 416
– perfluoropolyether (PFPE) 1003
– solid interface 832
– vapor interface 832
load
– critical 809
load contribution to friction 572–574
load dependence of friction 647
load-carrying capacity 824
load-controlled 610
load-controlled friction 571, 579
load-displacement 528
load-displacement curve 774, 778, 808, 809
loading curve 701
loading rate 483
local deformation 504, 518
local deformation of material 524
local density of states (LDOS) 422
local mechanical spectroscopy 664
local pinning 727
local shear stress 731
local stiffness 362, 1006
localization
– effect 122–125, 134
– theory 120
localized heat bonding 1116–1120
logarithmic-like friction–velocity behavior 885
logic
– devices 132
– gates 14, 15, 132
– protocols 15
Lognormal distribution 1127
London
– dispersion interaction 545, 550, 552
– penetration depth 431, 440
longitudinal
– magnetoresistance 120, 121
– piezo-resistive effect 356
– relaxation time (T_1) 305
long-range
– attraction 550, 562
– elastic deformation 722
– force 389, 398, 437
– tip-molecule force 406
long-term
– measurements 414
– memory 734
– reliability 1126
– stability 414, 948, 1102

– test 1125
loop-transfer recovery (LTR) 971
loss modulus 702
lot-to-lot variation 1072
low cost gyroscope 1094
low pressure chemical vapor deposition (LPCVD) 770
low temperature
– AFM/STM 332
– condition 60
– NC-AFM 402
– SFM (LTSFM) 417
low-cycle fatigue resistance 1032
lower lasing threshold 135
low-noise measurement 636
low-temperature microscope operation 414
low-temperature scanning tunneling spectroscopy (LTSTM) 419
low-temperature SPM (LTSPM) 414
LPCVD (low pressure chemical vapor deposition) 770, 962, 1034
– nitride sealing 1117
LTR
– loop-transfer recovery 971
LTSFM 417
LTSTM 419
lubricant 723
– atoms 732
– film 618
– meniscus 864
– spreading 864
– thickness 885
lubricated sliding 884
lubrication 717
– elastohydrodynamic 569, 581, 583
– intermediate or mixed 569, 570, 582, 583
– method 995
lumped
– parameter 1065
– parameter model 1063, 1073
Luttinger liquid 66
LVDT
– linear variable differential transformers 361

M

macrocyclic polyether 13, 25
macromolecular
– building block 747
– interaction 475
macromolecules as elastic rods 753
macroscale friction 842, 998
macroscopic
– building blocks 10
– friction 732
magnetic
– actuators 259
– applications 137
– dipole interaction 403
– disk 824
– disk drive 794
– information 138
– Ni 431
– ordering 129
– particles self-ordered arrays 914
– quantum flux 440
– recording 911, 922
– storage device 795, 831
– tape 466, 513, 524, 862
– thin-film head 825
– tip 465
magnetic field
– critical strength 122
– length 122
– microscopy (MFM) 114
– triggering 299
magnetic force 386
– gradient 338
– microscope (MFM) 338
– microscopy (MFM) 326, 439
magnetic resonance
– force microscopy (MRFM) 441
– imaging (MRI) 304
magnetically coated tip 490
magnetohydrodynamic (MHD) pumping 261
magnetomotive transduction 248
magneto-optics 128
magnetoresistance 120
– longitudinal 120, 121
– transverse 121
magnetoresistive RAM (MRAM) 899, 906
magnetostatic interaction 439
magnetron sputtered carbon 799
major histocompatibility complex (MHC) 311
manganites 431
manifold absolute pressure (MAP) 1091
manipulation of individual atoms 325
manually assembled MWNT tips 378
manufacturability 1087
manufacturing "risk" change 1093
MAP sensor 1094
marginal dimension 727
martensite 665
mask, high resolution stamp 186
mass
– effective 433
mass of cantilever 331
master–slave design 970
material
– for microfluidic devices 254
– hardness 765
– molecule-based 35
– property variation 1072
– structural 1028
Matthiessen rule 120
maximum likelihood estimator (MLE) 1126
MBE (molecular beam epitaxy) 152, 172
MBI (multiple beam interferometry) 547
mean time to failure (MTTF) 1127
mean-field theory 556
measurement of hardness 505
mechanical
– coupling 588
– dissipation in nanoscopic device 631
– instability 452, 609
– micropumps 260
– protection 1112
– relaxation 457, 848
– resonance 591
– spectroscopy 664
– stability 927
– surface relaxations 468
– wear 263
mechanical properties 764, 1023
– characterization 1036
– of bone 675
– of carbon nanotubes 668, 672
– of DLC coating 813
mechanically cut tips 382
mechanics of cantilevers 347
mechatronic device 952
media flatness 910
mediator molecules 304
medium
– effective 127
medium theory
– effective 127
melting freezing transition 894
melting point of SAM 853
membrane
– deflection method 1037
– fabrication of 314

– proteins 375
membrane-embedded machine 741
membranes as elastic media 754
memory
– cells and switches 116
– distance 588
– long-term 734
MEMS 147–171, 225, 270, 376, 606, 617, 764, 795, 831, 862, 901, 983, 1023, 1084
– accelerometer manufacturing 1091
– accelerometers 230
– applications 985
– capacitive switch 244
– components 1011
– device 247, 987, 990
– device operation 994
– fabrication 1111
– fiber optic switching 238
– industry 1093
– integration 1102
– manufacturing 1086
– microactuator (MA) 952, 971
– mounting condition 1105
– packaging 1111, 1116, 1128
– packaging requirement 1112
– passivation 1103
– post-packaging 1114, 1117
– pressure sensor production 1090
– production economics 1086
– resonators 244
– sensor 229
– standardization 1086
– storage device 907
– structure 1112
– surface 1104
– switch, metal-to-metal 244
– technology 233, 237, 249, 1034
– testing 1099
– tribological problems 989
– tunable capacitors 241
– vacuum packaging 1117
– wafer bonding 1095
– wafer protection 1097
MEMS/NEMS 530
MEMS-based
– microfluidic system 270
– storage concept 904
– storage device 922
meniscus
– bridge 831, 876
– force 516, 871
– of liquid 564
mercaptopropyltrimethoxysilane (MPTMS) 191
merocyanine 18
mesoporous molecular sieves 102
mesoscale osmotic actuator 269
messenger RNA 750
metabolic facility 282
metal 103
– catalyst 379
– cavity packaging 1099
– cluster agents 306
– electroplating 255
– evaporated (ME) tape 793
– matrix composites 81
– nanowires 69
– oxide 397, 400
– particle catalyst 54
– porphyrin (Cu-TBPP) 404
– seed layer 966
metal/insulator/metal (MIM) 197
metal-catalyzed
– chemical vapor deposition (CVD) 379
– polymerization method 287
metal-deposited Si surface 395
metallic nanotube 34
metallic SWNT 66
metal-organic compound 56
metal-particle (MP) tape 825
metal-to-metal MEMS switch 244
metastable lubricant systems 894
methylene stretching mode 842
methyl-terminated SAM 843
Meyer's law 708
MFM sensitivity 338
MgO tip 398
MgO(001) surface 403
Mg-terminated tip 398
MHD (magnetohydrodynamic) pumping 261
mica 102, 337, 615, 619, 640, 888
– films 100
– muscovite 615, 650
– surface 641
micelles 291, 292
micro/nanofabrication 147–150, 154
microactuator 990
– design 960
– fabrication requirements 962
– loop 972
– shape 965
microbially-derived antibiotics 293
microcantilever 994
microchannel 260
microcomponent 989
microcontact printing (μCP) 187, 834
microcrystalline graphite 803
microdevice 1026
microdispenser 266
microelectromechanical
– motor 226
– resonators 244
– switch 244
microelectronics 397, 749
micro-electro-rheological valve 260
micro-engine 908
microfabricated
– cantilever 372, 453
– silicon cantilevers 387
microfabrication 187, 226, 249
– methods 285
– techniques 285, 372
microfluidic 253
– applications 255
– channel 254, 269
– channels on silicon substrates 254
– control 260
– line 258
– mixers 265
– motherboards 270
– multiplexer 268
– network 268
– passive 262
– pumping 260
– structures 256
microfluidic device 269
– disadvantages 270
– materials for 254
– passive 257, 269
microfluidic dispenser
– passive 273
microfriction 511
micro-hinge technology 236
microindentation 689
micro-injection molding 273
micromachined
– high-Q resonators 245
– inductors 242
– notch 1032
– silicon 310
micromachining 147, 150, 151, 154, 155, 159–163, 166–168, 170
micromanipulator 988
micromechanical
– cantilevers 303
– device 1084
– structures 204
micromirror 992, 1040
– array 237
micromixer category 265

micromotor 1016
micron-sized channels 265
micropatterned SAM 848
micropipette aspiration 549
micropump 260
– passive 269
micro-Raman spectroscopy 693
microrelay 1062
micro-resonators 244
microscale
– diffusion 265
– friction 1006, 1010
– material removal 521
– scratching 503
– silicon gate 34
– wear 503, 520
– wear test 1008
microscanner 929
microscope eigenfrequency 348
microscopic friction 723
microscopic origin of friction 889
microscratch 813
– test 675
microscratching measurement 519
microsensors 229
microstrain gauge 1025, 1034
microstructure 1040
microsystem 249
micro-total analysis systems (μTAS) 257
microtransformer 197
microtriboapparatus 1011
microtubule 740
– bending 673
– buckling 673
microvalve 258
– design 260
microwear 813, 819
midplane 1046
– curvature 1048
– strain 1048, 1049
milligear system 989
millipede 901, 994
– chip 931
– system concept 921, 923, 943
MIM capacitors 197
MIMO design 971
miniaturized
– actuator 986
– connector 986
– gear 986
– motor 986
minimal detectable depression 374
minimally invasive screening 285
minimum-indentation pitch 937
mirror/yoke assembly 1089
misfit
– angle 638
– strain 1042
mismatch
– elastic 1051, 1056
mismatch of crystalline surfaces 566, 567, 593
mitochondria 747
mixed C-N nanotube 56
mixed lubrication *see* intermediate or mixed lubrication, 884
mixing performance 265
MMA air bag sensor 1092
Mn on Nb(110) 429
Mn on W(110) 432
m-nitrobenzylidene propanedinitrile (*m*-NBP) 916
MOCVD (metallorganic chemical vapor deposition) 152, 172
mode coupling gratings 199
model potentials 719
modeling
– at the nanoscale 752
– mechanics of systems 752
– tribological process 753
modular multifunctionality 280
modulation codes 946
modulus
– elastic 664, 698, 767, 770, 774, 805, 806, 809
modulus of elasticity 325, 516
molarity 749
molding 186
molecular
– AND gate 17
– assembly 26
– beam epitaxy (MBE) 988
– building blocks 12, 35, 747
– chain 839
– conduction band 33
– conformation 864
– cooperation 886
– dynamics (MD) 476, 718, 749
– dynamics (MD) calculation 639
– dynamics simulation (MDS) 488, 622, 631, 651, 753
– engineering 281
– interaction 879
– interconverting state 15
– logic gates 15
– lubrication 885
– machines 13
– mobility 894
– motor 13, 673, 740
– NOT gate 16
– placement by force microscopy 285
– precursor 58
– pump 741, 748
– recognition force microscopy (MRFM) 476
– recognition force spectroscopy principles 482
– resolution 421
– reversible transformation 15
– scale layer 833
– shape 558, 559, 586, 591, 592
– shuttle 14
– spring 536
– spring model 843, 848
– stiffness 849
– switch 14, 35
– theory of lubricated friction 890
– wires 289
molecule-based
– device 24, 28
– materials 35
– switches 15
moment of inertia 1030
monodispersity 286
monolayer 885
– switches 28
– thickness 855
monolithic accelerometers 231
monolithic high-Q MEMS 245
Monte Carlo simulation 848
MoO_3 film 614
Moore's law 1142
Morse potential 464
MoS_2 friction 640, 641, 668
motion
– unperturbed 390
motor
– actin-myosin 751
– flagellar 750
– micro 1016
– microelectromechanical 226
– miniaturized 986
– molecular 13, 752
– myosin 741, 745
– oil 884
– voice coil 952, 969
movement-constrained macromolecules 305
moving media model 908
moving-tips model 908
MRAM (magnetoresistive RAM) 899–907
– cell 906
MRFM

– magnetic resonance force microscopy 441
MRI contrast agents 305
MTTF (mean time to failure) 1126
multicomponent nanodevice 288
multi-input-multi-output (MIMO) 956
multilayer 709
– devices 197
– Fe-N/Ti-N 675
– structures 104
– thin-film 825
multimode AFM 338
multiple beam interferometry (MBI) 547
multiple cantilever 907
multiple therapeutic nanodevice 294
multiple-asperity contact 517
multiplexed sensor arrays 304
multiplication of dislocations 703
multivalent targeting strategy 298
multiwall carbon nanotubes (MWCNT) 43–49, 289, 342, 376, 667, 727
muscle contraction 740
muscovite mica 615, 650
μ-synthesis 974
μ-synthesis controller design 975
MWCNT
– multiwall carbon nanotubes 43–49, 289, 342, 376, 667, 727
MWNT (multiwall nanotube) 64
– based catalyst-support 76
– bc (bamboo-concentric) 44
– bh (bamboo-herringbone) 44
– bunches 58
– catalytically grown 673
– composites
MWNT-Al composite 81
– flexural modulus 67
– purified 378
– surface area 64
– tips
manually assembled 378
– Young's modulus 669
myosin motor 751

N

N acceptors 10
Na on Cu (111) 421
NaCl 397
– island on Cu(111) 655
– islands 468
– thin film on Cu(111) 398
NaCl(001) 397
NaCl(100) 641–643, 648
NaF 397, 640
NAND gate 15
nanites 1146
nanoasperity 515, 695
nanobeam 669
– array 766, 774
nanobiological
– applications 313
– design 312
– devices 281
– vaccine design 312
nanobiotechnological device 300
nanobiotechnological therapeutic device 313
nanochannel glass 100, 102
nanochemistry 984
nanocluster 400
nanocomposite
– coating 833
– materials 28
nanocrystallites 800
nanodeformation 864, 880
nanodevice 85, 281
– multiple therapeutic 294
nanoelectrodes 32
nanoelectromechanical systems (NEMS) 226, 246, 606, 764, 901, 984
nanoelectronic device 32, 34
nano-electronics 84
nanofabrication 147, 170, 175, 187, 326, 526
– techniques 253
nanofatigue 810
nanofiber 44, 53
nanofilament 44, 60
nanogaps 31
nanohardness 695, 804, 805, 1000
nanohelix 12
nanoimprint 171
nanoindentation 674, 688, 697, 805, 1024
– curve 677
– measurement 504
nanoindenter 765, 769, 818, 1028
nanolubrication 883
nanomachine 1147
nanomachining 325, 326
– techniques 205
nanomagnetism 431
nanomania 1146
nanomaterial 662
nanomechanical Si beams 247
nanometer resolution 186
nanometer servo precision 979
nano-objects 68
nanoparticle 49, 374
– contrast agents 305, 306
nanoparticulate carriers 310
nanophobia 1146
nanoporous membranes 285, 314
– microfabrication 286
nano-positioning 359
nanorheology 584
nanoscale
– antimicrobials 293
– architectural property 283
– assemblers 1146
– biomolecules 10, 12
– control of molecular events 310
– data bit 909
– drug delivery 291
– electrodes 23, 27
– FET 117
– film morphology 917
– interventions 285
– mechanical properties 661, 662
– molecular therapeutic device 308
– molecules 35
– Schottky barrier 34
– supramolecular assembly 407
– therapeutic device 281, 315
– therapeutic platforms 284
– wear 518
Nanoscope I 328
nanoscratch 522
– studies 765
nanostructures 147, 170, 171, 179
– containing proteins 307
– mechanical properties 764, 1023
nanoswitches 15
nanotechnological
– actions 755
– therapy 307
nanotechnology 280, 918, 1139
– definition 1136
– ethical issues 1136
– in the living world 740
– possible applications 1136
– recommendations 1138
– role of social sciences 1148
– social issues 1136
– unintended consequences 1143
nanotherapeutic 280, 288, 303
– contrast agents 304, 310
– delivery device 298
– design paradigms 282
– device application 307
– device components 285
– devices in oncology 307

– drug delivery 311
– triggering external stimuli 299
– vaccines 312
nanotransfer printing (nTP) 190
nanotransistor 32
nanotribology 717, 862
nanotube 12, 105, 133
– AFM tips 379
– buckling 377
– chemistry 67
– defects 47
– functionalization of wall 72
– functionalized 71, 72
– growth 45
– length 382
– morphology 49
– nucleation 45
– oxidation 72
– resistance 34
– strain 671
– surface energy 380
– tip fabrication 380
– tips, surface growth 380
– transistor 34
– volume fraction 83
– yield 49, 59
nanotube production 56
– efficiency 49
– electric-arc method 48
– laser ablation 46
– solar furnace 52
– techniques 57
nanotube-based
– emitter 74
– hybrid material 71
– sensor 75
– SPM tip 73
nanotube-Co-MgO 81
nanotube-containing
– composite 82
– electrical device 289
nanotube-Fe/Co-$MgAl_2O_4$ 81
nanotube-Fe-Al_2O_3 81
nanotube-polymer composite 83
nano-tweezer 85
nanowire 31
– applications 131, 138
– metal 69
– nucleation 110
– photodetectors 135
– semimetallic 118
– stress-induced crystalline 107
– superlattice 132
– Zn 120, 123
– ZnO 135
nanowire device
– two-terminal 132
naphthalene channel 21
National Nanotechnology Initiative (NNI) 1137
native oxide layer 868
Nb superconductor 423
NBIC convergence 1140
$NbSe_2$ 431, 649
NC-AFM (noncontact atomic force microscopy) 397, 433
near-field
– method 903
– scanning optical microscopy (NSOM) 128, 372
– technique 434
near-surface mechanical properties 665
necrotic domains 307
negative logic convention 26
negative magnetoresistance 121
neighbouring indents 678
NEMS 831, 983, 988, 1023
– devices 246
– nanoelectromechanical systems 226, 246, 606, 764, 901, 984
neocarzinostatin (NCS) 309
nested-arc structure 680
net electrostatic torque 1066
net fluid transport 261
newspaper
– electronic 196
Newton's first law 454
Newtonian flow 570, 580, 581, 583
Newton–Raphson method 753
n-hexadecane 888
N-hydroxysuccinimidyl (NHS) 478
Ni(001) tip on Cu(100) 653
Ni(111) tip on Cu(110) 653
Ni, Co catalyst 47
Ni/Y catalyst 49
Ni^{2+}-chelating 485
nickel-iron permalloy 259
Ni-Fe valve membrane 259
NIH3T3 fibroblast cell 489
NiO 402, 417
NiO(001) 397, 400, 402
NiO(001) surface 403
Ni-P
– beam 785
– cantilever 778
– film 770, 771
– microbeam 777, 785
NiTi 665
– transformation behavior 666
nitrogenated carbon 800
NMP (no moving part) valves 265
noble metal surfaces 424
n-octadecyltrichlorosilane ($n\text{-}C_{18}H_{37}SiCl_3$, OTS) 616, 842
n-octadecyltrimethoxysilane (OTE) 615
noise 387, 391, 453
– 1/f 387, 957
– electronic 333
– external 415
– performance 391
– signal 941
– source 355
– vertical 391
Nomarski-interferometer 352
nonconducting film 798
nonconductive
– materials 404
– sample 329
– surface 462
nonconforming contacts 890
noncontact
– dynamic force microscopy 654
– friction 438, 655
– imaging 331
– mode 373, 388, 909
noncontact atomic force microscopy (NC-AFM) 374, 385, 391, 393, 397, 404, 654
nondestructive contact-mode measurement 652
nondimensional curvature 1051
nonequilibrium interactions *see* dynamic interactions
nonequilibrium simulations 720
noninvasive device 283
nonlinear
– force 460
– geometry effect 1058
– I–V behavior 124
– material behavior 1058, 1059
– optical property 129
– spring 491
– strain-displacement relation 1055
nonliquidlike behavior 584
nonmagnetic Zn 431
non-Newtonian behavior 885
non-Newtonian flow 570, 581, 583
nonpolar group 838
nonspherical tip 647
nonsymmetrical passive valves 268
nontherapeutic nanobiological device 283
nonvolatile random access memories (NVRAM) 899–913
nonwetting 560
NOR gate 15

normal friction 569, 577, 579, 580
normal load 721, 866
– external 867
normalized frequency shift 463
no-slip boundary condition 729
NOT gate 15
N-succinnimidyl-3-(S-acethylthio)propionate (SATP) 478
NTA (nitrilotriacetate) - His_6 478
nTP (nanotransfer printing) 197
nuclear weapon 1140
nucleation of nanowires 110
nucleic
– acid (NA) 294, 299
– acid hybridization 294, 300
nucleotides 747
n-undecyltrichlorosilane ($n\text{-}C_{11}H_{23}SiCl_3$, UTS) 842

O

O acceptors 10
$(OCN)^-$ 398
octamethylcyclotetrasiloxane 888
ODD performance 903
odd-even effect 621
off-track error 958
OH-terminated tip 398
on-chip
– actuator 1031
– electronic detection 231
– integration 210
– microvalves 260
– spiral inductors 242
oncological nanotherapeutics 291
one-dimensional quantum effect 115
one-dimensional system 100
open loop architecture 1093
operation 387
opsonized particle 296
optical
– absorption 129
– applications 134
– beam deflection 607
– cavity 135
– data switching 237
– deflection systems 372
– detector 331
– disk drives (ODD) 899, 900
– gap behavior 127
– head 336
– head mount 337
– integration 236
– MEMS 233
– output 16
– property 126, 127
– reflection 127
– switch 135
– transmission 127
– trap 480
– tweezers (OT) 480, 549, 744
optical lever 353
– angular sensitivity 355
– deflection method 362
– optimal sensitivity 355
optimal beam waist 355
optimum sampling phase 946
optoelectronic
– components 198
– module 236
– switch 136
OR
– operation 23
– operator 15
order
– in-plane 567, 584, 585, 593
– long-range 570
– out-of-plane 557–559, 567, 570, 584, 585, 591
– parameter 570, 586
ordering nanowires 108
organelles 281
organic
– compounds 835
– device structure 917
– film 916
– inverter circuit 196
– technology 916
– transistor 193, 196
organics 915
– data storage 916
organized molecular arrays 23
organometallic vapor phase epitaxy (OMVPE) 988
Orowan strengthening 710
orthogonal conjugation 295, 296
orthotropic Young's modulus 1048
oscillating tip 334, 469
oscillation
– amplitude 389, 453, 454, 457, 461
– loop 462
oscillatory force 544, 545, 557–561, 567
osmosis response time 269
osmotic
– interactions 545, 555, 562, 564
– pressure 555, 562, 575
– stress technique 549
OTE (octadecyltrimethoxysilane) 615
OTS (octadecyltrichlorsilane) 848, 1075
OUM cell 906
out-of-plane stiffness 959
overall barrier height 887
overdamped SFM system 891
overdrive capability 1062
overwriting mechanism 937
Ovonyx 915
Ovonyx Unified Memory (OUM) 906
oxide layer 703
– amorphous 106
oxide-enhanced nanowire growth mechanism 107
oxide-sharpened tips 375
oxidized nanotube 72
oxygen
– content 826
– sensor 274

P

package level integration 1093
packaging 248, 249
– cap 1112
– equipment 1099
– induced thermal stress 1123
– reliability 1125
– uniform 1112
packing density
– chains 842
palladium nanowires 137
PAMAM (polyamido amine) dendrimer 286, 287
p-aminophenyl phosphate (PAPP) 271
paper-like displays 193
paraboloid load displacement 697
paraffins 835
parallel-plate
– actuator 955
– capacitor model 1064
– sensing 957
paramagnetic particles 305
parasitic charging 1077
particle
– contamination 1098
– control 1098
– opsonized 296
– track-etched mica films 103
particle model
– embedded 894
Paschen curve 1076
passive
– check valves 262

– linearization 361
– microfluidic device 257, 269
– microfluidic dispenser 273
– microfluidics 262
– micromixers 265
– micropump 269
– microvalves 257, 262
– structure 1025
– valve geometry 264
patient-monitoring systems 274
pattern of ink 186
pattern transfer 191
patterned
– conductive substrate 110
– film 1053
– self-assembled monolayer (SAM) 188
– silicon chip 25
patterning 1075
Pb on Ge(111) 428
PbI_2 filled SWNT 70
PDP (2-pyridyldithiopropionyl) 478
peak indentation load 810
peapods 71, 289
PECVD (plasma enhanced chemical vapor deposition) 833, 926, 1009
– carbon sample 800
PEG (polyethylene glycol)
– crosslinkers 479
pegylated molecules 313
Peierls instability 428
per molecule adhesion force 620
perfluorodecanoic acid (PFDA) 992
perfluoropolyether (PFPE) 530, 861
perfluoropolyether Z-DOL 614
performance of electromechanical actuator 1075
performance parasitics 1098
periodic
– force 725
– potential 637, 891
permanent dipole moment 407
perpendicular scan 344
persistence length 488
perturbation approach 390
perylene 404
PES (photoemission spectroscopy) 956
– generation 944
– loop compensator 973
PFPE (perfluoropolyether) 864, 1012
– lubricant 862, 868, 1003, 1016
– OH 892
pH of water environment 619, 621
pH sensor 137
phage display 297, 298
pharmacoeconomics 285
phase
– angle image 530
– breaking length 120
– curve 458
– images 470
– lag spectrum 666
– measurement 471
– signal change 458
– transformation 690, 705
phase change
– material 915
– medium 922
– RAM (PC-RAM) 899, 906
phenol 837
phenoxy-centered HOMOs 26
phonon
– confinement effects 125, 127, 131
– excitation 468
– scattering 120
phosphorus-doped polysilicon 206
photoacid 19
photoactive
– fragments 14
– solids 22
– stopper 14
photo-conductance 128
photo-crosslinker 478
photodetector sensitivity 766
photoemission spectroscopy (PES) 423
photogenerated hole 30
photoinduced electron transfer 16, 17
photoinduced proton transfer 19
photolithography 285
– near-field conformal 187
– proximity mode 186
photolithography patterns 186
photo-luminescence (PL) 127, 135
– optical imaging 128
photoresist 188, 190, 1095, 1121
photothermal effect 465
physical properties of nanowires 110
physical vapor deposition 104
physisorbed protein layer 476
physisorption 861
pick-up
– SWNT tips 382
– tip method 381
pico-slider 960
piezo
– ceramic material 360
– effect 359
– element 461
– excitation 454
– hysteresis 415
– relaxation 415
– stacks 361
– tube 359
piezoelectric
– actuation 259
– actuator 908
– drive 327
– leg 636
– positioning elements 414
– scanner 372, 415, 676
piezoresistive
– cantilever 355
– coefficients 356
– detection 355, 387
– pressure sensor 1089
– sensing 958
– sensor 227
piezoscanner 363, 382
piezotranslator 363
piezotube calibration 329
pigment particles 195
pile-up, nanoindent 693, 699, 700, 709, 711
– characteristic 935
piling-up behavior 674
pilot signal 945
pinning 441
– forces 727
pin-on-disk tribotester 841
PL (photoluminescence) NSOM imaging 135
plane-to-plane separation 12
plasma
– etch 286
– frequency 127
– resonance 127
plasma enhanced chemical vapor deposition (PECVD) 197, 339, 770, 795, 799
plasminogen activator (PA) 302
plasmon mode 421
plastic
– asperity model 884
– circuit 193
– contact regime 517
– deformation 522, 728, 812, 854
– electronics 197
– fluidic chips 273
– large area circuit 194
– packaging 1099
plasticity 1060
platelet nanofiber 60

platform 929
– fabrication process 930
platinum-iridium 330
– tip 29
PMMA (poly(methyl methacrylate)) 247, 257, 671, 924, 933, 936, 994
p-n
– diodes 116
– junction diode 132
– junctions 132, 135
pneumatic
– actuation 258
– damping 418
p-nitrophenolate 19
point probes 688
Poisseuille flow 885
Poisson statistics 620
Poisson's ratio 698, 1047
polar
– group 838
– lubricant 832, 861
polarization dependence 129
polishing 793, 1000
poly(amido) amine (PAMAM) dendrimer 286
poly(ethylene glycol) (PEG) 477
poly(ethyleneterephthalate) (PET) 192
poly(methyl methacrylate) (PMMA) 247, 257, 671, 924, 933, 936, 994
polyaromatic shells 63
polycarbonate (PC) 257
polycrystalline
– films 210
– graphite 801
– Si 226
– SiGe 210
polydimethylsiloxane (PDMS) 187, 834
polyethylene (PE) 257
polyethylene co-propylene 886
polyethylene terephthalate (PET) 524
poly-Ge
– mechanical properties 210
polymer 255, 610, 664, 910
– blend 471, 662
– brush 563, 575, 581
– chemistry 286, 291
– conjugation (pegylation) 296
– fluids 892
– intermediates 287
– matrix composites 82
– medium 932
– melt 886
– membranes 103
– microfabrication techniques 255
– microstructures 257
– mushroom 562
– nanojunction 32
– properties 263
– therapeutics 307
– transparent microcapsules 195
polymeric
– liquids (melts) 561, 581, 583
– magnetic tape 526
polypeptides 10
polypropylene 404, 471
polysilicon 963, 967, 1010, 1040
– beam 1044
– fatigue 1036
– film 229, 995, 1027
– fracture strength 1036
– fracture toughness 1035
– friction 1036
– layer 263
– mechanical properties 1034
– MEMS 206
– microstructure 1035
– residual stress 1034
– resonators 245
– strip 245
– surface micromachining 210
– Young's modulus 1035
polystyrene (PS) 257, 936
polysulfone 936
polytetrafluoroethylene (PTFE) 639
polyurethane 471
poor mixing 889
pop-ins 690
pop-outs 690
population explosion 1141
porous
– silica or alumina 120
– vycor glass 102
position
– accuracy 363
– error 972
– error signal (PES) 943, 968
– sensing resolution 957
positioning 109
positive logic convention 18, 26
post-packaging process 1118
potential barrier 887
potential energy
– total 720
power dissipation 237, 470, 654
power to weight ratio 752
PQ design method 970
Prandtl–Tomlinson model 725–727
pre-crack length 1027
precursor 56, 69, 70
– gases 210
preferential flow direction 261
preparation improvements 54
pressure 871
– injection 104
pressure sensor 227
– fabrication process 1090
– packaging 1099
pressure-induced phase reconstruction 888
primary minimum adhesion 544, 555, 559
printed
– coil 197
– DFB resonator 200
probability density functions 891
probe
– FIB-milled 342
probe storage 901, 907, 908
– data organization 909
– operating parameters 909
probe tip 372
– characterization 611
– performance 374
probe-sample distance 457, 462
probe-surface distance 455
process
– control 1088
– flow 1088
– gas 799
process coherence volume 887
processing conditions 44
prodrugs 298
product architecture 1088
production efficiency 47
programmable microfluidic systems (sPROMs) 266
programmed assembler 1146
projection mode photolithography 186
propagation of cracks 524
properties of a coating 795
protection of sliding surface 831
protein 477
– bioactivity 294
– coupled to nanomaterials 294
– folding 293
– primary structure 293
– secondary structure 293
– synthetic chemistry 294
– tertiary structure 293
protofilament 673
prototyping of biological nanodevices 282
protrusion force 545, 564

P-selectin glycoprotein ligand-1 (PSGL-1) 488
PSG (phosphorus-doped glass) 1117
Pt alloy tip 330
PTCDA 404
PTFE (polytetrafluoroethylene)
– coated Si-tip 640
– film 833
Pt-Ir
– tip 330
– wire cantilever 912
pull-hold-release cycle 485
pull-in 1072
– angle 1068
– charge 1064
– instability 956
– voltage 1062, 1064, 1067, 1068, 1071, 1075
pull-off force *see* adhesion force, 608, 614, 615, 617, 635, 864
pulse-etching 379
pumping chamber 261
purified MWNT 378
pyramidal
– AFM tip 615
– etch 373
pyrazoline 16
pyrolysis of hydrocarbons 59
pyrolytic method 58
PZT (led zirconate titanate)
– actuated suspensions 967
– scanner 500
– tube scanner 328, 334

Q

Q-Control 464, 465
Q-factor 388, 436, 455, 1126
quad photodetector 334
quadrant detector 354
quadrature signal 943–945
quality factor *Q* 339, 348, 389, 454, 911
quality system 1101
quantization energy 117
quantum
– corrals 427
– dot 134, 172, 173
– effects 126, 134
– Hall regime 438
– limit for the thermal conductance 125
– subbands 116
– well 421
– wire superlattices 123
– yield 16
quantum confinement 99, 126, 127, 129
– effect 116, 135, 138
quantum size
– effect 118, 134
– regimes 116
quantum-confined structures 133
quartz-crystal microbalance (QCM) 717
quasi-optical experiment 421
quasi-static
– bending test 775, 777
– mode 388
– strain rate 677

R

radial cracking 808
radio-frequency (RF) MEMS 239
radiotherapy 307
radius of gyration 561
Raleigh's method 348
Raman
– spectra 131, 802
– spectroscopy 800
ramped creep model 890
random access memory (RAM) 28
random access time 900
random force 891
randomly oriented polysilicon 1035
rapid thermal processing (RTP) 1118
raster scan 901
Rayleigh wave 692
Rayleigh–Ritz method 1071
RbBr 397
RC-oscillators 358
reaction kinetics 483
reactive
– lubrication 892
– sealing 1114
– spreading 189
reactive ion etching (RIE) 205, 247, 254
read
– channel 939
– time 908
readback
– rates 932
– signal 940, 941
readout
– electronics 354
– signal 944
receptor binding 295
receptor-ligand
– bond 482
– interaction 484, 545
– unbinding 483
– unbinding force 476
receptor-mediated uptake of synthetic nanomaterial 296
recognition force microscopy (RFM) 476
recognition force spectroscopy (RFS) 484
recognition imaging 489
recombinant protein therapeutics 285
rectangular cantilever 339, 610, 656
rectifying properties 132
recursive least square algorithm (RLS) 974
reduced modulus 690
reflection interference contrast microscopy (RICM) 550
reflectivity 129
relative humidity (RH) 613
– effects of 533, 873, 1004
relative position error signal (RPES) 956
relaxation
– mechanical 457, 848
relaxation time 570–574, 583, 587, 588, 593, 894
relaxation time (T_1)
– longitudinal 305
relaxation time (T_2)
– transverse 305
release
– long-term of drugs 994
– stiction 1095
– voltage 1062
reliability 831
– long-term 1126
– of MEMS 995
– testing 1125
relief on a surface 186
remanence 137
remote detection system 354
removable disk 901
repetitive IgG 484
replica molding 833
repulsive force 480
residual
– film stress 1023
– stress 692, 699, 709, 804, 809, 812
– thickness 855
resistive micro-heater 1119
resolution 233, 374
– vertical 331

resonance
– curve detection 339
– frequency 454
– mechanical 591
resonant
– enhancement effects 131
– frequency 373
Resovist 306
response time 467
rest time 869
retardation effect 550, 552
reticuloendothelial system (RES) 296
retractive slips 885
Reynolds hydrodynamic theory of lubrication 885
RF magnetron sputtering 795
rheological
– characterization 865
– model 469, 664
rhombhedral R-8 706
rigid
– biphenyl chain 853
– disk drive 862, 869
– surface asperities 723
Ritz procedure 1055
RLL (run-length-limited) code 947
RMS roughness 1010
RNA polymerase 747
robust manufacturing 1087
rolling friction 552, 568
root mean square 433
ropes (SWNT) 66
rotary microactuator 958, 993
rotating cylinder 907
rotating strain gauge 1026
rotaxane 13
rough macroscopic contact 885
roughness 558, 559, 566–568, 594
RTP (rapid thermal processing) MEMS vacuum packaging 1119
Ru(II)-trisbipyridine 13
rule of mixtures 708, 709
run-length-limited (RLL) 947
rupture forces 488

S

sacrificial film layers 1040
sales volume 1093
SAM (self-assembled monolayer) 165, 176, 178, 622, 1088
– coating 1075
– melting point of 853
– modified surface 620
sample holder 416
sandwich immunoassays 270
sandwiched molecules 26
saturated calomel electrode 24
saturated hydrocarbons 835
scale-dependent surface tension 614
scaling factor 975
scan
– area 335
– direction 345
– frequency 335
– head 417
– range 359, 360
– rate 329
– size 337
– speed 360
scanner, piezo 436
scanning
– acoustic microscopy (SAM) 663, 692
– capacitance microscopy (SCM) 326
– chemical potential microscopy (SCPM) 326
– electrochemical microscopy (SEcM) 326
– electron microscope (SEM) 110, 373, 521
– electrostatic force microscopy (SEFM) 326
– head 328
– ion conductance microscopy (SICM) 326
– Kelvin probe microscopy (SKPM) 326
– lateral range 929
– local-acceleration microscopy (SLAM) 663
– magnetic microscopy (SMM) 326
– nanoindentation (SN) 674
– nanoindenter 662
– near field optical microscopy (SNOM) 326
– polarization force microscopy (SPFM) 615
– speed 335
– system 359
– thermal microscopy (SThM) 114, 326
– tunneling microscope (STM) 385, 396, 450, 498, 922
– tunneling microscopy (STM) 114, 325, 371, 382, 397, 899
– tunneling probe 114
– velocity 519
scanning force
– acoustic microscopy (SFAM) 326
– microscopy (SFM) 326, 433, 663, 884
– spectroscopy (SFS) 433
scanning probe
– lithography 186
– methods 606
– microscope (SPM) 326
– microscopy (SPM) 371, 385, 612, 663, 899
schematic diagram of T-SLAM 663
Schottky diode 132
scission 622
scope of therapeutic nanodevices 285
scratch
– critical load 824
– damage mechanism 818, 819
– depth 771, 773, 1000
– depth profile 773
– drive actuator 1031
– profile 519
– resistance 765, 771, 853, 1001
– test 520, 814
scratch-induced damage 813
scratching measurement 325
SCREAM 161
screen printing 193
sealing 1112
second anodization 100
secondary signaling drug delivery 301
Seebeck coefficient 123, 133
seek
– control 976
– time 901
selected area electron diffraction (SAED) 111
selective electrodeposition 108
self-actuation voltage 1062
self-aligning optical system 236
self-assembled
– growth 420
– microscopic vesicles 834
– monolayers (SAMs) 27, 165, 176, 187, 476, 536, 606, 618, 703, 833, 839, 861, 1003, 1004
– nanoparticles 312
– nanotube bunches 58
– structures 291, 310
self-assembled monolayer (SAM)
– patterned 188
self-assembly 131, 147, 170, 173, 176, 178, 179, 288
– in biological systems 740
– of nanostructures 289
– of organic layers 23

self-enhancing instability 458
self-excitation
– modes 461
– scheme 464
self-lubrication 652, 653
self-regulation of a therapeutic device 302
self-replicating nanobot 1146
self-tuning control 974, 976
semibiological nanodevices 281
semiconductor 103, 391
– economics 1086
– quantum dots 428
– surface 392
– SWNT 74, 85
semimetallic nanowire 118
semimetal–semiconductor transition 118, 123
sensing
– channel 303
– element 227
– modalities of therapeutic nanotechnology 302
sensing devices
– biological 137
– chemical 137
sensitivity 356, 357, 390, 467, 479, 633, 972
sensitizer 30
sensor 233
– and effector communication 283
– applications 117
– noise 975
– systems 302
– transducer interface 304
sequential logic circuits 20
sequential microfluidic manipulation 268
servo
– demodulation 945
– field 943
– loop 943
– self-writing 945
servo control design 968
sessile-drop 832
SFA (surface forces apparatus) 884, 894, 1011
– measurement 613
SFM (scanning force microscope) 884, 888, 890, 894
– shear force experiments 893
shallow indentation measurement 679
shape memory alloys (SMA) 260
sharpened metal wire tip 382
shear
– energy 887
– force *see* kinetic friction, static friction
– melting 570, 586
– modulus 633, 933
– rate 866
– strength 646
– stress 850, 887
– thinning 581, 583, 584, 589
shear flow
– damping 1073
– detachment (SFD) 479
shear stress
– critical 571, 572, 577, 585, 586
– effective 648
shearing force 848
shearing interface 721
shear-melting transitions 734
SH-group 476
short-cut carbon nanotubes 428
short-range
– chemical force 437, 464
– chemical interaction 402
– contribution 386
– electrostatic attraction 399
– electrostatic interaction 398
– energy dissipation 468
– interatomic force 455
– magnetic interaction 397, 402, 403
shot noise 355
Shubnikov–de Haas quantum oscillatory effect 122
shuttle
– displacement 957
– finger 956
Si 373, 375, 619
– adatom 393
– cantilever 373, 458, 465
– MEMS 247, 985
– nanobeam 773, 781, 785
– stamp 190
– tip 374, 392
– trimer 395
– wafer 373, 459, 619
Si(001)2×1 394
Si(001)2×1-H surface 394
Si(100) 515, 520, 614, 818, 820, 862, 867–869, 875, 877, 878, 995, 1012
– wafer 863
Si(100)2×1:H monohydride 393
Si(111) 519, 527, 848, 850, 999
Si(111) surface 404
Si(111)$\sqrt{3}\times\sqrt{3}$-Ag 395, 396
Si(111)(7×7) 434, 468, 640
Si(111)(7×7) surface 389, 397, 462, 639
$Si(OH)_4$ film 1013
Si_3N_4 375, 501, 619
– based tip 619
– cantilever 373
– layer 373
– tip 341, 374, 531, 841, 848, 866, 871, 880
Si-Ag covalent bond 397
Si-based devices 764
SiC 995
– bulk properties 995
– film 770
signal
– distribution 1112
– transduction 20
silane (SiH_4) 206
silica
– thermally grown 840
silicon 332, 356, 388, 610, 693, 705, 1111
– AFM tip 478
– cantilever 387, 633
– chemistry 478
– crystalline 254
– dioxide (SiO_2) 254
– dioxide layer 31, 33
– FET 303
– fractures 777
– fusion 1118
– grain 1035
– membrane 263
– MEMS 1095
– micromachining 227
– micromotor 866
– oxide 106, 477
– surface 477
– tip 392, 434, 463, 635
– wafer 25, 27, 186, 189, 228, 286, 453, 614, 930
silicon nitride (Si_3N_4) 254, 332, 477, 610, 615, 913, 964, 968
– cantilever 372
– layer 1117
– tip 375, 635
silicon-based MEMS sensor 229
silicon-based mold masters 255
silicon-on-insulator (SOI) 205, 241, 964
– substrates 247
silicon-only wafer 965
silicon-on-silicon contact 1012
silicon-silicon bonding 1096
silicon-terminated tip 435

silicon-to-glass fusion bonding 1120
siloxane 833
simultaneous imaging 397
– of a TiO_2(110) surface 401
single asperity 646
– AFM tip 868
– contact 517, 1012
– size of 722
single bond 622
single crystal
– aluminum (100) 522
– nanowires 104
– Si(100) 770, 862
– silicon 504, 964, 995, 1024
– silicon cantilever 341, 611
– wafers 204
single crystalline iron 675
single domain magnetic 138
single magnetic flux quantum 122
single molecular motor 480
single molecule
– assays 743
– biology 743
– detection 477
– experiment 760
– studies 485
single particle model 894
single point sensor 1091
single protein therapeutics 281
single receptor-ligand pair 480
single stranded DNA (ssDNA) 746
single-input-multi-output (SIMO) 956
single-input-single-output (SISO) 969
single-layer
– beam 1044
– film 1050
single-particle wave function 422
single-stranded nucleic acid molecule (SSNA) 302
single-wall carbon nanotube (SWCNT) 34, 40, 289, 315, 342, 376, 668
single-wall nanocapsule 49
singularity in joint density of states 134
sink-in 693, 699, 700, 709, 711
– behavior 674
sintering 553
SiO_2 54, 373, 375, 531, 615
– film 770
– nanobeam 773, 774, 785
– thin films 205
SiOH groups 478
Si-SiO_2 stress formation 375
SISO 973
– design 969
site-specific drug delivery 298
Si-wafer 460
SLAM 663
slider 952
sliding
– contact 651
– direction 585
– distance 587
– friction 734
– of tip 632
– speed 889
– velocity 574, 585, 644, 884
SLIGA 168
s-like tip states 422
slip distance 732, 889
slipping and sliding regimes 586
SMA (shape memory alloys) 260
– actuation schemes 260
– driven micropumps 261
small angular deflection 1068
small specimens handling 1025
SMANCS (S-Methacryl-neocarzinostatin) 309
– albumin complexes 309
smart media card 905
smart plastic biochip 273
smooth sliding 585
snap-in process 609
SnO_2(110) 400
soap-like lubricants 887
social
– perception 1145
– services 1142
societal impacts 1136
socio-technical trends 1141
soft
– cantilever 453
– coatings 708
– lithography 833
– polymer stamps 190
SOI wafer 1096
solar furnace
– devices 52
– method 61
solder bonding 1115, 1118
solid boundary lubricated surfaces 571, 572, 576–578, 587, 588
solid solution catalyst 55
solid xenon 436
SOLID95 779
solidification 578, 582
solid-like
– behavior 593
– Z-DOL(BW) 866
solution-phase synthesis 107
solvation
– effect 621, 623
– forces 545, 557, 558, 561, 567
sonolubrication 513
soot 46
SP 815
spacer chain 839–841
– length 844
spark plasma sintering (SPS) 81
spatial variation curvature 1058
specific accumulation of nanoparticles 311
specificity 752
spectroscopic resolution in STS 419
speed-energy product 1073
spin
– casting 192
– coating 887
– density waves 431
– quantization 423
spin-on-glass 965
spiropyran 18
SPM (scanning probe microscopy) 73, 372
– methods 672
– tip 452
spreading
– dynamics 892
– profile 892
spring
– sheet cantilever 341
– system 327, 929
spring constant 377, 387, 390, 393, 433, 450, 451, 469, 480, 483, 611, 633, 671, 911, 956, 1030, 1067
– calculation 347
– changes 415
– effective 637, 645, 721
– lateral 338, 636
– measurement 346
– torsional 1068
– vertical 338
sputtered coatings
– physical properties 804
sputtering
– deposition 799
– power 805
sp^3 bonding 806
sp^3-bonded carbon 813
squeeze film damping 248, 1073
SRAM
– cells 1088

– CMOS static random access memory 235
SrF_2 398, 399
$SrTiO_3$ 402
$SrTiO_3$(100) 400, 401
stability
– mechanical 927
stabilization 1065
stable field performance 1102
stainless steel cantilever 813
stamp
– Au coated PDMS 192
– Au/Ti coated 191
– composite 188
– depth 190
– fabrication 186
– mechanical properties 187
– positioning 195
standardization 249
static
– advancing contact angle (SACA) 847
– AFM 388, 469
– charge 1103
– deflection AFM 450
– friction 571, 576, 579, 583, 585–587, 589, 590, 593, 718, 731, 1034
– friction force 722, 866, 1014
– interactions 544, 584
– mode 387, 433, 453
– mode AFM 354
statistical fluctuation model 889
statistical kernels 891
steady-state sliding *see* kinetic friction
steady-state velocity 733
step tilting 855
stepping motor 636
stereoelectronic properties 12
stick slip 592
sticking regime 586
stick–slip 571, 576, 579, 583–585, 587, 590, 876, 885, 886, 889, 894
– mechanism 636
– motion 732
stiction 579, 832, 885, 991, 1088, 1103
– phenomena in micromotors 1016
– suppression 1103
stiffness 390, 663, 819
– torsional 349
stiffness coefficient
– elastic 1048
stiffness measurement
– continuous (CSM) 810
STM (scanning tunneling microscope) 173–176, 325, 397
– cantilever 330
– cantilever material 339
– principle 327
– probe construction 330
– tip 325, 417
stoichiometry of compound nanowires 115
Stokes's equation 933
Stone–Wales defect 62, 68
Stoney equation 1024
storage
– capacity 946
– drive industry 902
– field 943
– modulus 702
– surface topography 942
strain
– energy 609
– energy difference 810
– engineering 147, 170
– theory 1144
– thermally induced 1125
– transformation 1041
strain–displacement relation 1049
stray capacitance 357
strength 1023
streptavidin 478
– mutants 484
stress 751, 930, 1023
– activation volume 887
– component 1043
– distribution 779, 1049
– field 1024
– gradient 1026
– maximum 1027
– measurement 1025
– relaxation 1060
– singularity 1043
– tensor 759
– transformation 1044
– uniaxial 1042
stresses within the capsid walls 758
stress–strain
– behavior 1028
– curve 671, 784
– relation 1048
stress–temperature response 1060
stretch modulus 751
Stribeck curve 569, 584, 884
strip domain 439
structural
– characterization 110
– entropy 886, 887
– forces 545, 557, 559, 566
– integrity 764
– material 1028
structurally programmable microfluidic system (sPROMs) 264
structure factor 730
structuring of simple liquids 886
SU-8 resist 929
sub-Angstrom deflections 372
subcontractor 1087
subculture theory 1145
sublimation 52
submicron
– electron-beam lithography 246
– ICs 247
– lithographic techniques 246
submonolayer
– coverage 864
– lubrication 731
subnanometer precision 10
sub-nanonewton precision 452
substrate 839, 840
substrate curvature technique 1024
substrate–slider interactions 727
subsurface water irrigation 676
super modulus 691
supercapacitor 85
superconducting
– gap 431
– magnetic levitation 327
– matrix 440
superconductivity 419
superconductors 103, 429
– type-I 431
– type-II 431
superficial nanotechnology aspects 1140
superlattice
– nanowire 103, 114
– structure 108
superlubricity 642
superparamagnetic materials 305
superparamagnetism 138
superstructures of nanowires 108
supramolecular
– assembly 17
– forces 35
surface
– adsorption 263
– amorphous 544
– asperity effect 513
– atom layer 388
– band 423
– barrier potential height 890
– charge 557, 621, 1074
– charge density 555–557

– chemistry 190, 296
– crystalline 544
– diffusion 190
– effects 129
– elasticity 529
– energy 619, 792, 1088
– energy (tension) 547, 552, 566, 567, 572, 574, 578
– forces 544
– forces apparatus (SFA) 479, 498, 546–548, 552, 576, 585, 588, 606, 717, 731, 884
– free energy 745, 746
– hydrophobic 544, 560
– micromachined accelerometers 231
– micromachining 163, 226, 229
– micromechanical properties 498
– microtribological properties 498
– mobility 865
– nanomechanical properties 498
– nanotribological properties 498
– passivation 1102
– potential 419, 555–557, 798
– potential measurement 503
– slope 844
– state lifetime 423
– stiffness 506
– structure 111, 400, 544, 558, 559, 566, 567, 570, 586, 591
– tension 265, 516, 618, 832, 846, 862
– terminal group 839
– topography 632, 778, 901
– transfer chemistry 190
– treatment 1104
surface roughness 383, 674, 678, 724, 773, 778, 844, 942, 997, 1009
– map 510
– measurement 500
surface-mounted biofilter 272
surfactant monolayers 547, 564, 568–570, 577, 580, 584, 585, 588
surfactants 103, 107
surgical resection 307
suspended membrane 1028
suspension 967
– vibration 959
suspension-level fabrication 967
sustainable internal pressure
– maximum 759
sustained drug delivery 269
switch
– two-state 19
switch design 1062
switches
– molecule-based 15
– monolayer 28
switching
– architecture 238
– devices 15
– energy 135
– speed 135
– time 1073
SWNT (single wall nanotube) 40, 52
– adsorption properties 63
– based FET 85
– based materials 64
– catalyst-based 61
– conductance 75
– electronic structure 66
– epoxy composite 82
– flexural modulus 67
– magneto-resistance 67
– matrix interaction 82
– production techniques 45
– rope 668, 670
– structure 41
– surface area 64
– tensile strength 67
– thermal conductivity 66
SWNT-PMMA composite 83
symmetric deformation mode 1054
sync pattern 946
synchronization 945
synthesis 100
– reactor 47
– yield 47
synthetic
– devices 281
– nanoporous membrane 313
– readback signal 941
system
– underdamped 894

T

tailoring of SAMs 841, 1006
tapping amplitude 529
tapping-mode 332, 373, 456, 501, 910
– AFM 456, 459
– etched silicon probe (TESP) 341
– resistance measurement 914
– tip assembly 507
targeting of nanomaterials 296
Taylor dispersion pattern 266
technical PVC 664
technological convergence 1139
Teflon layer 652
temperature
– annealing 227
– critical 423, 434, 1059
– dependence of friction 645
– domain 664
– effect 873
– regulation 720
– sensitive microcantilever 303
temperature-dependent resistance 118
template
– manufacturing 170, 176, 179
– synthesis 100
template-assisted synthesis 100
templates 103
templating technique 57
temporal scales 749
tensile
– load 1027
– maximum stress 783
– straining 1059
– stress 767, 1026, 1035, 1046
– test 1036
terabit 903
terabyte 901
test environment 824
tethering
– flexible 491
tetraarylmethane 13
tetracene 20
– channel 21
tetramethyl-aluminum hydroxide (TMAH) 204, 254
tetrapods 108
tetrathiofulvane (TTF) 916
T-filter 266
therapeutic
– bioactivity 312
– nanodevices 280
– nanomaterials 281
– nanotechnology 285
– photo-oxidation 299
thermal
– activation 894
– activation model of friction 884
– activation model of lubricated friction 887
– conductivity 125, 133
– drift 414, 417
– effect 642
– energy 481
– expansion coefficient 388, 1042
– expansion mismatch 1056
– fluctuation 633
– fluctuation forces 564
– fluctuations and velocity 890
– force 751

– frequency noise 415
– management 1112, 1122
– mismatch 1125
– noise 481, 891
– sensing 925
– vibration amplitude 377
thermocouples 123
thermodynamic forces 293
thermoelastic behavior 1040
thermoelectric
– applications 125, 133
– figure of merit 125, 133
– properties 122, 123
thermo-gravimetric analysis (TGA) 72
thermomechanical
– load 1040, 1058
– noise 433
– write/read 924
thermomechanics of multilayer film 1046
thermoplastics 257
thermopneumatic actuation 258
theta condition 561
thick film lubrication *see* elastohydrodynamic lubrication
thickness 1047
thickness of nanotubes 668
thin film 701, 707
thin oxide film 913
thin-film
– limit 1051
– lubrication 498
– microstructure 1040, 1041
thin-plate theory 1046, 1054
thiol 464, 837
three-axis accelerometer 233
three-digit input/output strings 20
three-dimensional force field spectroscopy 437
three-dimensional micromixers 265
three-input NOR gate 22
three-layer symmetric plate 1052
three-state molecular switch 18
three-terminal devices 31
three-way microvalve system 258
threshold potential 726
through wafer etching 1113
through-thickness
– average 1044
– cracking 807
thymine 10
Ti atom 405
TiAl alloy 675
TiC grains 508
tight-binding approximation 402
tilt angle of SAMs 855
time scales 721
timing field 943
timing recovery 945
TiO_2 400, 402, 425
– substrate 405
$TiO_2(100)$ 400
$TiO_2(110)$ 405
TiO_2-terminated layer 401
tip
– anion-terminated 399
– apex 435, 477, 655, 927
– artifact 392
– atom(s) 415
– atomic structure 383
– characterization 618
– deflection 503
– displacement 941
– Fe-coated 403
– ferromagnetic 441
– geometry 377
– hydrophilic 616
– load 933
– material 383
– mount 329
– oscillation 373
– oscillation amplitude 654
– performance 382
– preparation in UHV 635
– preparation method 330
– properties 382
– radius effect 339, 515, 813, 876, 878
– shape 611
– surface 477
– vibration 460
– vibration amplitude 664
tip-bound
– antigens 484
– biomolecule 303
tip-broadened image 374
tip–cantilever assembly 376
tip-induced atomic relaxation 435
tip-induced quantum dot 424
tip–liquid interface 531, 868
tip–molecule
– distance 419
– gap 406
– long-range force 406
tip–particle distance 29
tip–polymer interface 935
tip–sample 451
– contact 467
– dissipative interaction 468
– distance 388, 396, 399, 401, 435, 451, 459, 507
– electric field 656
– energy dissipation 389
– force 373, 388, 433, 436, 450, 453, 455, 460, 461
– force gradient 391
– gap 910
– interaction 391, 433, 441, 455, 456, 462, 654
– interaction force 608
– interaction potential 464
– interface 372
– junction 468
– potential 390
– separation 467, 608, 609
– system 664
tip–surface
– distance 406, 468
– interaction 490, 651
– potential 636
– separation 480
tip-terminated atom 400
tire pressure monitor 1094
tissue dendritic cells (DCs) 296
tissue engineering scaffolds 310
titin 485
TMR sources 969
Tomlinson model 631, 636
– finite temperature 642
– of friction 884
– one-dimensional 636
– two-dimensional 637
tooth enamel 675
top-down
– approach 10, 133
– materials 281
topographic
– AFM image 611
– images 331
topographical asymmetry 403
topography measurement 334
torsion mirror
– vertical 238
torsional plate 1068
torus model calculation 360
total internal reflection microscopy (TIRM) 550
touch mode architecture 228
Townsend electron avalanche theory 1077
toxicity monitoring 274
trabecular bone 679
track
– centerline 943, 945
– following 943
– mis-registration (TMR) 969
– pitch (TP) 943

track-etched
– polycarbonate membranes 102
– polymers 100
track-following controller 971
tracking servo loop 943
tracks per inch (TPI) 952
transducer component 302
transfer time 901
transformation
– strain 1041
– stress 1044
transistors 132
transition metal complexes 31
transition metal oxides 402
transitions between smooth and stick–slip sliding 585
translational microactuator 958
transmembrane channel 741
transmission 129
transmission electron microscopy (TEM) 111, 374, 521, 693
transplant rejection 311
transport properties of nanowires 115, 118
transversal piezo-resistive effect 356
transverse
– magnetoresistance 121
– midplane displacement 1055
– relaxation time (T_2) 305
traveling direction of the sample 346
Tresca criterion 695
triangular cantilever 339, 349, 354
triaryl 13
tribochemical
– oxidation 999
– reaction 531, 872
tribological
– characterization of coating 805
– computer simulations 721
tribological issues
– in MEMS 989
– in NEMS 989
tribological performance 996
– of coatings 822
tribological properties
– of SAM 841
– of silicon 996
tribometer 996
(tridecafluoro-1,1,2,2-tetrahydro-oct-1-yl) trichlorosilane ($n\text{-}C_6F_{13}CH_2CH_2SiCl_3$, FTS) 842
triethanolamine 14
triethoxysilane 644
trifluoromethyl-terminated SAM 843
triggering 288
– external 284
– strategy 298
trimer tip apex 452
true atomic resolution 389, 392, 434
true tip-sample interaction 608
truth table 17, 21
T-SLAM 663
tubulin monomer 740
tumor
– architecture 307
– associated antigens (TAAs) 307
– deposition 309
– properties 307
– targeting 307
– vasculature 307
tungsten 330
– sphere 633
– tip 330, 468, 614
tunneling
– current 327, 388, 389, 424
– detector 331
– junctions 25
– tip 431
turn-on voltage 117
two-body potentials 719
two-digit input strings 20
two-dimensional
– elasticity 754
– electron gas (2-DEG) 438
– electron system (2-DES) 428
two-layer beam 1045
two-state switch 19
two-terminal nanowire device 132

U

UHV environment 638
UHV-AVM 635
ultrahigh vacuum (UHV) 391, 418, 461, 607, 635
ultrasonic
– bonding 1119
– lubrication 513
– transducer 663
ultrasound triggering 299
ultrathin DLC coatings 795
umklapp processes 125
unbinding
– force 476, 481
– force distribution 485
– pathway 488
uncapped MEMS wafer 1097, 1098
uncoated monolithic cantilever 611
uncommon failure mode 1102
unconstrained binding 477
unidirectional electron transfer 28
unimolecular level 22
universality 752
unlubricated sample 879
unlubricated surfaces *see* dry surfaces
upward cantilever deflection 746
UV-LIGA lithography 267

V

vacuum encapsulation 1116, 1117, 1126
vacuum packaged MEMS 1118
vacuum packaging 231, 1116, 1121
van der Waals 607, 617, 621
– adhesion 616
– attraction 515
– contact 759
– force 372, 386, 398, 404, 406, 450, 544, 545, 550–554, 557, 558, 618
– interaction 58, 289, 402, 457, 843, 865
– surfaces 435
vanadium carbide 691
vapor deposition 104, 120
vapor grown
– carbon fiber 53
– carbon nanofiber 56
vapor-liquid-solid (VLS) mechanism 105
varactor diodes 241
variable capacitor 240
variable force mode 362
variable temperature STM
– setup 416
variable-temperature SLAM (T-SLAM) 663
vascular
– address system mapping 297, 298
– prosthesis 310
VCM loop compensator 972
vehicle stability 232
velocity
– critical 585, 587, 590, 591
velocity dependence of friction 644
velocity effect 871
Verlet algorithm 457
vertical coupling 329
vertical rms-noise 434
vertically aligned MWNT 59
Verwey transition temperature 439

very large-scale integration (VLSI) 922, 940
vibration 749
– external 329
vibratory rate gyroscope 232
vibromotor 236
Vickers
– hardness 674, 678, 695
– indentations 771
– indenter 765
vinylidene fluoride 404
viral life cycle 755
virgin single-crystal silicon 996
virtual jump distance 889
viruses 747
viscoelastic
– effect 468
– model 934
– properties 681
– relaxation time 731
– response 731
viscoelasticity 663, 702, 711
– mapping 505, 528
viscosity 730, 885
viscous
– damping 469
– force 569, 580, 584, 1061
VLS (Vapor-Liquid-Solid) 60
– growth method 116
– mechanism 60
– method 105, 108, 137
VLSI (very large scale integration) 922, 940
voice coil motor (VCM) 952
voltage bias 383
voltage-controlled oscillator (VCO) 239, 946
voltage-displacement function 1068
voltage-to-force gain 956
von Karman plate theory 1055
vortex in superconductor 440
V-shaped cantilever 341, 610, 634

W

W tip 330
wafer
– curvature technique 1034
– fabrication 1087
– saw dilemma 1097, 1098
wafer-level vacuum packaging 1122
wafer-to-wafer
– vacuum packaging 1114
– variation 1072
waste disposal concern 1090
water 613
– capillary force 847
– meniscus 461
water vapor 691, 695
– content 824
Watson-Crick base pairing 484
wear 498, *see* damage, 618, 793, 809
– contribution to friction 650
– control 832
– damage 824
– damage mechanism 823
– debris 521
– depths 1000
– measurement 325
– mechanical 263
– mechanism 522
– of tip 728
– performance 866
– process 822
– region 523
– resistance 822, 823, 853, 856, 879, 997, 1008
– study 879
– test 822
wearless
– friction 884
– static friction 726
Weibull
– distribution 774
– distribution function 1125
– statistics 1036
weight function 390
weighting functions for μ-synthesis 976
well-aligned nanowires 111
wet (chemical) etching 254
wet etching in HF 247
wettability 846
wetting 569
– lubricants 892
white-noise driven system 889
wire boundary scattering 121
wire cantilever 339
wireless communication 1094
work
– hardening 704, 709
– of adhesion 844, 847, 1011
– of indentation 701
write/read scheme 931
writing mechanism 932

X

xenon 417
XNOR 17
X-ray
– analysis 115
– diffraction (XRD) patterns 103
– lithographic techniques 286

Y

yield point 570, 571, 585
– load 703
yield stress 703
yielding 1055
Young's modulus 81, 346, 362, 377, 415, 611, 646, 668, 677, 779, 1023, 1024, 1044, 1071
– measurement 1028
Young–Dupre equation 846
Young–Laplace equation 612

Z

Z-15 530, 862, 874–876, 878, 1006
– film 531, 879
– lubricant 1003
– lubricant film 867
– properties 863
z-axis accelerometer 230
z-axis gyroscope 233
Z-DOL 530, 614, 862, 869
– partially bonded film 880
– properties 863
Z-DOL(BW) 868, 874, 877–879, 1006
– film 530, 867, 1003
zeolites 54
zeta potential 261
zig-zag axis 42
zig-zag-type SWNT 42
zinc 704
zipper actuator 1070
Zn nanowire 120, 123
ZnO nanowire 135