

EDITED BY

A. V.

NARLIKAR

Y. Y.

FU

≡ The Oxford Handbook of  
**NANOSCIENCE  
AND TECHNOLOGY**  
VOLUME III: APPLICATIONS

THE OXFORD HANDBOOK OF  
**NANOSCIENCE  
AND  
TECHNOLOGY**

*This page intentionally left blank*

# **The Oxford Handbook of Nanoscience and Technology**

Volume III of III  
Applications

*Edited by*  
A.V. Narlikar  
Y.Y. Fu

**OXFORD**  
UNIVERSITY PRESS

**OXFORD**  
UNIVERSITY PRESS

Great Clarendon Street, Oxford ox2 6DP

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide in  
Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press  
in the UK and in certain other countries

Published in the United States  
by Oxford University Press Inc., New York

© Oxford University Press 2010

The moral rights of the authors have been asserted  
Database right Oxford University Press (maker)

First published 2010

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
without the prior permission in writing of Oxford University Press,  
or as expressly permitted by law, or under terms agreed with the appropriate  
reprographics rights organization. Enquiries concerning reproduction  
outside the scope of the above should be sent to the Rights Department,  
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover  
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data  
Data available

Library of Congress Cataloging in Publication Data

The Oxford handbook of nanoscience and technology : frontiers and  
advances : in three volumes / edited A.V. Narlikar, Y.Y. Fu.  
p. cm.

Includes bibliographical references and index.  
ISBN 978-0-19-953306-0 (hardback)

1. Nanotechnology—Handbooks, manuals, etc. 2. Nanoscience—Handbooks, manuals, etc.  
I. Narlikar, A. V., 1940— II. Fu, Y. Y.  
III. Title: Handbook of nanoscience and technology.

T174.7.094 2010

620'.5-dc22 2009036761

Typeset by SPI Publisher Services, Pondicherry, India  
Printed in Great Britain  
on acid-free paper by  
CPI Antony Rowe, Chippenham, Wiltshire

-----  
ISBN 978-0-19-953306-0

1 3 5 7 9 10 8 6 4 2

# Preface

Wolfgang Pauli is known to have remarked, “God made solids, but surfaces were the work of the Devil.” This Handbook deals with the Devil’s work. As the size of the material is reduced, surfaces acquire increasing importance, and indeed override the bulk when one of the dimensions of the material shrinks to nanometers. Simultaneously, at the nanoscale, quantum effects come into play and the properties of matter confined to nanodimensions are dramatically changed. Nanoscience and nanotechnology are all about relating and exploiting the above phenomena for materials having one, two or three dimensions reduced to the nanoscale. Its evolution may be traced to three exciting happenings that took place in a short span from early to mid-1980s with the award of Nobel prizes to each of them. These were the discovery of the quantum Hall effect in two-dimensional electron gas, the invention of scanning tunnelling microscopy (STM) and the discovery of fullerene as the new form of carbon. The latter two, within a few years, further led to the remarkable invention of the atomic force microscope (AFM) and, in the early 1990s the extraordinary discovery of carbon nanotubes (CNT), which soon provided the launch pad for the present-day nanotechnology. The STM and AFM have emerged as the most powerful tools to examine, control and manipulate matter at the atomic, molecular and macromolecular scales and these functionalities constitute the mainstay of nanotechnology. Interestingly, this exciting possibility of nanolevel tailoring of materials was envisioned way back in 1959 by Richard Feynmen in his lecture, “There’s plenty of room at the bottom.”

During the last 15 years, the fields of nanoscience and nano-technology have expanded internationally and their growth has perhaps been more dramatic than in most other fields. They have been transformed into an intense and highly competitive research arena, encompassing practically all disciplines that include theoretical and experimental physics, inorganic, organic and structural chemistry, biochemistry, biotechnology, medicine, materials science, metallurgy, ceramics, electrical engineering, electronics, computational engineering and information technology. The progress made in all these directions is truly spectacular. In this edited Handbook of Nanoscience and Technology, we have attempted to consolidate some of the major scientific and technological achievements in different aspects of the field. We have naturally had to follow a selective rather than exhaustive approach. We have focused only on those topics that are generally recognized to have had a major impact on the field. Inherent in this selection process is the risk of some topics inadvertently getting overemphasized while others are unavoidably left out. This is a non-trivial problem especially in light of the great many developments that have

taken place in the field. However, a great diversity of important developments is represented in this Handbook and helps us overcome some of these risks.

The present Handbook comprises 3 volumes, structured thematically, with 25 chapters each. Volume I presents fundamental issues of basic physics, chemistry, biochemistry, tribology, etc. at the nanoscale. Many of the theoretical papers in this volume are intimately linked with current and future nanodevices, molecular-based materials and junctions (including Josephson nanocontacts) and should prove invaluable for further technology development. Self-organization of nanoparticles, chains, and nanostructures at surfaces are further described in detail. Volume II focuses on the progress made with a host of nanomaterials including DNA and protein-based nanostructures. This volume includes noteworthy advances made with the techniques of improved capability used for their characterization. Volume III highlights engineering and related developments, with a focus on frontal application areas like Si nanotechnologies, spintronics, quantum dots, CNTs, and protein-based devices, various biomolecular, clinical and medical applications. The other prominent application areas covered in this volume are nanocatalysis, nanolithography, nanomaterials for hydrogen storage, nano-field-emitters, and nanostructures for photovoltaic devices. This volume concludes the Handbook with a chapter that analyses various risks that are associated in using nanomaterials.

We realise that the boundaries separating a few of the topics of the above three volumes are somewhat shadowy and diffuse. Some articles of Volumes II and III could have also provided a natural fit with Volume I. For instance, some of the novel molecular devices of Volume III could have alternatively been included in the realm of basic studies that form a part of Volume I.

The three volumes together comprise 75 chapters written by noted international experts in the field who have published the leading articles on Nanoscience and Nanotechnology in high-profile research journals. Every article aims to bring out *frontiers and advances* in the topic that it covers. The presentation is technical throughout, and the articles in the present set of 3 volumes are not directed to the general and popular readership. The set is not intended as a textbook; however, it is likely to be of considerable interest to final-year undergraduates specializing in the field. It should prove indispensable to graduate students, and serious researchers from academic and industrial sectors working in the fields of nanoscience and technology from different disciplines like physics, chemistry, biochemistry, biotechnology, medicine, materials science, metallurgy, ceramics, electrical, electronics, computational engineering, and information technology. The chapters of the three volumes should provide readers with an analysis of the state-of-the-art technology development and give them an opportunity to engage with the cutting edge of research in the field.

We would like to thank all the contributors for their splendid and timely cooperation throughout this project. We are grateful to Dr Sonke Adlung for being most cooperative and considerate and for his important suggestions to help us in our efforts, and acknowledge with thanks the efficient assistance provided by April Warman, Ms Phaedra Seraphimidi and Mr Dewi Jackson. Special thanks are due to Mrs Emma Lonie and Ms Melanie Johnstone for

commendably coordinating the proof correction work with over 200 contributors. One of us (AVN) thanks the Indian National Science Academy, New Delhi for financial assistance in the form of a Senior Scientist fellowship and the UGC-DAE Consortium for Scientific Research, Indore, for providing infrastructural support. He thanks the Consortium Director, Dr Praveen Chaddah, and the Centre Director (Indore), Dr Ajay Gupta, for their sustained interest and cooperation. He further acknowledges with thanks the technical assistance provided by Mr Arjun Sanap, Mr D. Gupta, Dr N.P. Lalla, Mr Suresh Bharadwaj, and Mr U.P. Deshpande on many occasions. He is particularly grateful to his wife Dr Aruna Narlikar for her invaluable help, patience, and support throughout, and especially for her useful suggestions on many occasions during the course of the present project. He acknowledges the commendable technical support of his daughter Dr Amrita Narlikar at Cambridge, and also of Dr Batasha who remains a close and valued friend of the family. YYF extends his thanks to the National Natural Science Foundation of China (Contracts No. 60776053 and No. 60671021), and the National High Technology Research and Development Program of China (Program 863 and Contract No. 2007AA03Z311) for financial support. He remains indebted to his father, who passed away many years ago, for his invaluable guidance, advice and help to build his life and career, and to his mother, wife and son, for their sustained patience and support.

A.V. Narlikar

Y.Y. Fu

November 2008

*This page intentionally left blank*

# Contents

<b>List of Contributors</b>	<b>xvi</b>
<b>1 Role of computational sciences in Si nanotechnologies and devices</b>	<b>1</b>
<i>K. Shiraishi and T. Nakayama</i>	
1.1 Introduction	1
1.2 Present Si technology trend stimulated by scientific knowledge	2
1.3 Key knowledge for Si nanodevices obtained by computational science	3
1.4 Future Si technology trend predicted by computational science	39
1.5 Summary	43
Acknowledgments	43
References	43
<b>2 Few-electron quantum-dot spintronics</b>	<b>47</b>
<i>D.V. Melnikov, J. Kim, L.-X. Zhang, and J.-P. Leburton</i>	
2.1 Introduction and motivations	47
2.2 Two electrons in double quantum dots	50
2.3 Two electrons in quantum wire quantum dots	72
2.4 Few electrons in triple quantum dots	76
2.5 Conclusion	81
Acknowledgments	82
References	85
<b>3 Spintronics with metallic nanowires</b>	<b>90</b>
<i>J.-Ph. Ansermet</i>	
3.1 Introduction	90
3.2 Spin diffusion	96
3.3 Models for spin-polarized currents acting on magnetization	101
3.4 Current-induced magnetization switching	108
3.5 Current-driven magnetic excitations	114
3.6 Resonant-current excitation	118
3.7 Conclusion	123
References	124
<b>4 Molecular nanomagnets: Towards molecular spintronics</b>	<b>136</b>
<i>W. Wernsdorfer</i>	
4.1 Introduction	136
4.2 Overview of molecular nanomagnets	139

4.3	Giant spin model for nanomagnets	141
4.4	Quantum dynamics of a dimer of nanomagnets	152
4.5	Resonant photon absorption in Cr <sub>7</sub> Ni antiferromagnetic rings	155
4.6	Photon-assisted tunnelling in single-molecule magnet	160
4.7	Environmental decoherence effects in nanomagnets	161
4.8	Molecular spintronics using single-molecule magnets	166
4.9	Conclusion	173
	References	174
<b>5</b>	<b>Si/SiGe heterostructures in nanoelectronics</b>	<b>181</b>
	<i>D.J. Paul</i>	
5.1	Introduction	181
5.2	Growth of silicon-germanium alloys	181
5.3	Strain	183
5.4	Band structure	186
5.5	Mainstream nanoelectronic applications	189
5.6	Resonant tunnelling diodes	194
5.7	SiGe quantum cascade emitters	197
5.8	Conclusions	202
	References	202
<b>6</b>	<b>Quantum dots: Self-organized and self-limiting assembly</b>	<b>205</b>
	<i>Dimitri D. Vvedensky</i>	
6.1	Introduction	205
6.2	Methods of epitaxial growth	210
6.3	Self-organization in Stranski–Krastanov systems	212
6.4	Site control of quantum dots on patterned substrates	220
6.5	Nanophotonics with quantum dots	227
6.6	Arrays of quantum dots	234
6.7	Summary and outlook	236
	References	238
<b>7</b>	<b>Intersublevel quantum-dot infrared photodetectors</b>	<b>244</b>
	<i>E. Towe and D. Pal</i>	
7.1	Introduction	244
7.2	Infrared photon absorption	247
7.3	Some metrics for photon detectors	254
7.4	Experimenal single-pixel quantum-dot infrared photodetectors	260
7.5	Device characteristics	270
7.6	Toward quantum-dot focal plane array imagers	282
7.7	Challenges and prospects for high-performance detectors and arrays	288
	Acknowledgments	290
	References	290

<b>8 Nanoionics and its device applications</b>	<b>294</b>
<i>T. Hasegawa, K. Terabe, T. Sakamoto, and M. Aono</i>	
8.1 Introduction	294
8.2 Materials	295
8.3 Solid electrochemical reaction	296
8.4 Fundamentals of an atomic switch	299
8.5 New types of atomic switches	301
8.6 Applications of atomic switches	306
8.7 Summary and conclusion	309
References	310
<b>9 Molecular electronics based on self-assembled monolayers</b>	<b>312</b>
<i>D. Vuillaume</i>	
9.1 Introduction	312
9.2 Nanofabrication for molecular devices	313
9.3 Molecular tunnelling barrier	319
9.4 Molecular semiconducting wire	321
9.5 Molecular rectifying diode	323
9.6 Molecular switches and memories	326
9.7 Molecular transistor	330
9.8 Conclusion	331
Acknowledgments	332
References	332
<b>10 Self-assembly strategy of nanomanufacturing of hybrid devices</b>	<b>343</b>
<i>S. Hong, Y.-K. Kwon, J.S. Ha, N.-K. Lee, B. Kim, and M. Sung</i>	
10.1 Introduction	343
10.2 Direct patterning of nanostructures	343
10.3 Directed assembly of nanostructures	357
10.4 Characteristics of self-assembled hybrid nanodevices	369
10.5 Conclusion	380
Acknowledgments	380
References	380
<b>11 Templated carbon nanotubes and the use of their cavities for nanomaterial synthesis</b>	<b>386</b>
<i>T. Kyotani and H. Orikasa</i>	
11.1 Introduction	386
11.2 Synthesis of carbon nanotubes and carbon nano-test-tubes	387
11.3 Controlled filling of magnetic materials into carbon nano-test-tubes	392
11.4 Synthesis of water-dispersible and magnetically responsive carbon nano-test-tubes	397

11.5 Carbon nanotube cavities as a reaction field of hydrothermal synthesis	403
11.6 Conclusions	412
References	413
<b>12 Nanocatalysis</b>	<b>416</b>
<i>R.T. Vang, S. Wendt, and F. Besenbacher</i>	
12.1 Introduction	416
12.2 Surface characterization	419
12.3 Single-crystal surfaces	427
12.4 Changing the reactivity at the atomic scale: Design of new catalysts from first principles	433
12.5 Nanoparticles	437
12.6 TEM studies of nanoclusters on high surface area supports	458
12.7 Conclusions and outlook	464
References	465
<b>13 Bifunctional nanomaterials for the imaging and treatment of cancer</b>	<b>474</b>
<i>A. Burke, D. Carroll, F.M. Torti, and S.V. Torti</i>	
13.1 Introduction	474
13.2 Thermal ablative therapy in cancer	475
13.3 Nanomaterial applications	481
13.4 Gold nanoshells and nanorods	492
13.5 Iron-oxide nanoparticles	496
13.6 Conclusions and future directions	496
Acknowledgments	497
References	497
<b>14 Nanoparticles in medicine</b>	<b>503</b>
<i>D. Maysinger, P. Kujawa, and J. Lovrić</i>	
14.1 Introduction	503
14.2 Current problems with use of nanoparticles in medicine	513
14.3 Nanoparticle–cell interactions	519
14.4 Nanoparticles as imaging tools in animals and humans	525
14.5 Conclusions	530
Acknowledgments	530
References	531
<b>15 Nanostructured probes to enhance optical and vibrational spectroscopic imaging for biomedical applications</b>	<b>539</b>
<i>Anil K. Kodali and Rohit Bhargava</i>	
15.1 Introduction	539
15.2 Background	542
15.3 Theoretical modelling: NanoLAMPs	548
15.4 Design	564

15.5 Conclusion	566
References	567
<b>16 Protein-based nanodevices</b>	<b>570</b>
<i>P.P. Pompa and R. Rinaldi</i>	
16.1 Introduction	570
16.2 Protein fundamentals	572
16.3 Nanofabrication	574
16.4 Nanoelectronic devices based on proteins	580
16.5 Biophysical implications of protein-based nanobioelectronics	588
16.6 Nanodevices for biosensing	591
16.7 Conclusions	605
Acknowledgments	606
References	606
<b>17 Bioconjugated quantum dots for tumor molecular imaging and profiling</b>	<b>612</b>
<i>P. Zrazhevskiy and X. Gao</i>	
17.1 Introduction	612
17.2 Photophysical properties of quantum dots	615
17.3 Engineering of QD-based probes for biomedical applications	621
17.4 Tumor molecular imaging and profiling	629
17.5 Conclusions	636
Acknowledgments	636
References	637
<b>18 Modulation design of plasmonics for diagnostic and drug screening</b>	<b>641</b>
<i>C.-W. Lin, N.-F. Chiu, and C.-C. Chang</i>	
18.1 Introduction	641
18.2 Theoretical insights	642
18.3 Substrate effect (prism coupler, Ge-doped Si waveguide, grating, plasmonic)	649
18.4 Metallic effect (LRSPR, CMO adhesive layer)	657
18.5 Microfluidic parts	659
18.6 Biomolecular layer effect	661
18.7 Conclusions	669
Acknowledgments	670
References	670
<b>19 Carbon-nanotube field emission electron and X-ray technology for medical research and clinical applications</b>	<b>673</b>
<i>Sigen Wang, Otto Zhou, and Sha Chang</i>	
19.1 Introduction	673
19.2 Electron field emission from carbon nanotubes	674

19.3	Carbon-nanotube field emission electron and X-ray technologies in biomedical applications	677
19.4	Summary and conclusion	694
	References	696
<b>20</b>	<b>Theory of hydrogen storage in nanoscale materials</b>	<b>699</b>
	<i>Yufeng Zhao, Yong-Hyun Kim, S.B. Zhang, and Michael J. Heben</i>	
20.1	Introduction	699
20.2	Basic considerations	701
20.3	Hydrogen–material interaction	705
20.4	Internal interaction in HSMs	714
20.5	Structures of hydrogen sorbents	722
20.6	Required hydrogen-storage properties and design principles (DP)	725
20.7	Summary	731
	Acknowledgments	732
	References	732
<b>21</b>	<b>Electron cold sources: Nanotechnology contribution to field emitters</b>	<b>736</b>
	<i>Vu Thien Binh</i>	
21.1	Introduction	736
21.2	Driving forces for the evolution of cold cathodes	737
21.3	Single-atom emitters	739
21.4	Use of single-atom nanotip: The Fresnel projection microscope	742
21.5	Use of single-atom nanotip: The microgun	746
21.6	Material issues for field emitters: Carbon nanocompounds	748
21.7	Carbon-nanotube field emitters	749
21.8	Carbon-nanopearl field emitters	761
21.9	Applications and uses of carbon nanocompounds, CNTs and CNPs, as cold cathodes	765
21.10	Conclusions	769
	References	785
<b>22</b>	<b>Free-standing grid-like nanostructures assembled into 3D open architectures for photovoltaic devices</b>	<b>789</b>
	<i>X.Y. Kong, Y.C. Wang, X.F. Fan, G.F. Guo, and L.M. Tong</i>	
22.1	Introduction	789
22.2	Fabrication of photoelectrodes with 2D grid-like nanostructures by the biotemplating approach	791
22.3	Assembly and photophysics of grid-like nanostructures into 3D open architectures for the photocatalytic electrodes	794
22.4	Performance of DSSCs working with dye-sensitized TiO <sub>2</sub> stacked-grid array photoelectrodes	799

22.5 Characteristics and performance of DSSCs working with TiO <sub>2</sub> /NiO composite photoactive electrodes	802
22.6 Summary	805
Acknowledgments	806
References	806
<b>23 Nanolithography using molecular films and processing</b>	<b>808</b>
<i>C.L. McGuiness, R.K. Smith, M.E. Anderson, P.S. Weiss, and D.L. Allara</i>	
23.1 Introduction	808
23.2 Self- and directed patterning	814
23.3 Patterning via external tools	818
23.4 Directed self-masking via selective deposition on chemical patterns	831
23.5 Molecular rulers: A hybrid nanolithographic patterning method	842
23.6 Conclusion	848
References	848
<b>24 Laser applications in nanotechnology</b>	<b>860</b>
<i>M.H. Hong</i>	
24.1 Introduction	860
24.2 Pulsed laser ablation for nanomaterials synthesis	860
24.3 Laser as a heat source for device nanoprocessing	865
24.4 Laser surface nanopatterning with near-field and light-enhancement effects	868
24.5 Large-area parallel laser nanopatterning	879
24.6 Conclusions	884
References	885
<b>25 Evaluating the risks associated with nanomaterials</b>	<b>887</b>
<i>K. Thomas, N. Monteiro-Riviere, D. Warheit, and N. Savage</i>	
25.1 Introduction	887
25.2 Nanomaterials in consumer products	888
25.3 Characterization of nanomaterials	890
25.4 Hazard evaluation	891
25.5 Pulmonary exposure assessment	893
25.6 Dermal exposure assessment	894
25.7 Evaluating the risks associated with exposure to nanomaterials	896
25.8 Research priorities for the development of more refined estimates of nanomaterial risk	899
25.9 Conclusion	902
References	903
<b>Subject Index</b>	<b>905</b>

# List of Contributors

**Allara, D. L.** Departments of Chemistry and Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA. dla3@psu.edu

**Anderson, M.E.** Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA. mea151@psu.edu

**Ansermet, Jean-Philippe** Institut de Physique des Nanostructures (IPN), Ecole Polytechnique Federale de Lausanne, PHB-Ecublens station 3, 1015 Lausanne-EPFL, Switzerland. jean-philippe.ansermet@epfl.ch

**Aono, M.** International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. AONO.Masakazu@nims.go.jp

**Besenbacher, F.** Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark. fbe@inano.dk

**Bhargava, Rohit** Department of Bioengineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 N Mathews Ave, Urbana, IL 61801, USA. rxb@illinois.edu

**Binh, Vu Thien** Equipe Emission Electronique, Laboratoire de Physique de la Matiere Condensée et Nanostructures, Université de Lyon 1, UMR-CNRS 5586, Bat. L. Brillouin, 43 Boulevard du 11 November 1918, 69622, Villeurbanne, France. Vuthien.Binh@lpmcn.univ-lyon1.fr

**Burke, Andrew** Wake Forest University School of Medicine and Department of Cancer Biology, Wake Forest University, Winston-Salem, N.C. 27157, USA. aburke@wfubmc.edu

**Carroll, D.** Center for Nanotechnology and Molecular Materials and Department of Physics and Wake Forest University Comprehensive Cancer Center, Wake Forest University, Winston-Salem, N.C. 27109, USA. carroldl@wfu.edu

**Chang, C.-C.** Institute of Biomedical Engineering, National Taiwan University, No.1, Sec.4, Roosevelt Road, Taipei, 106, Taiwan, Republic of China. ccchang.ibme@gmail.com

**Chang, Sha** Department of Radiation Oncology, University of North Carolina, Chapel Hill, NC 27599, USA. sxchang@med.unc.edu

**Chiu, N.-F.** Institute of Biomedical Engineering, and Department of Electrical Engineering, National Taiwan University, No.1, Sec.4, Roosevelt Road, Taipei, 106, Taiwan, Republic of China. nfchiu@ntu.edu.tw

**Fan, X.F.** School of Materials Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, P.R. China.

**Gao, Xiaohu.** Department of Bioengineering, University of Washington, 1705 NE Pacific St., Foege N530M, Box 355061, Seattle, WA 98195, USA. xgao@u.washington.edu

**Guo, G.F.** School of Materials Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, P.R. China.

**Ha, J.S.** Department of Chemical Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul, 136-701, Korea. jeongsha@korea.ac.kr

**Hasegawa, T.** Atomic Electronics Group, International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. hasegawa.Tsuyoshi@nims.go.jp

**Heben, Michael, J.** Department of Physics and Astronomy, University of Toledo, Toledo, OH 43607, USA, and National Renewable Energy Laboratory, Golden, CO 80411, USA. michael.heben@utoledo.edu

**Hong, M.H.** Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, and Data Storage Institute, ASTAR, Singapore. elehmh@nus.edu.sg

**Hong, S.** Department of Physics, Seoul National University, Shillim-dong, Kwanak-gu, Seoul, 151–742, Korea. seunghun@snu.ac.kr

**Kim, B.** Department of Physics, Seoul National University, Shillim-dong, Kwanak-gu, Seoul, 151–742, Korea. etnblue@phya.snu.ac.kr

**Kim, J.** Department of Electrical and Computer Engineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. jihankim@uiuc.edu

**Kim, Yong-Hyun** National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401-3393, USA. yong\_hyun.kim@nrel.gov

**Kodali, Anil** Department of Mechanical Science and Engineering, and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 N Mathews Ave, Urbana, IL 61801, USA. kodali@illinois.edu

**Kong, X.Y.** School of Materials Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, P.R. China. xykong@sjtu.edu.cn

**Kujawa, P.** Labopharm, 480, bd. Armand-Frappier, Laval, Québec, H7V 4B4, Canada. pkujawa@labopharm.com

**Kwon, Y.-K.** Department of Physics, University of Massachusetts, 1 University Avenue, Lowell, MA 01854, USA and Department of Physics, Kyung Hee University, Hoegi-dong, Dongdaemun-gu, Seoul, 130-701, Korea. youngkyun.kwon@uml.edu

**Kyotani, T.** Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Katahira, Sendai 980-8577, Japan. kyotani@tagen.tohoku.ac.jp

**Leburton, J.-P.** Department of Electrical and Computer Engineering and Beckman Institute for Advanced Science and Technology, University of

- Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA. jleburto@illinois.edu
- Lee, N.-K.** Department of Physics, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul, 143-747, Korea. lee@sejong.ac.kr
- Lin, C.-W.** Institute of Biomedical Engineering, and Department of Electrical Engineering, and Institute of Applied Mechanics, and Center for Nanoscience and Technology, National Taiwan University, No.1, Sec.4, Roosevelt Road, Taipei, 106, Taiwan, Republic of China. cwlinx@ntu.edu.tw
- Lovric, Jasmina** Department of Pharmaceutics, Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovacica 1, 10000 Zagreb, Croatia. jlovric@pharma.hr
- Maysinger, Dusica** Department of Pharmacology and Therapeutics, McGill University, Montréal, Québec, Canada H3G 1Y6. dusica.maysinger@mcgill.ca
- McGuiness, C.L.** Plextronics Inc., 2180 William Pitt Way, Pittsburgh, PA15238, USA. cmcguiness@plextronics.com
- Melnikov, D.V.** Department of Electrical and Computer Engineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. dmm3@uiuc.edu
- Monteiro-Riviere, N.** Center for Chemical Toxicology Research and Pharmacokinetics, North Carolina State University, 4700 Hillsborough Street, Raleigh, NC 27606, USA. nancy\_monteiro@ncsu.edu
- Nakayama, T.** Department of Physics, Chiba University, 1-33 Yayoi, Inage-ku, Chiba 263-8522, Japan. nakayama@physics.s.chiba-u.ac.jp
- Orikasa, H.** Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Katahira, Sendai 980-8577, Japan.
- Pal, D.** Laboratory for Photonics, Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA15213, USA. debdas1@yahoo.com
- Paul, Douglas J.** Department of Electronics and Electrical Engineering, University of Glasgow, Oakfield Avenue, Glasgow. G12 8LT, UK. d.paul@elec.gla.ac.uk
- Pompa, P.P.** National Nanotechnology Laboratory, CNR-INFM, IIT Research Unit, ISUFI-Universita' degli Studi di Lecce, 73100 Lecce, Italy. piero.pompa@unile.it
- Rinaldi, R.** National Nanotechnology Laboratory, CNR-INFM, IIT Research Unit, ISUFI-Universita' degli Studi di Lecce, 73100 Lecce, Italy. Ross.rinaldi@unile.it
- Sakamoto, T.** Device Platforms Research Laboratories, NEC Corporation, Sagamihara 229-1198, Japan. t-sakamoto@dp.jp.nec.com
- Savage, N.** United States Environmental Protection Agency, 1200 Pennsylvania Avenue, N.W., Washington DC 20460, USA. savage.nora@epa.gov

**Shiraishi, K.** Graduate School of Pure and Applied Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8571, Japan. shiraishi@comas.frsc.tsukuba.ac.jp

**Smith, R.K.** Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA. rksmith@lbl.gov

**Sung, M.** Department of Physics, Seoul National University, Shillim-dong, Kwanak-gu, Seoul, 151–742, Korea. cameleon@snu.ac.kr

**Terabe, K.** Atomic Electronics Group, International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. terabe.kazuya@nims.go.jp

**Thomas, K.** Silicones Environmental, Health and Safety Council of North America, Herndon, VA 20171, USA. kthomas@sehsc.com

**Tong, L.M.** College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P.R. China.

**Torti, Frank** Wake Forest University School of Medicine and Department of Cancer Biology and Wake Forest University Comprehensive Cancer Center, Wake Forest University, Winston-Salem, N.C. 27157, USA. ftorti@wfubmc.edu

**Torti, S.V.** Wake Forest University School of Medicine and Department of Biochemistry and Wake Forest University Comprehensive Cancer Center, Wake Forest University, Winston-Salem, N.C. 27157, USA. storti@wfubmc.edu

**Towe, Elias** Laboratory for Photonics, Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. towe@cmu.edu

**Vang, R.T.** Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark. rtv@inano.dk

**Vuillaume, D.** Molecular Nanostructures and Devices Group, Institute for Electronics, Microelectronics and Nanotechnology, (IEMN)—CNRS, University of Lille, BP60069, Avenue Poincaré, 59652 Villeneuve d'Ascq Cedex, France. dominique.vuillaume@iemn.univ-lille1.fr

**Vvedensky, Dimitri** The Blackett Laboratory, Imperial College London, London SW7 2AZ, UK. d.vvedensky@imperial.ac.uk

**Wang, Sigen** Department of Radiation Oncology, and Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27599, USA. sgwang@email.unc.edu; swang88@yahoo.com

**Wang, Y.C.** College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P.R. China. ycw@pku.edu.cn

**Warheit, D.** DuPont Company, Wilmington, DE 19898, USA. david.b.warheit@usa.dupont.com

**Weiss, P.S.** Departments of Chemistry and Physics, Pennsylvania State University, University Park, PA 16802, USA. stm@psu.edu

**Wendt, S.** Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark.  
swendt@phys.au.dk

**Wernsdorfer, W.** Institut Néel, CNRS/UJF BP 166, 38042 Grenoble Cedex 9, France. wolfgang.wernsdorfer@grenoble.cnrs.fr

**Zhang, L.-X.** Department of Chemistry, Northwestern University, c/o Tamar Seideman, 2145 Sheridan Road, Evanston, IL 60208, USA. zhanglingxiao@hotmail.com

**Zhang, S.B.** Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, New York 12180-3590, USA. zhangs9@rpi.edu

**Zhao, Yufeng** National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401-3393, USA. Yufeng.Zhao@nrel.gov

**Zhou, Otto** Department of Physics and Astronomy, and Curriculum of Applied and Materials Sciences, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA. zhou@physics.unc.edu

**Zrazhevskiy, Pavel** Department of Bioengineering, University of Washington, 1705 NE Pacific St., Foege N530C, Box 355061, Seattle, WA 98195, USA. pavelz@u.washington.edu

# Role of computational sciences in Si nanotechnologies and devices

K. Shiraishi and T. Nakayama

## 1

1.1 Introduction	1
1.2 Present Si technology trend stimulated by scientific knowledge	2
1.3 Key knowledge for Si nanodevices obtained by computational science	3
1.4 Future Si technology trend predicted by computational science	39
1.5 Summary	43
Acknowledgments	43
References	43

## 1.1 Introduction

Due to the great success of the first-principles calculations, we can obtain and predict the physical properties of new materials by theoretical calculations. For example, surface reconstructions can well be reproduced at the atomistic level by the first-principles calculations (for example, Zhu *et al.* 1989). At present, first-principles calculations have become one of the most important tools for investigating various kinds of material surfaces.

This situation is crucial for modern Si nanotechnologies in which introduction of new materials are inevitable (Wilk *et al.* 2001). For example, Hf-based oxides have been used as alternative gate dielectrics of conventional SiO<sub>2</sub> and new channel materials such as Ge and GaAs are expected to be promising candidates of future Si nanodevices. Further, carbon nanotubes and Si nanowires are also expected as key building blocks of future LSIs. The merit of the first-principles calculations is that we can obtain realistic images of wavefunctions and detailed atomistic structures of new materials, which cannot be obtained through experiments alone. This knowledge, in addition to the experimental findings, enables us to construct useful concepts of modern nanotechnologies.

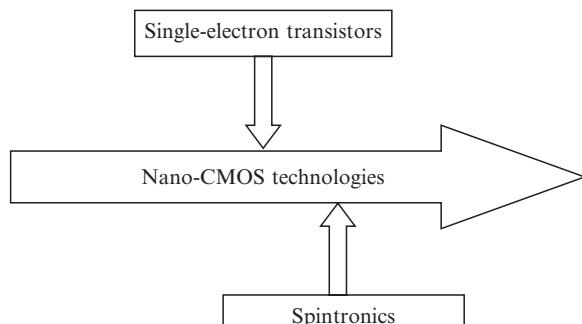
At the same time, recent Si nanotechnology requires the use of new materials to continue significant downscaling of device sizes, called Moore's law, which leads to higher computational speed, larger memory size and lower power consumption. The knowledge of new materials surely enriches our lives, since our modern society relies on the technologies developed in recent times. This means that the knowledge obtained by the computational science directly connects to our daily lives by way of Si nanotechnologies and devices.

It is also noted that Si nanotechnology gives us great knowledge, which leads to the new findings in nanoscience due to its precise control of sizes and shapes

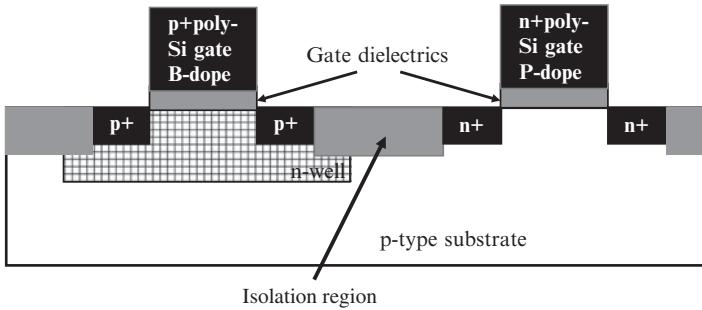
of various materials based on the highly developed fabrication techniques. In this chapter, we will show the new scientific findings that gives great insight into Si nanotechnologies.

## 1.2 Present Si technology trend stimulated by scientific knowledge

The present status of Si technology is facing a serious situation. Continuous and significant reduction of device size cannot be achieved by introducing new materials and new device concepts (Wilk *et al.* 2001; Robertson 2004). In Fig. 1.1, we show the schematic illustration of the present Si logic device trend. The main future technology is thought to be complimentary metal oxide semiconductor (CMOS) technology. As shown in this figure, however, CMOS technology will grow further stimulated by the advent of new physics and new device concepts like spintronics and single-electron transistors. For example, we should take into account the ballistic transport that has been extensively studied in mesoscopic physics, if the gate length of a Si metal oxide semiconductor field effect transistor (MOSFET) is shorter than 10 nm, which corresponds to the mean-free path of electrons in Si (Natori 1994; Datta *et al.* 1997). The second example is the importance of understanding the atomistic process of Si thermal oxidation. Recently,  $\text{SiO}_2$  thickness has become thinner than 2 nm, which contains 5–8 Si atomic layers, leading to the requirement of single atomic-layer control of Si oxidation (Wilk *et al.* 2001). To perform atomistic control, the atomistic understanding of Si oxidation is crucial. The third example is the necessity of gate electrode materials. Figure 1.2 gives the typical CMOS device structures (CMOS inverter). In CMOS devices, it is well known that we need two metals whose Schottky barrier heights are very different from each other. Conventionally, heavily B-doped poly-Si ( $p+$ -poly-Si) and heavily P-doped poly-Si ( $n+$ -poly-Si) are commonly used for high and low work function gate metals of CMOS because of its technological simplicities. For further downsizing, however, additional depletion layers formed in heavily doped poly-Si gates will become serious obstacles for the reduction of gate dielectric film thickness. Thus, metal gates alternative to conventional poly-Si gates should be developed to avoid the gate depletion (Wilk *et al.*



**Fig. 1.1** Schematic illustration of present technologies trend of Si nanotechnologies and devices.



**Fig. 1.2** Schematic illustration of a typical CMOS device structure. This example is a CMOS inverter structure.

2001; Robertson 2004). This situation, which corresponds to deep physical understanding of Schottky barrier heights, is extremely relevant to the future advancement of Si nanotechnologies and devices. As discussed above, clearly, new scientific knowledge with novel exploratory ideas would be essential for achieving continuous improvement and possible breakthroughs in Si nanotechnologies and devices.

In the following, we show several examples that computational science gives great insight into Si nanotechnologies.

## 1.3 Key knowledge for Si nanodevices obtained by computational science

### 1.3.1 Microscopic process of Si oxidation

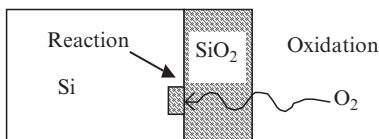
#### 1.3.1.1 Conventional understanding of si oxidation

Si nanodevices are fabricated by using many processes such as epitaxial growth, etching, ion implantation, etc. Among these processes, Si thermal oxidation is one of the most crucial processes in LSI technologies. This is because the interface state density at Si/SiO<sub>2</sub> interfaces obtained by the Si thermal-oxidation process is remarkably small, which amounts to the order of 10<sup>10</sup> cm<sup>-2</sup>. However, the detailed knowledge of Si thermal oxidation is still lacking, although Deal and Grove (1965) have reported pioneering works. Moreover, considering the fact that the SiO<sub>2</sub> thickness in Si nanodevices has become less than 2 nm, detailed understanding of Si thermal oxidation is really crucial for future Si nanodevices.

First, we introduce the conventional understanding of Si thermal oxidation by Deal and Grove (1965). The suggested steps consist of an oxidant diffusion process followed by an interfacial reaction process. When a Si substrate is put in an oxygen (O<sub>2</sub>) atmosphere at around 700–1100 °C, the substrate surface changes into Si oxide. This is the thermal-oxidation process. While this process can be described by a simple chemical reaction formula,



its microscopic picture is much more complicated. During the oxidation, O<sub>2</sub> molecules coming from the atmosphere must meet with the substrate Si to react. However, O<sub>2</sub> molecules must move through the Si oxide covering the



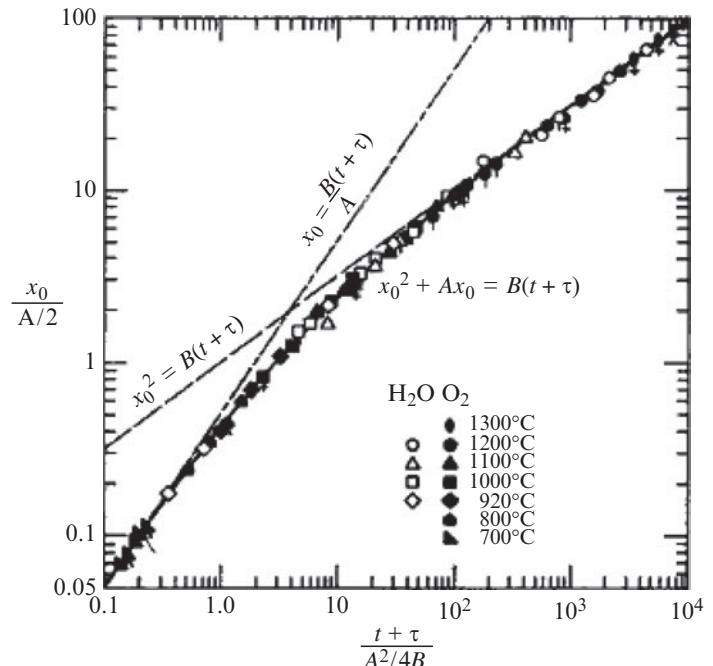
**Fig. 1.3** Schematic illustration of Si oxidation processes. Oxygen molecules should diffuse through an SiO<sub>2</sub> layer before reacting with Si at the Si/SiO<sub>2</sub> interface.

surface after the surface of the substrate is covered by Si oxide, as shown in Fig. 1.3. Deal and Grove suggested that the O<sub>2</sub> molecules move (the diffusion process) and arrive at the interface between the Si oxide and substrate Si, and react with the latter (the reaction process). By combining these two processes, they successfully explained the kinetics of oxide thickness growth. Experiments with O isotopes basically support this scheme (Hans *et al.* 1988; Lu *et al.* 1995). Isotope-labelled O atoms are located at Si/SiO<sub>2</sub> interfaces, which indicate that O<sub>2</sub> molecules diffuse through the SiO<sub>2</sub> layer and react at the Si/SiO<sub>2</sub> interface. Moreover, their model could reproduce the experimental oxidation rate, as shown in Fig. 1.4. The oxide thickness can be written as

$$X^2 + AX = B(t + \tau), \quad (1.2)$$

where,  $X$  and  $t$  are oxide thickness and oxidation time, respectively, and  $A$  and  $B$  are the constants related to oxidation reaction and O<sub>2</sub> diffusion through SiO<sub>2</sub>, respectively. The characteristic nature of their model is that there exist two regions called reaction-limited and diffusion-limited oxidation regions that correspond to thin and thick oxide thickness, respectively. In the case of reaction-limited oxidation, oxide thickness  $X$  is proportional to oxidation time ( $t$ ). On the other hand, in the diffusion-limited case,  $X$  is proportional to  $\sqrt{t}$ .

However, as Deal and Grove also pointed out in their original paper, this picture cannot explain several issues related to the thermal-oxidation process (Deal and Grove 1965; Deal 1988). In particular, it cannot explain the so-called initial enhanced oxidation, the rapid oxide growth when the oxide thickness is thinner than a few tens of nanometers. One important factor, which is not



**Fig. 1.4** Relation between oxide thickness ( $x$ ) and oxidation time ( $t$ ) (Deal and Grove 1965).

taken into account in the Deal–Grove scheme, is that the volume of newly formed oxide is more than 2 times larger than the volume of reacted substrate Si. Since the reaction occurs at the interface surrounded by the substrate and the surface oxide, the excess volume should cause a large compressive strain on the newly formed oxide. Such accumulated strain could be released by viscoelastic deformation of the Si oxide. Figure 1.5 schematically illustrates the Si density before and after thermal oxidation. As shown in this figure, Si density drastically decreases after Si thermal oxidation due to the volume expansion, indicating that Si species should move from the interface region to release accumulated strain. Therefore, it is natural to think that the transport of Si is also important during the thermal oxidation, though the Deal–Grove scheme considers only the transport of O. Thus, the movement of both O and Si should be clarified microscopically in detail to obtain a precise scheme of the thermal-oxidation process.

Another interesting phenomenon is layer-by-layer oxidation observed in Si thermal-oxidation processes. Watanabe *et al.* (1998) pointed out that Si oxidation proceeds in a layer-by-layer manner by the scanning reflection electron spectroscopy (SREM) observations (Watanabe *et al.* 1998). According to the fact that Si thermal oxidation is governed by the microscopic reaction where the O atom inserts into Si–Si bonds forming Si–O–Si bonds, this layer-by-layer oxidation cannot naturally be expected from the simple microscopic point of view. Nonetheless, these phenomena are crucial for the atomistic control of Si oxidation process involved in the fabrication process of Si nanodevices.

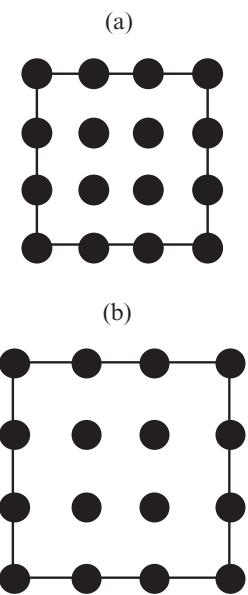
### 1.3.1.2 Atomistic si oxidation mechanism predicted by first-principles calculations

#### 1.3.1.2.1 Microscopic mechanism of Si oxidation

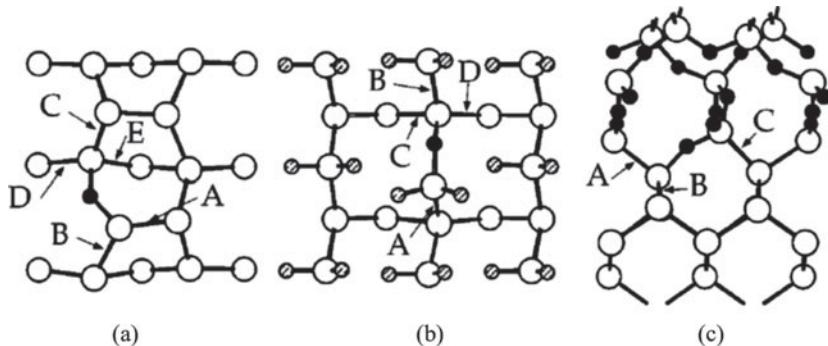
Although the Deal and Grove model can describe the important characteristics of thermal oxidation, it is agreed that it fails to explain the oxidation process especially in the thin-oxide case. Here, we introduce a more realistic scheme for the mechanism of the Si thermal-oxidation process based on the knowledge obtained by the computational sciences and related experiments (Kageshima and Shiraishi 1998; Kageshima *et al.* 1999; Uematsu *et al.* 2000, 2001, 2004; Fukatsu *et al.* 2003; Ming *et al.* 2006; Watanabe *et al.* 2006).

Kageshima and Shiraishi (1998) clearly show by first-principles calculations that the preferential growth direction of the oxide nucleus on the surfaces is vertical to the substrate, whereas at the interfaces it is lateral. Moreover, they have shown that Si atoms are inevitably emitted from the interface to release the stress induced during Si oxide growth (Kageshima and Shiraishi 1998).

First, the growth directions of an oxide nucleus on Si surfaces and at Si-oxide/Si interfaces are discussed. The surface growth direction is investigated by using the Si(100) surface with buckled dimers as the initial surface. For such a surface, the most stable adsorption site of the initial O atom is the back-bond of the lower dimer atom (Kato *et al.* 1998). The first O atom is placed at that site and another is added between the Si–Si bonds neighboring the first Si–O–Si bond. Second, the most stable adsorption site for the second O atom is determined by the first-principles results. The possible sites are



**Fig. 1.5** Schematic illustration of Si atom density before (a) and after (b) oxidation.



**Fig. 1.6** Atomic structures for studying the oxide nucleus growth. (a) Top view for the clean surface, (b) top view for the dihydride Si(100) surface model (Northrup 1991) as the initial surface, we also investigated the growth direction for a H-terminated surface. The most stable adsorption site for the initial O atom is the outermost Si–Si bond. The first O atom is placed at that site and another one between the Si–Si bonds neighboring the first Si–O–Si bond. Then, the most stable adsorption site of the second O atom can be determined. The calculated total energies are  $-0.12$ ,  $-0.22$ , and  $-0.59$  eV/unit cell, for the sites B, C, and D, respectively, relative to the total energy of the site A (Fig. 1.6(b)). These calculations indicate that the oxide nucleus on the (100) surface preferentially grows vertically into the substrate, being independent of the surface reconstruction.

Next, we discuss the oxide growth direction at Si/SiO<sub>2</sub> interfaces. For the investigation of the growth direction for interfaces, the quartz/Si(100) interface model is used as the initial interface (Fig. 1.6(c)) (Kagesima and Shiraishi 1998). While the real oxide layer formed by oxidation is amorphous, the oxide is modelled by a crystal of SiO<sub>2</sub> for simplicity. However, this crystal model certainly has a perfect bond network without large stress, which is an important feature of the real amorphous oxide interface. The first O atom is added to the interface and the stable structure is determined. Next, the second and third O atoms are introduced to the interface, assuming that all of the Si–O–Si bonds formed are connected. The calculations show that the structure, in which the second O atom is inserted into the site A, is energetically more stable (by 0.29 eV/unit cell) than the structure in which the second one is inserted into the site B. Moreover, the structure in which the second and third O atoms are inserted into the sites A and C is more stable (by 0.05 eV/unit cell) than the structure in which the second and third ones are inserted into the sites A and B. These results indicate that the oxide nucleus at the Si oxide/Si(100) interface preferentially grows laterally, parallel to the interface.

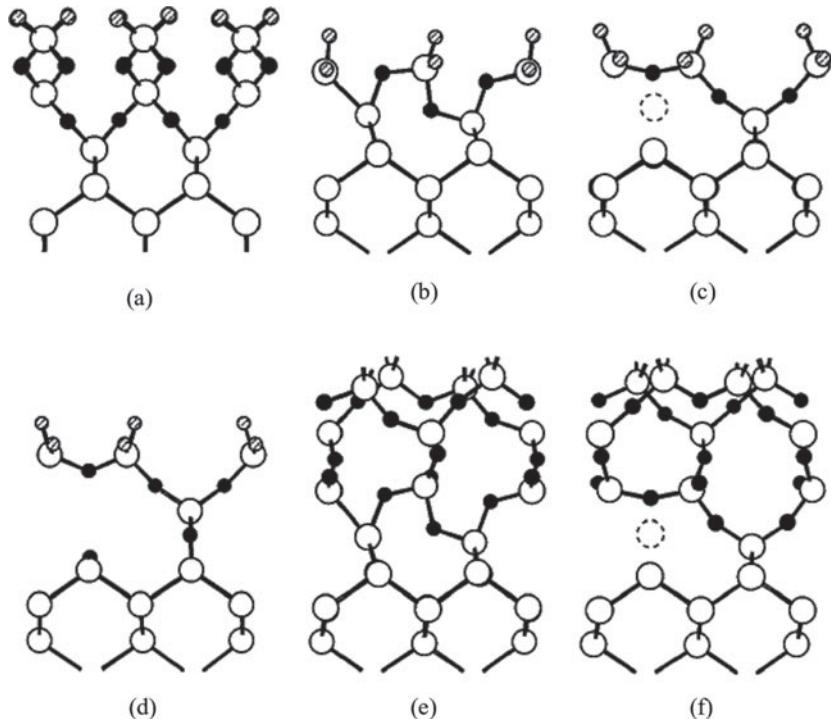
The preferential growth direction of the oxide nucleus for the (100) substrate is thus governed by the difference between “on the surface” and “at the interface.” Since the vertical oxide growth on the surfaces is independent of the surface reconstruction, the stress (rather than the bonding nature or the charge transfer) seems to control the growth direction. Actually, it is easy for

Si–O–Si bonds on the surfaces to expand vertically because the surface atoms in the Si region can move upwards with almost total freedom, while it is not easy for the bonds to expand laterally. Thus, the initial oxide nucleus on the surface should grow vertically in order to minimize the stress. In the case of the interface, the vertical expansion of Si–O–Si bonds is not easy because their movement is restricted by the covered oxide layer. Therefore, the energy gain due to the stress released by vertical growth is quite restricted. On the other hand, to minimize the interface energy, the initial oxide nucleus at the interface should grow laterally. These results have been confirmed by examining the stress distribution of the calculated atomic structures estimated from the shortening of the Si–Si bond lengths. These findings show the importance of the stress in determining growth direction. The simple phenomenological model that includes oxidation-induced compressive strain also reproduces the lateral oxide growth (Shiraishi *et al.* 2000). Recently, it has been pointed out that simple Monte-Carlo simulation that includes the diffusion of O species through SiO<sub>2</sub> region can reproduce layer-by-layer growth qualitatively (Watanabe *et al.* 2004).

The above considerations agree with the experimental results fairly well. A previous measurement, using scanning tunnelling microscopy (STM), of the oxide growth on a clean Si(111) surface (Ono *et al.* 1993) showed that oxide islands are formed in the initial stage at 600 °C. The depth of the islands reaches several atomic layers at the very initial stage. Furthermore, many experiments clearly show that the oxide grows atomically layer-by-layer at the Si oxide/Si(100) and (111) interfaces (Gibson *et al.* 1989; Ohishi *et al.* 1994; Komeda *et al.* 1998; Watanabe *et al.* 1998). These are consistent with the findings based on the computational results.

The results discussed above indicate that a uniform oxide layer can be obtained with any thickness by thermal oxidation once a uniform surface oxide layer is formed. Therefore, the preparation of the initial surface oxide is crucial for obtaining a uniform oxide layer with atomically controlled thickness, which gives an important guiding principle for modern Si nanotechnologies. Although first-principles results also indicate that the initial growth direction of the oxide nucleus on the surfaces is vertical into the substrate, this is true only when the thermodynamics govern the oxidation process. Actually, STM measurements have shown that oxidation does not form islands, but instead forms an atomically thin surface oxide layer from the very first stage at room temperature, where the oxidant cannot diffuse into the substrate easily (Ono *et al.* 1993). It has been reported that the O<sub>2</sub> adsorption in the second layer of the clean Si(100) surface has a non-zero barrier of about 0.3 eV, while the adsorption in the outermost layer is barrierless (Watanabe *et al.* 1998). Therefore, thermal oxidation at a lower oxidant pressure and lower temperature could result in the formation of a well-controlled atomically thin uniform oxide layer. The efficiency of these oxidation processes is supported by the experiments (Ohishi *et al.* 1994; Watanabe *et al.* 1998).

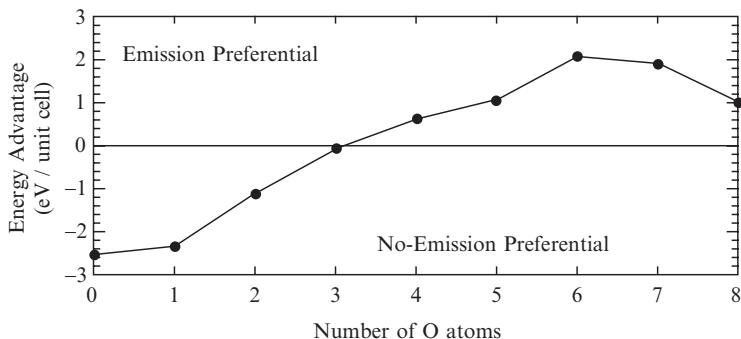
We now turn to the mechanism of how the accumulated stress during oxidation is released. Stress release was first investigated using a dihydrided Si(100) surface as the initial surface. We sequentially insert O atoms between Si–Si bonds of the surface, assuming atomical layer-by-layer oxide growth.



**Fig. 1.7** Side views of the atomic structures for studying the accumulation and the release of the stress. (a) The structure after sequential oxidation by two Si atomic layers; (b), (c) the structures before and after the Si-emission on the dihydride surface; (d) the structure with the Si-emission after oxidation by two Si atomic layers; (e), (f) the structures before and after the emission at the Si-oxide interface with the less-stressed quartz-like oxide. The broken circles indicate the position where the Si atom is emitted (Kageshima and Shiraishi 1998).

This assumption simplifies the analysis of the accumulated stress, as will be shown below. When eight O atoms per unit cell are introduced (Fig. 1.7(a)), the formed oxide has a Si–O–Si network similar to that of the cristobalite of crystal  $\text{SiO}_2$  (Wyckoff *et al.* 1963). However, the structure is highly compressed compared to that of the cristobalite. The *a*- and *b*-axes of the obtained oxidized region, which are parallel to the interface, are 23% shorter than the corresponding axes of the *a*-cristobalite. Despite the elastic theory, the *c*-axis, which is perpendicular to the interface, is only 20% longer than the corresponding axis of the *a*-cristobalite. Thus, the structure is largely compressed to about 3/4 of the volume of the *a*-cristobalite. This suggests the existence of a strain-release mechanism during the oxide growth. One possibility is the breaking, deformation, and rebonding of the formed Si–O–Si network, which would correspond to the viscous flow of the oxide. However, bond breaking and deformation after oxide formation require a lot of energy. Therefore, there must be some other mechanisms that work to release the stress before the compressed oxide is formed.

It is found that the atomic structure, when three O atoms per unit cell are introduced, is the key to the stress release (Fig. 1.7(b)) (Kageshima and Shiraishi 1998). In this structure, an O atom is quite close to a surface Si atom, which has only one Si–O bond. Thus, these two atoms can form a bond by breaking the bonds with the second-layer Si atom. Moreover, the second-layer Si atom, whose two bonds were broken, could be emitted from the surface because of laterally compressed stress on it (Fig. 1.7(c)). The calculated total energy of such a Si emitting structure indicates it to be only

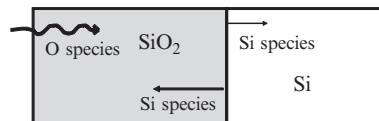


**Fig. 1.8** Energy advantage of the Si-emitting structures compared with the non-emitting structures as a function of the number of inserted O atoms per unit cell. The most stable structures for each case are compared assuming the atomical layer-by-layer oxide growth (Kagesima and Shiraishi 1998).

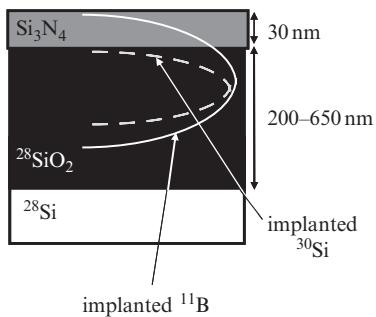
0.04 eV/unit cell higher than that of the non-emitting structure, though there remain two dangling bonds. This structure resembles the well-known *A* center (or the VO center) in bulk Si crystal (Pajot *et al.* 1994; Chadi 1996). In addition, when we sequentially insert O atoms, the total energies for all of the emitting structures are more stable than those for the corresponding non-emitting structures (Fig. 1.8). The energy advantage is up to 2 eV/unit cell. This is because the two remaining dangling bonds first form a weak bond by laterally compressed stress, and are finally terminated by forming a Si–O–Si bond. This also indicates that the Si-emission scarcely results in the creation of the interfacial-gap states. Moreover, when six O atoms per unit cell are introduced into the emitting structure (Fig. 1.7(d)), the resulting bond network resembles the quartz structure of SiO<sub>2</sub>. The *a*-axis of the obtained oxidized region is only 8% longer than the corresponding axis of the b-quartz. The *b*- and *c*-axes are only 1% and 0.2% shorter, respectively, than the corresponding axes of the b-quartz. Thus, the accumulated strain is successfully released by removing Si atoms during Si oxidation. In the real experimental situation, it is expected that the remaining stress in the formed oxide would be completely released after the Si-emission during oxide growth.

Silicon emission also occurs at the Si-oxide/Si interfaces. Si-emission from the interfaces has been considered using the quartz/Si(100) interface model (Kagesima and Shiraishi 1998). The total energy of the emitting structure (Fig. 1.7(f)) is more stable (by 0.41 eV/unit cell) than that of the non-emitting structure (Fig. 1.7 (e)), although two Si dangling bonds are formed after Si-emission. This means that, even at the oxide/Si interfaces, Si atoms are preferentially emitted during oxide growth. Moreover, although layer-by-layer oxidation is assumed as mentioned above, further calculations show that the Si-emission is independent of the oxide growth mode. Even after the initial vertical oxide growth on the surfaces, the emission can occur again when the oxide islands connect with each other. Since stress accumulation is inevitable in the Si-oxidation process, the release of this stress by Si-emission should be essential and universal.

As discussed above, the emitted Si atoms should play an important role in the oxidation process. Since the energy advantage of the Si-emission (up to 2 eV) is smaller than the formation energy of the Si interstitials (4.9 eV) (Car *et al.* 1984), which is thought to induce the oxidation-induced stacking faults (OSF) (Thomas 1963; Ravi *et al.* 1974; Hu 1975), oxidation-enhanced



**Fig. 1.9** Schematic illustration of Si oxidation processes. Oxygen molecules should diffuse through an SiO<sub>2</sub> layer before reacting with Si at the Si/SiO<sub>2</sub> interfaces.



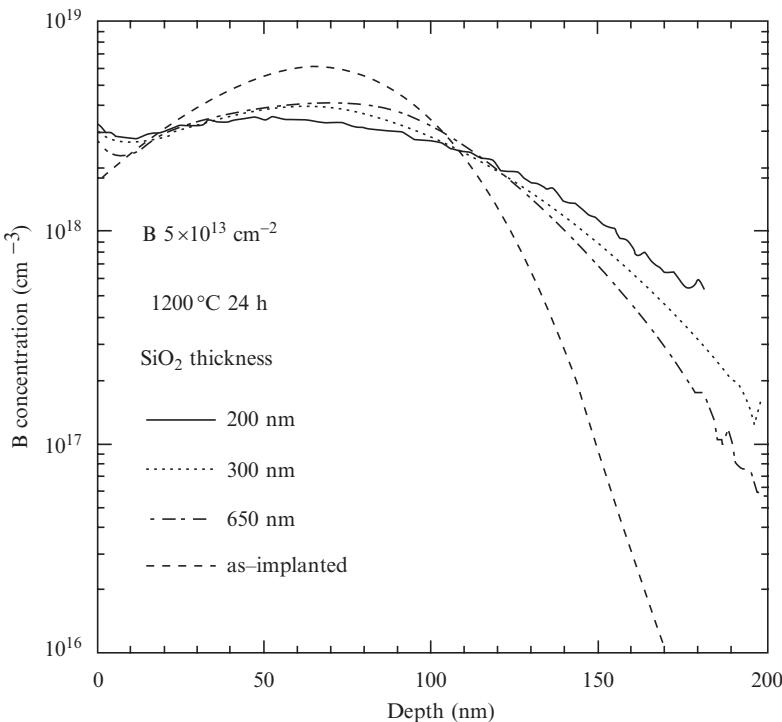
**Fig. 1.10** The sample structure employed for considering enhanced B diffusion near Si/SiO<sub>2</sub> interfaces (Uematsu *et al.* 2004).

diffusion (OED), and oxidation-reduced diffusion (ORD) (Mizuo *et al.* 1981; Tan and Gösele 1985). However, the energetically most stable way for emitted Si is that emitted Si backflow toward the SiO<sub>2</sub> region and react with O species as (Si + O<sub>2</sub> → SiO<sub>2</sub>). This is because the energy gain of this reaction amounts to 11.0 eV. Accordingly, the microscopic oxidation mechanism given by Deal and Grove (1965), which only considers the O diffusion, should be modified by taking into account the backflow diffusion of emitted Si species from the Si/SiO<sub>2</sub> interface. As a result, the oxidation process contains diffusion of both O and Si species, as illustrated in Fig. 1.9.

Next, we show the experimental findings of Si species backflow from a Si/SiO<sub>2</sub> interface to the SiO<sub>2</sub> region during oxidation. The first experimental example is that B and Si self-diffusion near the Si/SiO<sub>2</sub> interface is remarkably enhanced (Fukatsu *et al.* 2003; Uematsu *et al.* 2004). This is thought to be the effect of emitted Si species from the Si/SiO<sub>2</sub> interfaces. To investigate the B diffusion in SiO<sub>2</sub> by secondary ion mass spectroscopy (SIMS) analysis, the sample is prepared as follows. An isotopically enriched <sup>28</sup>Si epilayer was thermally oxidized in dry O<sub>2</sub> to form <sup>28</sup>SiO<sub>2</sub> of thicknesses 200, 300, and 650 nm. The samples were implanted with <sup>30</sup>Si at 50 keV to a dose of  $2 \times 10^{15} \text{ cm}^{-2}$  and capped with a ~30-nm thick silicon nitride layer. Subsequently, the samples were implanted with <sup>11</sup>B at 25 keV to a dose of  $3 \times 10^{15} \text{ cm}^{-2}$ . The final structure is shown in Fig. 1.10. The samples were pre-annealed at 1000 °C for 30 min to eliminate implantation damage, and were annealed in a resistively heated annealing furnace at various temperatures in the range of 1100–1250 °C. The diffusion profiles of <sup>11</sup>B and <sup>30</sup>Si were measured by SIMS.

Figure 1.11 shows the depth profiles of <sup>11</sup>B before and after annealing at 1250 °C for 6 h. As shown in Fig. 1.11, the profiles of <sup>11</sup>B become broader with decreasing thickness of the <sup>28</sup>SiO<sub>2</sub> layer, i.e. B diffusivity increases with decreasing distance from the Si/SiO<sub>2</sub> interface. If B diffusion is governed by a single process, the B diffusivity should have a distance dependence. However, it is physically unnatural. The distance dependence of diffusivity is also observed in the Si self-diffusion in SiO<sub>2</sub> (Fukatsu *et al.* 2003). In the case of Si self-diffusion, SiO molecules generated at the interface and diffusing into the oxide enhance Si self-diffusion (Fukatsu *et al.* 2003). These results indicate that SiO molecules also enhance B diffusion, because B diffusivity is higher near the interface where the SiO concentration is high. Moreover, first-principles calculations show that interstitial B–O complexes can diffuse through SiO<sub>2</sub> layers with a relatively low activation barrier (Otani *et al.* 2003). Considering that this interstitial B–O complex is stoichiometrically equivalent to the B<sub>Si</sub>–SiO complex, these first-principles results may indicate that the existence of SiO species also enhance B diffusion (Uematsu *et al.* 2006).

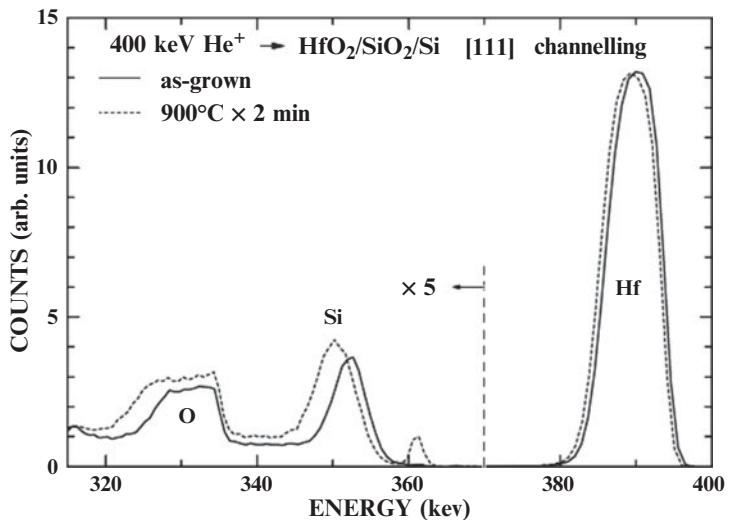
Taking into account the effect of SiO, coupled diffusion equations that include normal thermal B diffusion and SiO-assisted B diffusion can readily be constructed (Uematsu *et al.* 2004). As shown in Fig. 1.11, numerical simulations that includes the mechanism that SiO species generated from the Si/SiO<sub>2</sub> interface enhance B diffusion well reproduce the distance dependence of B diffusivities, although constant diffusivities are assumed.



**Fig. 1.11** Diffusion profiles of B in  $\text{SiO}_2$  with various thicknesses. Samples were implanted with B to a dose of  $5 \times 10^{13} \text{ cm}^{-2}$  and annealed at  $1200^\circ\text{C}$  for 24 h. The nearer the  $\text{Si}/\text{SiO}_2$  interface is, the broader the B profiles become (Uematsu *et al.* 2004).

Moreover, the time dependence of B diffusivities has also been reported. It is expected that the SiO concentration generated from the  $\text{Si}/\text{SiO}_2$  interface increases with longer annealing time. Considering the above discussions that SiO species enhance the B diffusion, B diffusivity is expected to be increased with longer annealing time. Actually, the clear enhancement in B diffusivity has been confirmed;  $1.5 \times 10^{-16} \text{ cm}^2/\text{s}$  and  $3.0 \times 10^{-16} \text{ cm}^2/\text{s}$  B diffusivities are obtained after 8 h and 24 h  $1200^\circ\text{C}$  anneals, respectively. This time dependence also supports the fact that B diffusion is assisted by SiO (Uematsu *et al.* 2006).

Next, we show much direct proof of Si species emission during Si oxidation (Ming *et al.* 2006). The experiments use the characteristic material properties of  $\text{SiO}_2$  in thin  $\text{HfO}_2$ . It is known that  $\text{SiO}_2$  and  $\text{HfO}_2$  reveal phase separation when  $\text{HfO}_2$  is thin enough. Thus, when the  $\text{HfO}_2/\text{SiO}_2/\text{Si}$  stacked sample is oxidized and Si substrate oxidation occurs, it is expected that the emitted Si species diffuse through  $\text{HfO}_2$  and segregate at the surface. Thus, surface-sensitive observation can detect the emitted Si species that segregate at the surface. Ming *et al.* (2006) performed high-resolution Rutherford backscattering (HRBS) measurements and confirmed the existence of surface Si component around 361 keV, in addition to the Si peak at the  $\text{Si}/\text{SiO}_2$  interface near 350 keV, only when interfacial  $\text{SiO}_2$  growth occurs, as clearly shown in Fig. 1.12. This experiment clearly confirmed that Si-emission during Si oxidation that was predicted by the first-principles calculations (Kageshima and Shiraishi 1998) is surely observed.

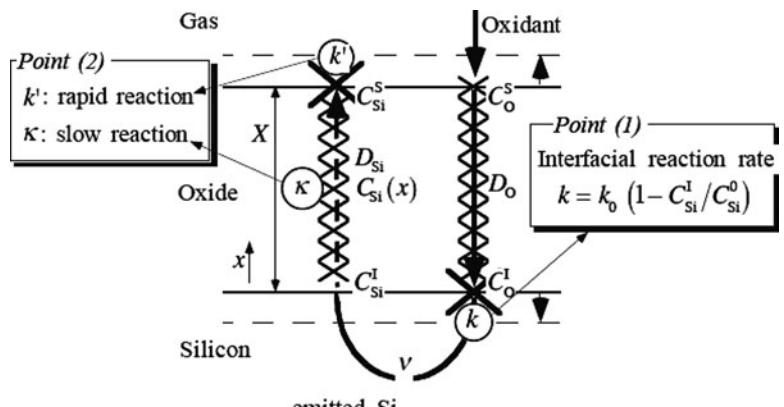


**Fig. 1.12** High-resolution Rutherford back-scattering spectra of as-grown and annealed  $\text{HfO}_2/\text{SiO}_2/\text{Si}$  (Ming *et al.* 2006).

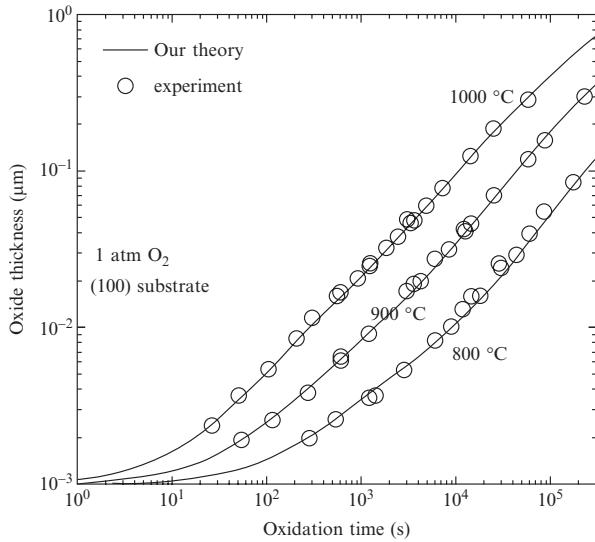
The knowledge of the above microscopic mechanism of Si oxidation is crucial for future Si nanotechnologies, since single-atomic-layer control of Si oxidation will be needed in the near future.

#### 1.3.1.2.2 Physical origin of initial enhanced oxidation

Here, we introduce the physical origin of initial enhanced oxidation that has been a mystery of the Si oxidation phenomenon for a long time. First, we introduce the Si-emission model (Kageshima *et al.* 1999). If the effect of such  $\text{SiO}$  interstitials as the so-called Si-emission model discussed previously is considered, the initial enhanced oxidation can systematically be reproduced (Kageshima *et al.* 1999; Uematsu *et al.* 2000). As per Deal and Grove, the oxide growth-rate equation can be derived from the reaction-diffusion equation, while newly considered  $\text{SiO}$  flow from the interface to the surface should be included besides the  $\text{O}_2$  flow from the surface to the interface (Fig. 1.13).



**Fig. 1.13** Schematic view of the reaction-diffusion kinetics in the Si-emission model (Kageshima *et al.* 2007).



**Fig. 1.14** Comparison of our theory on the oxide-growth kinetics with experimental data (Uematsu *et al.* 2000).

Due to the SiO flow, the interfacial reaction rate of the O<sub>2</sub> must be modified. The SiO should be much more easily oxidized on the surface than in the oxide because the oxidation of the SiO should be incorporated with the volume expansion. Then, in the thin-oxide limit, the combined reaction-diffusion equations are analytically solved as

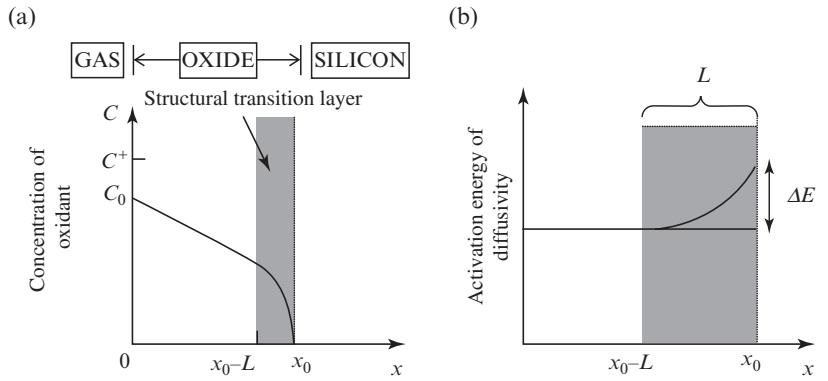
$$\frac{dX}{dt} = B/(A + 2X) + K \exp(-X/L), \quad (1.3)$$

where  $X$  is the oxide thickness and  $t$  is the oxidation time. This is quite similar to the empirical equation for the initial enhanced oxidation proposed by Massoud *et al.* (1985). The Si-emission model also indicates the physical meaning of the parameters,  $A$ ,  $B$ ,  $K$ , and  $L$ . For example,  $L$  is related to the diffusion length of SiO. By solving numerically the combined equations of the Si-emission model, the experimental results can be well reproduced, including the initial enhanced oxidation as well as the growth rate for thicker oxide as shown in Fig. 1.14.

Recently, it has been reported that the initial enhance oxidation can be reproduced by taking into account the diffusion barrier increase by the accumulated stress near Si/SiO<sub>2</sub> interfaces (Watanabe *et al.* 2006, 2007). They suppose a compressively strained oxide layer with a thickness of  $L$  localized in the proximity of the SiO<sub>2</sub>/Si interface, and that the oxidant diffusivity is suppressed in the strained layers, as shown in Fig. 1.15. As a result, the following model for the diffusivity  $D$  as a function of the depth  $x$  from the oxide surface is considered.

$$D(x) = \begin{cases} D_0 & 0 < x < x_0 - L \\ D_0 \exp\left[-\frac{\Delta E}{k_B T} \left(\frac{x - x_0 + L}{L}\right)^2\right] & x_0 - L < x < x_0, \end{cases} \quad (1.4)$$

where  $x_0$  is the total thickness of the oxide film,  $\Delta E$  is the incremental barrier of the diffusivity at the SiO<sub>2</sub>/Si interface, and  $D_0$  is the diffusivity



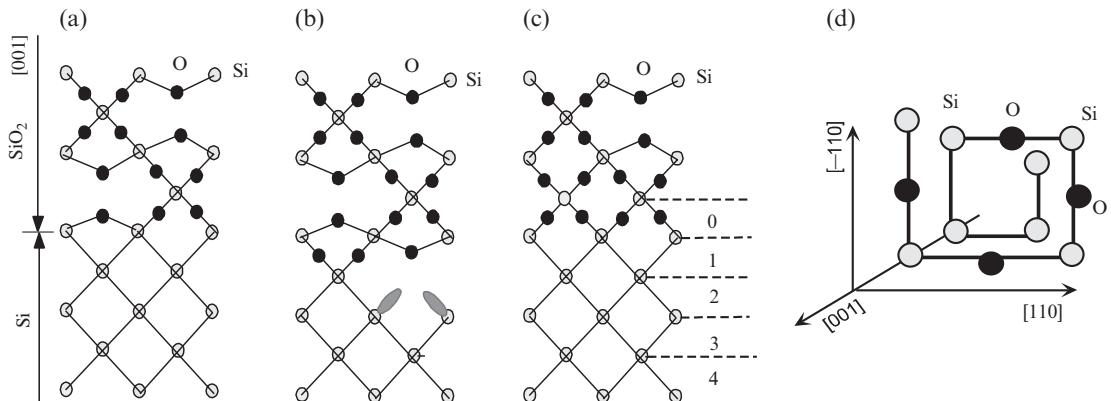
**Fig. 1.15** Depth profiles of oxidant concentration (a) and activation energy of diffusivity (b) in our oxidation model (Watanabe *et al.* 2007).

in the oxide film except for the interfacial strained region. This formula means that the activation energy of the oxidant diffusivity is raised monotonically in the strained region on approaching the  $\text{SiO}_2/\text{Si}$  interface. By considering the diffusion-barrier modification, the initial enhanced oxidation can also be reproduced (Watanabe *et al.* 2006, 2007).

#### 1.3.1.2.3 Characteristic reflectance difference spectroscopy (RDS) during Si thermal oxidation

In the downsizing trends of Si related nanoscale devices, the *in-situ* monitoring and thickness control of Si oxidation received an increasing demand as one of the key process technologies. As explained above, Watanabe *et al.* (1998) clearly demonstrated by scanning reflection electron microscopy (SREM) that the thermal oxidation of the Si(001) surface proceeds in a layer-by-layer manner. Since the SREM uses electrons as a probe, however, the observation is limited to the oxidation of ultrathin layers under ultrahigh-vacuum conditions. On the other hand, reflectance difference spectroscopy (RDS) is a powerful *in-situ* optical measurement to observe the electronic structures of semiconductor surfaces/interfaces and their time evolution (Murayama *et al.* 1998; Nakayama and Murayama 1999). Since the electromagnetic light penetrates deep into Si substrate, i.e. of the order of 1000 Å, the RDS can in principle detect the layer-by-layer thermal oxidation. Nakayama and Murayama (2000) proposed the use of RDS to detect the layer-by-layer thermal oxidation, by theoretically investigating the variation of RDS spectra of buried  $\text{SiO}_2/\text{Si}$  interfaces.

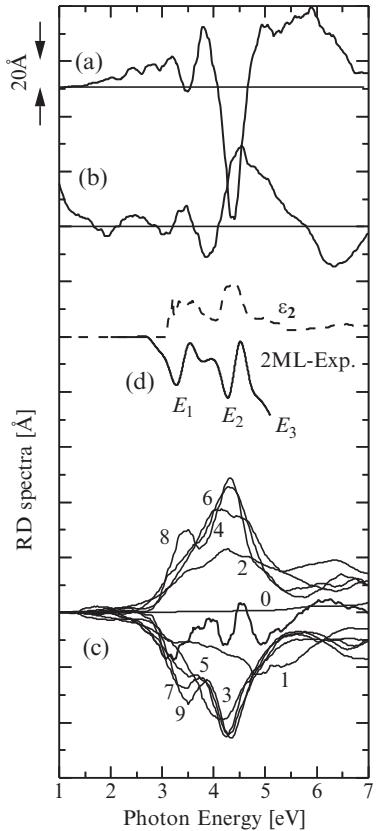
First, we explain why the RDS can detect the layer-by-layer oxidation. Figure 1.16(d) shows the schematic picture of  $\text{SiO}_2/\text{Si}$  interfaces viewed from the (001) direction. Roughly speaking, the oxidation corresponds to the insertion of the oxygen atoms into Si–Si bonds. Whenever the monolayer oxidation is completed in the layer-by-layer process, as seen in this figure, the edge Si–Si bonds terminated at  $\text{SiO}_2/\text{Si}$  interface alternately change the direction between  $[-110]$  and  $[110]$ . Therefore, since the RDS measures the reflectance difference between two perpendicular directions, i.e. the anisotropy of interface polarization originating from interface electronic structure, one can optically detect the change of the interface-bond direction and determine the advance of monolayer oxidation.



**Fig. 1.16** Schematic pictures of SiO<sub>2</sub>/Si (001) interfaces. (a) Flat interface A with the desorption of interface Si atoms, (b) flat interface B after one monolayer oxidation of A interface, and (c) flat interface with crystalline SiO<sub>2</sub> layer around the interface. (d) Bird's eye view of layer-by-layer oxidation of Si(001) surface displayed from the surface along [001].

Figures 1.17(a) and (b) show the calculated RDS spectra of two SiO<sub>2</sub>/Si interfaces, A and B, which are displayed in Figs. 1.16(a) and (b), respectively. It is noted that the interface B corresponds to the interface when the monolayer oxidation is completed starting from the interface A, and vice versa. Namely, these two interfaces alternately appear during the layer-by-layer oxidation. The detailed atomic positions at these interfaces are taken from the first-principles calculations (Kageshima and Shiraishi 1998). Here, let us concentrate on two features in Fig. 1.17: (i) It is noted that the spectra of the interfaces, A and B, have similar shapes and opposite signs. This result indicates that when the A and B interfaces appear alternately during the oxidation, the RDS signals oscillates between positive and negative values, which is similar to the reflectance high energy-electron diffraction (RHEED) oscillation in the epitaxial layer-by-layer growth. In particular, such an observation is apparently effective around 3.5/4.5 eV where the RDS signal is large. In this case, the period of oscillations corresponds to the bilayer oxidation and one can determine the number of oxidized layers by counting the oscillation. (ii) The other feature observed in Fig. 1.17 is the spectral shape of the large peaks around 3.5 and 4.5 eV, which, respectively, correspond to the  $E_1$  and  $E_2$  van Hove singularity energies of bulk Si shown in Fig. 1.17(d). These peaks have the energy-derivative-of- $\varepsilon_2$  shapes, which indicates, based on the general theory of spectral shapes, that the anisotropic RDS signals appear due to the modulation of bulk electronic structures around the interface (Nakayama and Murayama 1999). In fact, such a modulation can be examined by analyzing the layer contributions to the RDS, which are also shown in Fig. 1.17(c).

This theoretical prediction was confirmed by the RDS experiments by Yasuda *et al.* (2001). Figure 1.18 shows the spectral oscillation measured at around 3.5 eV, while the observed spectra after two-monolayer oxidation are shown in Fig. 1.17(d). The spectral oscillation is clearly seen as a function of the oxidization time, being in good agreement with the theoretical prediction. We note that the oscillation amplitude decreases as the oxidation proceeds. This is because both the interfaces, A and B, grow to coexist within the spot



**Fig. 1.17** Calculated RDS spectra of  $\text{SiO}_2/\text{Si}$  (001) interfaces as a function of photon energy, reported by Nakayama *et al.* (2000). (a) A interface shown in Fig. 1.16(a), (b) B interface shown in Fig. 1.16(b), and (c) the crystalline interface shown in Fig. 1.16(c). Observed RDS spectra after two monolayer oxidation are shown in (d), together with the imaginary parts of dielectric function of bulk Si (dashed line). In (c), the bold line corresponds to the RDS spectra, while the thin lines with the numbers  $n$ , respectively, correspond to the RDS spectra of the  $n$ th-layer bonds shown in Fig. 1.16(c).

size of the reflectance light, indicating some disordering of the layer-by-layer process. By analyzing the temperature and pressure dependences of the oxidation time, we can determine the microscopic mechanism of oxidation such as the oxidation speed and the activation energy of oxidation (Matsudo *et al.* 2002; Yasuda *et al.* 2003). With respect to the spectral shapes, the experiments showed the energy-derivative-of- $\epsilon_2$  shapes around 3.5/4.5 eV, being also in good agreement with the theoretical prediction in Fig. 1.17(a). However, it should be noted that there are some differences in the magnitude. Fig. 1.17(c) shows the calculated spectra of the interface depicted in Fig. 1.16(c), which has crystalline  $\text{SiO}_2$  structure around the interface. The existence of this kind of interface was suggested by Ikarashi *et al.* (2000) and Tu *et al.* (2002). The agreement of spectral shapes and magnitude is fairly good, indicating the appropriateness of  $\text{SiO}_2$  crystalline structures around the oxidation front interfaces.

In this way, the RDS measurement can observe not only the dynamical change of the interfaces but also the detailed electronic/atomic structures of interfaces. Moreover, from these studies, the *in-situ* control method of the Si oxidation is obtained. These studies have demonstrated a new direction in Si nanotechnology that the microscopic theoretical investigation is indispensable to the development of new technologies.

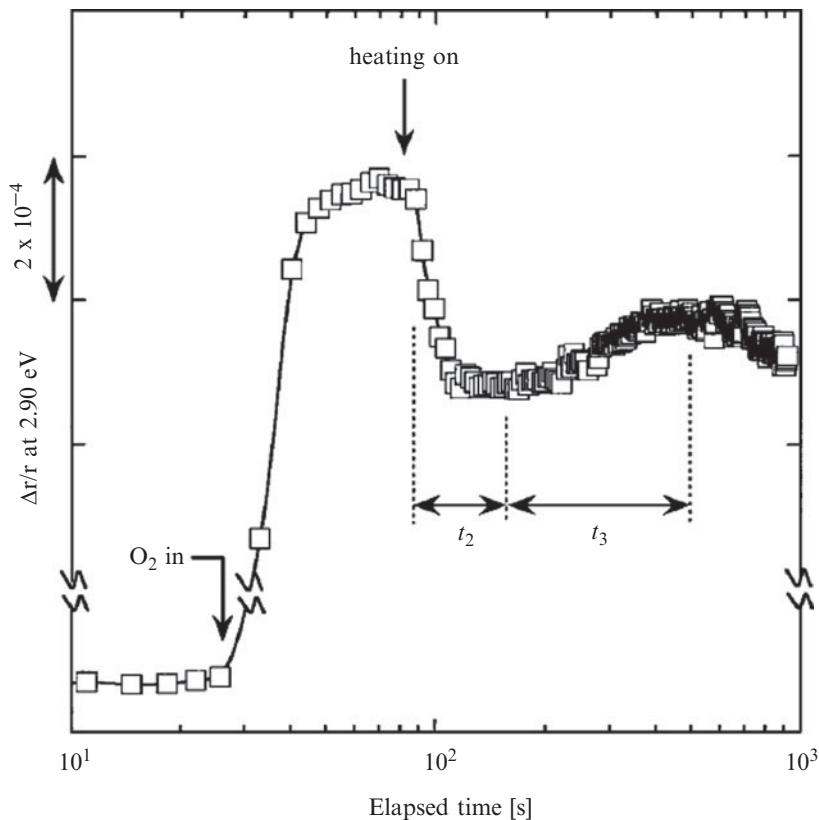
### 1.3.2 Curious properties of high- $k$ gate dielectrics

#### 1.3.2.1 Defects properties obtained by first-principles calculations

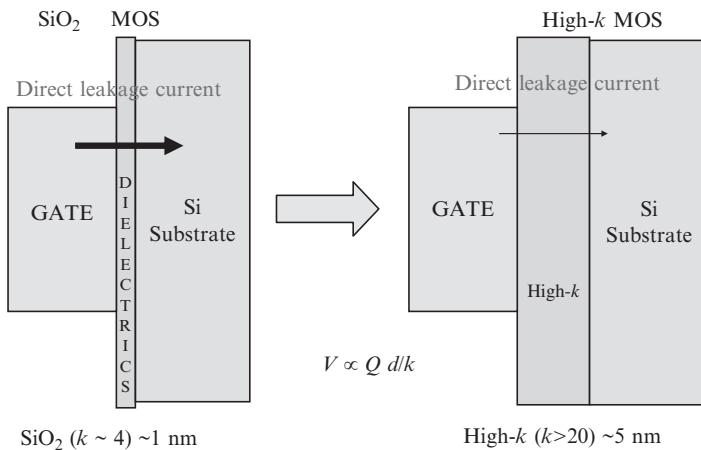
The considerable and ongoing downscaling of Si devices has brought about the requirement for a continuing need for significant reduction in gate dielectric film thickness, so much so that even current commercial-based large-scale integrated (LSI) circuits have  $\text{SiO}_2$  dielectrics of  $\sim 2$  nm thickness. As a result, increasing gate leakage current has become a serious issue for future LSI devices. Efforts to reduce the leakage current have given rise to the application of alternative high- $k$  gate dielectrics, which allows the use of physically thicker dielectric films, while maintaining the capacitance value  $C$ . This is because  $C$  is proportional to  $k/d$ , where  $k$  and  $d$  are the dielectric constant and physical thickness of a dielectric film. A schematic illustration of the necessity of high- $k$  gate dielectrics is shown in Fig. 1.19. Among these high- $k$  dielectrics,  $\text{HfO}_2$  and its related alloys ( $\text{HfSiO}_x$ , etc.), have been selected as alternative dielectrics in place of conventional  $\text{SiO}_2$ , due to their promising properties, as follows (Wilk *et al.* 1999, 2001):

- (1) sufficiently large band-offset value for both electrons and holes;
- (2) thermal stabilities when they are in contact with Si;
- (3) they have relatively stable amorphous phases;
- (4) they have large dielectric constant.

Although Hf-based dielectrics have been studied intensively, their physical properties are much different from the present conventional  $\text{SiO}_2$  dielectric. One such characteristic is their defect property.

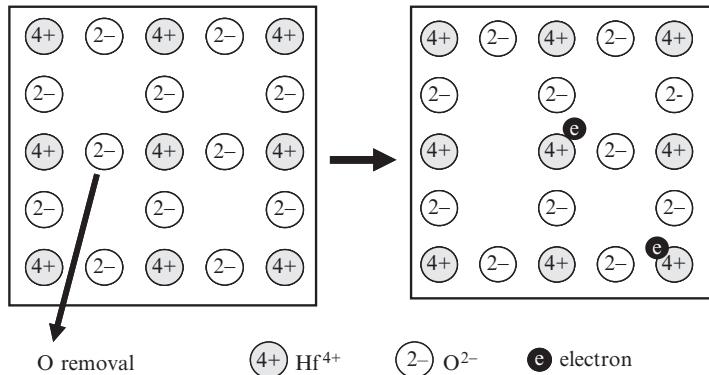


**Fig. 1.18** Observed RDS oscillation during the layer-by-layer oxidation of Si(001) surfaces, as a function of oxidation time, reported by Yasuda *et al.* (2001). The peaks around 80 and 500 s, respectively, correspond to the completion of one- and three-monolayer oxidation, while the dip around 150 s corresponds to that of two-monolayer oxidation.



**Fig. 1.19** Schematic illustration of the necessity of high- $k$  gate dielectrics. Direct tunnelling current can be greatly reduced by using high- $k$  gate dielectrics instead of conventional  $\text{SiO}_2$ .

It is well known that  $\text{HfO}_2$  contains a much higher concentration of oxygen vacancies ( $V_\text{o}$ ) than  $\text{SiO}_2$ , although Hf atoms can bind to O atoms much more strongly than do Si atoms; The heat of formation of  $\text{HfO}_2$  (11.6 eV) is much larger than  $\text{SiO}_2$  (9.4 eV). This is because  $\text{HfO}_2$  is a typical ionic crystal,



**Fig. 1.20** Schematic illustration of  $V_o$  formation in ionic  $HfO_2$  (Shiraishi *et al.* 2006).

whereas  $SiO_2$  has covalent-bond characteristics. In this section, we discuss the characteristic features of O vacancies in detail.

$HfO_2$  is an ionic crystal in which Hf and O atoms take the form of  $Hf^{4+}$  and  $O^{2-}$  ions as shown in Fig. 1.20. In ionic crystals, anion vacancies are revealed by a characteristic behavior. If a  $V_o$  is formed in  $HfO_2$  by removing an O atom from the crystal, two surplus electrons are generated, as shown in Fig. 1.20. First, we investigate the  $V_o$  behaviors when the O atoms are in an equilibrium condition with the gas phase. In other words, we suppose that the formation and elimination of  $V_o$  are balanced by the capture and release of O atoms between the crystal and the gas phase. In this situation, the behavior of the electrons in  $HfO_2$  is quite important in determining the  $V_o$  behavior.  $V_o$  formation in  $HfO_2$  generates two electrons that occupy the empty  $HfO_2$  conduction bands, and the generated electrons do not occupy a  $V_o$  level that is located energetically lower than the bottom of the  $HfO_2$  conduction band (Yamada 1986; Shiraishi *et al.* 2006). This originates from the increase in electron entropy due to the occupation of the empty  $HfO_2$  conduction bands.

The above interesting phenomena can be easily explained by comparing two reactions in an  $O_2$  atmosphere; one reaction is the formation of  $V_o^0$  in which two electrons are trapped, and the other is the formation of an empty  $V_o^{2+}$  and two conduction electrons;



$\Delta E_1$  and  $\Delta E_2$  are the energies needed to form the oxygen vacancies in reaction eqns (1.5) and (1.6), respectively. Since the  $V_o^0$  energy-level position  $E(V_o^0)$  is located in the gap region,  $\Delta E_2$  is usually larger than  $\Delta E_1$  by approximately  $2(E_c - E(V_o^0))$ , where  $E_c$  is the energy level at the bottom of the  $HfO_2$  conduction band.

According to the mass action law, the  $V_o^0$  concentration governed by reaction eqn (1.5) can be described as

$$N(V_o^0) \propto \exp(-\Delta E_1/kT). \quad (1.7)$$

On the other hand, the relationship between  $V_o$  and the electron concentration described by reaction eqn (1.6) can be expressed as

$$N(V_o^0) \times N(e)^2 \propto \exp(-\Delta E_2/kT). \quad (1.8)$$

By taking into account the relationships  $N(e) = 2N(V_o)$  and  $\Delta E_2 = \Delta E_1 + 2(E_c - E(V_o^0))$  eqn (1.8) becomes

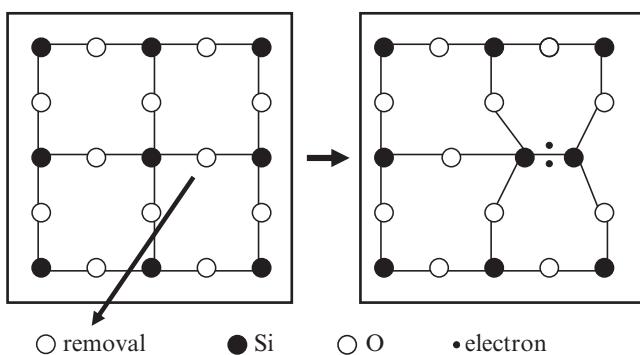
$$N(V_o^0) \propto \exp[-\{\Delta E_1 + 2(E_c - E(V_o^0))\}/3kT]. \quad (1.9)$$

As a result, the effective energies of formation of  $V_o$  in reactions (1.5) and (1.6) are  $\Delta E_1$  and  $\{\Delta E_1 + 2(E_c - E(V_o^0))\}/3$ , respectively. Therefore, if  $E_c - E(V_o^0) < \Delta E_1$ , reaction (1.6) becomes dominant. On the other hand, if  $E_c - E(V_o^0) > \Delta E_1$ , reaction (1.5) is the major reaction.

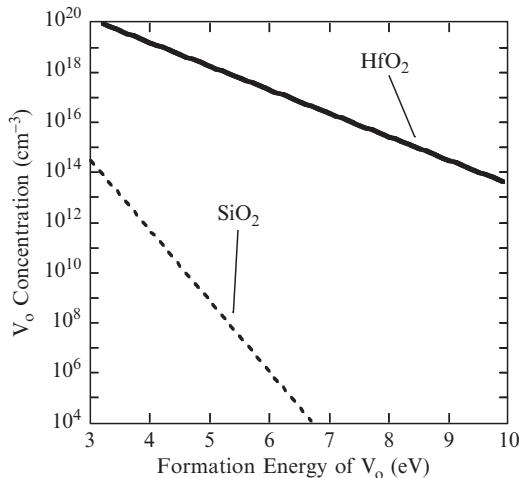
In  $\text{HfO}_2$ , the calculated  $V_o^0$  formation energy is about 6.4 eV (Scopel *et al.* 2004), and the experimentally observed energy level of  $V_o$  is about 1.2 eV below the  $\text{HfO}_2$  conduction band edge (Takeuchi *et al.* 2004). Thus, reaction (1.6) is the dominant reaction in the formation of  $V_o$  in  $\text{HfO}_2$ , and  $V_o^{2+}$  and two conduction electron generation occurs. The estimated effective energy required to form  $V_o$  in an  $\text{O}_2$  atmosphere is  $\sim 2.9$  eV.

The situation is quite different in covalent  $\text{SiO}_2$ . In  $\text{SiO}_2$ , the calculated  $V_o^0$  formation energy is about 5.2 eV (Scopel *et al.* 2004), and the  $V_o$  energy level is about 7 eV below the bottom of the  $\text{SiO}_2$  conduction band (Oshiyama 1998). This relatively lower position of  $V_o$  in  $\text{SiO}_2$  originates from the fact that formation of  $V_o$  induces a large lattice relaxation that enables generation of a new Si–Si bond, as illustrated in Fig. 1.21 (Oshiyama 1998). In  $\text{SiO}_2$ , as a result, reaction (1.5) is the major reaction and the estimated effective energy for the formation of  $V_o$  is  $\sim 5.2$  eV.

As discussed above, the effective formation energy of  $V_o$  is much lower in  $\text{HfO}_2$  than in  $\text{SiO}_2$ . Thus, the  $V_o$  concentration in  $\text{HfO}_2$  is much higher than in  $\text{SiO}_2$ , although the formation energy of  $V_o$  in  $\text{HfO}_2$  is much higher than that in  $\text{SiO}_2$ , as shown in Fig. 1.22. It has been reported recently that this higher concentration of  $V_o$  leads to the formation of electron leakage paths via the  $V_o$  levels (Torii *et al.* 2004). The higher concentration of  $V_o$  originates from the ionic nature of  $\text{HfO}_2$ . From a microscopic view point, the relatively higher energy level of  $V_o$  in  $\text{HfO}_2$  lowers the effective formation energy of  $V_o$ . This



**Fig. 1.21** Schematic illustration of  $V_o$  formation in covalent  $\text{SiO}_2$  (Shiraishi *et al.* 2006).



**Fig. 1.22** Oxygen vacancy ( $V_o$ ) concentration plotted as a function of the formation energy of  $V_o$ .

is why  $\text{HfO}_2$  has a much larger number of O vacancies than  $\text{SiO}_2$ , although Hf can bind O much more strongly than Si.

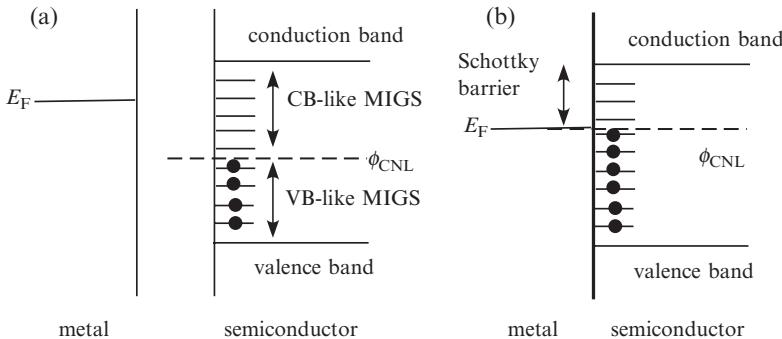
As for the precise position of the energy level of O vacancies, there are many first-principles calculations (Foster *et al.* 2002; Peacock *et al.* 2004; Umezawa *et al.* 2005; Hakala *et al.* 2006; Gavartin *et al.* 2007; Broqvist *et al.* 2008). The tendency that the position of O vacancy level is located at a relatively higher position of the  $\text{HfO}_2$  bandgap is common in these works, although the precise position is a little different. Recently, the amorphous effect on the position of the O vacancy level has been intensively studied (Broqvist *et al.* 2008). Their calculations show that the calculated  $V_o$  level position agrees well with the reported experiments by Takeuchi *et al.* (2004).

### 1.3.2.2 Schottky barrier heights and CMOS application

Schottky barrier heights at metal/insulator (semiconductor) interfaces are crucial physical properties for Si nanodevices. Among them the control of Schottky barrier heights are very important for constructing CMOS devices as mentioned before. For example, appropriate tuning of SBH of gate metals is indispensable to construct n and p-MOS, which are the building block of CMOS. In this circumstance, various gate stack structures containing gate metals on high- $k$  dielectrics have been intensively studied. In these studies, however, we encountered an unexpected behavior of SBH that can not be explained by the conventional SBH theories such as those based on the charge-neutrality-level concept. In this section, we first review the conventional theory of SBH, then explain the observed unusual behavior of SBH at high- $k$  interfaces and its origins, and introduce a new concept of SBH constructed from these studies.

#### 1.3.2.2.1 Conventional Schottky barrier-height theory

First, we consider how the bands (energy levels) of two materials align at their interface (Nakayama *et al.* 2006). In the case of metal/metal interface, the statistical mechanics teaches us that the movable carriers such as free



**Fig. 1.23** Schematic band alignment at metal/semiconductor interface (a) before and (b) after the connection. Reflecting this alignment, the Fermi energy of metal tends to move toward the charge neutrality level,  $\phi_{CNL}$ , by transferring electrons around the interface, which movement is also shown in Fig. 1.24.

electrons move across the interface to produce a dipole potential at the interface and equalize the Fermi energies of both metals with each other. Since there are no electric fields in metals in equilibrium, transferred carriers are often localized within a few atomic layers around the interface. In the case of a metal/semiconductor interface, if one can define Fermi energies for semiconductors, the same scenario as that of metal/metal interfaces applies to metal/semiconductor interfaces.

At the metal/semiconductor interface, in general, the semiconductor possesses electronic eigenstates that have eigenenergies within the bandgap of the bulk semiconductor. These states appear because the translational symmetry of semiconductor bulk crystals is broken at the interface perpendicular to the interface direction and thus the origin producing the bandgap in the energy spectrum and forbidding the existence of eigenstates within the bandgap disappears. These states are made of eigenstates in complex band structures of bulk materials and have complex wave numbers, thus being localized around the interface and are called evanescent-wave states. Such states are schematically shown in the energy diagram of Fig. 1.23(a). Louie *et al.* (1976) demonstrated by *ab-initio* calculations of electronic structures of Al/Si interface that these states really exist in semiconductors around metal/semiconductor interfaces. Since these states smoothly connect with wavefunctions in metals, they are sometimes called metal-induced gap states (MIGS) (Heine 1965).

Some of the MIGS are occupied from the bottom, corresponding to the number of electrons around the interface and the highest-occupied MIGS determines the effective Fermi energy of the semiconductor around the interface, as shown in Fig. 1.23(a). Since this Fermi energy is determined by the electron number to keep the semiconductor neutral, it is called the charge-neutrality level (CNL). Hereafter, we denote this CNL as  $\phi_{CNL}$ . When the semiconductor produces a contact with a metal, as in Fig. 1.23(b), electrons in the semiconductor and metal move across the interface to equalize their Fermi energies,  $\phi_{CNL}$  and  $E_F$ , which results in the final band alignment at the interface. The energy difference between  $E_F$  and the lowest conduction-band state of the semiconductor,  $E_c$ , or  $E_F$  and the highest valence-band one,  $E_v$ , acts as a potential barrier for electrons or holes, respectively, and is called the Schottky barrier (offset). In addition, the energy

position of  $E_F$  relative to the vacuum is sometimes called an effective work function (WF).

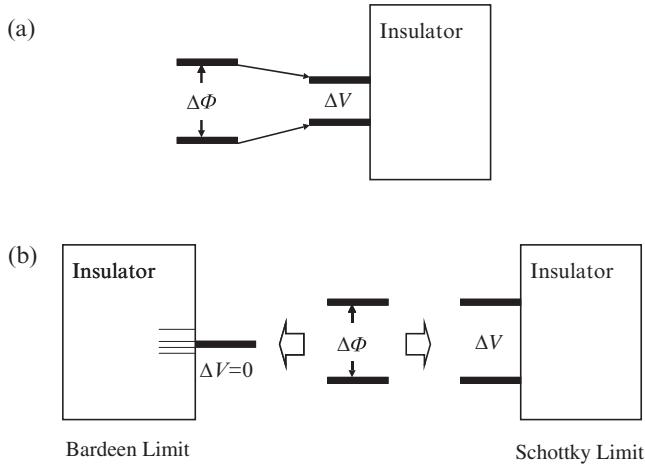
Tersoff (1984) proposed that the  $\phi_{\text{CNL}}$  is determined by the electronic structure of the bulk semiconductor. As explained above, the MIGS is the eigenstate in complex band structures having energies within the bandgap of the bulk system. Since the complex bands connect with real bands at the band edge of real bands, the MIGS eigenstates with lower energies possess valence-band character, while those having higher energies carry conduction-band character (see Fig. 1.23(a)). The effective Fermi energy,  $\phi_{\text{CNL}}$ , is defined as the boundary between these valence-band-like and conduction-band-like states, thus being roughly given by the sign boundary of the propagating Green function as

$$G(\mathbf{r}, \mathbf{r}', \phi_{\text{CNL}}) = \sum_n \int d^3k \frac{\varphi_{nk}^*(\mathbf{r})\varphi_{nk}(\mathbf{r}')}{\phi_{\text{CNL}} - \varepsilon_{nk}} = 0, \quad (1.10)$$

where  $\varphi_{nk}(\mathbf{r})$  and  $\varepsilon_{nk}$  are eigenwavefunction and eigenenergy of the bulk electronic states of the semiconductor, respectively, with the band index,  $n$ , and the Bloch wave number,  $\mathbf{k}$ . Using this  $\phi_{\text{CNL}}$ , he succeeded in reproducing SBH of various metal/semiconductor interfaces. The application of this  $\phi_{\text{CNL}}$  scenario to semiconductor/semiconductor interfaces was straightforward and also succeeded in predicting valence- and conduction-band offsets of many semiconductor/semiconductor interfaces.

Here, we would like to point out two important features in this kind of  $\phi_{\text{CNL}}$  theory. First, due to some ambiguity to determine the boundary between valance- and conduction-band-like MIGS, there exist several versions to define  $\phi_{\text{CNL}}$  (Spicer *et al.* 1980; Hasegawa *et al.* 1986; Cardona *et al.* 1987). Most of them present similar values of  $\phi_{\text{CNL}}$  for many semiconductors. This occurs because they assume that the full contact is realized between metals and semiconductors, details of which will be discussed in the following. Second, it should be noted that  $\phi_{\text{CNL}}$  is defined as intrinsic not to the interface but to the bulk properties of semiconductors. For example, assuming the simple band structure for the semiconductor,  $\phi_{\text{CNL}}$  is given by  $E_{\text{VB}} + E_{\text{G}}^* D_{\text{VB}} / (D_{\text{VB}} + D_{\text{CB}})$  (Cardona *et al.* 1987), where  $E_{\text{VB}}$  is the top of the valence band,  $E_{\text{G}}$  the bandgap energy,  $D_{\text{VB}}$  and  $D_{\text{CB}}$  are the density of states for valence and conduction bands, respectively.

In real metal/semiconductor (insulator) interfaces, the Fermi level of a metal does not match CNL. However, CNL concepts give us crucial information of SBH; According to the CNL concept, the Fermi levels of the metals move toward CNL when they are in contact with semiconductors (insulators), as shown in Fig. 1.24(a). Thus, the difference in SBH of two metals ( $\Delta V$ ) is always smaller than that in work functions ( $\Delta\phi$ ). Therefore, the slope parameter ( $S$ ) that is defined as  $S = \Delta V / \Delta\phi$ , should be  $0 < S < 1$ . The situations that correspond to  $S = 0$  and  $S = 1$  can be regarded as the intrinsic limits of Schottky barriers.  $S = 0$  means that SBH is always pinned at  $\phi_{\text{CNL}}$ , regardless of the metal species. This limit is called the Bardeen limit (Fig. 1.24(b), left). On the other hand,  $S = 1$  corresponds to the situation where the difference in vacuum work functions and in the Schottky barrier heights are the same. This limit is called the Schottky limit (Fig. 1.24(b), right).

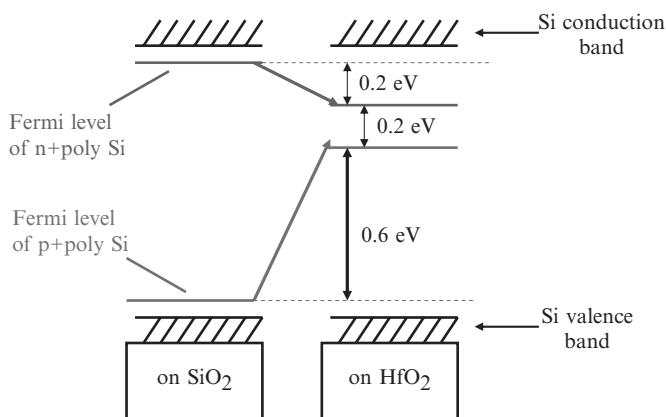


**Fig. 1.24** (a) Schematic illustration of conventional understanding of Schottky barrier height. (b) Schematic illustration of Bardeen and Schottky limit.

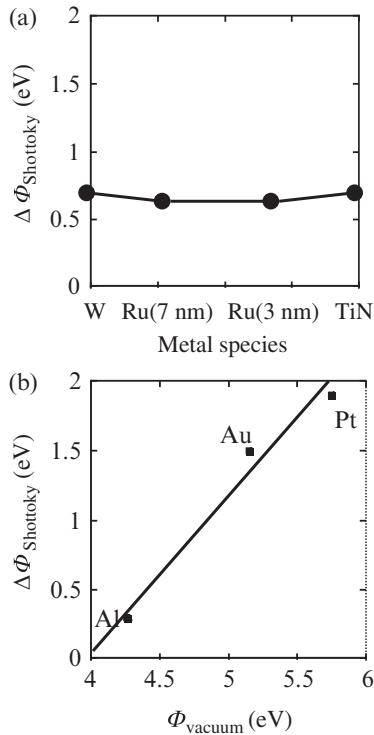
### 1.3.2.2 Curious behaviors of Schottky barrier heights at metal/HfO<sub>2</sub> interfaces

During developing HfO<sub>2</sub>-based high- $k$  dielectrics, we faced really unusual Schottky barrier height behavior that cannot be explained by the conventional interface physics concepts. The first example is that SBH at the p+poly-Si/HfO<sub>2</sub> interface is smaller than that of the n+poly-Si/HfO<sub>2</sub> interface by only 0.2 eV, although SBH at the p+poly-Si/SiO<sub>2</sub> interface is smaller than that at the p+poly-Si/SiO<sub>2</sub> interface by about 1 eV, reflecting the Si bandgap value of 1 eV. The schematic illustration is given in Fig. 1.25.

Moreover, we have also encountered an unusual Schottky barrier height behavior at metal/HfO<sub>2</sub> interfaces. After high-temperature anneal, SBH of p-metals are almost identical, irrespective of the metal species, as shown in Fig. 1.26(a) (Akasaka *et al.* 2006; Kadoshima *et al.* 2007). Since HfO<sub>2</sub> is a typical wide-gap insulator, SBH behavior is expected to be similar to the Schottky limit according to the conventional CNL concepts. However, the observed behavior rather resembles the Bardeen limit. Moreover, SBH are drastically changed without high-temperature treatment. As shown in Fig. 1.26(b), SBH



**Fig. 1.25** Schematic illustration of the unusual behaviors of the relative Schottky barrier heights at the p+poly-Si/HfO<sub>2</sub> interface compared with that at the n+poly-Si/HfO<sub>2</sub> interface. For comparison, well-known Schottky barrier height behavior at the poly-Si/SiO<sub>2</sub> interfaces is also shown.



**Fig. 1.26** Behavior of relative Shottky barrier height ( $\Delta\Phi_{\text{Schottky}}$ ) after (a) high-temperature treatment (Kadoshima *et al.* 2007), (b) low-temperature treatments (Koyama *et al.* 2004). Reported  $\Delta\Phi_{\text{Schottky}}$  are plotted as a function of (a) metal species and (b) vacuum work function ( $\Delta\Phi_{\text{vacuum}}$ ).

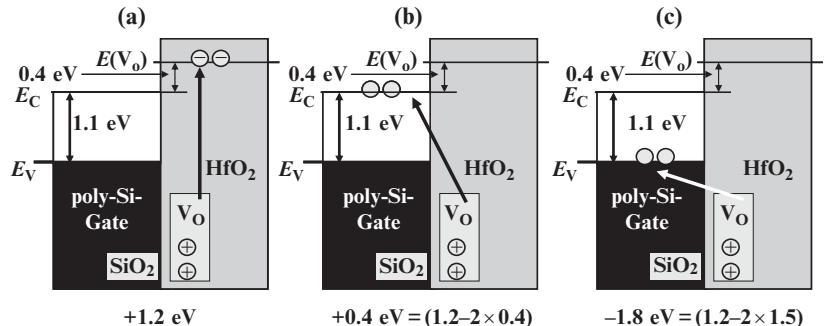
sensitively depends on the metal species (Koyama *et al.* 2004). Furthermore, the observed slope parameter is larger than unity, which means the breakdown of the Schottky limit. In the following sections, we show that new interface physics concepts readily explain these unusual results.

### 1.3.2.2.3 Schottky barrier heights at p+poly-Si/HfO<sub>2</sub> interfaces

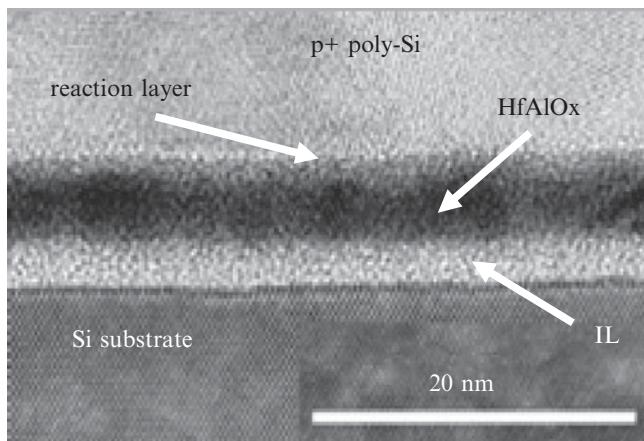
We first introduce an unexpected Schottky barrier height behavior at the interface between heavily B-doped Si (p+poly-Si) and HfO<sub>2</sub> (Shiraishi *et al.* 2004). The relatively higher energy level of V<sub>o</sub> in HfO<sub>2</sub> mentioned in the previous section causes a notable thermodynamic behavior of interfaces, when the HfO<sub>2</sub> is in contact with Si. Recent experiments indicate that the V<sub>o</sub> level is located about 0.4 eV above the bottom of the Si conduction band (about 1.2 eV below the bottom of the HfO<sub>2</sub> conduction band) (Takeuchi *et al.* 2004).

As mentioned previously, Hf atoms can bind much more strongly than Si atoms to O atoms. Actually, the formation enthalpy of HfO<sub>2</sub> is larger than SiO<sub>2</sub> by about 2.2 eV. Further, recent first-principles calculations show that the V<sub>o</sub> formation energy in bulk HfO<sub>2</sub> is larger than that in bulk SiO<sub>2</sub> by about 1.2 eV (Scopel *et al.* 2004). At first glance, the fact that Hf–O bonds are stronger than Si–O bonds indicates that the partial oxidation of poly-Si gates by pulling an O atom out of the HfO<sub>2</sub> dielectrics is an endothermic reaction with  $\sim 1.2$  eV energy loss, and this reaction occurs with difficulty. However, the situation changes completely if we take into account the electron behavior.

Fermi-level pinning (large  $V_{fb}$  shifts) in Hf-related high- $k$  gate stacks with p+poly Si gates can naturally be explained by taking into account the electron behavior as follows. In Figs. 1.27(a)–(c), the mechanism of V<sub>o</sub> formation in HfO<sub>2</sub> and subsequent electron transfer across the poly-Si/HfO<sub>2</sub> interface is schematically illustrated. First, let us assume that the poly-Si is partially oxidized by the formation of Si–O–Si bonds by O atoms being pulled out of the HfO<sub>2</sub>. As a result, Si–O–Si bonds in the poly-Si and V<sub>os</sub> in the HfO<sub>2</sub> are formed (Fig. 1.27(a)). An O atom in HfO<sub>2</sub> takes an O<sup>2-</sup> ion form, but an O atom in a Si–O–Si bond is neutral. Accordingly, two additional electrons are generated after one O atom is pulled out, and if these electrons remain inside the HfO<sub>2</sub>, they occupy the V<sub>o</sub> level in the HfO<sub>2</sub> (Fig. 1.27(b)). The assumption that two additional electrons remain in the HfO<sub>2</sub> corresponds to the same situation as the bulk calculation that give a  $\sim 1.2$  eV energy loss. However,



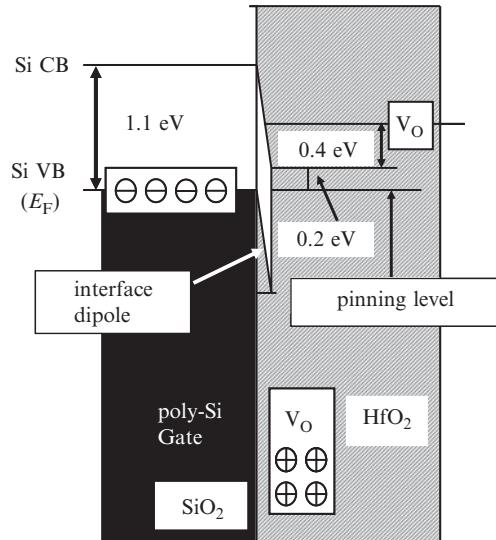
**Fig. 1.27** Schematic illustration of V<sub>o</sub> formation in HfO<sub>2</sub> with partial oxidation of poly-Si gate and subsequent electron transfer into the gate electrodes. (a) Partial poly-Si oxidation by pulling out an O atom from HfO<sub>2</sub>. (b) Energy loss in an n+poly-Si gate. (c) Large energy gain in a p+poly-Si gate (Shiraishi *et al.* 2006).



**Fig. 1.28** Cross-section of replacement  $p+$  gate  $HfAlO_x$  MISFET observed by TEM (Shiraishi *et al.* 2006).

since the  $HfO_2$  is in contact with the poly-Si gate, electrons have to transfer into the gate. This is because the  $V_o$  level is located above the poly-Si Fermi level (Fig. 1.27(c)). Now, we estimate roughly the two ultimate cases of energy loss (gain) when the  $HfO_2$  is in contact with the poly-Si gate: One is with an  $n+$ -poly-Si gate and the other is with a  $p+$ -poly-Si gate. For an  $n+$ -poly-Si gate, the two-electron transfer results in an energy gain of 0.8 eV ( $2 \times 0.4$  eV), and the total energy loss is reduced from 1.2 eV (bulk value) to 0.4 eV. Despite the energy reduction due to the electron transfer, the reaction for an  $n+$ -poly-Si gate is still endothermic. For a  $p+$ -poly-Si gate, on the other hand, the situation is quite different. The Fermi-level position of a  $p+$ -poly-Si gate is located about 1.5 eV below the  $V_o$  level in  $HfO_2$ . As a result, the two-electron transfer results in a total energy gain of 1.8 eV ( $2 \times 1.5$  eV – 1.2 eV). Surprisingly, the interface reaction, accompanied by  $V_o$  formation and subsequent electron transfer, becomes exothermic with an energy gain of 1.8 eV, when the  $HfO_2$  is in contact with a  $p+$ -poly-Si gate. Actually, the transmission electron microscope (TEM) observation shows that some interfacial reaction layers are observed in a  $p+$ -poly-Si gate  $HfAlO_x$  MISFET as shown in Fig. 1.28, and such reaction layers have not been observed in  $n+$ -poly-Si gate MISFETs. These results completely corroborate the above discussion based on the “oxygen vacancy model”.

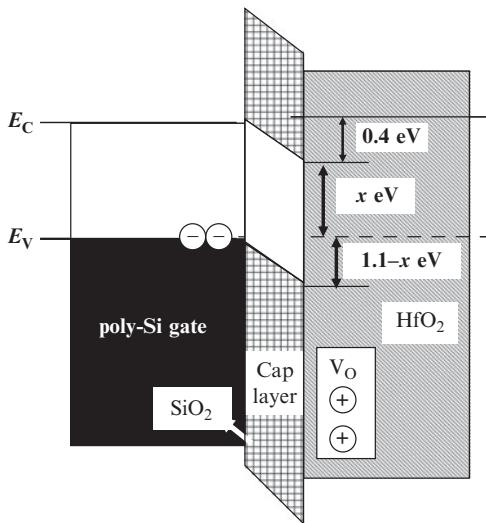
Now, we move on to the  $V_{fb}$  shift originating from the formation of  $V_o$  in  $HfO_2$ . For an  $n+$ -poly-Si gate, the interface reaction that induces electron transfer occurs with difficulty, since it is endothermic. For a  $p+$ -poly-Si gate, however, the interface reaction accompanied by the formation of  $V_o$  in  $HfO_2$  occurs easily, since this reaction has a large energy gain of 1.8 eV. At the same time, electron transfer from the  $HfO_2$  into the poly Si occurs. As a result, an interface dipole is formed as illustrated in Fig. 1.29. This dipole formation raises the position of the Fermi level of the  $p+$ -poly-Si gate and the flat-band voltage ( $V_{fb}$ ) decreases. It is noticeable that the energy gain of the interface reaction decreases when the Fermi level is elevated. A simple consideration indicates that the position at which the Fermi level is pinned corresponds to the energy level that makes the energy gain of the interfacial reaction zero. The final Fermi-level position satisfies the equation,  $1.2 - 2(0.4 + x) = 0$ , where



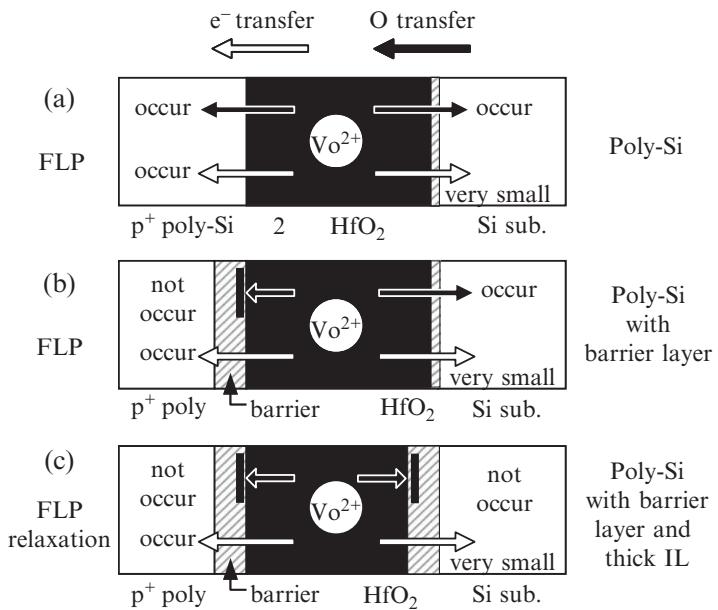
**Fig. 1.29** Schematic illustration of interface dipole formation and subsequent Fermi-level elevation toward the pinning level (Shiraishi *et al.* 2006).

$x$  is the final pinning position measured from the n+poly-Si Fermi level. The Fermi level position of a p+poly-Si gate obtained from this is about 0.2 eV below the n+poly-Si gate Fermi level, which is in fairly good agreement with the experimental results (Hobbs *et al.* 2003). As discussed above, the ‘‘oxygen vacancy model’’ can quantitatively reproduce the Fermi-level pinning position of p+poly-Si gate/HfO<sub>2</sub> interfaces.

Next, we comment on the effect of inserting a cap insulator between p+poly-Si and HfO<sub>2</sub> gate dielectrics. Recent experiments show that Fermi-level pinning of p+poly-Si gates cannot essentially be improved by inserting SiO<sub>2</sub> or SiN cap layers between p+poly-Si gates and high- $k$  Hf-related dielectrics (Cartier *et al.* 2004). The oxygen-vacancy model naturally explains the results of these experiments. The schematic illustrations are shown in Fig. 1.30. As shown in this figure, the pinning position is governed only by the energy position that balances the  $V_O$  formation energy loss and the electron-transfer energy gain. The final pinning position measured from the n+poly-Si Fermi level ( $x$ ) satisfies the equation  $1.2 - 2(0.4 + x) = 0$ , which is the same equation without a cap layer. Accordingly, if O atoms can penetrate through the cap layer until the system reaches thermal equilibrium, the final pinning position remains the same regardless of the existence of a cap layer. Systematic experiments have been reported by Kamimuta *et al.* (2005). They examined three geometries of poly-Si/HfO<sub>2</sub> gate stacks. The schematic illustrations are shown in Fig. 1.31. The usual gate stack structures, in which FLP is observed in p+poly-Si gates, is the gate stack structures with no barrier layer and thin IL, as shown in Fig. 1.31(a). At first, the effect of boron segregated near the Si/HfO<sub>2</sub> interfaces (Takayanagi *et al.* 2003) or interfacial Hf–Si bonds (Hobbs *et al.* 2003) were thought to be the cause of FLP of p+poly-Si gates. However, FLP cannot be avoided even a barrier layer is inserted between a p+poly-Si gate and HfO<sub>2</sub>, as described in Fig. 1.31(b), indicating that neither the B effect nor Hf–Si bonds can be the cause of FLP at p+poly-Si/HfO<sub>2</sub>.

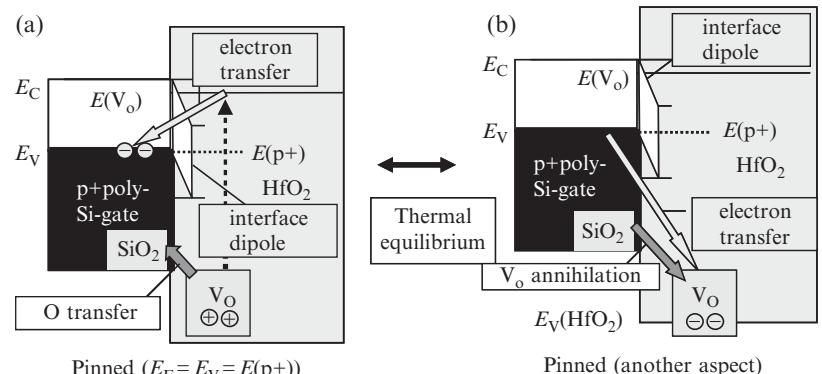


**Fig. 1.30** Schematic illustration of the cap-layer effect for Fermi-level pinning (Shiraishi *et al.* 2006).



**Fig. 1.31** Schematic illustration of three poly-Si/high- $k$  gate stack structures examined by Kamimuta *et al.* (2005). (a) Usual  $p^+$ -poly-Si gate stack structure. (b) Gate stack structure with barrier layer between a  $p^+$ -poly-Si gate and a high- $k$  dielectric. (c) Gate stack structure in which barrier layers are inserted both between a  $p^+$ -poly-Si gate and a high- $k$  dielectric and between a Si substrate and a high- $k$  dielectric (Akasaka *et al.* 2006).

interfaces. They have found that FLP relaxation can be achieved only when the gate stack structure contains both thick barrier layer and thick IL, as illustrated in Fig. 1.31(c). Their experimental finding indicates that interaction between  $\text{HfO}_2$  and a Si substrate is very important for FLP as well as that between poly-Si gate and  $\text{HfO}_2$ . In other words, FLP disappears only when the interaction between Si and  $\text{HfO}_2$  is weak enough. Further, it is expected that indirect interaction between Si and  $\text{HfO}_2$  can induce FLP. In short, the thermal



**Fig. 1.32** Schematic illustration of another understanding of Fermi-level pinning. (a)  $V_o$  generation and (b)  $V_o$  annihilation are balanced, and the system reaches thermal equilibrium.

equilibrium between  $V_o$  formation and annihilation reaction at the Si/HfO<sub>2</sub> interface given by the following equation determines the SBH (FLP position) at p+poly-Si/HfO<sub>2</sub> interface as shown in Fig. 1.32

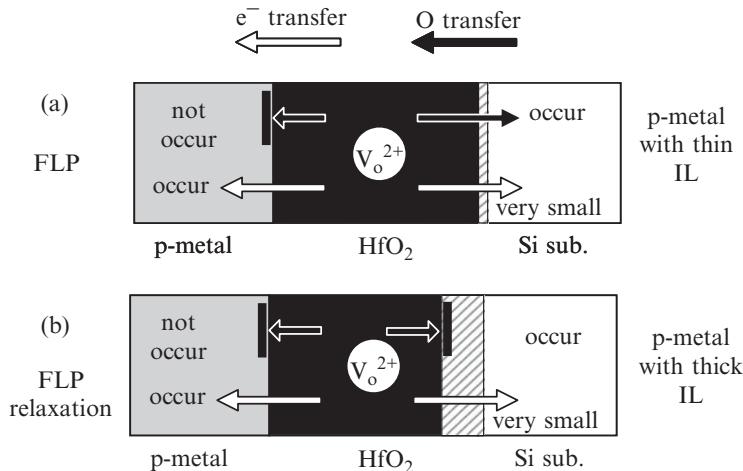


In the next section, we will consider the FLP observed at metal/HfO<sub>2</sub> interfaces after high-temperature treatment based on this concept.

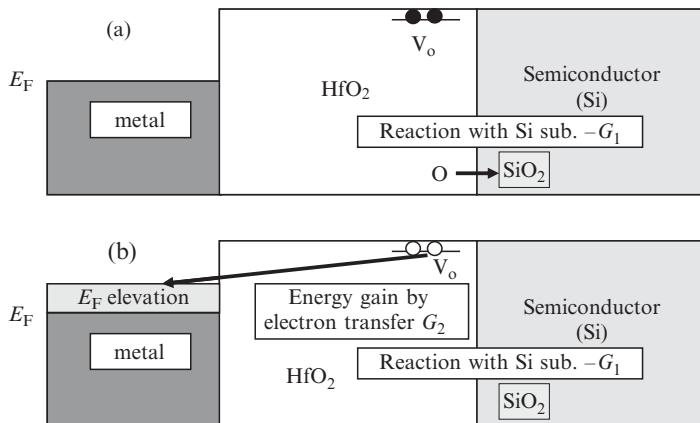
#### 1.3.2.2.4 Unusual Schottky barrier height behavior at metal/HfO<sub>2</sub> interface after high-temperature treatment

It has been reported that Schottky barrier heights of p-metal remarkably decrease revealing Fermi-level pinning (FLP) behavior when the interface layer (IL) is thin, after high-temperature annealing, whereas, FLP does not appear when IL is thick enough (Lee *et al.* 2006). This is called “ $V_{fb}$  roll-off”. The noticeable fact is that the energy position of FLP is similar to the pinning position of p+poly-Si gates (Hobbs *et al.* 2003) as mentioned in the previous section. This experimental fact indicates that the FLP of p-metal is also governed by a mechanism similar to FLP of p+poly-Si gates mentioned above (Akasaka *et al.* 2006).

Now, we move on to the mechanism of FLP of p-metal gates after high-temperature treatment. Since normal p-metals are non-reactive materials, the situation of p-metal/high-*k* gate stack can be schematically illustrated as in Fig. 1.33. O transfer hardly occurs from high-*k* dielectrics to p-metals due to the low reactivity of p-metals, whereas, O transfer is still possible if IL is thin (Fig. 1.33(a)). However, O transfer from high-*k* dielectrics is inhibited when IL is thick enough (Fig. 1.33(b)). As discussed in Fig. 1.33, the interface reaction between a high-*k* dielectric and Si substrate still occurs when IL is thin. Further, reaction with the Si substrate induces electron transfer from  $V_{os}$ s to p-metal gates. A schematic illustration of this situation is given in Fig. 1.34. It is noticeable that the net reaction between high-*k* dielectrics and the Si substrates is the same as that between poly-Si gates and high-*k* dielectrics, although electron transfer and O transfer directions are opposite to each other.



**Fig. 1.33** Schematic illustration of two typical p-metal/high- $k$  gate stack structures. (a) p-metal/high- $k$  gate stack structure with thin IL. (b) p-metal/high- $k$  gate stack structure with thick IL (Akasaka *et al.* 2006).



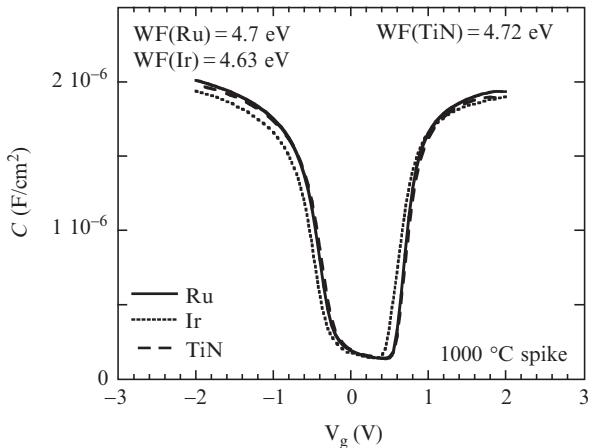
**Fig. 1.34** Schematic illustration of interface reaction with the Si substrate and subsequent electron transfer from V<sub>o</sub> to gate metals in p-metal/high- $k$  gate stacks. (a) O transfer into the Si substrate through thin IL. (b) Subsequent electron transfer from a V<sub>o</sub> level to a p-metal gate that induces gate Fermi-level elevation.

In fact, the reaction equation with the Si substrate can be described as



which is the same as in the case of p+poly-Si gates described in the previous section. According to Fig. 1.34, the pinning position corresponds to the energy at which the energy loss ( $G_1$ ) and energy gain ( $G_2$ ) by electron transfer from V<sub>o</sub> in a high- $k$  dielectric to a gate metal are cancelled by each other (i.e.  $G_1 - G_2 = 0$ ). This means that the FLP positions of p-metal gates are the same as that of the p+poly-Si gate irrespective of metal species. In fact, observed EWFs of p-metals are almost independent of metal species, as shown by the C-V curves in Fig. 1.35.

It is better to note that the V<sub>o</sub>-related mechanism is not the total cause of the  $V_{fb}$  shift. The V<sub>o</sub>-related mechanism determines the final position of FLP. For example, if other factors such as surface strain (Ikeda *et al.* 2006) or MIGS lower the EWFs of gate metals, a relatively small number of V<sub>o</sub> generation is sufficient to reach the FLP position. The thermodynamics of the interfacial



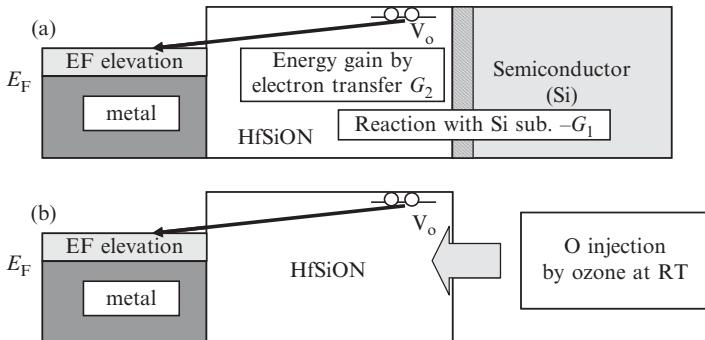
**Fig. 1.35** Observed  $C$ - $V$  curves of Ru, Ir and TiN. The estimated effective work functions are similar to that of the Fermi-level pinning position of the p+poly-Si gate (Akasaka *et al.* 2006).

reaction between high- $k$  dielectrics and the Si substrates that generates  $V_o$  and  $V_o$ -induced interface dipole, determines the final position of FLP, when the system can reach thermal equilibrium. Accordingly, the pinning position is independent of the process condition and the film quality. This is the main concept of the “oxygen vacancy model”.

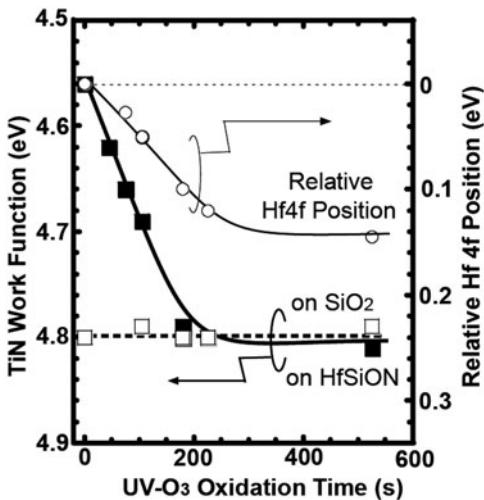
As discussed above, FLP of p-metals naturally occurs if the IL is thin, since the reaction between high- $k$  dielectrics and the Si substrate is inevitable. However, FLP can be avoided if the reaction between high- $k$  dielectrics and the Si substrate is suppressed. In order to suppress the corresponding reaction, there are two possibilities. One is insertion of a thick IL, and the other is the low-temperature process. The former leads to the remarkable increase in effective oxide thickness (EOT). Thus, it is not suitable for future LSI technologies. It is naturally expected that the latter (the low-temperature process) is the promising solution to avoid FLP if we use Hf-based high- $k$  gate dielectrics.

It is also noted that the interface dipole modulation between high- $k$  dielectrics and Si substrates is also effective, since this modulation does not change the thermodynamics of the interface reactions that generate O vacancies. In other words, the relative energy difference between the neutral  $V_o$  level and the FLP position of a gate metal does not change by this dipole modulation. F incorporation (Inoue *et al.* 2005) or counterdoping effects are categorized in this recipe, which modulates the dipole at IL/Si interfaces. It has also been proposed that Al and La incorporation into Hf-related oxides can modulate the dipole at high- $k$ /IL interfaces. This is also categorized into this recipe (Iwamoto *et al.* 2008).

Now, we show the experiments that confirm the validity of the oxygen-vacancy model. After the high-temperature treatment that causes FLP of p-metals, the Si substrate is removed. Next, oxygen atoms are injected from the substrate side into Hf-based high- $k$  dielectrics on room-temperature ozone treatment (Ohta *et al.* 2006). The schematic illustration of the experiments is given in Fig. 1.36. The obtained results are shown in Fig. 1.37. As is clearly



**Fig. 1.36** Schematic illustration of the experimental procedure that confirms our  $V_o$  model. (a) FLP occurs after high-temperature treatment. (b) After the removal of Si substrates, ozone is injected into Hf-based high- $k$  dielectrics at room temperature.



**Fig. 1.37** Observed TiN effective work function and relative Hf4f position as functions of UV-ozone oxidation times. Experiments are based on XPS measurements (Ohta *et al.* 2006).

shown in this figure, the EWF of TiN increases with the ozone-injection time, and it reaches the original position, releasing FLP. Thus, these experimental results clearly indicate that O deficiency is the main cause of FLP of TiN. It is also noticeable that the quantity of the Hf core-level shift does not match the total increase in EWF of TiN. This means that  $V_o$  generation in Hf-related high- $k$  dielectrics is not the whole cause of Fermi-level shift of TiN gates. Other factors also contribute to the shift of EWF. It is consistent with the oxygen-vacancy model that the thermodynamics of  $V_o$  generation determines the final position of FLP, although other factors also contribute to the  $V_{fb}$  shift (Ohta *et al.* 2006). As discussed above, the oxygen-vacancy model can naturally explain the FLP of p-metals observed when IL is very thin, and this model is experimentally confirmed.

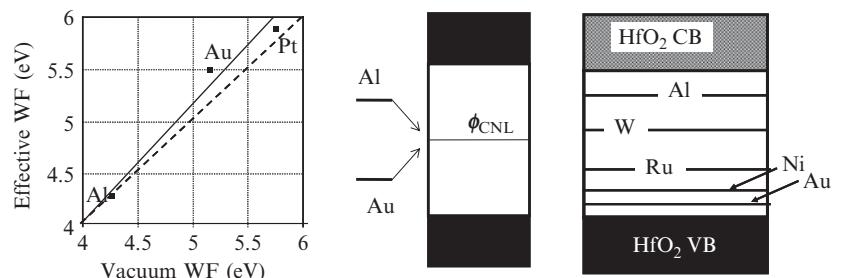
Finally, we mention the new finding of the physics of Schottky barriers. In the above mechanism of Schottky-barrier formation, SBH at a metal/HfO<sub>2</sub> interface is determined by the thermodynamics of the reaction at the other interface at HfO<sub>2</sub>/Si instead of the corresponding metal/HfO<sub>2</sub> interfaces. This is totally different from the conventional understanding of SBH that it is determined by the dipole at the corresponding interfaces.

### 1.3.2.2.5 Unusual behavior of SBH at High- $k$ interfaces after low-temperature treatment

Koyama *et al.* (2004) deposited Au, Pt, and Al metals on HfSiO substrate and measured their effective work functions (EWFs); their results are displayed in Fig. 1.38(a). They found that the effective EWFs of Au and Pt become large compared to WFs in vacuum, while the effective WF of Al is smaller than WF in vacuum. This result indicates that the  $S$  parameter becomes larger than unity, being beyond the conventional limit. This is because the conventional interface concept requires that the Fermi level of metals should move toward the  $\phi_{\text{CNL}}$ , as shown in Fig. 1.38(b). By using the CV techniques, the EWF shifts were observed on HfAlO<sub>x</sub> substrate for Al metal with a decrease of 0.36 eV, and for Ni and Au metals with increases of 0.20 and 0.22 eV, respectively (Shiraishi *et al.* 2005). However, it should be noted that the experimental results mentioned above have been reported based on the C–V measurement of the relatively complex gate stack structure of metal/HfO<sub>2</sub>/SiO<sub>2</sub>/Si. These findings are not direct proof of the SBH anomaly at metal/HfO<sub>2</sub> interfaces, although they provided great insights. In this section, we discuss whether such an anomaly is possible or not.

The change of EWF from in-vacuum to on-high- $k$ -dielectrics is schematically summarized by arrows in Fig. 1.38(c). The Fermi energies of p-metals such as Ni and Au, which have larger EWFs, are shifted toward the valence band of high- $k$  dielectrics, whereas those of n-metals such as Al having smaller EWFs move toward the conduction band. These EWF changes are quite different from the prediction of the conventional  $\phi_{\text{CNL}}$  theories. The conventional theories state that there is a single  $\phi_{\text{CNL}}$  level in the bandgap of a high- $k$  dielectric and the metal Fermi energies are aligned to this  $\phi_{\text{CNL}}$ , as shown in Fig. 1.38(c). Thus, the variation in Fig. 1.38(c) was beyond our ordinary understanding.

In order to clarify what happens at metal/high- $k$  interfaces, Shiraishi *et al.* (2005) investigated the electronic structures using the first-principles calculations, and found that two important pre-conditions implicitly assumed in conventional theories are broken at these interfaces. The first is concerned with the penetration length of MIGS from the interface into insulating materials. Figs. 1.39(a) and (b) show the charge densities of typical MIGS states around metal/Si(111) and metal/HfO<sub>2</sub>(110) interfaces, respectively. Here, the metal is represented by the jellied model. In the case of metal/Si interface, the MIGS penetrates deep into Si, about five atomic layers. Thus, the Si side of the

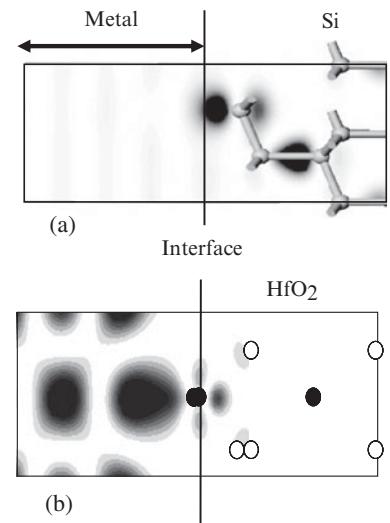


**Fig. 1.38** (a) Effective work functions of Al, Au, and Ni on HfSiON as a function of vacuum work function, reported by Koyama *et al.* (2004). (b) Schematic band alignment expected by the conventional theories based on the charge-neutrality level. (c) Observed band alignment of metals on HfO<sub>2</sub> (Shiraishi *et al.* 2005).

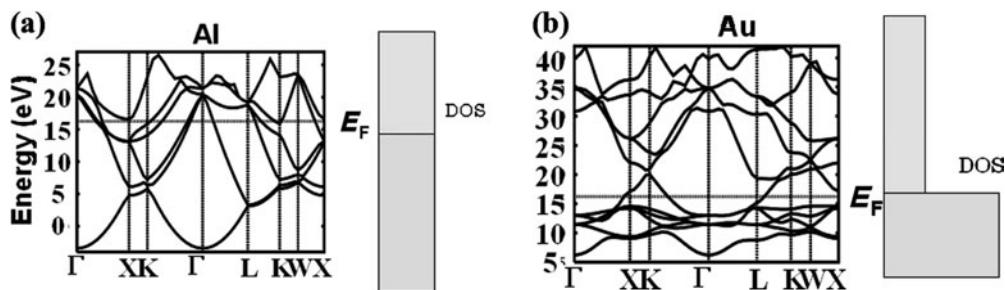
interface has movable electrons and looks like a metal. By moving electrons in these states, the band alignment is realized. Note here that, since the MIGS penetrates deep and touches a number of Si atoms around the interface, the detailed interface atomic structure does not manifest the electronic structure of the interface. In other words, one can say that the full contact of electronic states is realized at this metal/Si interface. On the other hand, in the case of metal/HfO<sub>2</sub>, the penetration of MIGS into HfO<sub>2</sub> is seen at most to one to two atomic layers, thus the full contact of electronic states is not realized at metal/HfO<sub>2</sub> interfaces. This occurs because of the high ionicity and large bandgap of HfO<sub>2</sub>. The present result indicates that the electronic structure at a metal/high- $k$  interface is very sensitive to the interface atomic structures and we had better start with the bonding picture of atoms to understand the electronic structure at this interface.

The other pre-condition is related to the individuality of metals. Figures 1.40(a) and (b) show the band structures of Al and Au metals around the Fermi energies, respectively. The schematic pictures of density of states are also shown on the right sides. As the extended  $s$  and  $p$ -orbital electrons are valence electrons, the Al metal has a featureless density of states around the Fermi energy. On the other hand, due to the localized  $d$ -orbital electrons, the Au metal has small-dispersion bands below the Fermi energy. Thus, the density of states is extremely large below the Fermi energy, while it is comparable to the Al case above the Fermi energy. As shown in the following, this kind of metal character is not considered in the conventional theory.

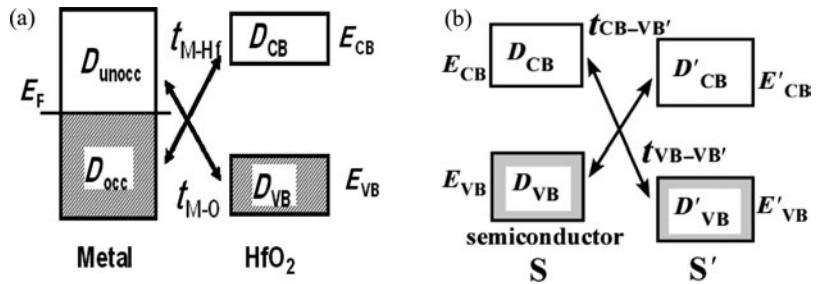
The full contact of electronic states at the interface and the featureless metal density of states are essential conditions in conventional theories. This is apparent because the  $\phi_{\text{CNL}}$  is defined using quantities intrinsic to bulk materials, such as  $E_{\text{VB}}$ ,  $E_{\text{G}}$ ,  $D_{\text{VB}}$  and  $D_{\text{CB}}$ , and does not include the interface information such as the atomic structures and characteristics of metals. In fact, in the case of the theory of disordered interface gap states (DIGS) (Hasegawa *et al.* 1986), for example, the interface is assumed to have random amorphous-like structures and some kinds of averages are implicitly taken to deduce  $\phi_{\text{CNL}}$ . However, as shown above, it is clear that these conditions are not satisfied at metal/high- $k$  interfaces.



**Fig. 1.39** Contour plots of MIGS wavefunctions at (a) metal/Si interface and (b) metal/HfO<sub>2</sub>.



**Fig. 1.40** Band structures of (a) Al and (b) Au bulk metals calculated by the first-principles method. Schematic pictures of density of states are also shown on the right of each panel (Shiraishi *et al.* 2005).



**Fig. 1.41** Schematic pictures of interface hybridization models to derive the generalized neutrality levels. (a) Metal/high- $k$  interface, (b) semiconductor/semiconductor interface.

### 1.3.2.2.6 A new theory of Schottky-barrier height

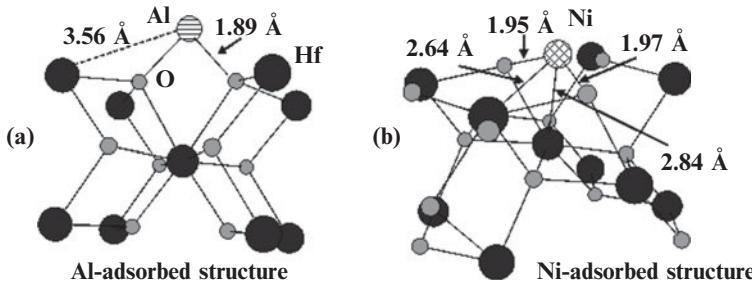
In order to understand electronic structures of metal/high- $k$ -dielectrics interfaces, we have to take into account the interface atomic structures and characteristics of metals. To simulate such interfaces, Shiraishi *et al.* (2005) adopted the effective-four-level tight-binding model of the interface. The schematic picture of this model is described in Fig. 1.41(a). They characterize a metal with the Fermi energy,  $E_F$ , and the effective local densities of states at the interface below and above  $E_F$ ,  $D_{\text{occ}}$  and  $D_{\text{unocc}}$ . The high- $k$  dielectric such as  $\text{HfO}_2$  is represented by the energies of valence-band top and conduction-band bottom,  $E_{\text{VB}}$  and  $E_{\text{CB}}$ , and the effective local densities of states of valence and conduction bands at the interface,  $D_{\text{VB}}$  and  $D_{\text{CB}}$ . Since  $\text{HfO}_2$  is an ionic material, the conduction bands are mainly made of Hf  $d$ -orbitals, while the valence bands are made of O  $p$ -orbitals.  $t_{M-\text{Hf}}$  is a transfer energy between occupied metal states and conduction-band states of  $\text{HfO}_2$ , while  $t_{M-\text{O}}$  is that between unoccupied metal states and valence-band states of  $\text{HfO}_2$ . It should be noted here that the charge transfer between a metal and  $\text{HfO}_2$  is realized only by such orbital hybridization between unoccupied and occupied states. Transfer energies between occupied states and those between unoccupied states never induce the charge transfer between the metal and  $\text{HfO}_2$ , thus they are not relevant in determining the band alignment and are thus excluded in the present model.

By applying the second-order perturbation theory of quantum mechanics, the charge transfer from a metal to  $\text{HfO}_2$  is written as

$$\Delta\rho \propto \frac{|t_{M-\text{Hf}}|^2 D_{\text{occ}} D_{\text{CB}}}{E_{\text{CB}} - E_F} - \frac{|t_{M-\text{O}}|^2 D_{\text{unocc}} D_{\text{VB}}}{E_F - E_{\text{VB}}}. \quad (1.13)$$

This charge transfer produces the dipole potential at the interface and increases the energies of  $E_{\text{VB}}$  and  $E_{\text{CB}}$  in  $\text{HfO}_2$ . Since the charge transfer should be completed by inducing a final dipole potential of  $\Delta V$ , the self-consistent equation of  $\Delta V$  becomes as

$$\frac{|t_{M-\text{Hf}}|^2 D_{\text{occ}} D_{\text{CB}}}{E_{\text{CB}} + \Delta V - E_F} - \frac{|t_{M-\text{O}}|^2 D_{\text{unocc}} D_{\text{VB}}}{E_F - E_{\text{VB}} - \Delta V} = 0. \quad (1.14)$$



**Fig. 1.42** Calculated metal-atom positions on  $\text{HfO}_2$  (110) substrate. (a) Al metal atom and (b) Ni metal atom (Shiraishi *et al.* 2005).

By solving this, the generalized charge-neutrality level of  $\text{HfO}_2$  that should match the Fermi energy of metal is obtained by  $\phi_{\text{CNL}}^G = E_F - \Delta V$  as

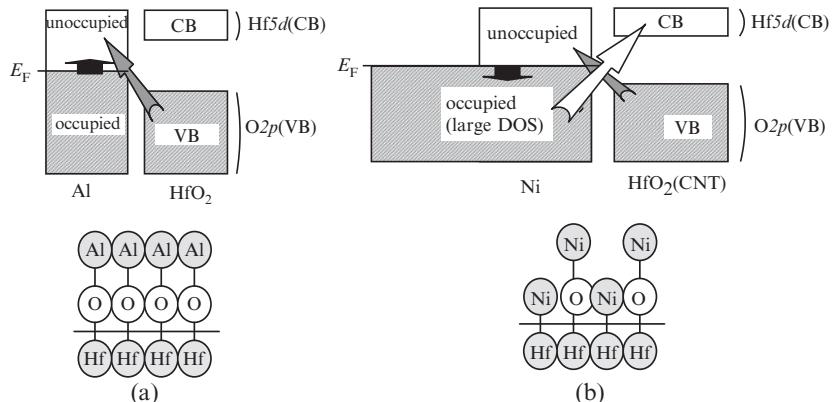
$$\phi_{\text{CNL}}^G = E_{\text{VB}} + E_G \frac{|t_{\text{M}-\text{O}}|^2 D_{\text{unocc}} D_{\text{VB}}}{|t_{\text{M}-\text{O}}|^2 D_{\text{unocc}} D_{\text{VB}} + |t_{\text{M}-\text{Hf}}|^2 D_{\text{occ}} D_{\text{CB}}}, \quad (1.15)$$

where  $E_G = E_{\text{CB}} - E_{\text{VB}}$  is the bandgap energy of  $\text{HfO}_2$ .

In the cases of ordinary metal/semiconductor interfaces, we can expect the full contact of electronic states,  $t_{\text{M}-\text{Hf}} = t_{\text{M}-\text{O}}$ , and the density of metal states is featureless,  $D_{\text{occ}} = D_{\text{unocc}}$ . Thus, we can reproduce the conventional  $\phi_{\text{CNL}}$  as  $E_{\text{VB}} + E_G * D_{\text{VB}} / (D_{\text{VB}} + D_{\text{CB}})$  (Cardona *et al.* 1987), which is the quantity intrinsic to a bulk semiconductor. This is why we call our  $\phi_{\text{CNL}}^G$  the generalized charge-neutrality level.

Next, we show how the new  $\phi_{\text{CNL}}^G$  theory explains the unusual behavior of work functions at metal/high- $k$  interfaces. As representative interfaces made of n and p metals, Al/ $\text{HfO}_2$ (110) and Ni/ $\text{HfO}_2$ (110) interfaces were investigated by the first-principles calculations. Figures 1.42(a) and (b) show the stable positions of Al and Ni atoms on the  $\text{HfO}_2$  substrate, respectively. It has been found that Al atoms prefer to locate on oxygen atoms and produce the connection only with oxygens. This occurs because the electronegativity of Al is small and thus the reactivity with oxygen becomes high. As a result, one can reasonably approximate  $t_{\text{M}-\text{Hf}} \ll t_{\text{M}-\text{O}}$ . Meanwhile the variation of Al density of states is monotonously continuous and featureless, as shown in Fig. 1.40(a), indicating  $D_{\text{occ}} = D_{\text{unocc}}$ . Therefore, the charge transfer is expected to occur from  $\text{HfO}_2$  to Al as schematically shown in Fig. 1.43(a). As a result, the  $\phi_{\text{CNL}}^G$  approaches the bottom of  $\text{HfO}_2$  conduction bands,  $E_{\text{CB}}$ , and the effective WF of Al decreases as shown in Fig. 1.38(c). With respect to Ni/ $\text{HfO}_2$  interface, Ni atom is located between Hf and O atoms on the  $\text{HfO}_2$  surface, thus  $t_{\text{M}-\text{Hf}} = t_{\text{M}-\text{O}}$  as expected. However, similar to Au in Fig. 1.40(b), the Ni density of states below the Fermi energy is much larger than that above the Fermi energy. Thus, we can approximate  $D_{\text{occ}} \gg D_{\text{unocc}}$ . Therefore, the charge transfer is expected to occur mainly from Ni to  $\text{HfO}_2$  as shown in Fig. 1.43(b),  $\phi_{\text{CNL}}^G$  approaches the top of  $\text{HfO}_2$  valence bands,  $E_{\text{VB}}$ , and the effective WF of Ni increases as shown in Fig. 1.38(c).

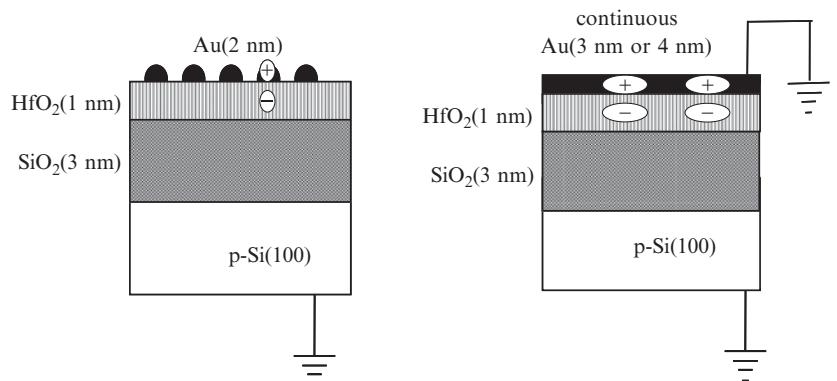
Now, we comment on the physical mechanism of the electron transfer across the interfaces. It should be noticed that, in spite of the large electronegativities of O and Au atoms, respectively, compared to Al and Hf, the charge transfer occurs from O to Al at the Al/ $\text{HfO}_2$  interface and from Au to Hf at the



**Fig. 1.43** Schematic views of the physical mechanism of electron transfer at metal/HfO<sub>2</sub> interfaces given with the interface structures. (a) Al/HfO<sub>2</sub> interfaces, (b) Ni/HfO<sub>2</sub> interfaces.

Au/HfO<sub>2</sub> interface. Roughly speaking, this is because O and Hf atoms are fully ionized in HfO<sub>2</sub> as O<sup>2-</sup> and Hf<sup>4+</sup>. Namely, O and Hf are not atoms but elements in bulk HfO<sub>2</sub>. Thus, there is no space to receive additional electrons around O and holes around Hf. Instead, once the Al/HfO<sub>2</sub> and Au/HfO<sub>2</sub> interfaces are grown and the Al–O and Au–Hf connections are produced, O and Au atoms partially present electrons to Al and Hf atoms, respectively. The schematic illustration of this mechanism is summarized in Fig. 1.43.

The above considerations based on  $\phi^G_{\text{CNL}}$  theory have directly been confirmed by photoemission spectroscopy by Miyazaki and colleagues (Shiraishi *et al.* 2005). Their technique called the “metal nanodot technique” is elegant. The basic principle of the “metal nanodot technique” is schematically illustrated in Fig. 1.44. The absolute value of the dipole at the interface between an electrically isolated metal nanodot and an insulator is essentially smaller than that dipole at the interface between an electrically isolated metal nanodot and an insulator is essentially smaller than that at the interface between a continuous metal and an insulator, although the interface dipole directions are the same. Thus, by comparing the core-level shifts of these two interfaces, we can determine the interface dipole direction at a metal/insulator interface.



**Fig. 1.44** Schematic illustration of the basic principle of the metal nanodot technique (Shiraishi *et al.* 2005).

Figure 1.45 shows the obtained Au-4f core level shifts of a Au nanodot and continuous Au. As shown in Fig. 1.45, the measured Au 4f binding energies of continuous films are larger than those of electrically isolated nanodots. These results directly show the electron transfer from Au to HfO<sub>2</sub> at a Au/HfO<sub>2</sub> interface is in good agreement with the above-mentioned C-V measurements (Koyama *et al.* 2004; Shiraishi *et al.* 2005). Accordingly, these results coincide with the  $\phi_{\text{CNL}}^G$  theory prediction, and also indicate the validity of  $\phi_{\text{CNL}}^G$  theory.

In this way, material properties intrinsic to both bulks and interfaces are essential to understand electronic structures of metal/high- $k$  interfaces. Looking at the  $\phi_{\text{CNL}}^G$  formula in eqn (1.14) once more,  $D_{\text{occ}}$  and  $D_{\text{unocc}}$  reflect the number of valence electrons in metals, while  $D_{\text{VB}}$  and  $D_{\text{CB}}$  reflect the number of cation and anion atoms in high- $k$  materials. On the other hand,  $t_{\text{M}-\text{Hf}}$  and  $t_{\text{M}-\text{O}}$  reflect the kind and strength of interface bonds, depending on the growth process. By designing these parameters, one can obtain desirable gate-metal offsets.

### 1.3.2.2.7 Universality of generalized charge-neutrality level

The present new concept of  $\phi_{\text{CNL}}^G$  explains not only the Schottky barriers at ordinary metal/semiconductor and unusual metal/high- $k$ -dielectrics interfaces but also the band offsets, i.e. band-edge discontinuity, at a variety of semiconductor/semiconductor interfaces by the same concept (Nakayama 1993). In the case of a semiconductor/semiconductor (S/S') interface, we have to start with the interface energy diagram in Fig. 1.41(b) instead of that in Fig. 1.41(a), where the notations are similar to those in Fig. 1.41(a) (Nakayama *et al.* 2006). When the charge transfer occurs from S' to S and the dipole potential increases the energy levels in S by  $\Delta V$ , the charge transfer from S' to S is given as

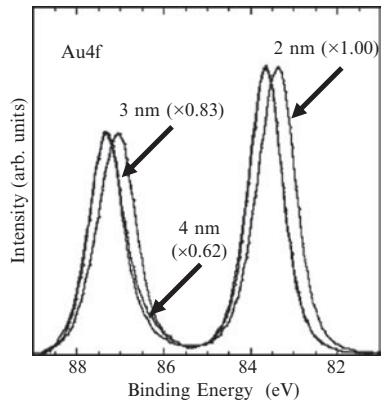
$$\Delta\rho(S' \rightarrow S) \propto \frac{|t_{\text{CB}-\text{VB}'}|^2 D_{\text{CB}} D'_{\text{VB}}}{E_{\text{CB}} + \Delta V - E'_{\text{VB}}} - \frac{|t_{\text{VB}-\text{CB}'}|^2 D_{\text{VB}} D'_{\text{CB}}}{E'_{\text{CB}} - E_{\text{VB}} - \Delta V}. \quad (1.16)$$

The band alignment is determined by equalizing this equation to zero. By simple arithmetic, we obtain

$$\Delta V + E_{\text{VB}} + E_{\text{G}} \frac{B}{A+B} = E'_{\text{VB}} + E'_{\text{G}} \frac{A}{A+B}, \quad (1.17)$$

where  $A = |t_{\text{CB}-\text{VB}'}|^2 D_{\text{CB}} D'_{\text{VB}}$  and  $B = |t_{\text{VB}-\text{CB}'}|^2 D_{\text{VB}} D'_{\text{CB}}$  are effective couplings at the interface. This equation represents the balance of two charge-neutrality levels. In fact, when the counterpart (S') of the interface is changed into metals such as  $A = |t_{\text{CB}-\text{M}}|^2 D_{\text{CB}} D_{\text{occ}}$  and  $B = |t_{\text{VB}-\text{M}}|^2 D_{\text{VB}} D_{\text{unocc}}$ , the third term of the left-hand side becomes  $|t_{\text{VB}-\text{M}}|^2 D_{\text{VB}} D_{\text{unocc}} / (|t_{\text{VB}-\text{M}}|^2 D_{\text{VB}} D_{\text{unocc}} + |t_{\text{CB}-\text{M}}|^2 D_{\text{CB}} D_{\text{occ}})$  and the sum of the second and third terms corresponds to  $\phi_{\text{CNL}}^G(S)$ . Similarly, when S of the interface is changed into metals such as  $A = |t_{\text{M}-\text{VB}'}|^2 D_{\text{unocc}} D'_{\text{VB}}$  and  $B = |t_{\text{M}-\text{CB}'}|^2 D_{\text{occ}} D'_{\text{CB}}$ , the sum of the second and third terms in the right-hand side becomes  $\phi_{\text{CNL}}^G(S')$ .

The present new concept of  $\phi_{\text{CNL}}^G$  is constructed using the effective-four-level tight-binding model of the interface. We used the energies,  $E_F$ ,  $E_{\text{VB}}$  and  $E_{\text{CB}}$ , and the densities of state,  $D_{\text{occ}}$ ,  $D_{\text{unocc}}$ ,  $D_{\text{VB}}$  and  $D_{\text{CB}}$ , to represent



**Fig. 1.45** Au 4f spectra taken from the Au/HfO<sub>2</sub>/SiO<sub>2</sub>/Si gate stacks with varying Au dot sizes (2–4 nm). The binding energy was calibrated with the Hf 4f 7/2 peak at 16.7 eV and no energy shift in O 1s was detected (Shiraishi *et al.* 2005).

properties intrinsic to bulk materials (Nakayama 1993). These quantities originate from bulk *band* structures. On the other hand, to describe the microscopic atomic structures of interfaces, the transfer energies,  $t_{M-Hf}$  and  $t_{M-O}$ , are introduced. These reflect the interface *bonds*. The present theory indicates that both band and bond pictures, which, respectively, correspond to the itinerant and localized characters of electrons, are necessary to describe electronic structures at the interface.

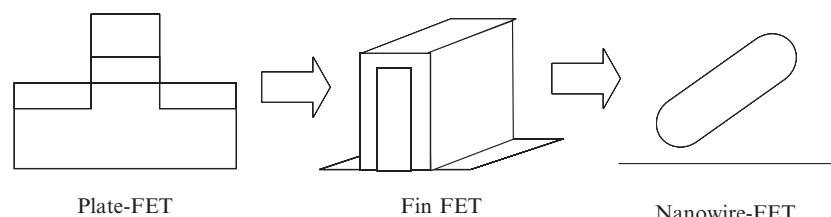
### 1.3.3 Channel engineering

In this section, we discuss the possibility of Si nanowire (SiNW) channels based on the knowledge obtained by the computational sciences as an example of computational-science-guided channel engineering. For the continuous development of LSI devices, the off-current increase becomes a serious problem because of the downsizing of the gate length. Thus, suppression of the off-current is one of the most inevitable problems. To reduce the off-current of FET, nanowire structure channels are desirable based on the electrostatic potential analysis. As a result, the FET channel structure will be changed from plate-FET, fin-FET and nanowire-FET as shown in Fig. 1.46.

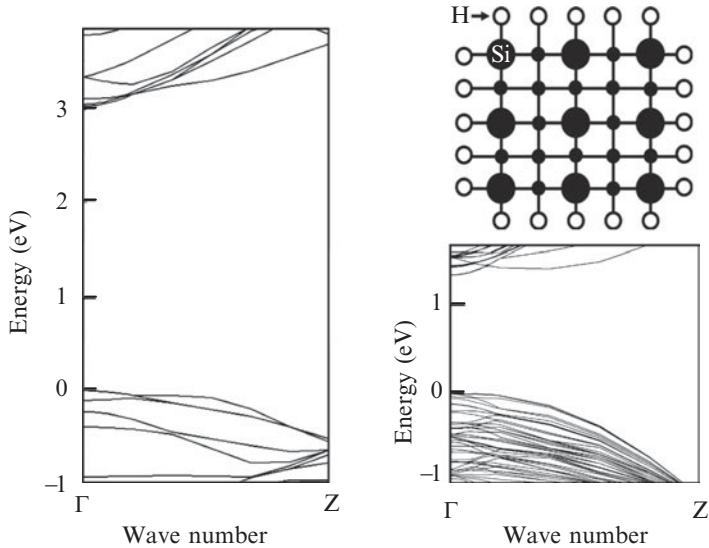
As discussed above, nanowire-shaped Si is a promising candidate for future LSI channel elements. However, the total current required for driving the circuit would be limited due to its narrower cross-section. One way to increase the transport current (on-current) at small cross-section is to achieve the ideal 1-dimensional transport limit (Landauer limit).

Here, the electronic structures of SiNW with different sizes derived by first-principles calculation are shown and a tradeoff model for  $I_{on}$  is proposed for SiNW MOSFET (Lee *et al.* 2008). The band structures of SiNWs with 0.77 and 2.30 nm are shown in Fig. 1.47. SiNWs with these dimensions have direct  $E_g$  at the gamma point, which are in contrast to the indirect band structure of bulk Si, reproducing the reported first principles calculations (Ohno *et al.* 1992).

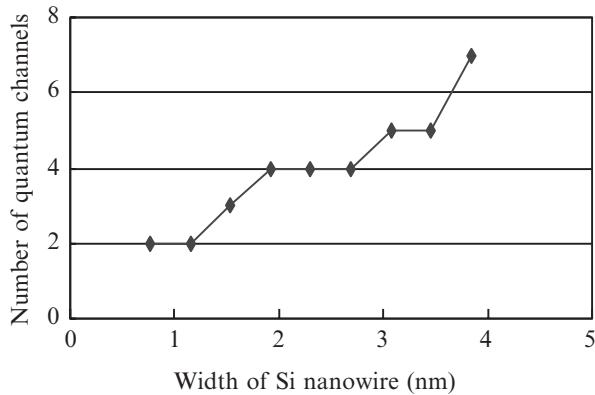
When one-dimensional ballistic conduction is achieved, the conductivity of a SiNW MOSFET is basically determined by the number of quantum channels near the conduction-band minimum (CBM) and valence-band maximum (VBM). Here, the numbers of quantum channels within 50 meV from the valence-band top (number of hole channels) are plotted in Fig. 1.46. The number of hole channels increases from 2 to 7, as the SiNW size increases from 0.77 nm to 3.84 nm. Therefore, larger SiNW size can achieve higher conductivity from the viewpoint of the number of quantum channels. However, as the



**Fig. 1.46** The concept of device scaling for future MOSFET.



**Fig. 1.47** The band structure of Si nanowires. The upper-right figure shows a cross-sectional image of the smallest Si nanowires. The band structure of Si nanowires. The upper-right figure shows a cross-sectional image of the smallest Si nanowires.



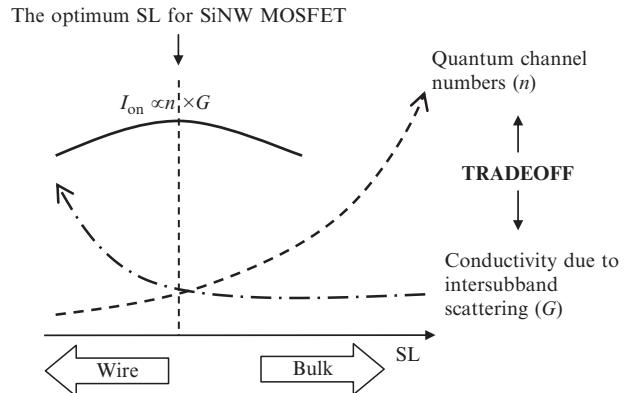
**Fig. 1.48** The number of quantum channels plotted as a function of Si nanowire width. The numbers of quantum channels within 50 meV from the valence-band top (number of hole channels) is given.

density of bands increases, the spacing of each band becomes narrow, which will allow intersubband scattering and eventually reduces the conductivity (Sakaki 1980). In the light of the above discussions, it may be concluded that there exists a tradeoff between the quantum channel numbers and intersubband scattering and an optimum size exists for MOSFET application (Fig. 1.49).

As discussed above, computational science gives an estimate of the number of quantum channels for different thickness in SiNWs. A tradeoff model for  $I_{on}$  is proposed for MOSFET with SiNW channels (Lee *et al.* 2008).

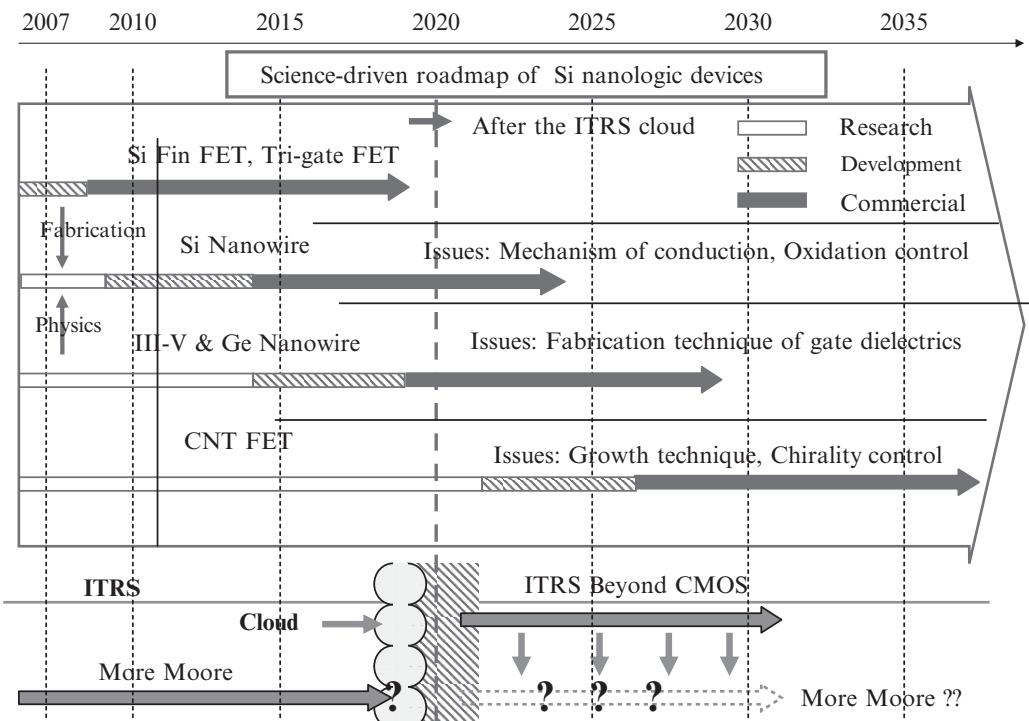
## 1.4 Future Si technology trend predicted by computational science

In this section, we show the future Si technology trend based on the discussion in this chapter. In the present ITRS road map, a cloud is described



**Fig. 1.49** The tradeoff model between the quantum channel numbers ( $n$ ) and the conductivity ( $G$ ) with intersubband scattering. The arrows of the  $n$  and  $G$  curves point in a direction to obtain a high  $I_{on}$ . An optimum Si nanowire size should exist (Lee *et al.* 2008).

after 2020 as shown in the lower part of Fig. 1.50 (Iwai 2008). This corresponds to the period at which gate length shall become shorter than the Si mean-free path (5–10 nm). In this case, downscaling results in neither high performance nor low per consumption. This is called the “end of scaling”. However, even after the “end of scaling”, we should continue to increase the quality of Si nanodevices. In this sense, Si nanotechnology should progress in a



**Fig. 1.50** Schematic illustration of science-driven roadmap of Si nanologic devices (Iwai 2008).

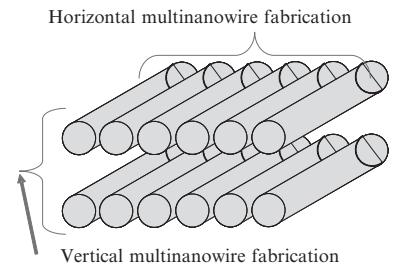
science-driven manner instead of the present scaling-driven manner. To achieve science-driven Si nanotechnologies, computational science plays a crucial role.

Now, we will show a part of the science-driven roadmap of future Si nanodevices summarized in Fig. 1.48. As discussed in Section 3.3, future channel elements of Si nanologic devices will be nanowire-like structure due to the requirement of suppressing off-current. However, nanowire structures have difficulty in obtaining higher on-current, since each nanowire channel can carry a small current density; If the number of quantum channels is one, the maximum conductance is  $G_0 = (h/2e) = 77.8 \mu\text{S}$ . In the near future (around 2020),  $3 \text{ mA}/\mu\text{m}$ . is required at the supply voltage of 0.6 V. Therefore, multinanowires should be used as channels in future LSIs as shown in Fig. 1.51. Further, it is noted that at least  $\sim 65$  multinanowires should be fabricated in the region of  $1 \mu\text{m}$  to meet the criterion of the Si roadmap even when the ideal conductance  $G_0$  is achieved.

To achieve the above criterion, there are several factors that should be overcome:

- (1) achievement of ideal one-dimensional conductance  $G_0$ ;
- (2) increase in the number of nanowires by the fabrication techniques (Fig. 1.51);
- (3) increase in the number of quantum channels of each nanowire.

There are several candidates for nanowire FETs. Si nanowires, III-V nanowires, including GaAs nanowires (for n-FET), Ge nanowires (for p-FET) and carbon nanotubes are typical promising candidates. However, these materials have both merits and demerits. We discuss the characteristics of each channel material as follows.



**Fig. 1.51** Schematic illustration of multinanowire fabrication method.

### 1.4.1 Si nanowire

A Si nanowire has one great merit: It can be fabricated by top-down fabrication based on the lithography techniques. First, we prepare larger nanowires by lithography. Next, we oxidize them in order to make thinner nanowires. It is well known that thermally oxidized Si/SiO<sub>2</sub> interfaces have excellent quality for nanodevices. Top-down lithography insures the controllability of the positions and numbers of nanowires. Further, thermal oxidation insures the excellent interface quality between the channel nanowire and gate dielectrics. However, there are also demerits. The number of Si nanowire channels is limited by the resolution of lithography. This is a drawback of top-down technologies. Another demerit is that it is difficult for Si nanowires to achieve ideal one-dimensional conductivity because of the relatively smaller effective Bohr radius and relatively shorter mean-free path compared with GaAs. In other words, thinner nanowires are necessary for one-dimensional conduction.

### **1.4.2 GaAs nanoWire**

GaAs nanowires for n-FET have merits. One is that ideal one-dimensional conductivity can be obtained even when the nanowire size is large because of the large effective Bohr radius of 10 nm and longer mean-free path of electrons. However, it has difficulty also in making corresponding dielectrics. The excellent quality of GaAs/insulator interfaces with very thin insulator thickness should be necessary for realizing a GaAs nanowire FET. The fabrication of an excellent quality GaAs/insulator interface is the most important issue of GaAs nanowire devices. This problem is common to other nanowire FETs such as InSb n-FET, InAs n-FET, and Ge p-FET.

### **1.4.3 Carbon nanotubes**

Carbon nanotubes have several merits for nanowire FETs. In principle, they do not have interface roughness, which is inevitable for semiconductor nanowires, since they are an ideal nanowire material. Thus, achievement of ideal one-dimensional conductivity is much easier than in other semiconductor nanowires. This is the crucial merit of carbon nanotubes. However, they have serious demerits.

One is the fabrication technique. It is well known that carbon nanotubes can be categorized into metal and semiconductor nanotubes, depending on their chiralities. For FETs, semiconductor nanotubes should be used. To achieve this requirement, highly selective growth of semiconductor nanotubes should be realized. Usually, the bandgap of carbon nanotubes is around 0.5 eV. Therefore, the supply voltage will be limited to less than 0.5 V. Carbon nanotubes always have two quantum channels due to the degeneracy of their band structures. Thus, at least 40 multicarbon nanotubes should be fabricated in the region of 1  $\mu\text{m}$  by the bottom-up spontaneous growth techniques. Moreover, it should be noted that in real LSIs, conventional Si CMOS will be included in the low-performance regions, although carbon-nanotube channels are used in the high-performance regions. At present, carbon-nanotube growth is done after electrode formation at about 700  $^{\circ}\text{C}$ . This high-temperature carbon-nanotube growth can deteriorate the Si MOS region, which usually suffer from a 400  $^{\circ}\text{C}$  forming-gas anneal. As mentioned above, precise chirality and position control and low-temperature growth of is urgently required for carbon-nanotube FETs.

### **1.4.4 Future trend of si nanotechnologies**

As discussed above, a nanowire-shaped channel is highly desired to suppress the off-current. At present, there are many candidates of nanowire channel materials. In the near future, many materials will be discarded due to technical or intrinsic limitations, as shown in Fig. 1.50. In Fig. 1.50, the roadmap after the “cloud” (after 2020) is clearly shown, by considering the above discussion.

## 1.5 Summary

We have shown several examples where computational science can give crucial knowledge to recent Si nanotechnologies and nanodevices. This is because present Si nanotechnology has faced the situation where science-driven innovations can only improve the quality of devices. For example, present Si nanotechnology requires the development of new materials, leading to the necessity of acquiring new material concepts. In such circumstances, computational science has emerged as one of the most powerful approaches to meet the challenge. In fact, an enormous knowledge gained through the computational science has yielded valuable insights into the Si nanotechnologies including devices. At the same time, it should be noted that recent science-driven Si nanotechnologies have proved unique for the growth of nanoscience owing to their amazing capabilities of controlling the size and shape of various nanomaterials and nanointerfaces on the atomic scale and have played a pivotal role in deciding the frontiers and advances in nanoscience and technology.

## Acknowledgments

This work is our researches related to the interface physics of semiconductor and insulator interfaces in “real” Si devices over more than ten years. Researches of Si oxidation could not have been performed without the long-term collaboration with Dr. Hiroyuki Kageshima and Dr. Masashi Uematsu. Researches related to the high- $k$  dielectrics have arisen from the research project of “high- $k$  net” conducted by Prof. Keisaku Yamada. We could not have performed this study without the collaboration with the members of the “high- $k$  net” project. In particular, Prof. Seiichi Miyazaki, Prof. Heiji Watanabe, Dr. Yasuo Nara, Dr. Tsunetoshi Arikado, Dr. Yasushi Akasaka, Dr. Kazuyoshi Torii, Dr. Naoto Umezawa, Prof. Kikuo Yamabe and Dr. Toyohiro Chikyow have given us a lot of experimental and calculation data. During construction of the science-driven roadmap of Si nanotechnologies, we are indebted to the collaboration with Prof. Hiroshi Iwai, Prof. Kenji Natori and Prof. Kuniyuki Kakushima. We are grateful for financial support over the years from the Semiconductor Leading Edge Technology Inc. (Selete) through the “high- $k$  net” project.

## References

- Akasaka, Y., Nakamura, G., Shiraishi, K., Umezawa, N., Yamabe, K., Ogawa, O., Lee, M.B., Amiaka, T., Kasuya, T., Watanabe, H., Chikyow, T., Ootsuka, F., Nara, Y., Nakamura, K. *Jpn. J. Appl. Phys.* **45**, L1289 (2006).
- Broqvist, P., Alkauskas, A., Pasquarello, A. *Appl. Phys. Lett.* **92**, 132911 (2008).
- Car, R., Kelly, P.J., Oshiyama, A., Pantelides S.T. *Phys. Rev. Lett.* **52**, 1814 (1984).
- Cardona, M., Christensen, N.E. *Phys. Rev. B* **35**, 6182 (1987).
- Chadi, D.J. *Phys. Rev. Lett.* **77**, 861 (1996).

- Cartier, E., Narayanan, V., Gusev, E.P., Jamison, P., Linder, B., Steen, M., Chan, K.K., Frank, M., Bojarczuk, N., Copel, M., Cohen, S.A., Callegari, A., Zafar, S., Gribelyuk, M., Chudzik, M., Cabral, Jr. C., Carruthers, R.A., D'Emic, C., Newbury, J., Lacey, D., Guha, S., Jammy, R., Technical Digest of 2004 Symposium on VLSI Technology, Honolulu, USA, June, p. 44 (2004).
- Datta, S., Assad, F., Lundstrom, M. *Superlatt. Microstruct.* **23**, 771 (1997).
- Deal, B.E., Grove, A.S. *J. Appl. Phys.* **36**, 3770 (1965).
- Deal, B.E. *The Physics and Chemistry of SiO<sub>2</sub> and the Si–SiO<sub>2</sub> Interface*, (eds) Helms, C.R. and Deal, B.E. (Plenum Press, New York, USA, 1988), p. 5.
- Fukatsu, S., Takahashi, T., Itoh, K.M., Uematsu, M., Fujiwara, A., Kageshima, H., Takahashi, Y., Shiraishi, K., Gösele, U. *Appl. Phys. Lett.* **83**, 3897 (2003).
- Foster, A.S., Gejo, F.L., Shluger, A.L., Nieminen, R.M. *Phys. Rev. B* **65**, 174117 (2002).
- Gavartin, J.L., Schluger, A.L. *Microelectron. Eng.* **84**, 2412 (2007).
- Gibson, J.M., Lanzerotti, M.Y. *Nature* **340**, 128 (1989).
- Hakala, M.H., Foster, A.S., Gavartin, J.L., Havu, P., Puska, M.J., Nieminen, R.M. *J. Appl. Phys.* **100**, 043708 (2006).
- Hans, C.J., Helms, C.R. *J. Electrochem. Soc.* **135**, 1824 (1988).
- Hasegawa, H., Ohno, H. *J. Vac. Sci. Technol. B* **4**, 1130 (1986).
- Heine, V. *Phys. Rev.* **138**, A 1689 (1965).
- Hobbs, C., Fonseca, L., Dhandapani, D., Samavedam, S., Taylor, B., Grant, J., Dip, L., Triyoso, D., Hegde, R., Gilmer, D., Garcia, R., Roan, D., Lovejoy, L., Rai, R., Hebert, L., Tseng, H., White, B., Tobin, P. Technical Digest of 2003 Symposium on VLSI Technology, Kyoto, Japan, June, p. 9 (2003).
- Hu, S.M. *Appl. Phys. Lett.* **27**, 165 (1975).
- Ikarashi, N., Watanabe, K., Miyamoto, Y. *Phys. Rev. B* **62**, 15989 (2000).
- Ikeda, M., Kresse, G., Kadoshima, M., Nabatame, T., Satake, H., Toriumi, A. Extended Abstracts of the 2006 Conference on Solid State Device and Materials, Yokohama, Japan, September, p. 222 (2006).
- Inoue, M., et al. Technical Digest of IEEE International Electron Devices Meeting, Washington D.C., USA, December, p. 425 (2005).
- Iwai, H. Extend Abstract of International Workshop of Junction Technology, 2008, Shanghai, China, May, p. 1 (2008).
- Iwamoto, K., et al. *Appl. Phys. Lett.* **92**, 132907 (2008).
- Kadoshima, M., Sugita, Y., Shiraishi, K., Watanabe, H., Ohta, A., Miyazaki, S., Nakajima, K., Chikyow, T., Yamada, K., Aminaka, T., Kurosawa, E., Matsuki, T., Aoyama, T., Nara, Y., Ohji, Y. Technical Digest of 2007 Symposium on VLSI Technologies, Kyoto, Japan, June, p. 66 (2007).
- Kageshima, H., Shiraishi, K. *Phys. Rev. Lett.* **81**, 5936 (1998).
- Kageshima, H., Shiraishi, K., Uematsu, M. *Jpn. J. Appl. Phys.* **38**, L971 (1999).
- Kageshima, H., Uematsu, M., Akiyama, T., Ito, T. *ECS Trans.* **6**, 449 (2007).
- Kamimuta, Y., Koyama, M., Ino, T., Sekine, K., Sato, M., Eguchi, K., Takayanagi, M., Tomita, M., Nishiyama, A. Extended Abstracts of the 2005 Conference on Solid State Device and Materials, Kobe, Japan, September, p. 24 (2005).
- Kato, K., Uda, T., Terakura, K. *Phys. Rev. Lett.* **80**, 2000 (1998).

- Kameda, T., Namba, K., Nishioka, Y. *Jpn. J. Appl. Phys.* **37**, L214 (1998).
- Koyama, M., Kamimuta, Y., Ino, T., Kaneko, A., Inumiya, S., Eguchi, K., Takayanagi, M., Nishiyama, A. Technical Digest of IEEE International Electron Devices Meeting, San Francisco, USA, December, p. 499. (2004).
- Lee, B.H., Oh, J., Tseng, H.H., Jammy, R., Huff, H. *Mater. Today* **32**, (2006).
- Lee, Y., Nagata, T., Kakushima, K., Shiraishi, K., Iwai, H. Extend (ed.) Abstracts of the 2008 International Workshop on Dielectric Thin Films for Future ULSI Devices: Science and Technology, November, Tokyo, Japan, p. 83. (2008).
- Louie, S.G., Cohen, M.L. *Phys. Rev. B* **13**, 2461 (1976).
- Lu, H.C., Gustafsson, T., Gusev, E.P., Garfunkel, E. *Appl. Phys. Lett.* **67**, 1742 (1995).
- Massoud, H.Z., Plummer, J.P., Irene, E.A. *J. Electrochem. Soc.* **132**, 2685.
- Matsudo, T., Ohta, T., Yasuda, T., Nishizawa, M., Miyata, N., Yamasaki, S., Shklyaev, A., Ichikawa, M. *J. Appl. Phys.* **91**, 3637 (2002).
- Ming, Z., Nakajima, K., Suzuki, M., Kimura, K., Uematsu, M., Torii, K., Kamiyama, S., Nara, Y., Yamada, K. *Appl. Phys. Lett.* **88**, 153516 (2006).
- Mizuo, S., Higuchi, H. *Jpn. J. Appl. Phys.* **20**, 739 (1982).
- Murayama, M., Shiraishi, K., Nakayama, T. *Jpn. J. Appl. Phys.* **37**, 4109 (1998).
- Nakayama, T. *Physica B* **191**, 16 (1993).
- Nakayama, T., Murayama, M., *Jpn. J. Appl. Phys.* **38**, 3497 (1999).
- Nakayama, T., Murayama, M., *Appl. Phys. Lett.* **77**, 4286 (2000).
- Nakayama, T., Shiraishi, K., Miyazaki, S., Akasaka, Y., Nakaoka, T., Torii, K., Ohta, A., Ahmet, P., Ohmori, K., Umezawa, N., Watanabe, H., Chikyow, T., Nara, Y., Iwai, H., Yamada, K. *ECS Trans.* **3**, 129 (2006).
- Natori, K. *J. Appl. Phys.* **76**, 4879 (1994).
- Northrup, J.E. *Phys. Rev. B* **44**, 1419 (1991).
- Ohishi, K., Hattori, T. *Jpn. J. Appl. Phys.* **33**, L675 (1994).
- Ohno, T., Shiraishi, K., Ogawa, T. *Phys. Rev. Lett.* **69**, 2400 (1992).
- Ohta, A., Miyazaki, S., Akasaka, Y., Watanabe, H., Shiraishi, K., Yamada, K., Inumiya, S.Y. Nara, Extended Abstracts of 2006 International Workshop on Dielectric Thin Films for Future ULSI Devices—Science and Technology, Kawasaki, Japan, November, p. 61. (2006).
- Ono, Y., Tabe, M., Kageshima, H. *Phys. Rev. B* **48**, 14 291 (1993).
- Oshiyama, A. *Jpn. J. Appl. Phys.* **37**, L232 (1998).
- Otani, M., Shiraishi, K., Oshiyama, A., *Phys. Rev. B* **68**, 184112 (2003).
- Pajot, B. *Oxygen in silicon*, Semiconductors and Semimetals Series, vol. 42, (ed.) F. Shimura (Academic Press, San Diego, USA, 1994) p. 191.
- Peacock, P.W., Robertson, J. *Phys. Rev. Lett.* **92**, 057601 (2004).
- Ravi, K.V.C.J., Varker, C.J. *J. Appl. Phys.* **45**, 263 (1974).
- Robertson, J. *Eur. Phys. J. Appl. Phys.* **28**, 265 (2004).
- Sakaki, H. *Jpn. J. Appl. Phys.* **19**, L735 (1980).
- Scopel, W.L., da Silva, A.J.R., Orellana, W., Fazzio, A. *Appl. Phys. Lett.* **84**, 1492 (2004).
- Shiraishi, K., Kageshima, H., Uematsu, M. *Jpn. J. Appl. Phys.* **39**, L1263 (2000).

- Shiraishi, K., Yamada, K., Torii, K., Akasaka, Y., Nakajima, K., Konno, M., Chikyo, T., Kitajima, H., Arikado, T. *Jpn. J. Appl. Phys.* **43**, L1413 (2004).
- Shiraishi, K., Akasaka, Y., Miyazaki, S., Nakayama, T., Nakaoka, T., Nakamura, G., Torii, K., Furutou, H., Ohta, A., Ahmet, P., Ohmori, K., Watanabe, H., Chikyow, T., Green, M.L., Nara, Y., Yamada, K. Technical Digest of IEEE International Electron Devices Meeting, Washington D.C., USA, December, (2005) p. 43.
- Shiraishi, K., Yamada, K., Torii, K., Akasaka, Y., Nakajima, K., Konno, M., Chikyo, T., Kitajima, H., Arikado, T., Nara, Y. *Thin Solid Films* **508**, 305 (2006).
- Spicer, W. E., Lindau, I., Skeath, P., Su, C.Y. *J. Vac. Sci. Technol.* **17**, 1019 (1980).
- Takayanagi, M., Watanabe, T., Iijima, R., Kaneko, A., Inumiya, S., Hirano, I., Sekine, K., Nishiyama, A., Eguchi, K., Tsunashima, Y., Extended Abstracts of International Workshop on Gate Insulator, Tokyo, Japan, November (2003) p. 174.
- Takeuchi, H., Ha, D., King, T.J. *J. Vac. Sci. Technol. A* **22**, 1337 (2004).
- Tan, T.Y., Gösele, U. *Appl. Phys. A* **37**, 1 (1985).
- Tersoff, J. *Phys. Rev. Lett.* **52**, 465 (1984).
- Thomas, D.J.D. *Phys. Status Solidi* **3**, 2261 (1963).
- Torii, K., Shiraishi, K., Miyazaki, S., Yamabe, K., Boero, M., Chikyow, T., Yamada, K., Kitajima, H., Arikado, T. Technical Digest of 2004 IEEE International Electron Device Meeting, San Francisco, USA, December, p. 129 (2004).
- Tu, Y., Tersoff, J. *Phys. Rev. Lett.* **89**, 086102 (2002).
- Uematsu, M., Kageshima, H., Shiraishi, K. *Jpn. J. Appl. Phys.* **39**, L699 (2000).
- Uematsu, M., Kageshima, H., Shiraishi, K. *J. Appl. Phys.* **89**, 1948 (2001).
- Uematsu, M., Kageshima, H., Takahashi, Y., Fukatsu, S., Itoh, K.M., Shiraishi, K. *Appl. Phys. Lett.* **85**, 221 (2004).
- Uematsu, M., Kageshima, H., Fukatsu, S., Itoh, K.M., Shiraishi, K., Otani, M., Oshiyama, A. *Thin Solid Films* **508**, 270 (2006).
- Umezawa, N., Shiraishi, K., Ohno, T., Watanabe, H., Chikyow, T., Torii, K., Yamabe, K., Yamada, K., Kitajima, H., Arikado, T. *Appl. Phys. Lett.* **86**, 143507 (2005).
- Watanabe, H., Kato, K., Uda, T., Fujita, K., Ichikawa, M., Kawamura, T., Terakura, K. *Phys. Rev. Lett.* **80**, 345 (1998).
- Watanabe, T., Tatsumura, K., Ohdomari, I. *Appl. Surf. Sci.* **237**, 125 (2004).
- Watanabe, T., Tatsumura, K., Ohdomari, I. *Phys. Rev. Lett.* **96**, 196102 (2006).
- Watanabe, T., Ohdomari, I. *J. Electrochem. Soc.* **154**, G270 (2007).
- Wilk, G.D., Wallace, R.M. *Appl. Phys. Lett.* **74**, 2854 (1999).
- Wilk, G.D., Wallace, R.M., Anthony, J.M. *J. Appl. Phys.* **89**, 5243 (2001).
- Wyckoff, R.W.G. *Crystal Structure* (Interscience, New York, USA, 1963).
- Yamada, K. Extended Abstracts of 1986 International Conference on Solid State Devices and Materials, Tokyo, Japan, August, p. 257 (1986).
- Yasuda, T., Yamasaki, S., Nishizawa, M., Miyata, N., Shklyaev, A., Ichikawa, M., Matsudo, T., Ohta, T. *Phys. Rev. Lett.* **87**, 037403 (2001).
- Yasuda, T., Kumagai, N., Nishizawa, M., Yamasaki, S., Oheda, H., Yamabe, K. *Phys. Rev. B* **67**, 195338 (2003).
- Zhu, Z., Shima, N., Tsukada, M. *Phys. Rev. B* **40**, 11868 (1989).

# Few-electron quantum-dot spintronics

D.V. Melnikov, J. Kim, L.-X. Zhang, and J.-P. Leburton

2

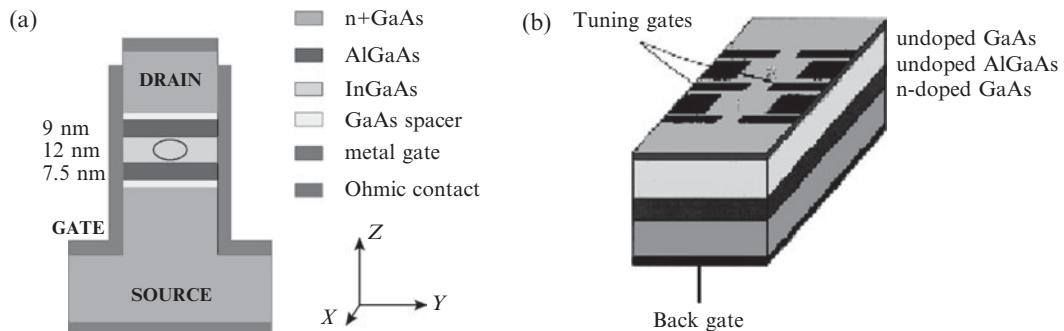
## 2.1 Introduction and motivations

In the last decade semiconductor quantum dots (QDs) have attracted much attention (Ashoori 1992; Kastner 1993; Goldhaber-Gordon 1998; Ciorga *et al.* 2000; Reimann and Manninen 2002; van der Wiel *et al.* 2002) because these nanostructures are promising candidates for novel logic and optoelectronic devices (Likharev 1999). They also exhibit many features characteristic of real atomic (molecular) systems and as such, are frequently called artificial atoms (molecules). Among several types of QDs, the so-called “gated” quantum dots in which controlled single-electron charging and atomic-like shell structure were first observed (Tarucha *et al.* 1996) are of special interest. In these man-made systems conduction electrons are held together in a finite region of space by the quantum-confinement potential that is created by the heterostructure barriers and/or the electrostatic potential of remote dopant charge distributions modulated by external gate voltages. The gated QDs usually fall into two different device configurations:

(1) In vertical QDs (Tarucha *et al.* 1996), a quasi-two-dimensional (quasi-2D) electron island (quantum dot) is formed in a mesa-structure pillar between two heterostructure barriers (Fig. 2.1(a)). The size of the dot (or the number of electrons in the QD) can be changed electrostatically by applying a voltage to the gate wrapped around the pillar. The current flows vertically (hence the term “vertical QD”) through the heterostructure in response to a bias applied between the source and drain contacts on top and bottom of the mesa-structure. The direction of the current is thus perpendicular to the QD 2D plane. This QD device has three gates (terminals), which resembles a field effect transistor where the effect of the gate terminal is analogous to the action of the side gate controlling the single-electron current through the heterostructure.

(2) Another popular method of QDs fabrication is to use lithographic patterning of gates, i.e. to deposit metal electrodes on top of a heterostructure (Ciorga *et al.* 2000). By properly biasing these electrodes, the two-dimensional (2D) electron gas formed at the heterointerface between different materials

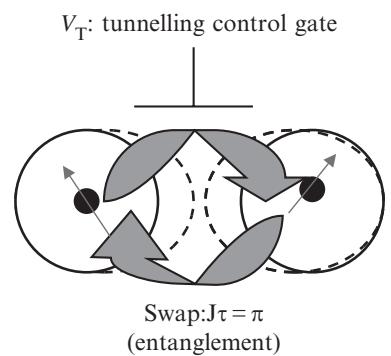
2.1 Introduction and motivations	47
2.2 Two electrons in double quantum dots	50
2.3 Two electrons in quantum wire quantum dots	72
2.4 Few electrons in triple quantum dots	76
2.5 Conclusion	81
Acknowledgments	82
Appendix A: Variational Monte Carlo	82
Appendix B: Heitler–London approach	85
References	85



**Fig. 2.1** Schematic representation of (a) the vertical QD structure and (b) planar QD system.

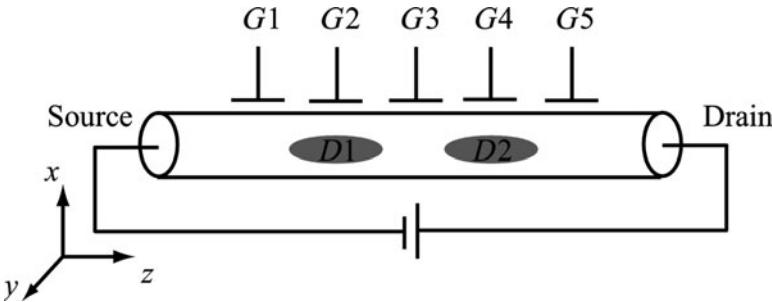
(such as AlGaAs and GaAs) can be depleted, thereby creating an island of non-zero electron density that can be further fine tuned by nearby gates (Ciorga *et al.* 2000; Elzerman *et al.* 2003). This device is called a lateral or planar QD (see Fig. 2.1(b)) because the current flows in the plane of the 2D electron gas (2DEG).

As natural extensions of the concept of artificial atoms, coupled QDs can be compared to molecules where many-body effects can be tailored by electrically varying the coupling barrier or distance between the dots (Waugh *et al.* 1995; Livermore *et al.* 1996; Holleitner *et al.* 2002; Elzerman *et al.* 2003; Chan *et al.* 2004) as well as by detuning the individual QDs with appropriate gates (Hanson *et al.* 2007). The control of individual electron spins and their interaction are fundamental physical ingredients for realizing a quantum logic gate (Loss and DiVincenzo 1998). Fundamental quantum logic operations such as control-NOT (C-NOT) operation can be achieved in coupled QDs by controlling the entanglement between two electron spins by means of external electromagnetic fields (Hatano *et al.* 2005). In this operation, the exchange energy, which is the energy difference between the lowest triplet and the singlet states of the two-electron system, quantifies the entanglement between the spins or qubits (Fig. 2.2). The comprehensive theoretical analysis of this quantity is the focus of the present chapter. Recently, a coherent control cycle of state preparation, interaction and projective read-out has been realized in laterally coupled QDs (Petta *et al.* 2005). In this and other similar experiments (Koppens *et al.* 2007), the precise control of the exchange energy that drives the Rabi oscillations, and hence ascertains the value of the  $\sqrt{\text{SWAP}}$  operation time required for the exchange of information between the two qubits (Loss and DiVincenzo 1998), is mandatory. However, measuring and controlling the exchange energy between two spins still remains a formidable task in experimental coupled QDs. In this context numerical simulations provide a valuable tool to gain insights into the complexity of the basic physical effects at play in the QD, and their manipulation in a device environment.



**Fig. 2.2** Schematic representation of the entanglement between two spins (thin straight arrows) in double-QD system. Gate  $V_T$  controls the interaction (entanglement) between the spins.

Very recently, a new approach to fabricate a system of coupled QDs with controlling gates adjacent to a quantum wire (QW) was proposed (Fasth *et al.* 2005; Fuhrer *et al.* 2007; Hu *et al.* 2007). In these structures (Fig. 2.3),



**Fig. 2.3** Sketch of two coupled QWQDs  $D_1$  and  $D_2$  formed in a quantum wire. Gates  $G_1$  and  $G_5$  define the outer barriers of the two QDs,  $G_3$  controls the interdot coupling,  $G_2$  and  $G_4$  are plungers tuning the confinement in each QD. Charging current flows along the wire from source to drain.

electrons are confined in the plane perpendicular to the axial direction of the wire by its physical dimensions (QW diameters are tens of or even a few nm) while longitudinal confinement along the wire axial direction is maintained by the electrostatic potential barriers achieved by the local controlling gates (Björk *et al.* 2005). The local gate width and separation  $\sim 10\text{--}100$  nm result in smaller effective dot sizes and interdot separations than in 2DEG-based QDs. In these so-called quantum-wire quantum-dot (QWQD) systems, the distance between the controlling gates and the QD region ( $\sim 25$  nm) is also smaller than that in 2DEG-based dots ( $\sim 100$  nm) (Elzerman *et al.* 2003) potentially leading to better local electrostatic control of the charge (spin) states in the coupled dots. Furthermore, QWQD structures offer linear scalability (with the linear grid of the controlling gates) instead of the 2D scalability resulting from top (Elzerman *et al.* 2003; Hanson *et al.* 2007) or side gate patterning (Hatano *et al.* 2005) in 2DEG-based QDs.

Beyond the concept of two coupled QDs, three coupled QDs arranged either in a linear array or in a triangular formation provide extra functionality. Triple QD structures have recently been proposed to work as a solid-state entangler (Saraga and Loss 2003), a charging rectifier (Vidan *et al.* 2004), a coded qubit (Gaudreau *et al.* 2006; Schröer *et al.* 2007), and quantum gates (Sasakura *et al.* 2004) providing incentives to investigate these systems (Kim *et al.* 2006). From a fundamental physical point of view, the triple-QD system presents great interest as an artificial triatomic molecule where the interplay between geometrical confinement, interdot coupling, and many-body effects offer a rich variety of new phenomena, which have not yet been extensively investigated.

In this work we provide theoretical analysis of spin and charge properties of the double and triple QDs populated with just a few electrons. We focus on laterally coupled QDs as they offer most control over the spin states. In Section 2.2 we describe the theoretical approach utilized throughout this work—exact diagonalization method—on the example of the two-electron system in coupled QDs that are modelled as two parabolas (Section 2.2.1). Then we proceed with the discussion of the general characteristics of the two-electron double-QD structure and limitations of the approximate methods frequently used for its theoretical description (Section 2.2.2). Next, we discuss the stability diagram for two circular dots and investigate how its features are affected by the QD elliptical deformations (Section 2.2.3). Then we consider the behavior of the two-electron system in the realistic double-dot confinement potentials

of the experimental planar laterally coupled QD structures (Section 2.2.4). The properties of the two-electron system confined in novel fully three-dimensional (3D) quantum wire QDs are presented in Section 2.3, while a few-electron system in triple QDs is discussed in Section 2.4. Finally, Section 2.5 contains concluding remarks.

## 2.2 Two electrons in double quantum dots

In the double-QD system the confinement potential in the lateral plane is the sum of the individual dot potentials and the electrostatic potential barrier separating them (Zhang *et al.* 2004). In general, tunnelling through the barrier as well as Coulomb interaction between the two electrons control the magnitude of the exchange coupling. Earlier theoretical approaches based on the approximate two-level Heitler–London method in a model double-dot potential (Burkard *et al.* 1999) have shown that the exchange energy can be tuned with applied magnetic fields all the way down to zero to become negative. Increasing magnetic field further causes the exchange energy to go to zero as electrons localize in individual dots with vanishing overlap. More elaborate approaches such as the Hund–Mullikan method (Burkard *et al.* 1999; Hu and DasSarma 2000) confirmed these results, albeit producing smaller values of the exchange coupling under the same conditions. The problem of two interacting particles in the double-dot confinement potential was also tackled by the generalized Heitler–London method with the broken-symmetry Hartree–Fock orbitals forming single-particle basis wavefunctions (Yannouleas and Landman 2002). For two very strongly coupled QDs analytical variational calculations have also been performed that not surprisingly showed a behavior of the exchange energy that is very close to that in a single dot (Dybalski and Hawrylak 2005).

In order to establish the validity and limitations of the approximate methods, full-scale numerical simulations have to be conducted. Fully numerical simulations of QDs based on the density-functional theory (DFT) are well suited for quantum device modelling (Macucci *et al.* 1993; Stopa 1996; Matagne *et al.* 2002). However, in the two-electron system in double QDs, this approach suffers from a fundamental flaw as it predicts negative exchange energy in the weak interdot coupling regime at zero magnetic fields (Wensauer *et al.* 2000; Saarikoski *et al.* 2002), thereby contradicting basic many-body quantum mechanics (Ashcroft and Mermin 1976) according to which this quantity should always remain positive. Application of the numerically “exact” diagonalization (or the configuration-interaction (CI)) approach to the solution of the two-electron Schrödinger equation (Harju *et al.* 2002; Szafran *et al.* 2004) effectively corrected this problem, albeit producing smaller values of the exchange coupling in magnetic fields.

### 2.2.1 Exact diagonalization approach

To obtain the energy spectrum and other quantities of interest of an  $N$ -electron system, in this work we use mostly the method of the numerically exact

diagonalization of the many-electron Schrödinger equation. The Hamiltonian  $\hat{H}$  of such a system is represented as the sum of the single-electron Hamiltonians  $\hat{h}_i$  and the interelectron Coulomb interaction:

$$\hat{H} = \sum_{i=1}^N \hat{h}_i + \sum_{i < j} \frac{e^2}{\epsilon |\mathbf{r}_i - \mathbf{r}_j|}, \quad (2.1)$$

$$\hat{h}_i = -\frac{\hbar^2}{2m^*} \left( \nabla_i - \frac{ie}{\hbar c} \mathbf{A}_i \right)^2 + V_{\text{conf}}(\mathbf{r}_i) \pm \frac{1}{2} g \mu_B B, \quad (2.2)$$

where  $m^*$  and  $\epsilon$  stand for the electron effective mass and dielectric constant for the QD electrons, respectively.  $\mathbf{A} = (1/2)(Bx, -By, 0)$  is the vector potential in the symmetric gauge for the magnetic field  $B$  oriented perpendicular to the QD confinement plane along the  $z$ -direction. The term  $\pm \frac{1}{2} g \mu_B B$  accounts for Zeeman splitting with  $g = -0.44$  being the electron  $g$ -factor in GaAs. In this equation the confinement potential  $V_{\text{conf}}(\mathbf{r})$  is assumed for simplicity to be two-dimensional as the width of the QD is usually much smaller than the lateral extension of the electron “puddle” (Melnikov and Leburton 2006). The coupling between orbital and spin degrees is neglected as in GaAs it is rather small in absolute value<sup>1</sup> and usually affects the electronic properties of QDs only close to degeneracy points in the energy spectrum (Stano and Fabian 2006; Chaney and Maksym 2007; Clemente *et al.* 2007).

We diagonalize the above Hamiltonian eqn (2.1) by expanding the  $N$ -electron wavefunction for the  $\alpha$ th state in terms of  $N \times N$  Slater determinants (Mikhailov 2002):

$$\Psi_\alpha(N) = \sum_{i \dots n} c_{i \dots n}^\alpha \begin{vmatrix} \varphi_i(\mathbf{r}_1, s_{z1}) & \dots & \varphi_n(\mathbf{r}_1, s_{z1}) \\ \vdots & & \vdots \\ \varphi_i(\mathbf{r}_N, s_{zN}) & \dots & \varphi_n(\mathbf{r}_N, s_{zN}) \end{vmatrix}. \quad (2.3)$$

Here, the basis wavefunction  $\varphi_i(\mathbf{r}_j, s_{zj})$  is the product of a 2D anisotropic harmonic oscillator eigenfunction (with frequencies being adjustable parameters) and a spin-wave function. Each quantum number  $i = (n_x, n_y, s)$  corresponds to the set of the 2D harmonic oscillator ( $n_x, n_y$ ) and spin  $s$  quantum numbers. The summation is carried over all permutations  $\{i \dots n\}$  available for the particular spin state, and the coefficients  $c_{i \dots n}^\alpha$  are determined after the minimization of  $\langle \Psi_\alpha | \hat{H} | \Psi_\alpha \rangle$ , which leads to the generalized eigenvalue problem with dense Hermitian matrices. We found that the harmonic oscillator frequencies smaller than the confinement strengths work best due to the fact that the Coulomb interaction tends to flatten out the effective potential (Reimann *et al.* 2000). In the case of a 2D circular confinement, the Coulomb matrix elements can be evaluated analytically yielding a four-fold series. For the anisotropic 2D potential (or in the 3D case), the matrix elements are also expressed through the four- (six- in 3D) fold series but the auxiliary one-dimensional exponential integral evaluated numerically by means of the Gauss–Kronrod quadrature has to be used to compute their final numerical values (Drouvelis *et al.* 2004).

<sup>1</sup>In InAs quantum wire QDs spin-orbit coupling plays a somewhat larger role in the energy spectrum (Fasth *et al.* 2007).

The above method of numerically exact diagonalization of the many-electron Hamiltonian eqn (2.1) also known as the full CI approach yields very accurate values for the eigenenergies and eigenfunctions provided that the single-particle basis set eqn (2.3) is chosen to be large enough (Chakraborty 1999). The downside of this method resides in the size and complexity of the Hamiltonian matrix that grows very quickly with the number of electrons ( $\propto M^N$  with  $M$  being the number of single-particle basis states). As a result, it is usually used for QDs with highly symmetric confinement potential (Rontani *et al.* 2001; Abolfath and Hawrylak 2006) containing a small number of electrons (typically, less than  $\sim 10$ ). The most crucial step in this approach is the construction of the many-particle wavefunction that can be carried out by using either (1) the single-particle wavefunctions of the corresponding Hamiltonian or (2) analytic wavefunctions of an elementary confinement potential such as a 2D (3D) harmonic oscillator (as used in this work) or Fock–Darwin states (Reimann *et al.* 2000; Kyriakidis *et al.* 2002; Saarikoski and Harju 2005) or (3) multicenter expansion of the single-particle wavefunctions (Szafran *et al.* 2004). Computations using method (1) are usually performed fully numerically on a grid (Bellucci *et al.* 2004) and can in principle be applied to various complex confinement potentials (Stopa and Marcus 2006). The disadvantage of this approach is the absence of a systematic way to build up an adequate basis set that ensures fast convergence in the computed energy values and the inherent loss of accuracy and dramatic increase in the computational time required for the numerical evaluation of the Coulomb integrals as the basis set gets larger. A hybrid Poisson-CI approach in which the single-particle contributions to the Hamiltonian with the real self-consistent confinement potential  $N = 0$  were computed numerically with the harmonic oscillator basis set, while the Coulomb integrals were evaluated analytically, has been also applied for simulation of realistic double-QD systems (see Section 2.2.4).

### 2.2.2 The double-parabola potential model: State mixing and the Heitler–London limit

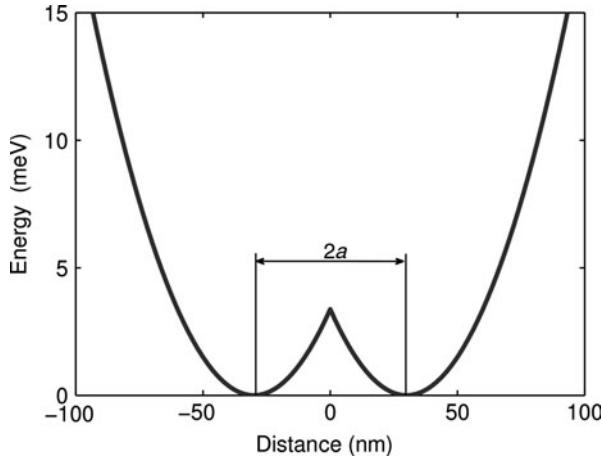
In this section we model the double-dot electron-confinement potential as a superposition of two parabolas, i.e.

$$V_{\text{conf}}(\mathbf{r}) = \frac{1}{2}m^*\omega^2 \left\{ \min \left[ (x - a)^2, (x + a)^2 \right] + y^2 \right\}, \quad (2.4)$$

where the confinement energy is taken to be  $\hbar\omega = 3$  meV and  $2a$  is the interdot separation (Fig. 2.4). This form of the confinement potential is the simplest approximation allowing one to gain insight into the basic interplay between single-particle and Coulomb effects in coupled QDs.

In order to better understand the behavior of the exchange energy in magnetic fields, the two-electron wavefunction obtained after numerically exact diagonalization of the Schrödinger equation is expanded in terms of the eigenfunction products of the single-particle (SP) Hamiltonian eqn (2.2):

$$\Psi_{S(T)}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{i < j} \alpha_{ij}^{S(T)} \psi_i(\mathbf{r}_1) \psi_j(\mathbf{r}_2), \quad (2.5)$$

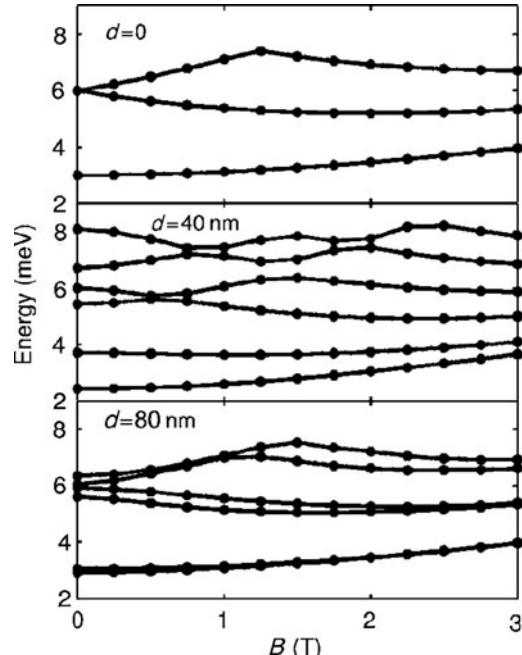


**Fig. 2.4** Double-parabola confinement potential (eqn (2.4)) in the  $x$ -direction for  $2a = 60$  nm,  $y = 0$ .

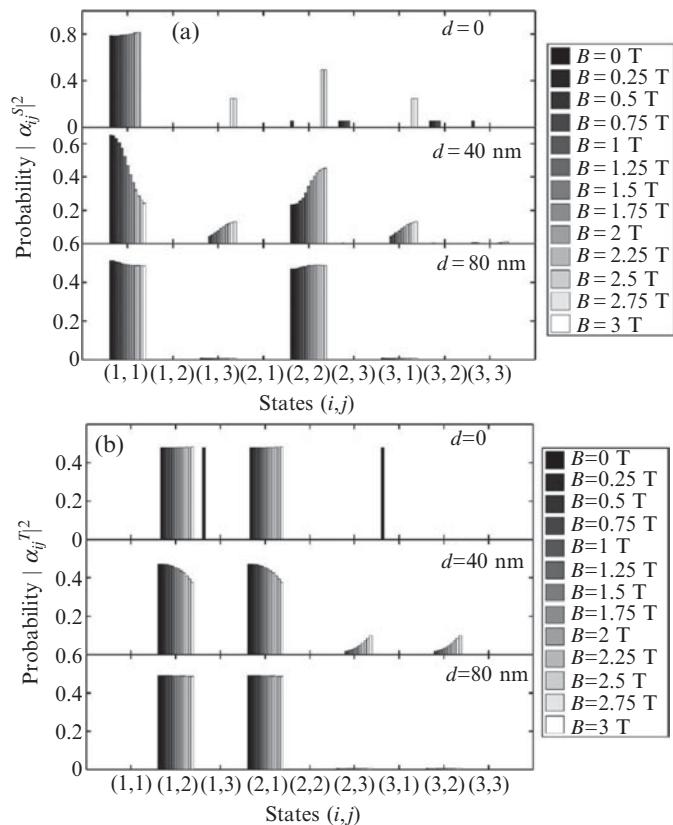
where  $|\alpha_{ij}^{S(T)}|^2$  is the probability of finding the two electrons in the  $(i, j)$  state pair (spectral function) with indices  $i$  and  $j$  labelling the eigenlevels of the Hamiltonian eqn (2.2) starting from the ground state  $i = j = 1$  (Wójs and Hawrylak 1997; Melnikov and Leburton 2006). Here, the SP orbitals  $\psi_i(\mathbf{r})$  are constructed from the wavefunctions of a 2D harmonic oscillator similar to the above two-particle wavefunction eqn (2.3).

The dependence of the SP energies on the magnetic field is presented in Fig. 2.5 for three different interdot separations:  $d = 0$  corresponding to a single circular QD,  $d = 40$  nm and  $d = 80$  nm. In a single dot (top plot) the eigenenergies form the familiar Fock–Darwin energy spectrum of a 2D circular harmonic oscillator (Reimann and Manninen 2002). Upon increasing the interdot distance to 40 nm (middle plot), the Fock–Darwin sequence among the eigenenergies is completely destroyed: There are no obvious orbital degeneracies (no shell structure), but rather avoided crossings (anticrossings) among the energy curves appear throughout the plot. This is the regime of strong molecular coupling when two dots strongly affect each other (Dybalski and Hawrylak 2005). Further increase in the separation between the dots to  $d = 80$  nm (bottom plot) recovers the Fock–Darwin spectrum but the six levels shown in the plot are clearly grouped into three pairs corresponding to the lowest three levels of the single dot (see top plot) indicating that the two QDs are weakly coupled (or equivalently the QD molecule is on the verge of dissociation).

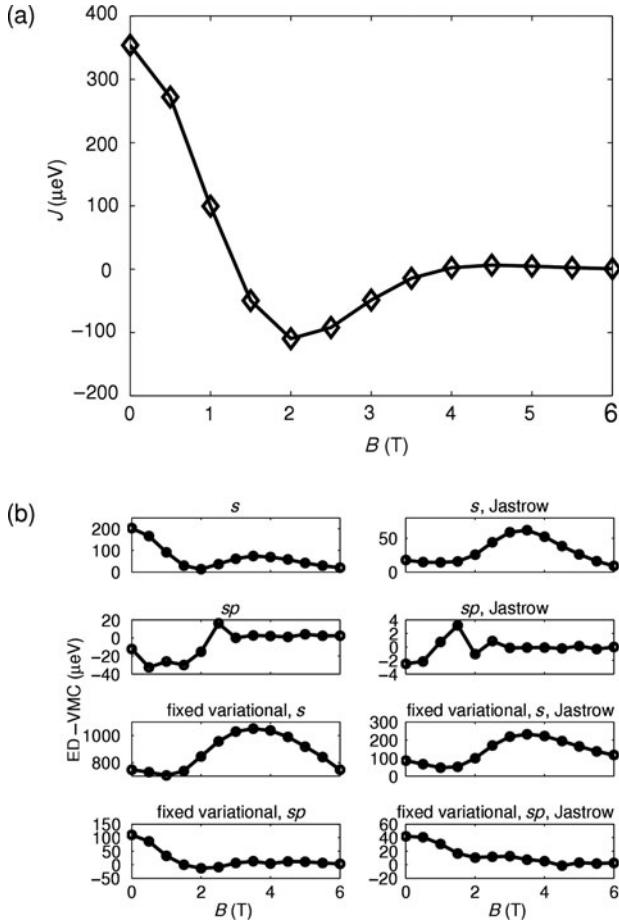
The transition from the single QD populated with two electrons to the two uncorrelated QDs with one electron each can be analyzed in greater detail by plotting the spectral function (SF)  $|\alpha_{ij}^{S(T)}|^2$  for the same three interdot distances as a function of the applied magnetic field (Fig. 2.6). One can see that in the single-dot case (Fig. 2.6, top plots) at zero magnetic field the system is mostly in the lowest, allowed by the Pauli principle, SP states, forming a  $(1, 1)$  pair for the singlet and  $(1, 2), (2, 1)$  equivalent pairs for the triplet. With increasing magnetic fields, the  $(1, 1)$  pair remains dominant up until  $\sim 2.7$  T when the transition to the state with higher total orbital momentum  $L$  (Wagner *et al.* 1992) in the singlet state takes place (at this magnetic



**Fig. 2.5** Single-particle energies in magnetic fields for  $d = 0$  (top part),  $d = 40 \text{ nm}$  (middle part), and  $d = 80 \text{ nm}$  (bottom part). In the top part, only three lowest states are shown, while in the middle and bottom parts six lowest states are plotted. (Reprinted with permission from Melnikov and Leburton (2006). Copyright 2006 by the American Physical Society.)



**Fig. 2.6** (a) SF coefficients  $|\alpha_{ij}^{S(T)}|^2$  for  $1 \leq (i, j) \leq 3$  in the magnetic field for the singlet state for  $d = 0$  (top),  $d = 40 \text{ nm}$  (middle part), and  $d = 80 \text{ nm}$  (bottom). (b) Same as in (a) but for the triplet state. The contributions from  $(1, 3)$  and  $(3, 1)$  for  $d = 0$  case appear because the triplet state is twofold degenerate at  $B = 0$ . (Reprinted with permission from Melnikov and Leburton (2006). Copyright 2006 by the American Physical Society.)



**Fig. 2.7** Exchange energy vs. magnetic fields in strongly coupled QDs (Kim *et al.* 2008): (a) ED method, (b) difference between the energies obtained from ED and VMC trial wavefunctions.

field there is an absolute minimum in the exchange energy, see Fig. 2.7). At higher magnetic fields the electron system redistributes itself almost equally among the three lowest single-particle states, i.e. the probabilities of finding both electrons in the excited states sharply increase. Similar SF behavior is expected in the triplet state as well, with the difference that the transition to the state with higher total orbital momentum occurs at around 4 T; as such, both electrons remain in the (1, 2)–(2, 1) pairs in the studied range of the magnetic fields. Note that within the intervals with defined  $L$ , SF weakly depends on the magnetic field, thereby reflecting the conservation of the total angular momentum and parity of the system.

On the other hand, when  $d = 40$  nm, SF shows a strong dependence on the magnetic field (Fig. 2.6, middle plots), though the general features of the single dot are clearly recognizable. One can see that at zero magnetic field there is already a large contribution from the lowest excited-state pair (2, 2) in the singlet state that increases with magnetic field and becomes dominant at about 1.7 T when the exchange energy reaches its minimum (see Fig. 2.7). At the same time, the probability of finding the electron pair in the ground-state pair quickly decreases. One also observes that the relative

weight of the (1, 3) and (3, 1) pairs compared to (1, 1) and (2, 2) pairs in the two-particle state is smaller than in the single-dot case, indicating the onset of electron localization in each dot, i.e. the onset of the Heitler–London regime. As for the triplet state, its SF smoothly decreases with increasing magnetic field, while the contributions from other single-particle states gradually increase. Overall, the dependences of the probabilities on the magnetic field in this case are smooth as the system no longer has a definite angular momentum.

For the weakly coupled QDs when  $d = 80$  nm (Fig. 2.6, lower part), the singlet (triplet) state consists mostly of the (1, 1) and (2, 2) [(1, 2) and (2, 1)] equally weighted pairs of the SP states, i.e. they correspond to the Heitler–London states (see, e.g. Golovach and Loss 2004 and Appendix B). However, even in this limiting case, the effect of other SP states is still visible, albeit small, as the SF probabilities still show a slight dependence on the magnetic field. One would expect that only in the limit of  $d \rightarrow \infty$ , do these coefficients become independent of the magnetic field.

The redistribution of the electrons among different SP states may also have interesting consequences for approximate theories that rely on the analysis of the single-particle orbitals, such as the DFT. The standard DFT formulation assumes that the density can be represented by the single ground state. However, the above-considered mixing of the SP states would require utilization of their properly weighted sum to obtain the correct density distribution (for the discussion, see Ullrich and Kohn 2001, and references therein). This means, in particular, that the standard DFT approach will be a very good approximation in a single circular dot at low magnetic fields, while its application at higher fields, or in the QD structures where mixing of several states occurs, may be difficult. This, in particular, may explain why the so-called “broken-symmetry” solution of the Hamiltonian, i.e. the states that do not possess the original symmetry of the Hamiltonian, can sometimes lead to surprisingly good results; for a comprehensive discussion of the “broken-symmetry” concept and its repercussions for various physical systems see Yannouleas and Landman (2007).

We can also explicitly compare (Kim *et al.* 2008) the above dependence of  $J$  on the magnetic field with the results obtained from the variational Monte Carlo (VMC) calculations (see Appendix A). Here, we can gauge the effect of different terms in the Hamiltonian eqn (2.1) on the exchange energy by using the VMC trial wavefunction, eqns (2.16) and (2.17), with different approximations: (1) by using VMC with the Jastrow term and  $s$ -like orbitals only, (2) by using VMC with  $s$ -like orbitals but setting the Jastrow term  $j(r_{ij}) = 1$ , (3) by using VMC with the Jastrow term and  $sp$ -like orbitals, (4) by using VMC with  $sp$ -like orbitals without the Jastrow term, (5) by using VMC with the Jastrow term,  $s$ -like orbitals, and fixed variational parameters (eqn (2.20)), (6) by using VMC with  $j(r_{ij})$ ,  $s$ -like orbitals and fixed variational parameters, (7) by using VMC with the Jastrow term,  $sp$ -like orbitals, and fixed variational parameters, and (8) by using VMC with  $sp$ -like orbitals with  $j(r_{ij}) = 1$ , and fixed variational parameters. In Fig. 2.7(a), we display  $J$  from the exact diagonalization (ED) calculations at various magnetic fields for the separation between the two QDs in eqn (2.4) along the  $x$ -axis  $2a = 30$  nm (strongly coupled case).

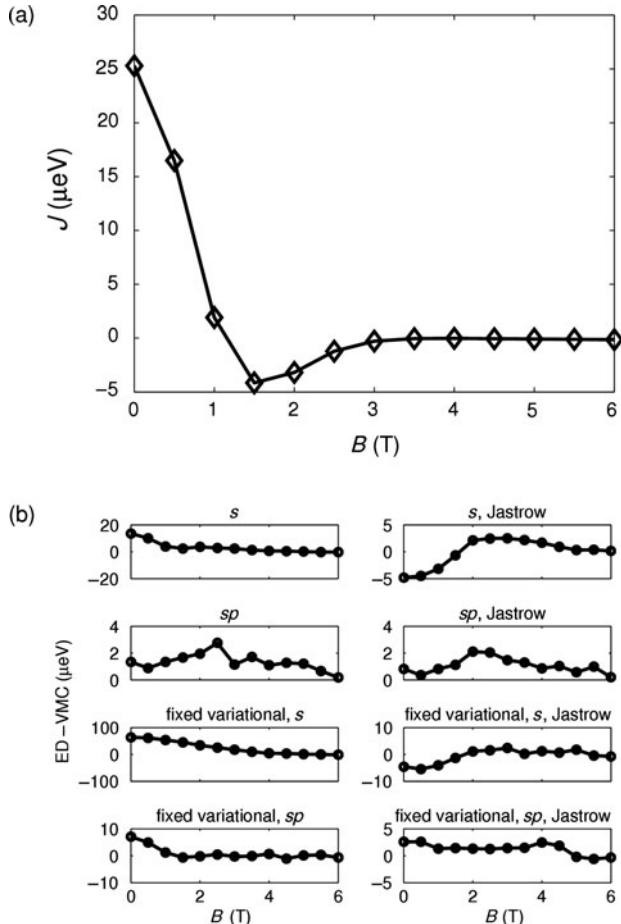
In Fig. 2.7(b), we show the difference in exchange energies obtained from various VMC trial wavefunctions and the exact diagonalization results. One can see the usage of  $sp$ -like orbitals gives the most accurate results as expected, while using fixed variational parameters in the trial wavefunction, for which it corresponds to the Heitler–London case, gives very inaccurate results. One can also see that the presence of the Jastrow term improves the accuracy of the VMC technique but at higher magnetic field values the improvement is less pronounced or even negligible, depending on the specific regime and the specific trial wavefunctions. This can be easily explained by noting that since  $J(r_{ij})$  accounts for correlation effects between electrons, it has a weaker effect at high magnetic fields where the electron orbitals are more compressed and the two electrons are more decoupled. In Fig. 2.7(b), we also notice that in the two  $s$ -like orbital curves (top row) the error increases when magnetic fields exceed 2 T with a largest error at 3.5 T. From the SF analysis (Fig. 2.6), we know that in this interval higher-energy SP orbitals contribute significantly to the singlet two-electron wavefunctions. Because such states are omitted in the  $s$ -like orbital VMC trial wavefunctions, the VMC exchange energy decays to zero at higher magnetic fields than the ED curve. Upon including the  $p$ -like orbitals (second row), the accuracy of VMC results improves significantly.

In Fig. 2.8(a), we plot the ED exchange energy for a weakly coupled QD system, for which we set  $2a = 60$  nm. Similar to the strongly coupled QD case, we also plot the difference in total energies obtained from various VMC trial wavefunctions from the ED results in Fig. 2.8(b). In Fig. 2.8(b) at  $B = 0$  T, we notice the exchange energies are about one order of magnitude smaller than in strongly coupled QDs (Fig. 2.7), which is expected since the wavefunction overlap between the two electrons exponentially decreases with the dot separation. Comparing Figs. 2.7(b) and 2.8(b), we see closer agreement between all of the VMC and the ED results for weakly coupled QDs with noticeable improvements in the  $s$ -like orbital VMC wavefunction cases. These results confirm the above SF analysis and indicate that (1) in the strong-coupling case the exchange energy approaches zero not because of the vanishing overlap between the electron wavefunctions (as predicted by the HL method and that requires a large magnetic field) but rather due to the involvement of the excited SP states that raise the singlet energy with respect to the triplet one, and (2) the Heitler–London approach can be trusted only when the coupling between the dots is weak, which implies small exchange energy values. As such, one should be careful when utilizing approximate methods for evaluating exchange coupling in the two-electron system confined in the double-dot molecule (Pedersen *et al.* 2007).

## 2.2.3 The coupled gaussian potential model

### 2.2.3.1 Charge-stability diagram of coupled circular dots

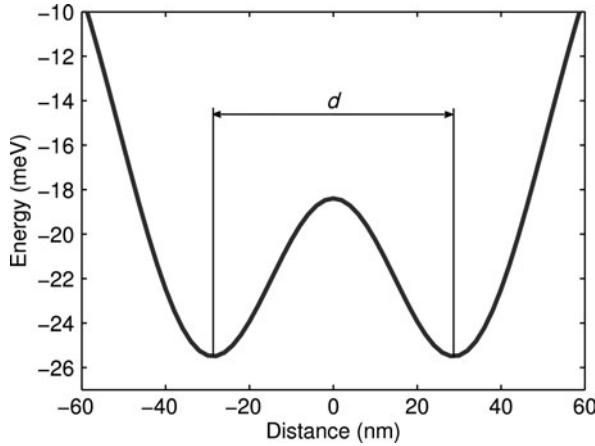
While extensive theoretical work focuses on the dependence of  $J$  on the system parameters such as the interdot separation, the tunnelling barrier between the QDs, and the external magnetic field (Burkard *et al.* 1999; Hu and DasSarma



**Fig. 2.8** Same as in Fig. 2.7 but for the weakly coupled QDs (Kim *et al.* 2008).

2000; Harju *et al.* 2002; Szafran *et al.* 2004), the charge stability diagram of coupled QDs (van der Wiel *et al.* 2002) has received far less attention. Meanwhile, recent advances in experimental techniques have made it possible to study coupled QDs in the few-electron regime when each QD contains only one conduction electron (Qin *et al.* 2001; van der Wiel *et al.* 2002; Ancilotto *et al.* 2003; Elzerman *et al.* 2003; Hayashi *et al.* 2003; Hatano *et al.* 2005; Petta *et al.* 2005). In this case the stability diagram becomes a powerful tool to study interdot coupling and electronic transport through double QD systems. Analysis of the stability diagram and its evolution in magnetic fields also allows one to estimate the values of the exchange energy as was demonstrated recently in the case of the two laterally coupled vertical QDs (Hatano *et al.* 2005).

In general, in the stability diagram the boundaries between distinct stable charge states, i.e. between the states with fixed number of electrons  $N_1$  and  $N_2$  in each of the coupled dots, are represented as functions of the two controlling gate biases, one for each dot (van der Wiel *et al.* 2002). These equilibrium charges are determined from the condition that the chemical potential of the



**Fig. 2.9** Double-Gaussian confinement potential (eqn (2.7)) in the  $x$ -direction ( $y = 0$ ) for  $d = 60$  nm,  $V_L = V_R = 25$  meV and  $R_x = R_y = 30$  nm.

QD structure  $\mu(N_1 + N_2)$  defined as (van der Wiel *et al.* 2002):

$$\mu(N_1 + N_2) = E_G(N_1 + N_2) - E_G(N_1 + N_2 - 1), \quad (2.6)$$

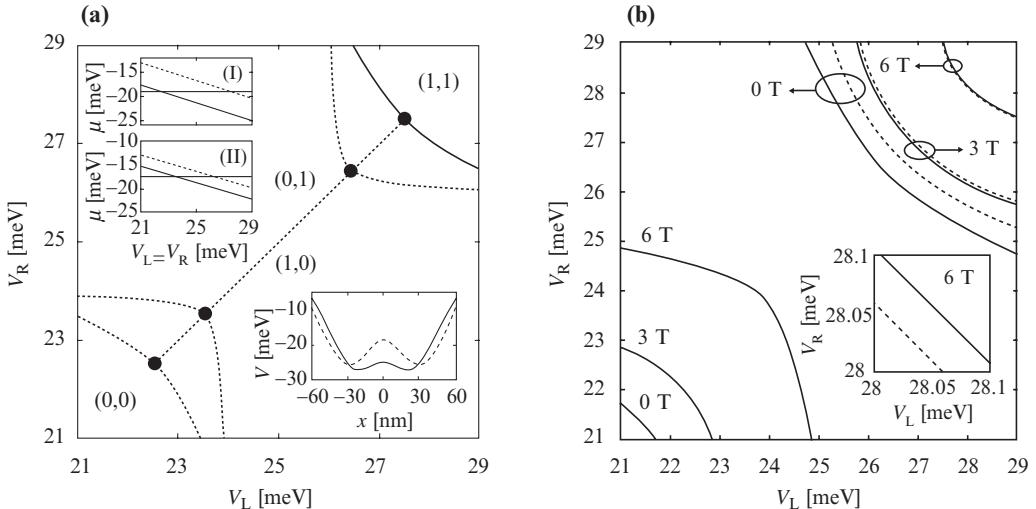
where  $E_G(N)$  is the ground state energy of the  $N$ -electron state, is less than that of the leads (source and drain).

In this section, we study the stability diagram in coupled QDs with  $N_1 + N_2 \leq 2$  electrons in external magnetic fields, and investigate its properties for different interdot coupling strengths (Zhang *et al.* 2006). The confinement in double QDs is simulated by the following model potential (Hu and DasSarma 2000; Szafran *et al.* 2004; Zhang *et al.* 2006):

$$V_{\text{conf}}(\mathbf{r}) = -V_L \exp \left[ -\frac{(x + d/2)^2}{R_x^2} - \frac{y^2}{R_y^2} \right] - V_R \exp \left[ -\frac{(x - d/2)^2}{R_x^2} - \frac{y^2}{R_y^2} \right], \quad (2.7)$$

where  $V_L$  and  $V_R$  are the “depths” of the left and right dots (equivalent to the QD gate voltages in experimental structures) that can be independently varied,  $d$  is the interdot separation, and  $R_x$ ,  $R_y$  are the QD “radii” (Fig. 2.9). Compared with the two-parabola potential eqn (2.4) used in the previous section, this form of the confinement potential allows for easy study of the electron relocalization from one QD to another; it also has a “smooth” barrier in between the dots, making it more comparable to the realistic QD potentials (see Section 2.2.4).

In order to simulate the stability diagram for two circular QDs we put  $R_x = R_y = 30$  nm and also consider both strongly ( $d = 50$  nm) and weakly ( $d = 60$  nm) coupled dots. Figure 2.10(a), main panel, displays the contour plots of the chemical potentials  $\mu(1)$  (lower branches) and  $\mu_S(2)$  (upper branches) for the charging of the first and the second (label “S” stands for the singlet) electrons, respectively, in the  $V_L - V_R$  plane for  $d = 50$  nm at the constant value of  $\mu(1) = \mu_S(2) = -19.07$  meV and for  $d = 60$  nm at  $\mu(1) = \mu_S(2) = -17.54$  meV, which in both cases is the chosen reference value of the chemical



**Fig. 2.10** (a) Contour plots of  $\mu(1)$  and  $\mu_S(2)$  in  $V_L$ - $V_R$ -plane for  $d = 50$  nm (solid) and  $d = 60$  nm (dashed). The turning points on the contour lines are indicated by solid dots and the dotted line is a guide for the eye along the main diagonal ( $V_L = V_R$ ). Top insets: chemical potentials  $\mu(1)$  (solid) and  $\mu_S(2)$  (dashed) as functions of  $V_L = V_R$  for (I)  $d = 50$  nm and (II)  $d = 60$  nm; the gray line indicates the value of the chemical potential corresponding to the contours in the main panel. Bottom inset: the  $x$ -direction profile of the confinement potential (eqn (2.7)) for  $d = 50$  nm (solid) and  $d = 60$  nm (dashed). (b) Contour plots of the chemical potentials  $\mu(1)$  (lower branches, solid),  $\mu_S(2)$  (upper branches, solid), and  $\mu_T(2)$  (upper branches, dashed) at  $d = 50$  nm for different magnetic fields. Inset is an enlarged region of the main plot showing the separation between  $\mu_S(2)$  and  $\mu_T(2)$  contour lines at  $B = 6$  T. (Reprinted from Melnikov *et al.* (2007). Copyright 2007 with permission from Elsevier.)

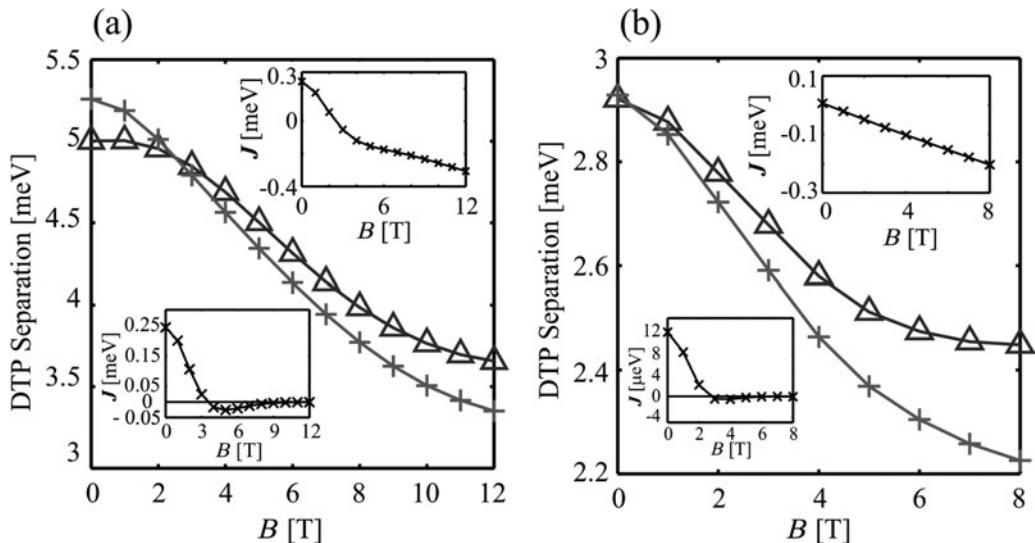
potential in the source/drain of the QD device. The reference values are indicated in the top insets by horizontal lines and selected such that the turning points (shown by solid dots in the main panel) of the lower and upper branches for each dot separation are symmetric with respect to the center of the diagram ( $V_L = V_R = 25$  meV).

In this stability diagram we recognize four different regions separated by the chemical-potential contour lines and the diagonal line,  $V_L = V_R$  (dotted line), as the stable charge states with different electron numbers in each dot given in parenthesis. At the turning point on each branch, three different stable charge states coincide, as they correspond to the same value of the total energy. The distance between the turning points on the lower and upper branches is the so-called double-triple point (DTP) separation (van der Wiel *et al.* 2002; Hatano *et al.* 2005). From the data shown we note that the DTP separation for  $d = 50$  nm along each voltage axis is  $\Delta V_L = \Delta V_R = 5.00$  meV, and is larger than the DTP separation value 2.86 meV for  $d = 60$  nm. Furthermore, the curvature of the branches with respect to the DTPs for  $d = 50$  nm is smaller than for  $d = 60$  nm and it also takes different values for the lower and upper branches (Zhang *et al.* 2006). According to the general theory (van der Wiel *et al.* 2002), a smaller DTP separation (or a larger curvature of the chemical potential contour lines) suggests a smaller interdot coupling, which is consistent with our results.

In Fig. 2.10(b), we plot the chemical-potential contour lines for  $\mu(1)$  (lower branches, solid),  $\mu_S(2)$  (upper branches, solid) and  $\mu_T(2)$  (upper branches, dashed) in the  $V_L$ - $V_R$  plane at  $B = 0$ , 3 and 6 T. Here, the contours are drawn

for  $\mu(1) = \mu_S(2) = \mu_T(2) = -18$  meV, which is the reference value for the chemical potential of the QD device source/drain. As the magnetic field increases, the chemical-potential contour lines shift from the lower left corner ( $V_L = V_R = 21$  meV) to the upper right corner ( $V_L = V_R = 29$  meV). This is due to the fact that the single-particle eigenenergies in coupled QDs increase with the magnetic field. The same happens for  $\mu(2)$ , only at a smaller pace due to the decrease in the Coulomb repulsion with increasing magnetic field as the electrons localize in different dots and the overlap between them diminishes; as a result of this, the DTP separation becomes smaller. The magnetic localization of electrons also causes the curvatures of the chemical-potential contours to increase for both singlet and triplet states, thereby indicating the increasing decoupling of the two dots in the magnetic field. We observe that at  $B = 0$  and 3 T, the  $\mu_T(2)$  contour is above the  $\mu_S(2)$ , while at  $B = 6$  T the situation is reversed. This can be correlated with the variation of the exchange coupling  $J$  as a function of the magnetic field:  $J > 0$  and decreases from  $B = 0$  through  $B = 3$  T, in which range  $\mu_S(2) < \mu_T(2)$ ;  $J < 0$  and assumes a smaller value for  $B > 3.6$  T so that  $\mu_T(2) < \mu_S(2)$ .

Figures 2.11(a) and (b) show the extracted DTP separation along  $V_L$  (or  $V_R$ ,  $V_L = V_R$ ) axis as a function of magnetic fields for  $d = 50$  nm and 60 nm interdot separations, respectively. Note that at  $B = 0$  the DTP separation for the singlet state is smaller than that for the lowest triplet state because the singlet is the ground state, while at larger magnetic fields, the lowest triplet state becomes the ground state and the order of the DTP separations is reversed. In both (a) and (b), the DTP separation for the lowest triplet state decreases faster with magnetic fields than that for the singlet state. This is because the DTP separation is proportional to  $\mu(2) - \mu(1) = E_G(2) - 2E_G(1)$  for a fixed



**Fig. 2.11** DTP separation along  $V_L = V_R$  as a function of the magnetic field for (a)  $d = 50$  nm and (b)  $d = 60$  nm. The data for the singlet and the lowest triplet states are labelled by “ $\Delta$ ” and “ $+$ ”, respectively. The upper (lower) inset in each figure shows the exchange energy  $J$  as a function of the magnetic field with (without) the Zeeman effect. For the data in the insets  $V_L = V_R = 25$  meV. (Reprinted with permission from Zhang *et al.* (2006). Copyright 2006 by the American Physical Society.)

$V_L = V_R$  on the main diagonal of the stability diagram (see Fig. 2.10). For the singlet state,  $E_G(2)$  does not change with the  $B$  field, while  $E_G(1)$  decreases with the  $B$  field due to the Zeeman effect, therefore the Zeeman contribution to  $\mu(2) - \mu(1)$  increases with the magnetic field. For the triplet state, the Zeeman contributions to  $E_G(2)$  and  $2E_G(1)$  cancel out, and  $\mu(2) - \mu(1)$  is not affected. The decrease of the DTP separation in the magnetic field was also recently observed experimentally (Hatano *et al.* 2005). The upper (lower) inset in each figure shows the corresponding exchange energy  $J$  as a function of the magnetic field with (without) the Zeeman effect. In both cases, the Zeeman effect induces a linear dependence of  $J$  on  $B$ . However, in (a) given the strong coupling between the dots, the orbital contribution to  $J$  dominates at low magnetic fields before being overcome by the Zeeman-induced decrease at higher field; in (b),  $J$  is totally dominated by the Zeeman contribution, which decreases linearly with the applied magnetic field. Comparison of the DTP separation and exchange energy in the absence of the Zeeman effect shows that the latter saturates at much lower values of the magnetic field than the former. This is because the DTP separation is determined by the Coulomb interaction between electrons that decreases as the electrons become localized by the magnetic field in individual dots (within the Heitler–London approximation, this decrease is proportional to  $B^{-1}$ , (Melnikov and Leburton 2006)), while the exchange energy in the absence of the Zeeman effect approaches zero much faster than the Coulomb interaction since it is proportional to the overlap between the individual electron wavefunctions that decays exponentially fast in strong magnetic fields (Burkard *et al.* 1999; Melnikov and Leburton 2006).

### 2.2.3.2 Exchange energy: Exact diagonalization vs. the Hubbard model

It is also interesting to compare the exact values of the exchange energy (see the insets in Fig. 2.11) with those extracted from the stability diagrams in magnetic fields using the Hubbard model (Burkard *et al.* 1999; Hatano *et al.* 2005). According to this model,  $J_{\text{est}} = 4t^2/(V_{\text{intra}} - V_{\text{inter}})$  where  $2t$  is the tunnel (symmetric-asymmetric) splitting,  $V_{\text{intra}}$  and  $V_{\text{inter}}$  are the intradot and interdot Coulomb interactions. The interdot Coulomb interaction  $V_{\text{inter}}^{50(60)} \approx 3.4$  (2.0) meV for  $d = 50$  (60) nm is given by the DTP separation (for the lowest triplet state) in the limit of large magnetic fields (Fig. 2.11). These numbers are in good agreement with the corresponding expectation values of the Coulomb interaction matrix (3.5 and 2.2 meV, respectively) obtained from direct calculations, thereby confirming electron localization and interdot decoupling. Since at zero magnetic field, the DTP separation is equal to  $2t + V_{\text{inter}}$  (van der Wiel *et al.* 2002), we obtain  $2t^{50(60)} \approx 1.6$  (0.7) meV, which is consistent with the computed difference between the two lowest single-particle levels of 1.9 (0.4) meV.  $V_{\text{intra}} \approx 8$  meV is given by the electron addition energy in one QD (this is the distance between the “corners” of the linear region where single-electron relocalization occurs from one dot to the other in the  $N = 2$  energy diagram) so that the estimated values of the exchange energy become  $J_{\text{est}}^{50(60)} \approx 0.6$  (0.08) meV. These numbers are of the same order as the numerically exact values of 0.24 (0.012) meV, but they both significantly *overestimate* the computed data, and therefore, can only be used as a general guideline to gauge the

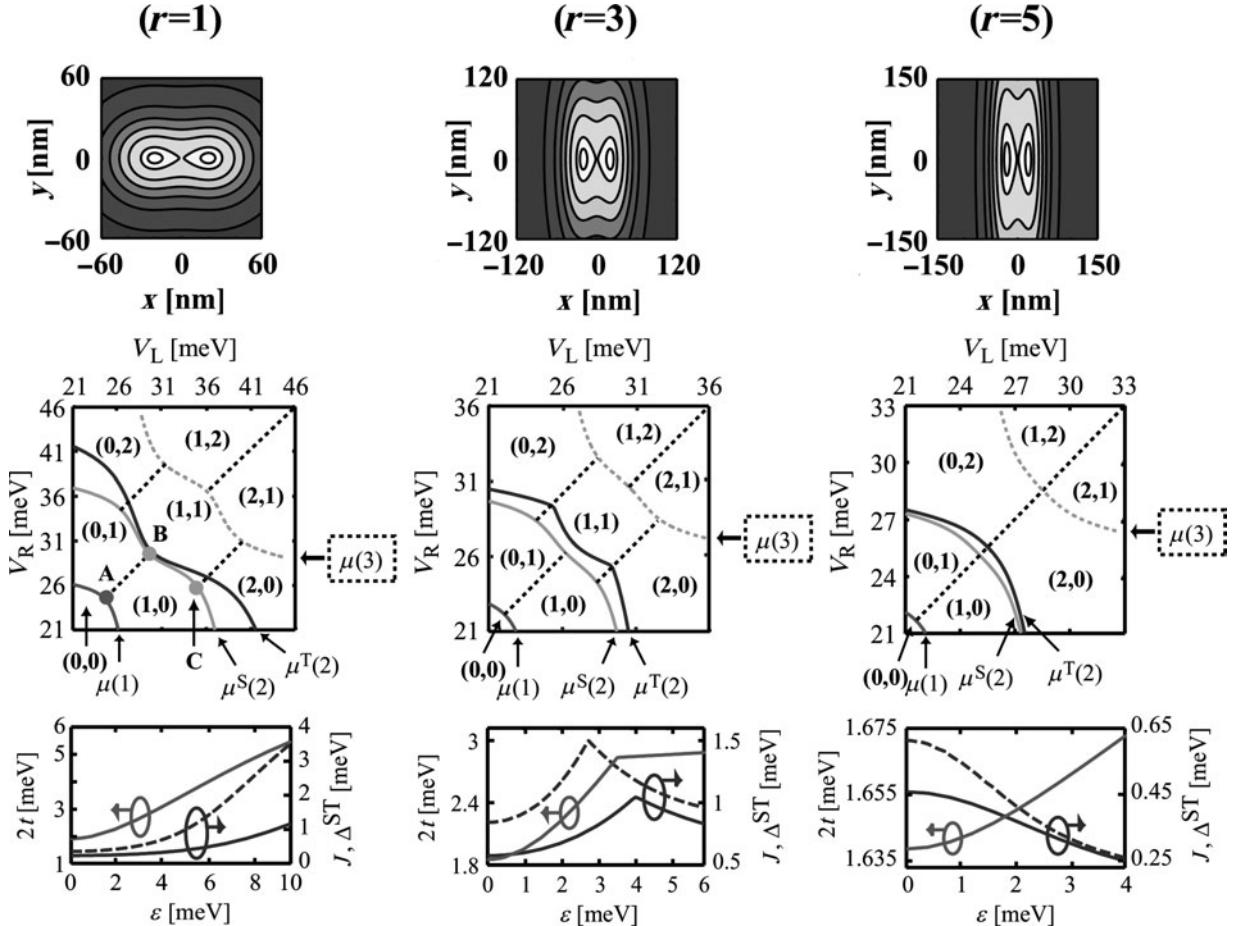
magnitude of the exchange coupling in double QDs. The overestimation is due to the mixing of the higher-energy SP orbitals that lowers the exchange energy (Section 2.2.1) and is not taken into account in the simple Hubbard model.

### 2.2.3.3 The von Neumann–Wigner Theorem in coupled quantum dots

Elliptic deformations in single QDs occur frequently, even if the QD structure is designed to be perfectly circular (Fujisawa *et al.* 2002). In double QDs, this problem becomes even more pronounced as the interdot coupling and resulting energy spectrum strongly depend on the asymmetry of individual dots (Hatano *et al.* 2004; Ravishankar *et al.* 2004; Szafran *et al.* 2004). To perform a general study of the geometric effects on the properties of the two-electron system in the QD molecule, we again use a model Gaussian potential eqn (2.7), where we fix  $d = 50$  nm,  $R_x = 30$  nm and introduce a new parameter—the QD aspect ratio  $r = R_y/R_x$ —which accounts for the dot elliptic deformation ( $r = 1$  corresponds to a circular QD).

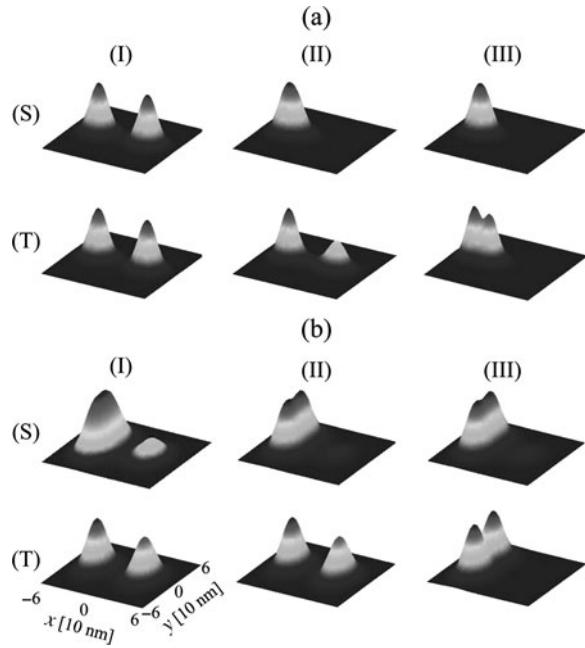
In Fig. 2.12, middle row, we plot the stability diagrams (van der Wiel *et al.* 2002; Zhang *et al.* 2009) of the coupled QDs with  $r = 1$  (left),  $r = 3$  (middle), and  $r = 5$  (right). The corresponding profiles of the confinement potential are shown in the top row. The solid curves indicated by arrows show the computed contours where chemical potentials of the first electron [ $\mu(1)$ ], the second electron in the singlet state [ $\mu^S(2)$ ], and second electron in the triplet state [ $\mu^T(2)$ ] are equal to the reference value  $-21$  meV. According to the general shape of the stability diagram for coupled QDs (van der Wiel *et al.* 2002), we place dotted straight lines on the diagrams to separate different charge states (*e.g.* (0, 1) means zero electrons in the left dot and one electron in the right dot). Specifically, the boundaries between the (1, 1) and (0, 2) [or (2, 0)] states are taken extending from the point on the  $\mu^S(2)$  curve at which the curvature is the largest for  $V_L \neq V_R$  (*e.g.* point C on the upper left panel) and parallel to the main diagonal. In the absence of magnetic field ( $B = 0$ ), the  $\mu^S(2)$  curve is the boundary between one and two electrons in the system (in the linear transport regime wherein the source and drain chemical potentials are nearly the same). Analogously to the boundaries between the (1, 1) and (0, 2) states, we can extrapolate to get the boundary between two- and three-electron states (dotted curve indicated by  $\mu(3)$ ). As  $r$  increases, the crossing points of the  $\mu(1)$ ,  $\mu_S(2)$  and  $\mu_T(2)$  curves with the main diagonal shift to smaller  $V_L = V_R$  values because the single-particle energies decrease as  $r$  increases. Due to the same reason, the DTP separation, *i.e.* the separation between the crossing points of  $\mu(1)$  and  $\mu(2)$  curves with  $V_L = V_R$ , decreases with  $r$ . For the singlet (triplet) state, the DTP separation measured in  $\Delta V_L = \Delta V_R$  is 5.18 (5.27), 4.13 (4.73) and 3.47 (3.91) meV for  $r = 1, 3$  and 5, respectively. Interestingly, as  $r$  increases, the (1, 1) stability region shrinks and finally vanishes. This is because as  $r$  increases it becomes energetically more favorable (Coulomb repulsion is smaller) to localize both electrons in the left or right dot rather than keep them apart in individual dots (Zhang *et al.* 2009).

The lower panels in Fig. 2.12 show the separation between the  $\mu^S(2)$  and  $\mu^T(2)$  curves and tunnel coupling strength  $2t$  projected along the main diagonal as a function of the interdot detuning  $\epsilon = |V_L - V_R|$  (note that



**Fig. 2.12** Top row: Potential-contour plots of coupled QD with  $r = 1$  (left),  $r = 3$  (middle) and  $r = 5$  (right) for  $V_L = V_R = 25$  meV. Darker regions correspond to higher potential. Middle row: Stability diagrams for  $r = 1$  (left),  $r = 3$  (middle) and  $r = 5$  (right). In each diagram, the solid curves are computed contour lines for which the chemical potential  $\mu(1)$ ,  $\mu^S(2)$  and  $\mu^T(2)$  is equal to the reference value  $\mu_{\text{ref}} = -21$  meV. The dashed lines are a guide for the eye separating different stable charge states. Note that the exact locations of the  $\mu(3)$  curve and  $(1, 2)$ ,  $(2, 1)$  regions are not computed. In the left panel, we also indicate the double-triple points A and B. Point C is where the  $\mu^S(2)$  curve has the largest curvature for  $V_L \neq V_R$ . Bottom row:  $2t$  and  $J$  (solid curves) as functions of interdot detuning  $\epsilon$  from the center of the  $(1, 1)$  region for the same  $r$  as in top panels. The dashed curves show the separation ( $\Delta ST$ ) between the contour lines of  $\mu^S(2)$  and  $\mu^T(2)$  projected along the main diagonal of the stability diagram above.

experimentally,  $J$  can also be extracted from the spin blockade regime in the stability diagram (Johnson *et al.* 2005)). At zero detuning, i.e. when  $V_L - V_R = \epsilon = 0$ ,  $J$  at  $r = 3$  is larger than that at  $r = 1$  and  $r = 5$ . As  $\epsilon$  increases, for  $r = 1$ , the exchange energy first maintains a small value and then increases as was recently observed both experimentally (Petta *et al.* 2005) and theoretically (Stopa and Marcus 2006). For  $r = 3$ ,  $J$  increases (decreases) monotonically for  $\epsilon$  smaller (larger) than 2.79 meV at which a sharp *kink* followed by a local minimum in  $J$  occurs. With increasing  $r$ , electron localization in the left dot occurs at progressively smaller detuning, as it becomes easier for the system to lower its energy in a configuration where the two electrons are located at the opposite “ends” of the elliptical QD (Fig. 2.13(b)) to minimize



**Fig. 2.13** (a) Two-dimensional density plots for  $r = 1$  at zero magnetic field. Columns (I), (II) and (III) correspond to  $V_L = 21$ , 29 and 42 meV, respectively. First (second) row, labelled S (T), is for the lowest singlet (triplet) states. (b) Same as (a) but for  $r = 3$ . Columns (I), (II) and (III) correspond to  $V_L = 25$ , 25.45 and 25.47 meV, respectively.

their Coulomb interaction. As a result, both singlet and triplet electrons move into the left dot at close  $V_L$  values. Depending upon the difference between the  $V_L$  required to localize singlet and triplet states into one dot, the crossing between the two lowest triplet states associated with the electron density rotation manifests itself as a kink ( $r \sim 1$ ) or a sharp maximum ( $r > 1$ ) in the exchange coupling at intermediate values of detuning (Zhang *et al.* 2008b). For  $r \gg 1$ ,  $J$  decreases monotonically with  $\epsilon$  due to the involvement of several SP states in the formation of the singlet state (analogously to the strongly coupled QDs, see Section 2.2.2) and decreased difference between electron–electron interactions in the singlet and triplet states in a single QD, i.e. in this case the exchange energy is smaller in one very long QD than in two coupled QDs.

The detuning effects on the exchange energy can be further illustrated by inspecting the electron-density behavior. In Fig. 2.13(a) we see that for coupled circular QDs ( $r = 1$ ) with increasing  $\epsilon$  electrons gradually move into the lower (left) QD for both singlet and triplet states, albeit more quickly for the singlet state. When both electrons are in the left dot (Fig. 2.13(a), column (III)), the singlet density has a single maximum corresponding to the state with nearly zero angular momentum (Wagner *et al.* 1992), while the triplet state exhibits two maxima along the QD coupling direction ( $x$ ) as the constituent QDs are slightly more extended in the  $x$ -direction than in the  $y$ -direction. For  $r = 3$ , as  $\epsilon$  increases, the singlet density gradually localizes into the left dot with the electron density showing two maxima in the  $y$ -direction for  $V_L > 25$  meV because of the relaxed confinement in that direction (see Fig. 2.13(b) columns (I) and (II)). For the triplet state, however, it is seen that the electron density abruptly changes from being spread over the two QDs with a higher peak in the left dot ( $V_L = 25.45$  meV, column (II) in Fig. 2.13(b)) to occupying only

the left dot with two peaks of equal height in the  $y$ -direction ( $V_L = 25.47$  meV, column (III) in Fig. 2.13(b)). This abrupt transition (or “rotation”) of the electron density occurs at  $V_L = 25.46$  meV.

In order to better understand the difference in the dependences of the energy levels on  $\epsilon$  between the QD configurations with  $r = 1$  and  $r = 3$ , we compute the expectation values of the parity operator

$$\langle \hat{P} \rangle = \langle \Psi(\mathbf{r}_1, \mathbf{r}_2) | \Psi(-\mathbf{r}_1, -\mathbf{r}_2) \rangle, \quad (2.8)$$

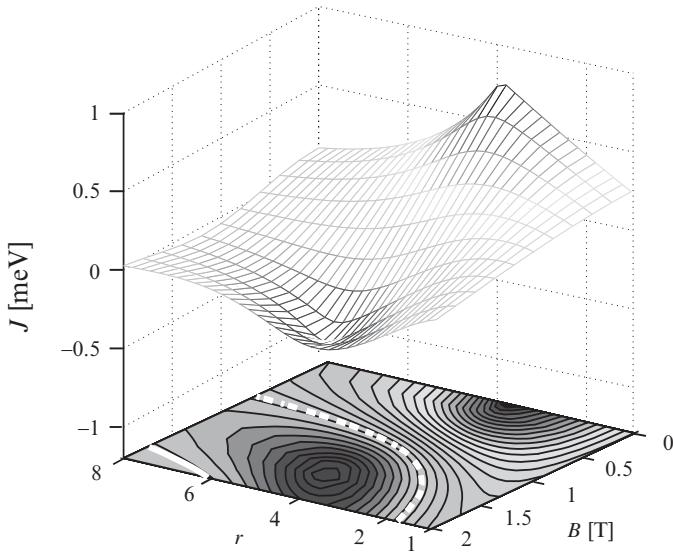
and of the parity operator with respect to the  $y$ -axis

$$\langle \hat{P}_y \rangle = \langle \Psi(x_1, y_1, x_2, y_2) | \Psi(x_1, -y_1, x_2, -y_2) \rangle. \quad (2.9)$$

Here,  $\mathbf{r}_i = (x_i, y_i)$ ,  $i = 1, 2$  are the electron coordinates. In the investigated detuning range, for the singlet state  $0 < P < 1$  and  $P_y = 1$  for all  $r$  while for the triplet state at  $r = 1$   $P < 0$  and  $P_y = 1$ , as expected from the symmetry of the Hamiltonian. However, in the triplet state for  $r = 3$  at  $V_L = 25.46$  meV, our calculation shows that the value of  $\langle \hat{P} \rangle$  sharply increases from  $-0.98$  to  $-0.02$  while  $\langle \hat{P}_y \rangle$  changes sign from  $1$  to  $-1$ . This indicates that the parity of the singlet state with respect to the  $y$ -axis remains even, while the triplet wavefunction parity changes from even to odd, thereby changing its symmetry along the  $y$ -direction. In fact, we find that as the interdot detuning is increased, the two lowest singlet states at any  $r$  have the same symmetry, and their energy levels anticross. The same is true for the two lowest triplet states at  $r = 1$ , whereas the energy levels of the two lowest triplet states at  $r > 1$  cross due to their different symmetries, in agreement with the von Neumann–Wigner theorem (von Neumann and Wigner 1929; Landau and Lifshitz 1977). According to this theorem, energy levels cross for states that bear different symmetries, while they anticross for states with the same symmetry. This fundamental theorem has found successful applications in the spectroscopy of alkaline salts (Bron and Wagner 1966), the interpretation of the Zeeman spectrum of hydrogen molecules (Lichten 1971), and more recently in the explanation of the magnetic anisotropy in thin metallic films (Pick and Dreyssé 1993).

Near the crossing point of the two lowest triplet energy levels, the analysis of the spectral function (Melnikov and Leburton 2006) reveals that the first excited single-particle state localized in the left QD has the  $p_y$ -character. Before the level crossing, the  $p_y$ -like orbital is unoccupied so that the electronic states form a  $sp_x$ -pair consisting of the  $s$ -like state mostly localized in the left QD and the lowest antisymmetric  $p_x$ -like state spreading over both QDs. After the level crossing, it becomes energetically favorable for the triplet electrons to be in the same QD as an  $sp_y$ -pair, even though the expectation value of the Coulomb interaction in the triplet state increases from  $2.0$  meV before the transition  $sp_x \rightarrow sp_y$  to  $2.9$  meV afterwards.

In Fig. 2.14, we plot the exchange energy  $J$  as a function of the QD aspect ratio  $r$  and the magnetic field  $B$  applied perpendicular to the  $xy$ -plane (Helle *et al.* 2005). At fixed  $r$ , as the magnetic field increases,  $J$  first decreases and then increases again. It is seen that at intermediate  $r \sim 4$ ,  $J$  changes much faster with magnetic fields than at very small or large aspect ratios. We note that the strong magnetic field effect on the exchange energy at intermediate  $r$



**Fig. 2.14** Exchange energy  $J$  as a function of QD aspect ratio  $r$  and the magnetic field  $B$ . The thick white curves (solid and dashed) separate regions with  $J > 0$  and  $J < 0$ . In the investigated ranges of aspect ratios and magnetic fields,  $J$  assumes a maximum (minimum) value of 0.77 ( $-0.37$ ) meV at  $r \approx 3.9$ ,  $B \approx 0$  T ( $r \approx 4.4$ ,  $B \approx 1.6$  T).

is associated with the shape of the individual QD confinement potential, i.e. near  $r = 4$  the single-particle level separations in the  $x$ - and  $y$ -directions are comparable. With increasing  $r$ , the relative change of the exchange energy is small for  $B \sim 1$  T and much larger for  $B \sim 0$  T or  $B \sim 2$  T. The kink in  $J$  at  $B = 0$  due to the crossing of two lowest triplet levels (Fig. 2.12) does not exist for  $B \neq 0$  T because the magnetic field destroys the symmetry of the Hamiltonian (Zhang *et al.* 2008b). We note that the first singlet-triplet transition point (shown by the thick white dashed curve in Fig. 2.14), at which the exchange energy crosses zero, decreases as  $r$  increases and it does this more rapidly at smaller  $r$  ( $r < 4$ ) than at  $r > 4$ . At higher magnetic fields and larger aspect ratios,  $J$  changes sign again (thick solid white curve in Fig. 2.14) so that the ground state is a singlet (triplet) outside (in between) the two white curves.

## 2.2.4 Realistic double quantum dots

### 2.2.4.1 Hybrid multiscale approach

In this section we apply the exact diagonalization (CI) method of Section 2.2.1 to compute the exchange coupling  $J$  in a realistic two-electron double-dot system with a confinement potential  $V_{\text{conf}}(\mathbf{r})$  obtained from the 3D multiscale simulation (hybrid multiscale approach) of the planar laterally coupled QD molecules (Elzerman *et al.* 2003). We evaluate the confinement potential assuming zero electron occupation ( $N = 0$ ) of the QD region because this potential is produced only by external sources outside the QD region (such as gate voltages, ionized dopants, etc.). The interaction among the electrons in the QD is taken into account exactly during the diagonalization of the two-electron Hamiltonian eqn (2.1) in the second stage. Performing the two-stage calculations in this way, we neglect image charges and the polarization effects of the dot electrons in the outside regions. However, these effects are likely to

be very small for the two-electron case as the QD device usually operates in the regime of the strong depletion so that there is a large spatial separation between the charges in the QD and the external environment (Bruce and Maksym 2000; Melnikov and Leburton 2006).

To this end, we divide our simulation domain into the QD region where the electron dynamics is described fully quantum mechanically by the exact diagonalization of the corresponding many-particle Schrödinger equation (Section 2.2.1) and the outer region where the total charge density is determined from the semi-classical Thomas–Fermi electron  $n(\mathbf{r})$  and hole  $p(\mathbf{r})$  densities screened by the ionized donors  $N_D^+$  and acceptors  $N_A^-$  (Melnikov *et al.* 2006):

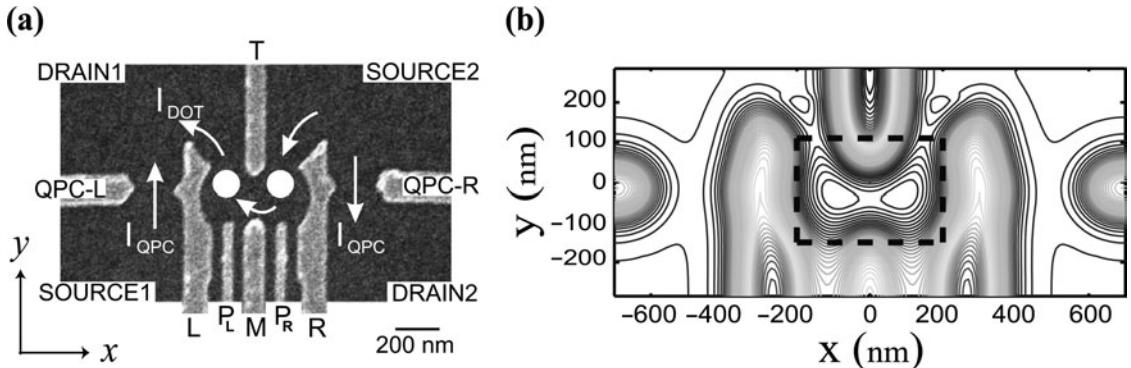
$$\rho(\mathbf{r}) = -e [N_D^+(\mathbf{r}) - N_A^-(\mathbf{r}) + p(\mathbf{r}) - n(\mathbf{r})]. \quad (2.10)$$

In order to extract the electron confinement potential, we solve the Poisson equation for the whole QD device:

$$\nabla \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = 4\pi \rho(\mathbf{r}), \quad (2.11)$$

where  $\rho(\mathbf{r})$  is the charge density defined above. The electrostatic potential  $\phi(\mathbf{r})$  consists of the external potential due to the applied gate voltage, and the screening potential arising from the ionized impurities in the structure (Melnikov *et al.* 2006).

Afterwards, eqn (2.1) is solved with the confinement potential  $V_{\text{conf}}(\mathbf{r}) = \phi(\mathbf{r})$  in the QD region. Note that for the lateral double QD structure, we obtain our confinement potential by taking a 2D cut of the total 3D electrostatic potential  $\phi(\mathbf{r})$  through the maximum of the wavefunction in the  $z$ -direction. Instead of the analytical expressions of the model confinement potentials, this potential and the associated single-particle matrix elements are computed numerically on the grid, while the rest of the matrix (the Coulomb matrix elements) is formed in the same way as discussed in the Section 2.2.1. The use of a 2D potential instead of a full 3D approach reduces considerably the computation time. On the other hand, the value of  $J$  is usually somewhat underestimated due to the enhancement of the Coulomb interaction in 2D geometry (Yannouleas and Landman 2002; Melnikov and Leburton 2006). However, the error is expected to be small as the extension of the electron wavefunction in the  $z$ -direction ( $\sim 20$  nm) is much smaller than the interelectron separation  $\sim 100$  nm (Zhang *et al.* 2004). As this 2D potential is obtained from the full 3D calculations, it already incorporates effects due to the images of the external charges in the gates. On the other hand, gate-screening effects (Bruce and Maksym 2000) on the electrons localized in the QDs are not considered (i.e. there is no self-consistency at this stage (Friesen *et al.* 2003)) since the screening length is larger than the distance between the electric gates and the electronic distribution in the QDs (the QD is sufficiently well isolated from the environment). Note also that the influence of this correction on  $J$  is in general strongly reduced as constant background effects tend to cancel out in energy differences between the two spin states.



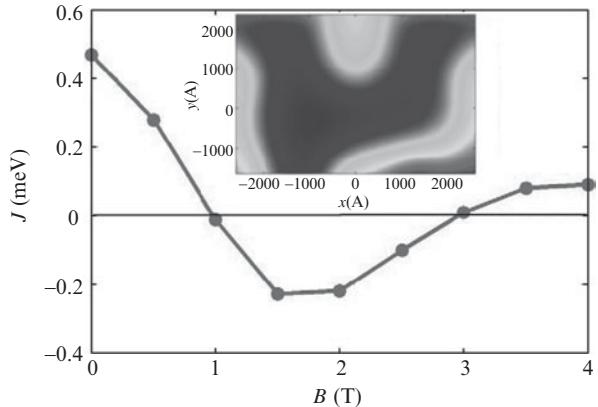
**Fig. 2.15** (a) SEM of the top gates of the planar coupled QD device. Light gray areas show the gate pattern, circles show the dots, curved arrows show the possible charging current paths, and straight arrows show the QPC currents. (Reprinted with permission from Elzerman *et al.* (2003). Copyright 2003 by the American Physical Society.) (b) Contour plot in the  $xy$ -plane at the 2DEG interface where the dashed rectangle shows the dots' location. (Reprinted with permission from Zhang *et al.* (2004). Copyright 2004 by the American Physical Society.)

#### 2.2.4.2 Coulomb localization and exchange energy in coupled QDs

We apply our multiscale method described above to study the two-electron system in the experimental structure consisting of two laterally coupled QDs (LCQD) (Qin *et al.* 2001; Elzerman *et al.* 2003; DiCarlo *et al.* 2004; Petta *et al.* 2004, 2005). Figure 2.15(a) shows a scanning electron micrograph (SEM) view of the metal gates on top of the LCQD circuit (Zhang *et al.* 2004). The two QDs, shown by white circles, are defined by properly biasing the L, R, T and M gates. The coupling between the two dots is controlled by the T and M or L and R gates. The PL and PR gates, called the “plungers,” are used for fine tuning the potential energy in the two dots, thus controlling the single-electron charging in the LCQD. Two quantum point contact (QPC) detectors are formed by the tips between the QPC-L (QPC-R) and L(R) gates. The charging paths into the two dots are shown by the curved arrows, and the QPC currents are shown by the straight arrows. In this structure the 2D electron gas is formed 90 nm below the top gates, while a 50-Å thick n-type GaAs layer ( $N_D = 1.5 \times 10^{18} \text{ cm}^{-3}$ ) is located right below the surface (Elzerman *et al.* 2003; Zhang *et al.* 2004; Stopa and Marcus 2006).

Using the above method we first performed calculations characterizing individual QDs (Melnikov and Leburton 2007) comprising the experimental double-dot structure by changing the right plunger voltage  $V_{PR}$  and keeping the rest of the voltages constant, and found a reasonable agreement with available experimental measurements (Elzerman *et al.* 2003): Our calculations gave the electron addition energy to the individual QD of 4.2 meV and SP excitation energy of 1.7 meV. The exchange energy as a function of the magnetic field is shown in Fig. 2.16 where one can see that  $J \approx 0.5$  meV at zero magnetic field while the singlet-triplet transition occurs at about 1 T, again in close agreement with experiments. The exchange energy also becomes positive at larger magnetic fields analogous to the behavior of  $J$  found in circular QDs (Wagner *et al.* 1992).

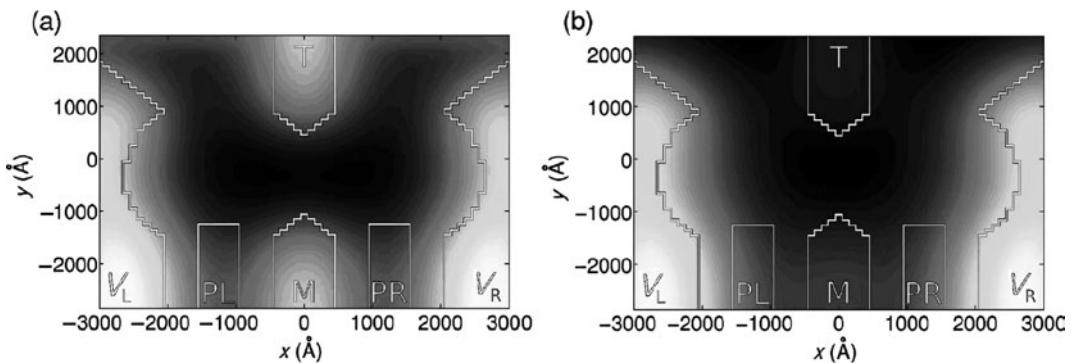
Next, by varying the L and R gate voltages  $V_L = V_R$ , while keeping the rest of the gates at fixed voltages (T and M gates are at  $-0.6$  V, while both plungers



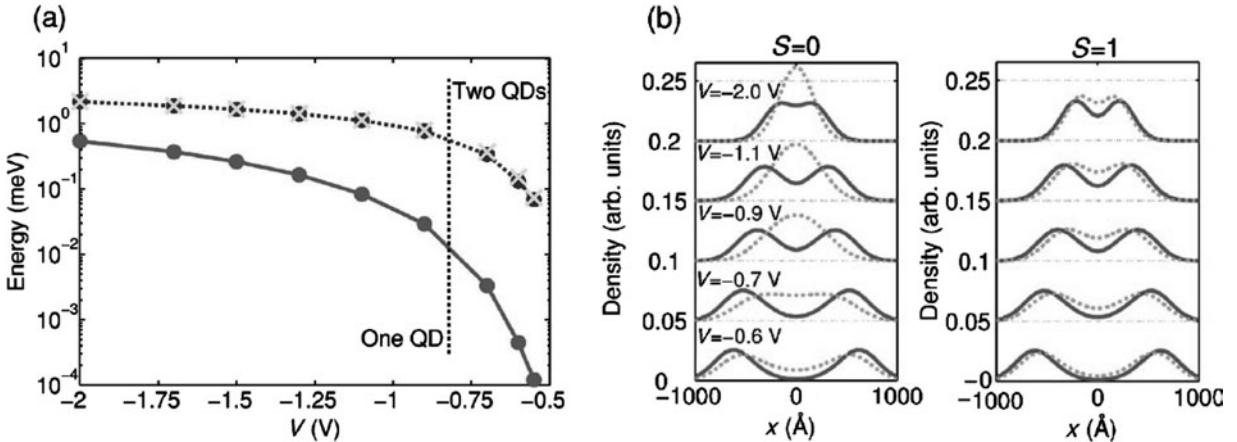
**Fig. 2.16**  $J$  vs. magnetic field  $B$  in a single QD for the L/R gate biases  $V_L = V_R = -0.5$  V, PL is at  $V_{PL} = -0.15$  V, PR is at  $V_{PR} = -2$  V, and T and M gates are at  $-0.6$  V. Inset shows the 2D profile of the confinement potential with the absolute minimum of the energy at  $x \sim -100$  nm and  $y \sim -20$  nm. (From Melnikov and Leburton (2007). Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

are set to zero), the system is changed from a large single QD formed in the center of the structure at  $V_L = V_R = -2.0$  V to two coupled smaller QDs with an interdot separation of about 150 nm and a tunnel barrier of  $\sim 1.5$  meV at  $V_L = V_R = -0.5$  V (Fig. 2.17). At  $-2.0$  V (Fig. 2.17(a)), the potential has an approximately elliptic shape with confinement strengths  $\hbar\omega_x \sim 2.1$  meV and  $\hbar\omega_y \sim 2.8$  meV and the minimum located in the center of the structure around  $x \sim 0$  and  $y \sim 0$ . The electrons are prevented from leaking into the outside regions by the L, R gates and by constrictions between the L, R and top gates. When the L, R gate voltages are made more positive, the tunnel barrier between the two QDs finally emerges at about  $V_L = V_R \sim -0.8$  V, and the overall confinement potential profile in the QD region achieves a bowtie (or butterfly) shape with the inversion symmetry in the  $x$ -direction only (Fig. 2.17(b)).

Such significant changes in the confinement potential geometry exert a large effect on the exchange energy as well as other quantities of interest, as can be evidenced from the data obtained at zero magnetic field and shown in Fig. 2.18. In Fig. 2.18(a) as  $V_L = V_R$  is swept from  $-2.0$  to  $-0.55$  V, the



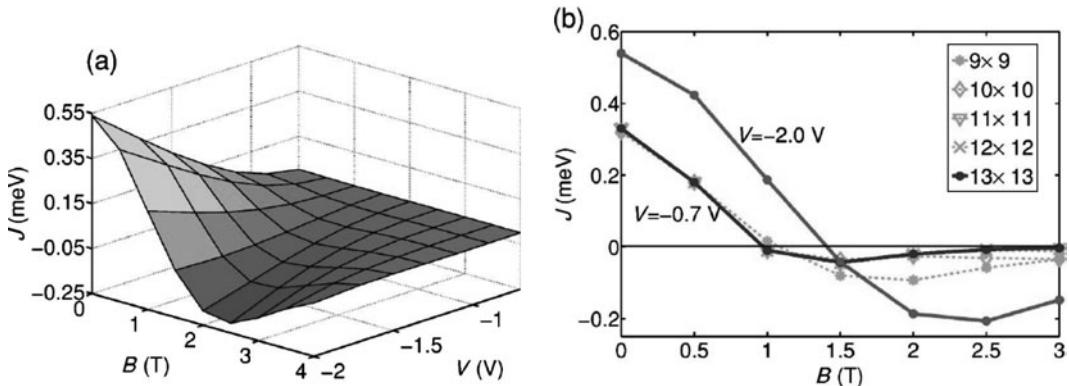
**Fig. 2.17** Confinement potential in the QD region at  $V_L = V_R = -0.5$  V (a),  $-2.0$  V (b). The white contours show the discretized top gate layout. Symbols T, M, PL, and PR stand for top, middle, and left and right plunger gates, respectively. (Reprinted with permission from Melnikov *et al.* (2006). Copyright 2006 by the American Physical Society.)



**Fig. 2.18** (a) Exchange energy  $J$  (solid) and the tunnel coupling  $2t$  (dashed curve and symbols) as functions of the L/R gate bias  $V = V_L = V_R$  at zero magnetic field. The dashed curve is the result of the present hybrid approach, while the symbols give the values of  $2t$  from the 3D calculations. (b) Electron-density profile along the  $x$ -direction for the interacting (solid) and non-interacting (dashed) electrons in the singlet ( $S = 0$ ) and triplet ( $S = 1$ ) states for  $V = V_L = V_R = -2.0, -1.1, -0.9, -0.7$ , and  $-0.6$  V. (Reprinted with permission from Melnikov *et al.* (2006). Copyright 2006 by the American Physical Society.)

exchange energy  $J$  (solid curve) changes dramatically, by more than four orders of magnitude, from  $\sim 0.5$  meV to about  $120$  neV. At the same time, the difference between the two lowest single-particle states,  $2t$ , which would give the exchange energy in the non-interacting system of two electrons, exhibits a much slower decrease from  $2$  meV to  $28\ \mu\text{eV}$  (dashed curves). Note that the two curves for the SP energy separation in Fig. 2.18(a) obtained from the 2D hybrid and full-scale 3D DFT approaches (Zhang *et al.* 2004; Melnikov *et al.* 2005) are very close to one another, thus validating our assumption about the use of a 2D potential. 3D calculations also give rise to a decrease in the electron–electron interaction that is more pronounced in the triplet state. This effectively leads to a 15% increase in the exchange energy, consistent with our results for single QDs (Melnikov and Leburton 2006). The computed interval of  $2t$  values overlaps with the energy range found experimentally for the interdot coupling strengths (van der Wiel *et al.* 2002; Hayashi *et al.* 2003; Petta *et al.* 2004; Hüttel *et al.* 2005; Pioro-Ladrière *et al.* 2005). The large disparity between the values of  $2t$  and  $J$  demonstrates the rising importance of the Coulomb repulsion in the system when the interdot distance is increased and electrons in both singlet and triplet states become localized in the individual QDs. This is also confirmed by inspection of the electron-density plots (Fig. 2.18(b)), where one notices that the distance between the two density maxima is significantly smaller (or the overlap is larger) in the case of two non-interacting electrons. It is also clearly seen in these plots that electrons in the triplet state are farther from each other, which is due to the Pauli principle that tends to pull same-spin particles apart.

The exchange energy in magnetic fields is plotted in Fig. 2.19 for different voltages. It is seen that with increasing voltage (decreasing interdot coupling), the values of the exchange energy decrease as expected, and  $J$  becomes negative at about  $1.5$  T in the single-dot regime (Fig. 2.19(a)). The so-called



**Fig. 2.19** (a) Exchange energy  $J$  as a function of the L/R gate bias  $V = V_L = V_R$  in the magnetic field  $B$ ; (b)  $J$  vs.  $B$  at  $V = V_L = V_R = -2.0$  V and  $-0.7$  V. For the latter case, the values of  $J$  are multiplied by 100 to bring them to scale. For this case several curves computed with different numbers of basis states (dashed) are also shown to demonstrate the convergence in the exchange energy values with increasing basis set. (Reprinted with permission from Melnikov *et al.* (2006). Copyright 2006 by the American Physical Society.)

singlet–triplet transition gradually shifts to  $\sim 1$  T when the voltage becomes more positive and the QDs decouple. The decrease can be explained by the rising influence of the Coulomb interaction compared to the SP effects (Section 2.2.1). From the results displayed in Fig. 2.19(b), one also deduces that the ratio of  $J(B = 0)$  to the exchange energy (magnitude) at its minimum (negative) value is much larger in the double-dot regime ( $\sim 8$  at  $V_L = V_R = -0.7$  V vs.  $\sim 3$  at  $V_L = V_R = -2.0$  V), i.e. the negative- $J$  region becomes suppressed, similarly to the situation in single-elliptic QDs (Melnikov and Leburton 2006) and model double-dot systems (Melnikov and Leburton 2006).

Including Zeeman interaction in the two-electron Hamiltonian eqn (2.1) leads to a lowering of the triplet energy by  $\sim 25$   $\mu\text{eV}/\text{T}$ , thereby strongly affecting our results only in the limit of the weak interdot coupling where the singlet–triplet transition shifts from 1 T to about 20 mT. However, the exchange energy values obtained for  $V_L = V_R = -0.55$  V even at small magnetic fields are comparable to the Zeeman splitting induced by nuclear field  $B_{\text{nuc}} \sim 3$  mT (Petta *et al.* 2005) which effectively mixes singlet and triplet states, and, as such, direct measurement of this energy in this voltage interval can be difficult.

## 2.3 Two electrons in quantum wire quantum dots

### 2.3.1 Variational Heitler–London method

In this section, we compute the electronic structure of coupled QWQDs containing two electrons with a variational Heitler–London (VHL) method. We first construct a 3D model confinement potential for the QWQDs and introduce three variational parameters in the HL wavefunctions (see Appendix B) that account for the specific 3D confinement profile. The material under consideration is InAs, for which we use the electron effective mass  $m = 0.023 m_0$

(Hansen *et al.* 2005) and dielectric constant  $\varepsilon = 14.6$ . We assume a parabolic confinement potential in the  $xy$ -plane  $V_{\text{conf}}(\rho) = m\Omega_\rho^2(x^2 + y^2)/2$ , where we take  $\Omega_\rho = \hbar/m(D/2)^2$  with  $D$  being the nominal value of the wire diameter (see Fig. 2.3). In the  $z$ -direction along which the QDs are coupled, the confinement potential is modelled by a linear combination of three Gaussians:

$$V_{\text{conf}}(z) = -V_0 \left\{ \exp \left[ -\frac{(z-d)^2}{l_z^2} \right] + \exp \left[ -\frac{(z+d)^2}{l_z^2} \right] \right\} \\ + V_b \exp \left( -\frac{z^2}{l_{bz}^2} \right), \quad (2.12)$$

where  $V_0$  gives the depth of two Gaussian wells describing the confinement of the two individual QDs (we fix  $V_0 = 20$  meV),  $V_b$  controls the barrier height between the two dots ( $V_b = 0$  except otherwise specified),  $l_z$  is the radius of each QD,  $2d$  is the nominal separation between the two QDs, and  $l_{bz}$  denotes the radius of the tunnelling barrier between the QWQDs.

For the single-particle wavefunctions in eqn (2.25) we use

$$\varphi_{L/R}(\mathbf{r}) = \left( \frac{m\omega_\rho}{\pi\hbar} \right)^{\frac{1}{2}} \exp \left[ -\frac{m\omega_\rho}{2\hbar}(x^2 + y^2) \right] \\ \times \left( \frac{m\omega_z}{\pi\hbar} \right)^{\frac{1}{4}} \exp \left[ -\frac{m\omega_z}{2\hbar}(z \pm a)^2 \right]. \quad (2.13)$$

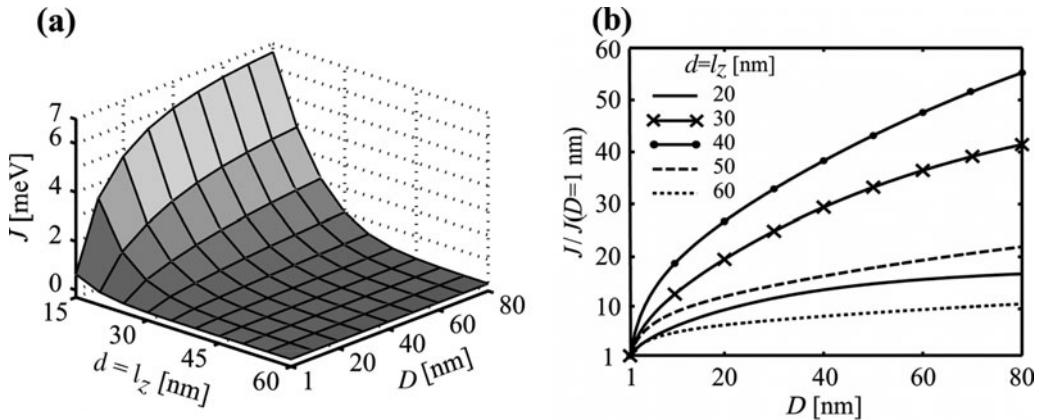
In our variational HL (VHL) approach, we use the effective inplane confinement strength  $\omega_\rho$ ,  $z$ -direction confinement strength  $\omega_z$  and effective half-interdot separation  $a$  as variational parameters to minimize the system energies (Zhang *et al.* 2008a). By fixing these variational parameters equal to their nominal values  $\omega_\rho = \sqrt{\Omega_\rho^2 + \omega_c^2}$  with  $\omega_c = eB/mc$ ,  $\omega_z = \Omega_z = \sqrt{2V_0/ml_z^2}$  and  $a = d$ , we recover the results from the conventional HL method. Using both HL and VHL methods, we calculate the exchange coupling  $J = E^T - E^S$  with  $E_T$  and  $E_S$  being the expectation values of the two-particle Hamiltonian (see Section 2.2 and Zhang *et al.* 2008a).

As an inherent drawback of the original HL method, our variational scheme breaks down when the overlap between the localized states is large, which occurs for small inter-dot separations. For example, in our calculations of the system energies, the VHL method fails for  $(d = l_z) < 12$  nm independent of  $D$ . A signature of the VHL approach breakdown at small  $d$  is that the variational parameter  $a$  becomes zero during the minimization process. This numerical behavior stems from the fact that at small  $d$  a global minimum in the system energies does not exist for the physical range of  $a$ , given the variational wavefunction eqn (2.25). We note that this shortcoming in the HL method is seemingly not present in the conventional HL approach, and as long as  $(d = l_z) > 0$ , one can still use the HL method (without variational analysis) to calculate the system energies even though the obtained results are likely to be not very accurate.

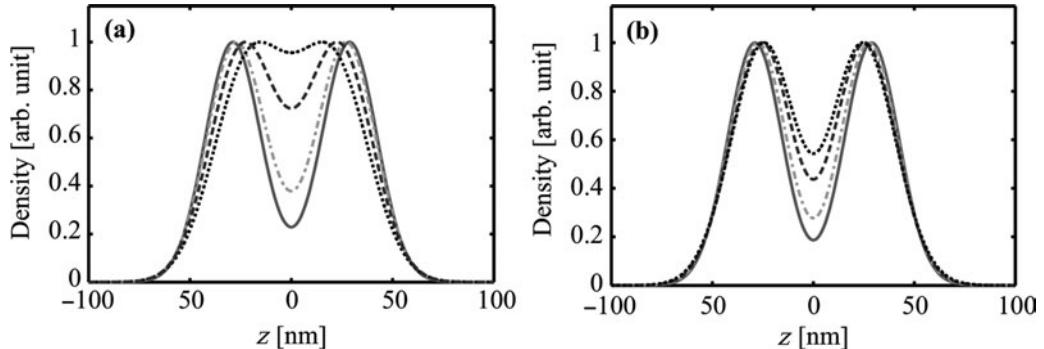
It was pointed out (Calderón *et al.* 2006) that the HL method breaks down as the parameter  $c = \sqrt{\pi/2}(e^2/\epsilon a_B)/\hbar\omega_0$  ( $a_B = \sqrt{\hbar/m\omega_0}$ ) becomes larger than 1.95, 2.8, and 5.8 for coupled QDs with harmonic oscillator confinement of strength  $\hbar\omega_0$  in each direction for 1D, 2D and 3D potentials, respectively. Below, we considered  $l_z$  from 15 to 60 nm corresponding to  $c$  ranging from 0.44 to 0.87 that are uniformly smaller than the smallest breakdown value  $c = 1.95$ . However, as a consistency check, we also extended our calculation to very large value of  $d = l_z$  and found that for  $c = 1.723$  ( $D = 20$  nm,  $(d = l_z) = 200$  nm and  $\hbar\omega_z = 1.553$  meV), which is close to the above 1D limit of  $c = 1.95$ ,  $J \sim 10^{-14}$  meV bearing no practical interest.

### 2.3.2 Size modulation of the exchange energy

Figure 2.20(a) displays the exchange coupling  $J$  as a function of both the wire diameter  $D$  and the half-separation ( $d$ ) between the two QDs. Here, we set  $d = l_z$  because in experiments coupled QWQDs are defined by a linear gate grid (Fasth *et al.* 2005), which produces approximately the same effective QD size and interdot separation. For the confinement potential given by eqn (2.12), this configuration leads to a constant effective barrier height of 5.68 meV, independent of the value of  $d = l_z$ . The nominal confinement strength for a single Gaussian well ( $V_0 = 20$  meV) with  $l_z = 15$  nm and 60 nm is  $\hbar\Omega_z = 24.27$  and 6.07 meV, respectively. For a wire diameter  $D = 1$  nm, the nominal confinement strength in the  $xy$ -plane is  $\hbar\Omega_\rho = 1.33 \times 10^4$  meV, which physically corresponds to the *quasi-1D limit* for the systems with aspect ratio  $(\lambda_\rho/\lambda_z = \sqrt{\Omega_z/\Omega_\rho}) < 0.05$  in the investigated range of  $d = l_z$  from 15 to 60 nm. In the opposite limit, where  $D = 80$  nm,  $\hbar\Omega_\rho = 2.07$  meV and the aspect ratio  $\lambda_\rho/\lambda_z > 1.71$ . In Fig. 2.20(a) we can see that at fixed  $D$  the exchange energy exhibits rapid decay with decreasing  $d = l_z$ . It is also observed in this figure that  $J$  decreases with decreasing  $D$  at fixed  $d = l_z$ . This trend is shown explicitly in Fig. 2.20(b) for different  $d = l_z$ . For comparison,



**Fig. 2.20** (a) Exchange coupling  $J$  as a function of wire diameter  $D$  and half-separation between the QDs  $d$ , which is set equal to QD radius  $l_z$  ( $d = l_z$ ). (b)  $J$  as a function of  $D$  for different  $d = l_z$  values (shown in the figure). In each curve  $J$  is normalized to its value at  $D = 1$  nm. For  $(d = l_z) = 20, 30, 40, 50, 60$  nm,  $J(D = 1 \text{ nm}) = 2.33 \times 10^{-1}, 2.47 \times 10^{-2}, 3.53 \times 10^{-3}, 1.37 \times 10^{-3}, 4.81 \times 10^{-4}$  meV, respectively.



**Fig. 2.21** Electron density in the  $z$ -direction for (a) singlet and (b) triplet states at  $d = l_z = 30$  nm. In each plot, density is plotted at  $D = 1$  nm (solid),  $D = 10$  nm (dashed-dotted),  $D = 40$  nm (dashed) and  $D = 80$  nm (dotted). For each  $D$ , the density is normalized to its peak value.

the data on each curve are normalized to the value of  $J$  at  $D = 1$  nm. At fixed  $d = l_z$ , as  $D$  is decreased from 80 nm,  $J$  decreases, and the rate with which it decreases becomes larger as  $D$  approaches 1 nm, which is the quasi-1D limit. The faster rate of decrease in  $J$  near  $D = 1$  nm is due to the large value of  $\Omega_\rho \propto 1/D^2$  leading to a stronger effect of the variation in  $\Omega_\rho$  on values of  $J$  in the interval of small  $D$  (through the enhanced Coulomb interaction that diverges in the 1D case (Das Sarma and Lai 1985)). We also note that although the general trend of  $J$  is to decrease as  $D$  is made smaller, the decreasing rates are much larger for intermediate  $d = l_z$  values than for small or large values.

These effects of the wire-diameter variation on the exchange coupling are rather unexpected as they show that  $J$  depends on the wire confinement perpendicular to the coupling direction. In fact, we find that the variation of  $D$  not only changes  $\omega_\rho$  but also induces significant changes in  $\omega_z$  and  $a$ , both of which minimize the singlet and triplet state energies. One can directly visualize such changes by inspecting the electron density variation with respect to the wire diameter. In Fig. 2.21, we plot the electron density for different values of  $D$  ( $d = l_z = 30$  nm) for the singlet and triplet states. For both states, as  $D$  decreases, the separation between the two density peaks becomes larger (the Coulomb repulsion is enhanced when the system approaches the 1D limit) and the width of each peak becomes smaller. Consequently, the overlap between the two electrons is reduced and with it, the exchange energy.

### 2.3.3 Comparison with experiments

In recent experiments on InAs QWQDs, values of the exchange energy  $J$  from 2.8 to 3.2 meV were reported for a single QD formed in a wire with effective harmonic confinement strength  $\hbar\Omega_z = 6.3$  meV (corresponding to a confinement length  $2\lambda_z = 2\sqrt{\hbar/m\Omega_z} = 46$  nm) and  $\hbar\Omega_\rho = 40$  meV ( $2\lambda_\rho = 2\sqrt{\hbar/m\Omega_\rho} = 18$  nm) (Fasth *et al.* 2005; Fuhrer *et al.* 2007). By fitting these values in our model ( $D = 18$  nm,  $V_0 = 41.6$  meV,  $V_b = 0$  meV,  $d = 0$  nm and  $l_z = 117.9$  nm), we obtain  $J = 3.51$  meV, which is comparable to the experimental result. This indicates that our model is capable not only of qualitative but quantitative predictions as well.

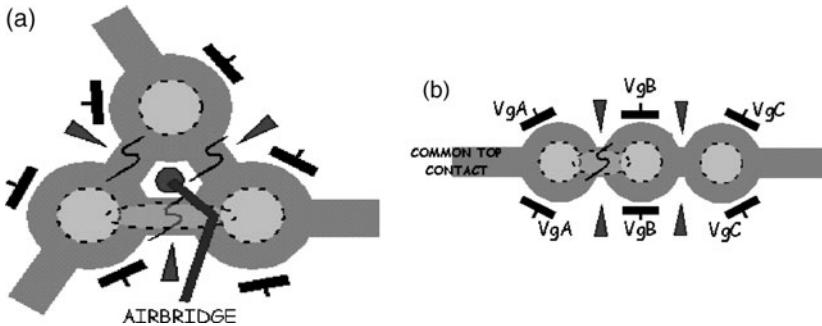
We note that  $J \sim 3$  meV as obtained above is actually the result for a single QD with a potential minimum at  $z = 0$  (Sako and Diercksen 2007; Agarwal *et al.* 2008). For double QDs with realistic parameters  $D = 20$  and  $d = l_z = 30$  nm (Fasth *et al.* 2005), we obtain  $J \sim 0.5$  meV, which translates into a characteristic SWAP time of  $(\tau_J = \hbar/J \sim 1.3$  ps. The value of  $\tau_J$  is on the same order as the reported spin decoherence time  $T_2 = 0.5 - 1$  ps in InAs QWQDs (Hansen *et al.* 2005) and is much smaller than the reported spin-dephasing time  $T_2^* = 50 - 500$  ps in self-assembled InAs QDs (Merkulov *et al.* 2002; Braun *et al.* 2005). This indicates that the experimental structures based on InAs QWQDs may benefit from the relatively large exchange coupling when utilized as building blocks for quantum computing applications.

## 2.4 Few electrons in triple quantum dots

As mentioned in Section 2.1, triple QD (TQD) structures, that can also be regarded as artificial triatomic molecules, have not yet received a lot of attention but a number of possible applications (Saraga and Loss 2003; Sasakura *et al.* 2004; Vidan *et al.* 2004; Schröer *et al.* 2007) have already been envisioned for them. We model the TQD structure with a Gaussian-shaped confinement potential that is similar to the one used in Section 2.2.3 (eqn (2.7)):

$$V_{\text{conf}}(\mathbf{r}) = - V_L \exp \left[ -\frac{(x + d_x/2)^2 + (y + d_{yL})^2}{R_L^2} \right] \\ - V_R \exp \left[ -\frac{(x - d_x/2)^2 + (y + d_{yR})^2}{R_R^2} \right] \\ - V_C \exp \left[ -\frac{x^2 + (y - d_{yC})^2}{R_C^2} \right], \quad (2.14)$$

where  $V_L$ ,  $V_C$ , and  $V_R$  determine the potential depths of the three QDs,  $(-d_x/2, -d_{yL})$ ,  $(0, d_{yC})$ ,  $(d_x/2, -d_{yR})$  indicate the  $(x, y)$  coordinates of the potential minima corresponding to the left, the center, and the right dot, and  $R_L$ ,  $R_C$ , and  $R_R$  specify the confinement strengths of the respective the dots. The interdot separation,  $d_x = 50$  nm in all of our simulations with  $N = 2$ . We focus on two configurations of the TQDs: (1) triangular TQDs in which the QDs are in an equilateral triangular formation along the  $xy$ -plane (Fig. 2.22(a)), and (2) linear array TQDs in which the centers of the QDs are located along a single line (Fig. 2.22(b)). For the triangular TQDs  $d_{yL} = d_{yR} = d_x/\sqrt{3}$ , and  $d_{yC} = 2d_{yL}$  making the three QDs equal in size, whereas for the linear array TQDs,  $d_{yL} = d_{yC} = d_{yR} = 0$ . In both configurations, we fix  $V_L = V_R = 25$  meV and vary  $V_C$  from 0 to 40 meV, thereby introducing assymetry in the TQD system. We keep  $R_L = R_C = 30$  nm and set  $R_C = 20, 25, 30$ , and 35 nm in the triangular TQDs providing additional deformation to the system. For the linear array QDs, we consider the symmetrical case of three equal dots with  $R_L = R_C = R_R = 30$  nm.

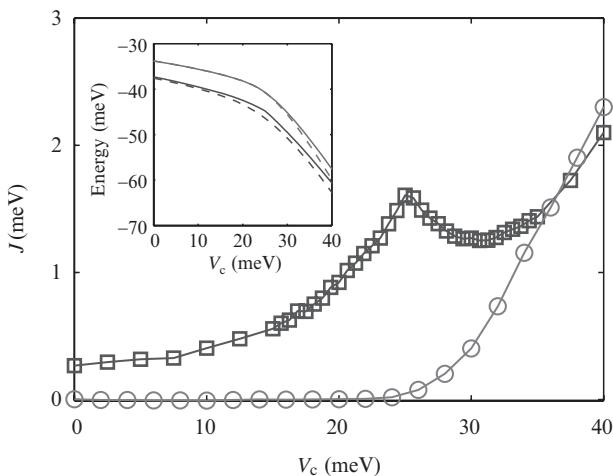


**Fig. 2.22** Schematic representation of (a) triangular and (b) linear array TQDs (courtesy of D.G. Austing).

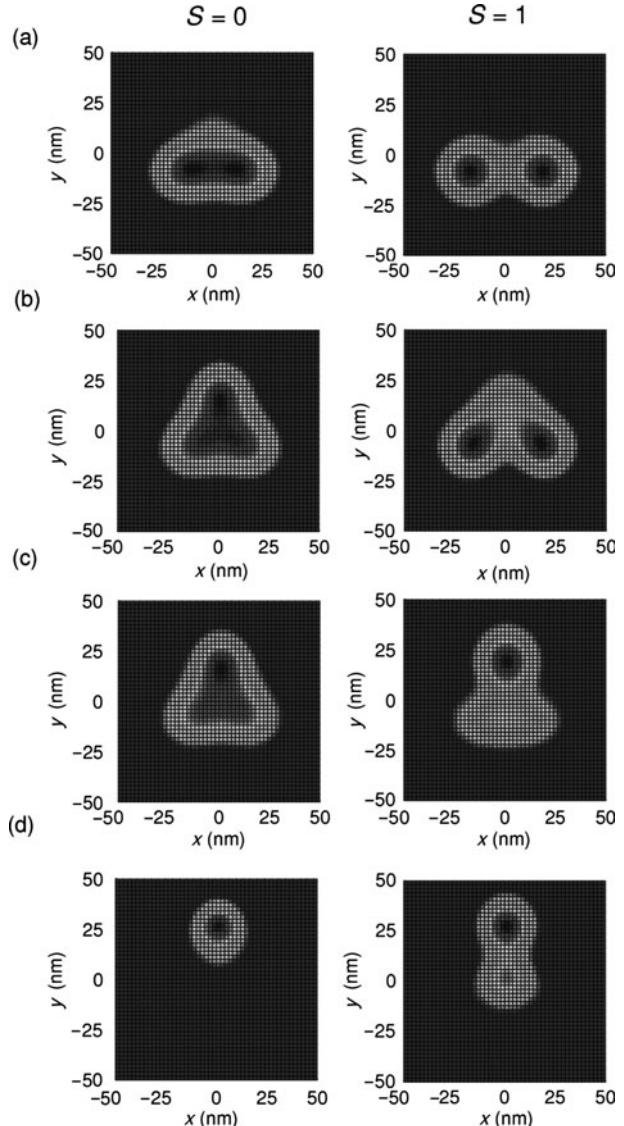
### 2.4.1 $N = 2$

First, we consider the situation when only two electrons are confined in the TQD. In Fig. 2.23, we plot the exchange energies  $J$  for both the triangular and the linear configurations with respect to  $V_C$  as obtained from the VMC calculations (Appendix A). We analyze the triangular system case first. At  $V_C \sim 0$  meV, the system resembles a double quantum dot with a relatively small value of  $J$  since the effective elimination of the center dot decouples the two electrons by pushing them into the left and right QDs. At  $V_C \sim 25$  meV, the three dots are similar to one another, thus making the system a “true” TQD. For small  $V_C$ , there exists strong localizations in the left and right dots for both the singlet and the triplet densities. With increasing  $V_C$  electrons gradually move into the center QD and exchange energy increases without reaching any saturation.

Compared to the double-dot configuration, it is difficult to a priori predict the electron configuration in the TQD. For  $V_C < 25$  meV, the exchange energy increases monotonically as the electron occupation of the center dot gradually increases. However, at around  $V_C = 25$  meV, there is a sharp kink in  $J$  (or a discontinuity in the  $J$ -derivative); the reason behind this kink is an abrupt



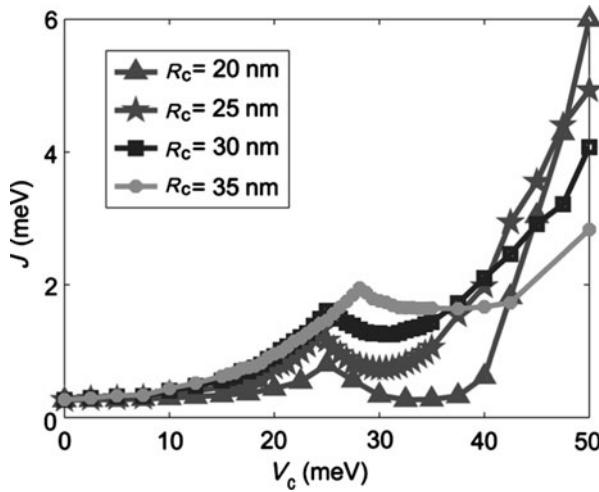
**Fig. 2.23** Exchange energy for triangular array TQD ( $\square$ ) and linear array TQD ( $\circ$ ) for  $N = 2$  as a function of the center dot depth  $V_C$ . The inset shows singlet (triplet) energy in dashed (solid) line for both the triangular (upper curves) and the linear (lower curves) TQD.



**Fig. 2.24** Two-dimensional density plots in triangular TQD for  $R_C = 30$  nm. (a),(b),(c), and (d) correspond to  $V_C = 20, 25, 25.625$ , and  $40$  meV, respectively. The left (right) column shows the singlet (triplet) electron densities.

change in the slope of the triplet energy as opposed to the smooth change in the singlet energy in  $V_C \sim 25$  meV region (see inset to Fig. 2.23).

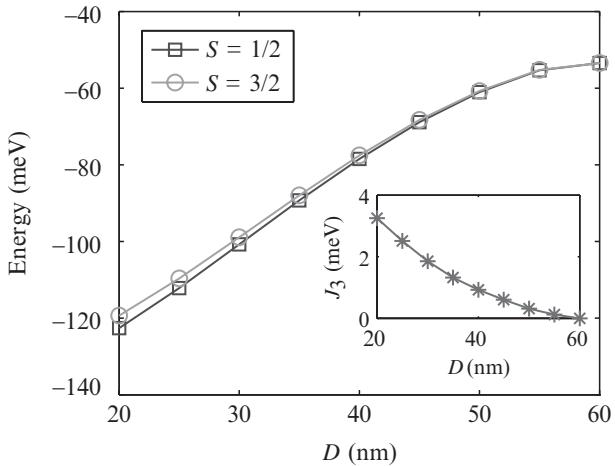
In order to further illustrate the cause of the kink occurrence, we plot the electron densities (Fig. 2.24) for singlet (triplet) states for  $V_C = 20, 25, 25.625$ , and  $40$  meV in the left (right) column. We keep other external parameters in the model potential constant. At  $V_C = 20$  meV (Fig. 2.24(a)), the electrons are pre-dominantly localized in the left and the right dots. As expected, the overlap between the singlet electrons is larger than the triplet electrons due to the Pauli principle. At  $V_C = 25$  meV (Fig. 2.24(b)), the singlet electron density is spread over all three dots, while the triplet electrons still reside mainly in the left and the right dot (at this point, due to symmetry in potential, the triplet is



**Fig. 2.25**  $J$  vs.  $V_C$  for  $R_C = 20, 25, 30$ , and  $35\text{ nm}$ .

actually triply degenerate with left-center and right-center occupations being the other two possible states). At  $V_C = 25.625\text{ meV}$  (Fig. 2.24(c)), the singlet electron density deviates only slightly from the  $V_C = 25\text{ meV}$  case, while for the triplet state, an abrupt transition occurs as the majority of the density can now be found in the center dot. The abrupt density transition phenomenon is similar to a situation observed in the double dot (see Section 2.2.3). For higher values of  $V_C$ , the system becomes effectively a single dot as the center-dot potential becomes much deeper than the potentials of the left or the right dots. Exchange energy is largest in this region since triplet electrons move away from the lowest orbital to occupy a higher-energy orbital (see Section 2.2.1). At  $V_C = 40\text{ meV}$  (Fig. 2.24(d)), both the singlet and the triplet electrons are in the center dot. We observe that the triplet densities are more elliptical than the singlet one due to the larger effect of potential deformation on the higher energy SP state occupied in the triplet state.

Another way to produce a deformation in the QD system is to change  $R_C$ . Figure 2.25 shows the exchange energy curves for  $R_C = 20, 25, 30$ , and  $35\text{ nm}$  in the triangular TQD. For large (small) values of  $R_C$ , both the depth of the center-dot potential well and the energy barrier between the center and the side dots decrease (increases). As a result, for large  $R_C$ , the system effectively becomes a single dot, whereas for small  $R_C$ , the system is a combination of two large and one small single QD. From Fig. 2.25, we see that for  $V_C = 0$ , all exchange-energy curves converge to the same value since at this point the potential eqn (2.14) is independent of  $R_C$ . Similarly to the above case, the discontinuity in the  $J$ -derivative persists for all values of  $R_C$  with  $J = 0.80, 1.31, 1.61$ , and  $1.95\text{ meV}$  at the kink for  $R_C = 20, 25, 30$ , and  $35\text{ nm}$ , respectively. We also observe that the  $V_C$  corresponding to the kink is the largest for  $R_C = 35\text{ nm}$  ( $V_C = 28.125\text{ meV}$ ) from all studied cases ( $V_C = 25, 24.375, 25\text{ meV}$  for  $R_C = 20, 25$ , and  $30\text{ nm}$ , respectively). This phenomenon can be explained by noting that due to a shallower potential in the center dot at large  $R_C$ , electron relocalization there requires larger  $V_C$ . Finally, for large  $V_C$ , the system becomes effectively a single dot and exchange is larger for small  $R_C$ .



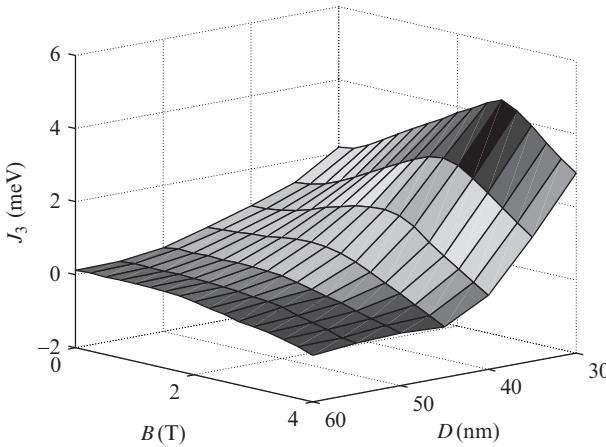
**Fig. 2.26** Doublet ( $S = 1/2$ ) and quadruplet ( $S = 3/2$ ) energies as functions of the interdot separation  $D$ . The inset shows  $J_3$  vs.  $D$ .

For small QDs (small  $R_C$ ), the harmonic confinement becomes strong and the SP energy spacing increases, leading to a stronger exchange coupling in that case.

#### 2.4.2 $N = 3$

Next, we turn our attention to the case of three electrons in the TQD. For the quantum-dot potential, we use the same form and the parameters as in the previous section, except that in our VMC calculations here we vary the interdot separation. For all results, we keep the rotational symmetry along the  $xy$ -plane intact by maintaining equal dot sizes. In the VMC trial wavefunctions eqn (2.21, 2.22), we include double and triple occupation terms for completeness, but the contributions of these terms are negligible due to large total energies associated with them. For  $N = 3$ , the two lowest states have total spins  $S = 3/2$  (quadruplet) and  $S = 1/2$  (doublet); the difference in total energy between the two states we define as exchange energy  $J_3$ , in analogy to the two-electron case.

Energies for  $S = 3/2$  and  $S = 1/2$  states are plotted in Fig. 2.26 with respect to the interdot separation ranging from 20 nm (single dot) to 60 nm (isolated triple dots). As can be seen from the inset curve, exchange energy decreases monotonically with respect to separation from 3.3 meV at 20 nm to  $\sim 0$  meV at 60 nm as expected since the electrons become more decoupled with separation. Similar to the two-electron case in double dots (Burkard *et al.* 1999), exchange energy  $J_3$  decays quasi-exponentially with respect to the interdot separation. Unlike the two-electron case where the  $S = 0$  state is always the ground state in zero magnetic field (Ashcroft and Mermin 1976), it remains unclear whether the lowest spin state is the ground state for three electrons at zero magnetic field. Thus far in our simulations, we have observed that  $S = 1/2$  is always the ground state.



**Fig. 2.27** Exchange energy  $J_3$  vs. magnetic field  $B$  and interdot separation  $D$ .

Figure 2.27 shows the exchange energies  $J_3$  for interdot separation varying from 30 to 60 nm for magnetic fields ranging  $B$  from 0 to 4 T. For interdot separation values less than 45 nm when the system can still be considered to be a single QD, we observe a local maximum at non-zero magnetic field. The magnetic field corresponding to the local maximum decreases with increased separation ( $B = 2.75$  T for 30 nm and  $B = 1.75$  T for 45 nm separation). For interdot separation values where the three dots become distinct and form a TQD (for separations greater than 45 nm), the local maximum disappears, and  $J_3$  decreases monotonically at low magnetic fields. The magnetic field at which a doublet–quadruplet transition occurs ( $J_3$  changes sign) is smaller for larger interdot separations ( $B = 2.7$ , 2.3, and 2.0 T for interdot distances of 50, 55, and 60 nm, respectively).

Since with increasing interdot separation the TQD system effectively changes from one to three QDs, this leads to two distinct  $J_3$  vs.  $B$  characteristics. In the single-dot regime, at non-zero magnetic field there is a crossing of two quadruplets for  $N = 3$  (Drouvelis *et al.* 2004) which introduces a non-monotonic behavior in the triplet energy leading to a local maximum in the exchange energy curves. However, in the triple-dot region, the energy levels do not cross one another in the studied interval of magnetic fields, and a monotonic decrease in the exchange energy curve is observed as a result. For high magnetic field, one expects the exchange energies to asymptotically go to zero for both the singlet and the triplet regimes due to decoupling of the electrons.

## 2.5 Conclusion

Artificial QD molecules offer high potential for quantum electronics and continue to be a fast-growing area of research. They provide excellent opportunities for investigating the basic physical properties of interacting many-body system. In this work the electronic properties of few-electron laterally coupled quantum dots were reviewed. The many-body problem is solved via several complementing approaches—variational Heitler–London, exact

diagonalization and variational Monte Carlo methods. Strong emphasis is placed on the investigation of a two-electron system confined in two coupled QDs. This particular problem attained large interest recently in view of its potential applicability for solid-state quantum computation. First, the properties of two interacting electrons in a model confinement potential are studied. It is shown that the standard Heitler–London approach does not work reliably for this problem, primarily because in order to realize a sizable coupling between the two electrons, a large overlap between electron orbitals is required that necessitates full-scale numerical simulations including many configurations of single-particle states. Next, we investigated the charge-stability diagram of the two coupled QDs and found the computed exchange energies to be significantly smaller than the values estimated from the stability diagram. Also, by analyzing the chemical-potential variation with external biases and magnetic fields, it becomes possible to distinguish between the weak and strong interdot couplings. The stability diagram of two elongated (in the direction perpendicular to the coupling) quantum dots reveals rich underlying physics. In particular, we observed density “rotations” with electron transfer to a single dot that give rise to sharp variations of the exchange coupling between electrons as a function of interdot detuning. In general, the behavior of exchange energy in this case is governed by the von Neumann–Wigner theorem that defines the variations of the energy spectrum from the symmetries of the molecular states during the detuning (or equivalently, QD elongation) process. We also studied the exchange coupling in realistic, state-of-the-art coupled QD devices and found that in these systems the electron coupling is dominated by the Coulomb repulsion of particles rather than tunnelling through a barrier separating QDs. This repulsion is responsible for extremely small values of exchange interaction and weak electron–electron correlations observed in realistic double-dot structures. Moving beyond two QDs, we considered triple QDs populated with two and three electrons and found that here the peculiar electronic properties can also be understood within the von Neumann–Wigner theorem. Finally, our study of two qubits in coupled QDs formed in a quantum wire performed within the variational Heitler–London approach revealed that decreasing the wire diameter leads to a fast decay in exchange coupling. However, even in this limit the exchange energies are still of the order of 1 meV, which makes this system an interesting candidate for practical realization of quantum gates.

## Acknowledgments

We are grateful to Dr. D.G. Austing for helpful discussions. We also acknowledge support from the NSF through the MCC and the University of Illinois Research Board.

## Appendix A: Variational Monte Carlo

One of the modelling approaches we utilize in this work is the quantum Monte Carlo technique that also offers simple yet accurate ways to simulate many-particle quantum systems (McMillan 1965; Ceperley and Alder 1977). In particular, the variational Monte Carlo method (VMC) relies on

a set of trial wavefunctions with adjustable parameters to derive by the variational method the total energy of the system; the latter being estimated by random-number generation. Compared with the CI method, the benefits of using VMC include (1) relatively simple simulation code implementation, (2) low memory requirement in the numerical method, and (3) easy transition to parallel code due to the inherent parallel nature of Monte Carlo method. The obvious disadvantage of VMC is that it can find only the lowest (ground-state) energy of the system. Here, we use VMC to compute the exchange energy of two and three electrons in coupled QD systems in the presence of magnetic fields, thereby extending the previous works on single QDs (Harju 2005; Ghosal *et al.* 2006) and double QDs in zero magnetic fields (Saarikoski *et al.* 2002).

In VMC the most commonly used trial wavefunction has a Slater–Jastrow form (Foulkes *et al.* 2001), which is a product of Slater determinants consisting of single-particle orbitals for spin-up and spin-down electrons ( $D^\uparrow$  and  $D^\downarrow$ , respectively) combined with a Jastrow term,  $J(r_{ij})$ :

$$\Psi_\alpha(N) = D^\uparrow D^\downarrow \prod_{i < j}^N J(r_{ij}), \quad (2.15)$$

where  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ .

We focus first on the two-electron case and set the two-body Jastrow term to be  $J(r_{ij}) = \exp\left[\frac{a_1 r_{12}}{1 + b_1 r_{12}}\right]$  with  $a_1$  being fixed by the “cusp” condition (Kato 1957). We also neglect three-body and higher correlation terms in the Jastrow term and replace the Slater determinants with the following expressions for the two-electron trial wavefunctions for the lowest singlet (S) and the triplet (T) states, respectively:

$$(D^\uparrow D^\downarrow)_S = \sum_{i < j} c_{ij}^S [\phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_2) + \phi_j(\mathbf{r}_1)\phi_i(\mathbf{r}_2)], \quad (2.16)$$

$$(D^\uparrow D^\downarrow)_T = \sum_{i < j} c_{ij}^T [\phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_2) - \phi_j(\mathbf{r}_1)\phi_i(\mathbf{r}_2)]. \quad (2.17)$$

For the single-particle states  $\phi_i(\mathbf{r}_k)$ , here we use the three lowest Fock–Darwin orbitals (Fock 1928) localized in individual dots (Hu and DasSarma 2000):

- *s*-like orbitals:

$$\phi_j(\mathbf{r}_k) = \exp\left\{-\alpha_x^2 \left[x_k + (-1)^{j-1} \frac{d}{2}\right]^2 - \alpha_y^2 y_k^2\right\} \exp\left\{i(-1)^j \alpha_b y_k\right\}. \quad (2.18)$$

- *p*-like orbitals:

$$\begin{aligned} \phi_j(\mathbf{r}_k) = & [(x_k + d_2) + i(-1)^{j-1} y_k] \\ & \times \exp\left\{-\alpha_x^2 \left[x_k + (-1)^{j-1} \frac{d}{2}\right]^2 - \alpha_y^2 y_k^2\right\} \exp\left\{i(-1)^j \alpha_b y_k\right\}, \end{aligned} \quad (2.19)$$

where  $c_{ij}$ ,  $\alpha_x$ ,  $\alpha_y$ ,  $\alpha_b$ ,  $d$  and  $d_2$  are treated as variational parameters, and  $\mathbf{r}_k = (x_k, y_k)$ . For comparison purposes, in some simulations, we only vary  $c_{ij}$ , while keeping other variational parameters fixed at their nominal values (Reimann and Manninen 2002):

$$\alpha_x = \alpha_y = \frac{(eB/2)^2 + (m^*c\omega)^2}{\sqrt{\hbar c}}, \quad \alpha_b = \frac{ae^2B^2}{2\hbar c}, \quad d = d_2 = a. \quad (2.20)$$

For the singlet wavefunction eqn (2.16) with just the  $s$ -like orbitals eqn (2.18), coefficients  $c_{11}$  and  $c_{22}$  determine the electron double-occupation probability, i.e. the probability of two electrons occupying the same QD (Golovach and Loss 2004). It is seen that when  $J(r_{ij}) = 1$ , the double-occupation probability is zero for  $c_{11} = c_{22} = 0$  as the wavefunction represents a pure Heitler–London state. The VMC triplet wavefunction with  $s$ -like orbitals only formally coincides with the Hund–Mulliken wavefunction when  $J(r_{ij}) = 1$  (Golovach and Loss 2004).

In the case of  $N = 3$  electrons, we use the following expressions for the two lowest ground-state wavefunctions with  $S = 1/2$  and  $S = 3/2$  total spin:

$$(D^\uparrow D^\downarrow)_{S=1/2} = \sum_{ijk} c_{ijk}^{S=1/2} \begin{vmatrix} \varphi_i(\mathbf{r}_1) & \varphi_j(\mathbf{r}_1) \\ \varphi_i(\mathbf{r}_2) & \varphi_j(\mathbf{r}_2) \end{vmatrix} \varphi_k(\mathbf{r}_3), \quad (2.21)$$

$$(D^\uparrow D^\downarrow)_{S=3/2} = \sum_{ijk} c_{ijk}^{S=3/2} \begin{vmatrix} \varphi_i(\mathbf{r}_1) & \varphi_j(\mathbf{r}_1) & \varphi_k(\mathbf{r}_1) \\ \varphi_i(\mathbf{r}_2) & \varphi_j(\mathbf{r}_2) & \varphi_k(\mathbf{r}_2) \\ \varphi_i(\mathbf{r}_3) & \varphi_j(\mathbf{r}_3) & \varphi_k(\mathbf{r}_3) \end{vmatrix}. \quad (2.22)$$

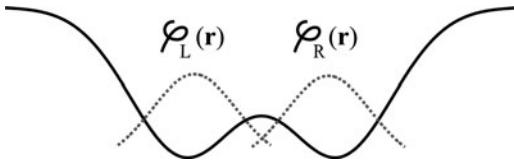
To optimize the expectation value of the energy, we use the steepest descent method (Casula *et al.* 2004) in which the set of variational parameters  $\alpha_k$  in many-particle wavefunctions eqns (2.16, 2.17) is updated as  $\alpha_k \rightarrow \alpha_k + \Delta t f_k$  at each numerical iteration with

$$\begin{aligned} f_k = -\frac{\partial E}{\partial \alpha_k} &= -\frac{\langle \Psi_\alpha | H O_k + O_k^* H + \partial_{\alpha_k} H | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \\ &+ \frac{\langle \Psi_\alpha | O_k | \Psi_\alpha \rangle \langle \Psi_\alpha | H | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} + \frac{\langle \Psi_\alpha | O_k | \Psi_\alpha \rangle \langle \Psi_\alpha | O_k^* | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle}, \end{aligned} \quad (2.23)$$

where  $O_k = \frac{\partial}{\partial \alpha_k} \ln \Psi_\alpha$ ,  $O_k^* = \frac{\partial}{\partial \alpha_k} \ln \Psi_\alpha^*$  and  $E = \frac{\langle \Psi_\alpha | H | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle}$  is the total energy. The step  $\Delta t$  is chosen such that after an adequate number of changes in the  $f_k$  sign,  $f_k$  becomes inversely proportional to the total number of iterations, allowing the variational parameters to converge numerically. After that, by using a sufficiently large number of randomly generated sets of electron coordinates (“random walkers”)  $N_w = 10^7$ , we solve for the expectation values of the energy (or any other observable quantity) by utilizing the Monte Carlo integration:

$$E = \frac{\langle \Psi_\alpha | H | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \approx \frac{1}{N_w} \sum_l E_l \left[ \mathbf{r}_1^l, \mathbf{r}_2^l \right], \quad (2.24)$$

where  $E_l$  is the local energy and  $(\mathbf{r}_1^l, \mathbf{r}_2^l)$  are the electron positions sampled from the Metropolis algorithm (Foulkes *et al.* 2001).



**Fig. 2.28** Confinement potential (solid) and electron orbitals (dashed) along the interdot coupling direction.  $\varphi_L(\mathbf{r})$  and  $\varphi_R(\mathbf{r})$  denote the localized states in the left and right QDs, respectively.

## Appendix B: Heitler–London approach

Among all theoretical approaches considered in this work, the Heitler–London (HL) technique (Heitler and London 1927) is probably the simplest in its conceptual methodology to extract the exchange coupling values in coupled QDs (Burkard *et al.* 1999; Hu and DasSarma 2000): Its validity has been discussed for QD systems of various dimensions (Calderón *et al.* 2006; Pedersen *et al.* 2007; Saraiva *et al.* 2007), and some effort has been made to improve the accuracy of the method by integrating variational parameters in the approach (Burkard *et al.* 2000; Koiller *et al.* 2004).

Within the HL approach, the following forms for the spatial part of the two-electron wavefunctions of the singlet (S) and triplet (T) states,  $\Psi_{S(T)}$ , are used:

$$\Psi_{S(T)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\varphi_L(\mathbf{r}_1)\varphi_R(\mathbf{r}_2) + (-)\varphi_L(\mathbf{r}_2)\varphi_R(\mathbf{r}_1)}{\sqrt{2(1 \pm S^2)}}, \quad (2.25)$$

where  $S = \langle \varphi_L | \varphi_R \rangle$  is the overlap between the lowest single-particle orbitals  $\varphi_L(\mathbf{r})$  and  $\varphi_R(\mathbf{r})$  localized in the left and right QDs (Fig. 2.28), respectively. Using these wavefunctions, the expectation values of the Hamiltonian eqn (2.1) for the singlet and triplet energies can be computed as  $E_{S(T)} = \langle \Psi_{S(T)} | \hat{H} | \Psi_{S(T)} \rangle$ .

The above wavefunction eqn (2.25) assumes that two electrons are always localized in separate QDs and the double occupation of each QD in the singlet state is neglected (Ashcroft and Mermin 1976). As such, the HL method is expected to work well only for the large interdot distances (or strong magnetic fields that “squeeze” orbitals, thereby effectively decoupling electrons) when the two lowest single-particle states are quasi-degenerate while for smaller distances (weaker magnetic fields) other higher-energy single-particle states have to be explicitly taken into account.

## References

- Abolfath, R.M., Hawrylak, P. *Phys. Rev. Lett.* **97**, 186802 (2006).
- Agarwal, S., Melnikov, D.V., Zhang, L.-X., Leburton, J.P. (unpublished).
- Ancilotto, F., Austing, D.G., Barranco, M., Mayol, R., Muraki, K., Pi, M., Sasaki, S., Tarucha, S. *Phys. Rev. B* **67**, 205311 (2003).
- Ashcroft, N.W., Mermin, N.D. *Solid State Physics* (Thomas Learning, 1976).
- Ashoori, R.C. *Phys. Rev. Lett.* **68**, 3088 (1992).
- Bellucci, D., Rontani, M., Troiani, F., Goldoni, G., Molinari, E. *Phys. Rev. B* **69**, 201308(R) (2004).

- Björk, M.T., Fuhrer, A., Hansen, A.E., Larsson, M.W., Fröberg, L.E., Samuelson, L. *Phys. Rev. B* **72**, 201307(R) (2005).
- Braun, P.-F., Marie, X., Lombez, L., Urbaszek, B., Amand, T., Renucci, P., Klevich, V.K., Kavokin, K.V., Krebs, O., Voisin, P., Masumoto, Y. *Phys. Rev. Lett.* **94**, 116601 (2005).
- Bron, W.E., Wagner, M. *Phys. Rev.* **145**, 689 (1966).
- Bruce, N.A., Maksym, P.A. *Phys. Rev. B* **61**, 4718 (2000).
- Burkard, G., Loss, D., DiVincenzo, D.P. *Phys. Rev. B* **59**, 2070 (1999).
- Burkard, G., Seelig, G., Loss, D. *Phys. Rev. B* **62**, 2581 (2000).
- Calderón, M.J., Koiller, B., Das Sarma, S. *Phys. Rev. B* **74**, 045310 (2006).
- Casula, M., Attaccalite, C., Sorella, S. *J. Chem. Phys.* **121**, 7110 (2004).
- Ceperley, D.M., Alder, B.J. *Phys. Rev. B* **16**, 3081 (1977).
- Chakraborty, T. *Quantum Dots, A Survey of the Properties of Artificial Atoms* (North Holland, Amsterdam, 1999).
- Chan, I.H., Fallahi, P., Vidan, A., Westervelt, R.M., Hanson, M., Gossard, A.C. *Nanotechnology* **15**, 609 (2004).
- Chaney, D., Maksym, P.A. *Phys. Rev. B* **75**, 035323 (2007).
- Ciorga, M., Sachrajda, A.S., Hawrylak, P., Gould, C., Zawadzki, P., Jullian, S., Feng, Y., Wasilewski, Z. *Phys. Rev. B* **61**, R16315 (2000).
- Climente, J.I., Bertoni, A., Goldoni, G., Rontani, M., Molinari, E. *Phys. Rev. B* **75**, 081303(R) (2007).
- Darwin, C. *Proc. Cambridge Philos. Soc.* **27**, 86 (1930).
- Das Sarma, S., Lai, W.-Y. *Phys. Rev. B* **32**, R1401 (1985).
- DiCarlo, L., Lynch, H.J., Johnson, A.C., Childress, L.I., Crockett, K., Marcus, C.M., Hanson, M.P., Gossard, A.C. *Phys. Rev. Lett.* **92**, 226801 (2004).
- Drouvelis, P.S., Schmelcher, P., Diakonos, F.K. *Phys. Rev. B* **69**, 155312 (2004).
- Dybalski, W., Hawrylak, P. *Phys. Rev. B* **72**, 205432 (2005).
- Elzerman, J.M., Hanson, R., Greidanus, J.S., van Beveren, L.H.W., De Franceschi, S., Vandersypen, L.M.K., Tarucha, S., Kouwenhoven, L.P. *Phys. Rev. B* **67**, 161308(R) (2003).
- Fasth, C., Fuhrer, A., Björk, M.T., Samuelson, L. *Nano. Lett.* **5**, 1487–1490 (2005).
- Fasth, C., Fuhrer, A., Samuelson, L., Golovach, V.N., Loss, D. *Phys. Rev. Lett.* **98**, 266801 (2007).
- Fock, V. *Z. Phys.* **47**, 446 (1928).
- Foulkes, W.M.C., Mitas, L., Needs, R.J., Rajagopal, G. *Rev. Mod. Phys.* **73**, 33 (2001).
- Friesen, M., Rugheimer, P., Savage, D.E., Lagally, M.G., van der Weide, D.W., Joynt, R., Eriksson, M.A. *Phys. Rev. B* **67**, 121301(R) (2003).
- Fuhrer, A., Fasth, C., Samuelson, L. *Appl. Phys. Lett.* **91**, 052109 (2007).
- Fujisawa, T., Austing, D.G., Tokura, Y., Hirayama, Y., Tarucha, S. *Nature* **419**, 278 (2002).
- Gaudreau, L., Studenikin, S.A., Sachrajda, A.S., Zawadzki, P., Kam, A., Lapointe, J., Korkusinski, M., Hawrylak, P. *Phys. Rev. Lett.* **97**, 036807 (2006).

- Ghosal, A., Guclu, A.D., Umrigar, C.J., Ullmo, D., Baranger, H.U. *Nature Phys.* **2**, 336 (2006).
- Goldhaber-Gordon, D. *Nature* **391**, 156 (1998).
- Golovach, V.N., Loss, D. *Phys. Rev. B* **69**, 245327 (2004).
- Hansen, A.E., Björk, M.T., Fasth, C., Thelander, C., Samuelson, L. *Phys. Rev. B* **71**, 205328 (2005).
- Hanson, R., Kouwenhoven, L.P., Petta, J.R., Tarucha, S., Vandersypen, L.M.K. *Rev. Mod. Phys.* **79**, 1217 (2007).
- Harju, A., Siljamäki, S., Nieminen, R.M. *Phys. Rev. Lett.* **88**, 226804 (2002).
- Harju, A. *Condens. Matter* **140**, 0505053 (2005).
- Hatano, T., Stopa, M., Yamaguchi, T., Ota, T., Yamada, K., Tarucha, S. *Phys. Rev. Lett.* **93**, 066806 (2004).
- Hatano, T., Stopa, M., Tarucha, S. *Science* **309**, 268 (2005).
- Hayashi, T., Fujisawa, T., Cheong, H.D., Jeong, Y.H., Hirayama, Y. *Phys. Rev. Lett.* **91**, 226804 (2003).
- Heitler, W., London, F. *Z. Phys.* **44**, 455 (1927).
- Helle, M., Harju, A., Nieminen, R.M. *Phys. Rev. B* **72**, 205329 (2005).
- Holleitner, A.W., Blick, R.H., Hüttel, A.K., Eberl, K., Kotthaus, J.P. *Science* **297**, 70 (2002).
- Hu, X., Das Sarma, S. *Phys. Rev. A* **61**, 062301 (2000).
- Hu, Y., Churchill, H.O.H., Reilly, D.J., Xiang, J., Lieber, C.M., Marcus, C.M. *Nature Nanotech.* **2**, 622 (2007).
- Hüttel, A.K., Ludwig, S., Eberl, K., Kotthaus, J.P. *Phys. Rev. B* **72**, 081310(R) (2005).
- Johnson, A.C., Petta, J.R., Marcus, C.M., Hanson, M.P., Gossard, A.C. *Phys. Rev. B* **72**, 165308 (2005).
- Kastner, M.A. *Phys. Today* **46**, 24 (1993).
- Kato, T. *Comm. Pure. Appl. Math.* **10**, 151 (1957).
- Kim, J., Melnikov, D.V., Leburton, J.-P., Austing, D.G., Tarucha, S. *Phys. Rev. B* **74**, 035307 (2006).
- Kim, J., Melnikov, D.V., Leburton, J.-P. *Open Condens. Matter Phys. J.* **1**, 1 (2008).
- Koiller, B., Capaz, R.B., Hu, X., Das Sarma, S. *Phys. Rev. B* **70**, 115207 (2004).
- Koppens, F.H.L., Klauser, D., Coish, W.A., Nowack, K.C., Kouwenhoven, L.P., Loss, D., Vandersypen, L.M.K. *Phys. Rev. Lett.* **99**, 106803 (2007).
- Kyriakidis, J., Pioro-Ladrière, M., Ciorga, M., Sachrajda, A.S., Hawrylak, P. *Phys. Rev. B* **66**, 035320 (2002).
- Landau, L.D., Lifshitz, E. *Quantum Mechanics, Non-relativistic Theory* (Pergamon Press, Oxford, 1977).
- Lichten, W. *Phys. Rev. A* **3**, 594 (1971).
- Likharev, K.K. *Proc. IEEE* **87**, 606 (1999).
- Livermore, C., Crouch, C.H., Westervelt, R.M., Campman, K.L., Gossard, A.C. *Science* **274**, 1332 (1996).
- Loss, D., DiVincenzo, D.P. *Phys. Rev. A* **57**, 120 (1998).
- Macucci, M., Hess, K., Iafrate, G.J. *Phys. Rev. B* **48**, 17354 (1993).

- Matagne, P., Leburton, J.P., Austing, D.G., Tarucha, S. *Phys. Rev. B* **65**, 085325 (2002).
- McMillan, W.L. *Phys. Rev.* **138**, A442 (1965).
- Melnikov, D.V., Matagne, P., Leburton, J.-P., Austing, G., Yu, W., Tarucha, S., Fettig, J., Sobh, N. *Phys. Rev. B* **72**, 085301 (2005).
- Melnikov, D.V., Leburton, J.-P. *Phys. Rev. B* **73**, 085320 (2006).
- Melnikov, D.V., Leburton, J.-P., Taha, A., Sobh, N. *Phys. Rev. B* **74**, 041309(R) (2006).
- Melnikov, D.V., Leburton, J.-P. *Phys. Rev. B* **73**, 155301 (2006).
- Melnikov, D.V., Leburton, J.-P. *Phys. Status Solidi. C* **4**, 578 (2007).
- Melnikov, D.V., Zhang, L.-X., Leburton, J.-P. *Curr. Opin. Solid State Mater Sci.* **10**, 114 (2007).
- Merkulov, I.A., Efros, A.I., Rosen, M. *Phys. Rev. B* **65**, 205309 (2002).
- Mikhailov, S.A. *Phys. Rev. B* **65**, 115312 (2002).
- Pedersen, J., Flindt, C., Mortensen, N.A., Jauho, A.-P. *Phys. Rev. B* **76**, 125323 (2007).
- Petta, J.R., Johnson, A.C., Marcus, C.M., Hanson, M.P., Gossard, A.C. *Phys. Rev. Lett.* **93**, 186802 (2004).
- Petta, J.R., Johnson, A.C., Taylor, J.M., Laird, E.A., Yacoby, A., Lukin, M.D., Marcus, C.M., Hanson, M.P., Gossard, A.C. *Science* **309**, 2180 (2005).
- Pick Š., Dreyssé H. *Phys. Rev. B* **48**, 13588 (1993).
- Pioro-Ladrière, M., Abolfath, R., Zawadzki, P., Lapointe, J., Studenikin, S.A., Sachrajda, A.S., Hawrylak, P. *Phys. Rev. B* **72**, 125307 (2005).
- Qin, H., Holleitner, A.W., Eberl, K., Blick, R.H. *Phys. Rev. B* **64**, 241302 (2001).
- Ravishankar, R., Matagne, P., Leburton, J.-P., Martin, R.M., Tarucha, S. *Phys. Rev. B* **69**, 035326 (2004).
- Reimann, S.M., Koskinen, M., Manninen, M. *Phys. Rev. B* **62**, 8108 (2000).
- Reimann, S.M., Manninen, M. *Rev. Mod. Phys.* **74**, 1283 (2002).
- Rontani, M., Troiani, F., Hohenester, U., Molinari, E. *Solid State Commun.* **119**, 309 (2001).
- Saarikoski, H., Räsänen, E., Siljamäki, S., Harju, A., Puska, M.J., Nieminen, R.M. *Eur. Phys. J. B* **26**, 241 (2002).
- Saarikoski, H., Harju, A. *Phys. Rev. Lett.* **94**, 246803 (2005).
- Sako, T., Diercksen, G.H.F. *Phys. Rev. B* **75**, 115413 (2007).
- Saraga, S., Loss, D. *Phys. Rev. Lett.* **90**, 166803 (2003).
- Saraiva, A.L., Calderon, M.J., Koiller, B. *Phys. Rev. B* **76**, 233302 (2007).
- Sasakura, H., Adachi, S., Muto, S., Usuki, T., Takatsu, M. *Semicond. Sci. Technol.* **19**, S409 (2004).
- Schröer, D., Greentree, A.D., Gaudreau, L., Eberl, K., Hollenberg, L.C.L., Kotthaus, J.P., Ludwig, S. *Phys. Rev. B* **76**, 075306 (2007).
- Stano, P., Fabian, J. *Phys. Rev. B* **74**, 045320 (2006).
- Stopa, M. *Phys. Rev. B* **54**, 13767 (1996).
- Stopa, M., Marcus, C.M. *Cond-mat/0604008* (2006).
- Szafran, B., Peeters, F.M., Bednarek, S. *Phys. Rev. B* **70**, 205318 (2004).
- Szafran, B., Peeters, F.M., Bednarek, S., Adamowski, J. *Phys. Rev. B* **69**, 125344 (2004).

- Tarucha, S., Austing, D.G., Honda, T., van der Hage, R.J., Kouwenhoven, L.P. *Phys. Rev. Lett.* **77**, 3613 (1996).
- Ullrich, C.A., Kohn, W. *Phys. Rev. Lett.* **87**, 093001 (2001).
- van der Wiel, W.G., De Franceschi, S., Elzerman, J.M., Fujisawa, T., Tarucha, S., Kouwenhoven, L.P. *Rev. Mod. Phys.* **75**, 1 (2002).
- von Neumann, J., Wigner, E. *Z. Phys.* **30**, 467 (1929).
- Vidan, A., Westervelt, R.M., Stopa, M., Hanson, M., Gossard, A.C. *Appl. Phys. Lett.* **85**, 3602 (2004).
- Yannouleas, C., Landman, U. *Int. J. Quantum. Chem.* **90**, 699 (2002).
- Yannouleas, C., Landman, U. *Rep. Prog. Phys.* **70**, 2067 (2007).
- Wagner, M., Merkt, U., Chaplik, A.V. *Phys. Rev. B* **45**, 1951 (1992).
- Waugh, F.R., Berry, M.J., Mar, D.J., Westervelt, R.M., Campman, K.L., Gossard, A.C. *Phys. Rev. Lett.* **75**, 705 (1995).
- Wensauer, A., Steffens, O., Suhrke, M., Rössler, U. *Phys. Rev. B* **62**, 2605 (2000).
- Wójcik, A., Hawrylak, P. *Phys. Rev. B* **56**, 13227 (1997).
- Zhang, L.-X., Matagne, P., Leburton, J.P., Hanson, R., Kouwenhoven, L.P. *Phys. Rev. B* **69**, 245301 (2004).
- Zhang, L.-X., Melnikov, D.V., Leburton, J.P. *Phys. Rev. B* **74**, 205306 (2006).
- Zhang, L.-X., Melnikov, D.V., Agarwal, S., Leburton, J.-P. *Phys. Rev. B* **78**, 035418 (2008a).
- Zhang, L.-X., Melnikov, D.V., Leburton, J.-P. *Phys. Rev. B* **78**, 085310 (2008b).
- Zhang, L.-X., Melnikov, D.V., Leburton, J.-P. *J. Phys.: Condens. Matter* **21**, 095502 (2009).

# 3

# Spintronics with metallic nanowires

J.-Ph. Ansermet

3.1 Introduction	90
3.2 Spin diffusion	96
3.3 Models for spin-polarized currents acting on magnetization	101
3.4 Current-induced magnetization switching	108
3.5 Current-driven magnetic excitations	114
3.6 Resonant-current excitation	118
3.7 Conclusion	123
References	124

## 3.1 Introduction

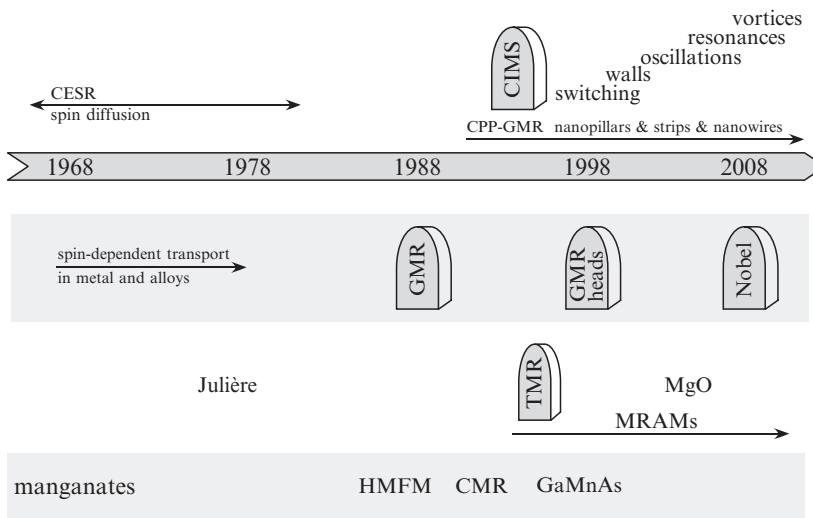
### 3.1.1 Highlights of spintronics research

The emerging field of spintronics thrives on the many possibilities of using the spin of electrons in fast, nanosized electronic devices. The first successes had to do with the effect of magnetization on the flow of conduction electrons, leading to giant magnetoresistance (Baibich *et al.* 1988; Binasch 1989), its commercial use in reading heads, or to tunnel magnetoresistance (Moodera *et al.* 1995) and its promising use in magnetic memories (Tehrani *et al.* 2000). Following a theoretical prediction (Berger 1996; Slonczewski 1996), the converse effect of spin-polarized current on magnetization was confirmed experimentally (Tsui *et al.* 1998a; Myers *et al.* 1999; Wegrowe *et al.* 1999a), with the possibility of writing selectively magnetoresistive memory bits (Slonczewski 1997; Sun 1998; Wegrowe *et al.* 1999b) or of forming radically new gigahertz oscillators (Kiselev *et al.* 2003; Rippard *et al.* 2005).

Giant magnetoresistance in metallic multilayers is at the root of spintronics. It is an amazing success to consider when pondering on the interplay of science and society. IBM feels this way and asks on its web site<sup>1</sup> “what is 10 years to you?” with the answer:“—To us, it is enough time for a revolution!” In the early 1980s, it had become possible to make superlattices of metals. Physicists investigated what becomes of the oscillatory response of non-magnetic metals about magnetic impurities (the so-called RKKY interaction), if instead of isolated magnetic atoms in a normal metal, layers of magnetic materials were used. A magnetic coupling was found, indeed, alternating as a function of spacer thickness between ferromagnetic and antiferromagnetic interactions. As this research in magnetism was in full swing, the group of Prof. A. Fert in Paris and the group of Prof. Grünberg in Aachen measured the resistance of the thin film composed of alternating magnetic and non-magnetic layers.

The resistance was found to vary by as much as 80%. This was considered a very large change. Permalloy, an alloy of iron and nickel, was known to have its resistance change with magnetic field by about 4%. This material was used in the reading heads of hard disks in the computers of the time. These

<sup>1</sup><http://www.research.ibm.com/research/gmr.html>



**Fig. 3.1** Milestones of spintronics: the discovery of giant magnetoresistance (GMR), their widespread use in reading heads of hard disks only 10 years later, the Nobel prize to Prof. A. Fert and P. Gr  nberg, the prediction of “current-induced magnetization switching” (CIMS) by Berger (1996) and Slonczewski (1996), with a string of confirmations and discoveries to be discussed here, stemming out of the research on “current-perpendicular-to-the-plane GMR” (CPP-GMR); the discovery of tunnel magnetoresistance by Juli  re (1976), the onset of intensive R&D on tunnel magnetoresistnace (TMR) (Miyazaki and Tezuka 1995; Moodera 1995), the discovery of MgO for tunnel barriers, the early studies of manganates in the 1950s, rediscovered as colosal magnetoresistance (CMR) materials; the idea of spin injection (Johnson and Silsbee 1985), in particular with half-metallic ferromagnets (HMFM) (van Son *et al.* 1987), the discovery of the magnetic semiconductor GaMnAs (Ohno *et al.* 1996); studies on conduction electron spin resonance (CESR) yielding electron spin scattering cross-section and working out s-d interactions at zero applied current.

new measurements were carried out in a research environment, at 4 K and with fields of several Tesla. It took research and development worldwide only 10 years to transform this discovery into a useful material (Fig. 3.1).

In multilayers, one way of getting the required antiparallel alignment of the magnetization of adjacent layers is to adjust the spacer layer so as to obtain an antiferromagnetic RKKY coupling. In a spin-valve, one layer has its magnetization pinned, while the other is freer to be reversed. In granular materials with ferromagnetic ultrafine grains in a non-magnetic matrix, GMR was also observed GMR (Berkowitz *et al.* 1992; Xiao *et al.* 1992; Holody *et al.* 1994) owing to the random orientation of the magnetization in adjacent grains.

Experiments in which the current is perpendicular to the interfaces have been particularly useful in shedding light on the fundamental mechanisms of GMR (Gijs and Bauer 1997). Perpendicular transport has also practical advantages over transport parallel to the interfaces: layers can be thicker when the current is perpendicular rather than parallel to the interfaces. This is relevant for practical applications, in which one must consider the ease of manufacturing and the reliability of the materials. When the current is parallel to the layers, the electrons experience both layers only if their separation is of the order of the electron mean-free path. On the contrary, when the current is perpendicular to the layers, the spins undergo a diffusion/relaxation process that extends over distances much longer than the electron mean-free path, since the spin-flip events are rare. Electron spin resonance experiments of the 1960s and the 1970s already determined that electrons can travel in metals like Ag and Cu a long way without flipping their spin (Fig. 3.1).

Much of the research on spin-dependent transport with current perpendicular to the interfaces has been carried out with nanostructures: nanopillars or nanowires. The reason for using nanostructures is the following. A small square cut out of a thin film of a magnetic multilayer, of a size in the sub-millimeter range, has a resistance in the nanoohm range! Its measurement has been carried out by a few worker (Pratt *et al.* 1991; Dauguet *et al.* 1996). Others

have relied on advanced lithographic techniques in order to reduce the surface area so as to bring the resistance to more accessible values (Gijs *et al.* 1993). This is quite an extensive process (Vavra *et al.* 1995). Furthermore, when the columns are very thin, the current density is no longer even throughout the column. Another issue with small pillars is that their field sensitivity is reduced considerably.

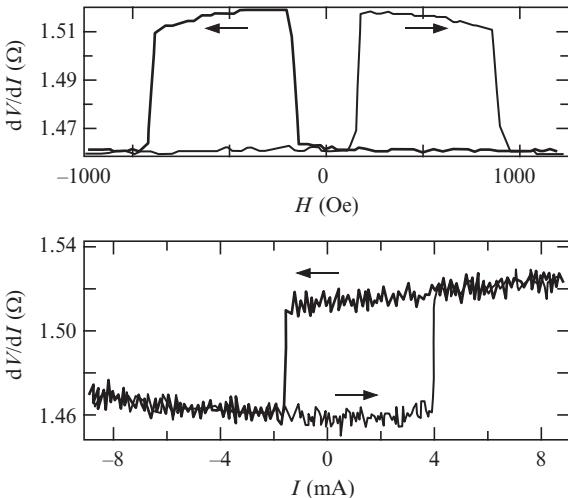
Instead of using advanced lithography in order to obtain a thin column, some groups (including this author's group) produced multilayers in the form of wires by electro-deposition of multilayers in pores of nanoporous membranes (Blondel *et al.* 1994; Piraux *et al.* 1994; Liu *et al.* 1995). Others used grooved substrates and a straight deposition so that they measured GMR with current-at-an-angle (CAP) (Ono and Shinjo 1995) or they deposited the layers sideways, so as to have a structure in which the current was mostly perpendicular to the layers (Gijs *et al.* 1995).

This chapter is intended to review spintronics with metals. However, in this introductory survey of the highlights of spintronics, mention must be made of the very important developments accomplished with tunnel junctions. Around the time of a suggestion by Slonczewski (1989), a major effort was launched to make spin-valves with tunnel barriers. In fact, Jullière (1975), in the early days of tunnel-barrier studies, had demonstrated that tunnelling between two magnetic layers ought to depend on the relative orientation of the magnetization in the two layers. Meservey and Tedrow (1994) were the champions of tunnelling studies, in particular between a magnet and a superconductor. Moodera *et al.* (1995), and Moodera and Kinder (1996) managed to make excellent aluminum oxide barriers and found sharp transitions and a large relative effect, called tunnel magnetoresistance, TMR. The early work was done almost exclusively with aluminum-oxide barriers. Following calculations of Buttler's group (Buttler *et al.* 2001), the case of MgO as a tunnel barrier overtook Al<sub>2</sub>O<sub>3</sub> as the material of choice (Yuasa *et al.* 2004). The breakthrough arose when it was recognized that some electronic states of the tunnel barrier do or do not match with a spin subband of the ferromagnetic electrode. Likewise, the large tunnel resistance observed with an STM (Rusponi *et al.* 2005) is ascribed to surface states modified by adsorbates (Hofer *et al.* 2008). Granular materials composed of ferromagnetic metal particles in an insulating matrix present also a form of TMR. There are also the half-metallic ferromagnets in which, as is generally thought, one spin subband only crosses the Fermi level. CrO<sub>2</sub> is an example. TMR was found when such clusters are embedded in a Cr<sub>2</sub>O<sub>3</sub> matrix (Coey *et al.* 1998).

### 3.1.2 Effect of current on magnetization

Thus far, effects such as GMR or TMR refer to the effect of magnetic configuration on the ability of a material to transport electric current. Now, we highlight major results concerning the converse effect, the effect of spin-polarized currents on magnetization.

In 1996, Berger (1996) and Slonczewski (1996), independently, predicted that a spin-polarized current, injected into a magnetic layer, would flip the



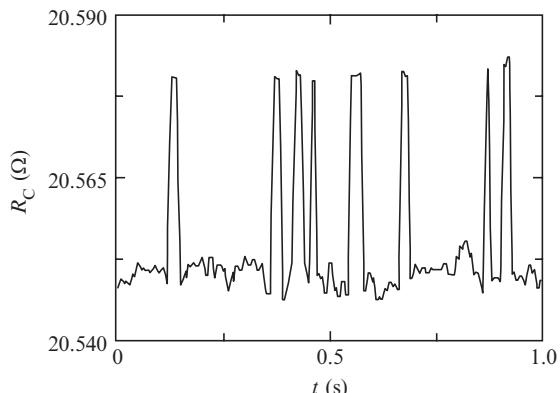
**Fig. 3.2** Magnetoresistance of a Co-Cu-Co nanopillar, differential resistance as a function of field (top) or current (bottom). Albert *et al.* (2002), copyright American Physical Society.

magnetization of this layer, provided the current exceeds some threshold, estimated by both at about  $10^7$  A/cm $^2$ . Experimental confirmations appeared soon after (Tsoi *et al.* 1998a; Myers *et al.* 1999; Wegrowe *et al.* 1999a). This effect was coined “current-induced magnetization switching” (CIMS). The first observations were of a quasi-static sort. Either the magnetization following an impulse of current was detected (Wegrowe *et al.* 2001a; Kelly *et al.* 2003), or hysteresis in current was demonstrated (Albert *et al.* 2002), as shown in Fig. 3.2.

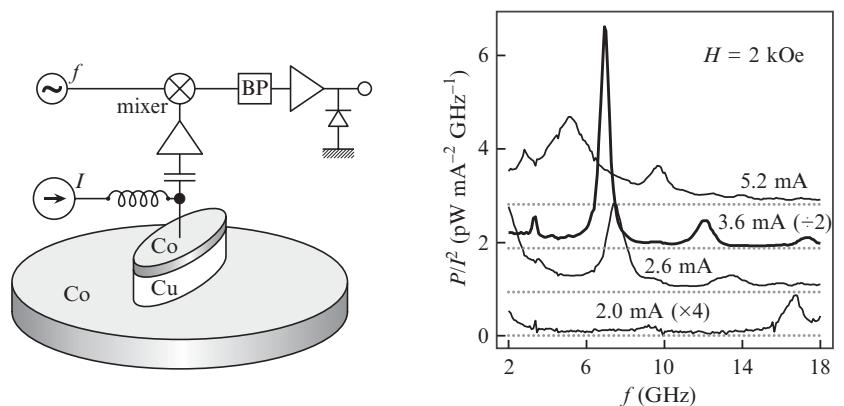
This was immediately perceived as a very important new field of research for two reasons. First, insuring a physical description of the effective underlying mechanism was a challenge to our basic understanding of spin-dependent transport. Second, applications to magnetoresistive random access memories immediately came to mind (Slonczewski 1997; Sun 1998; Wegrowe *et al.* 1999b). In the early years of the discovery of this new effect, Tsoi *et al.* (1998a) also discovered, using point contacts, peaks in the differential resistance  $dV/dI$ , that were ascribed to be magnetic excitations. In the same work, two-level fluctuations of the magnetization were detected (Fig. 3.3).

Two-level fluctuations were also observed in spin-valves (Fabian *et al.* 2003). Some groups were able to collect data over a large time range (Urazhdin *et al.* 2003; Pufall *et al.* 2004). Later, it became progressively clearer that dc currents could excite *oscillations* at frequencies near that of ferromagnetic resonance. The first report showed broad resonances in spin-valve nanopillars (Kiselev *et al.* 2003) (Fig. 3.4). Later publications demonstrated much sharper resonances, reaching a quality factor  $Q = \omega/\Delta\omega$  of 1400 (Krivorotov *et al.* 2004). The improvement came in part by perfecting the nanostructuring of the spin-valves (Rippard *et al.* 2005).

Instead of the dc excitation as above, it has also become possible to excite resonantly the free layer of a spin valves (Sankey *et al.* 2006; Tulapurkar *et al.* 2005). This last report (Tulapurkar *et al.* 2005) is truly remarkable, as the high current density is driven through a tunnel barrier, furthermore, as the barrier



**Fig. 3.3** Fluctuation of the resistance  $R_C$  as evidence for excitation at a point contact to a magnetic multilayer. Tsoi *et al.* (1998), copyright American Physical Society.



**Fig. 3.4** Principle of detection (left) and dissipation spectrum signalling the excitation of the free layer of a spin-valve by a dc current. Kiselev *et al.* (2003), copyright Nature Publishing Group.

is made of MgO, the relative change of magnetoresistance (used to detect the resonance) is of the order of 300%!

### 3.1.3 Meeting the challenges of spintronics with nanowires

Nanowires can be thought of as a new variety of samples with various advantages much like thin films had advantages over bulk crystals. There might arise an age of nanowire physics, like there has been on thin-film physics. In this chapter, the ability of achieving large current densities with proper thermal management and the ability to approach single-domain magnetic configurations constitute essential design parameters in exploring the effect of spin-polarized currents on magnetization. Thus, several key results were obtained either with nanopillars or with nanowires.

While nanowires constitute a radical geometrical confinement compared to thin films, analogies are many, as far as strategies for doing physics are concerned. Sometimes, thin films are of interest just because they are more accessible than growing bulk crystals of complex crystalline structure. Likewise, nanowires may be simpler to make. For example, the plating of a

nanowire amounts to plating a film of a few tens of nanometers in lateral size, so that columnar growth is not an issue. As a consequence, it is possible to form, without many artful developments, multilayers with hundreds of repeat units. Sometimes, the physics may change with reduced dimensionality. Thin films may simply enhance surface effects. For example, surface magnetocrystalline anisotropies can be evidenced by studying the thickness dependence of the coercivity. Likewise, magnetic nanowires present very large switching fields, and this is of interest for magnetic recording. Thin films may give rise to new physics, such as quantum size effects. Likewise, evidence for quantization of spin-waves has been discussed and quantization of electronic states in metals such as bismuth could be seen in transport and optical measurements. Some thin films are produced one atomic layer at a time. The most noted case is that of MBE, but it is not the only one. Indeed, nanowires of dimensions down to the atom have been studied, in ultrahigh-vacuum setups, on surfaces designed for nanostructure growth, or by pulling on a nanocontact or even by controlled growth in an aqueous environment.

Nanowires can be produced by a variety of means. Below, when addressing spintronics with domain walls, we will see that normal planar lithography plays a role. However, when studying spin-dependent transport with current forced to flow perpendicular to the interfaces, a compromise has to be found between two challenges: either one defines with simple lithography a contact of say, 1 micrometer in size, and then the resistance is in the nanoohm range, or some means is developed in order to restrict the cross-section where the current flows, so as to have a resistance more easily measured. In the latter case, one may either use point contacts, perform lithography in the sub-100-nm range in order to make nanopillars, or grow nanowires.

The author's group has shown various options of making nanowires by electrodeposition in nanostructured templates. The method was coined “template synthesis” by Charles Martin who explored extensively this technique for making nanostructures, in particular, nanotubules (Martin 2002). The template can be made of anodized aluminum, so-called alumite structures, that were used by the group of Moskovitz (see, e.g. AlMawlawi *et al.* 1991), or it can be formed by etching the tracks left by high-energy ions (Spohr 1990). The GSI group of Spohr *et al.* pushed this technique to the point of forming a single pore, and filling it in various ways (Man Leo *et al.* 2007). A great variety of materials were produced by electrodeposition in templates: ZnO (Abid *et al.* 2006b), Fe<sub>3</sub>O<sub>4</sub> (Abid *et al.* 2006a), Co-ferrites (Carlier and Ansermet 2006), CdTe (Ohgai *et al.* 2005), even Co nanowires exchanged biased with a sputtered layer of GdCo (Wegrowe *et al.* 2000).

Free-standing nanowires were grown by the technique known as “VLS”, for “vapor-liquid-solid”. A catalyst of nanometric dimensions sits on the top of a growing nanowire in the presence of the appropriate vapor of the starting materials. The nanowires can be connected to one another, even forming transistors (see, e.g. Huang *et al.* 2002; Appell 2002). Carbon nanotubes can be considered as a whole class of materials by themselves and the spintronics effects in them are reviewed in several places, see for example Cottet *et al.* (2006) and references therein.

The current in nanopillars may not be homogeneous, as the electrons flow from a very large lead into the relevant nanoconstriction. The extensive review of Gijs and Bauer (1997) on spin-dependent transport alludes to this. At the time, the concern was about making sure the current is driven perpendicular to the interfaces. Nowadays, inhomogeneities in the field produced by the current are considered in setting up the oscillation of the magnetization driven by a dc current. The temperature of a nanopillar can be assessed by analyzing the parabolic form of the  $dV/dI$  data. In nanowires, in general the relevant magnetic nanostructure is deep into a nanopore, so the current density is uniform by the time the electrons reach it. It turns out that nanowires of about 40 nm in diameter do not heat up very much, because of a strong heat-sinking effect of the embedding matrix. As the diameter reaches 100 nm or more, a marked temperature rise, of the order of tens of Kelvin, were observed, by resistive measurements. The question arose as to the possibility of having a small hot spot somewhere in the nanowire, which could contribute trivially to the magnetization switching. Numerical integration of the Fourier equation in the cylindrical geometry (Fabian *et al.* 2006) helped us realize that, either the time of observation is so short that the defect does not have time to heat up, or on a longer timescale, the heat propagates throughout the nanowire. Switching was observed over this longer timescale.

## 3.2 Spin diffusion

### 3.2.1 Spin-dependent conductivities

In ferromagnetic metals, the idea of spin-polarized currents can be traced back to the so-called Mott picture (Mott 1935, 1964), according to which charge carriers have spins either aligned or antialigned with the magnetization, the current is due mostly to electrons of “s” character, and the most frequent scattering event is from “s” electrons to “d” electrons. Hence, if the “d” states of either spin directions are differently occupied, one must expect *spin-dependent conductivities*.

The consequence of this spin dependence for conduction was explored most notably by Fert and Campbell. They interpreted the departure of the conductivity of nickel from Matthiessen’s rule in terms of spin-mixing, the non-dissipative process by which electrons flip their spins but do not relax their momentum (Fert and Campbell 1968). Similar considerations in alloys lead to estimates of the ratio  $\alpha$  of the spin-dependent conductivities (Fert 1969).

Giant magnetoresistance occurs when conduction takes place in nanostructured materials where magnetic materials alternate with normal metals over distances shorter than the spin-diffusion length (see below). When the current is perpendicular to the interfaces, the effect of spin-dependent conductivities on the overall transport as a function of the magnetic configuration is intuitively clear. If the magnetizations of two adjacent layers are antiparallel, then each spin channel has to go through a combination of a path with a large conductivity and a path with a small one. If the layers have parallel magnetizations, then one spin orientation has a high conduction path, thus lowering the overall

resistivity of these two layers. Various theoretical approaches can be used to model this. Theoretical and experimental results have been reviewed in numerous places. See, for example, Zutic *et al.* (2004).

A major consequence of these spin-dependent conductivities is the ability of generating a spin-polarized current by passing a current in between two ferromagnets (Johnson and Silsbee 1985, 1987; Johnson and Byers 2003). Johnson and Silsbee proposed spin transistors that are spintronics devices in the true sense of the term. Various forms of spin transistors have been demonstrated on a laboratory scale (Monsma *et al.* 1998; Dennis *et al.* 2003).

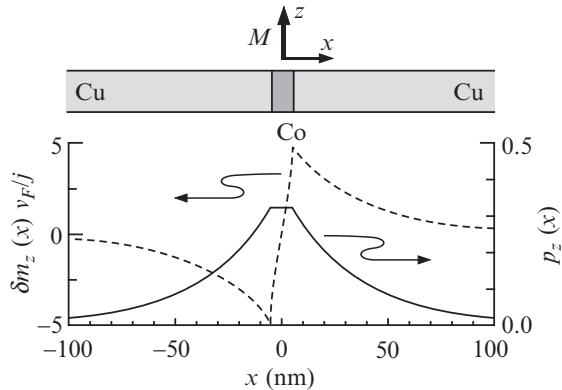
There are materials that are thought to be half-magnetic ferromagnets. We refer to Dowben and Skomski (2004) for a critical analysis of this concept. Typical candidates are materials such as the Heusler alloys, magnetite,  $\text{CrO}_2$  or manganites. It was envisaged that such materials could constitute spin injectors (van Son *et al.* 1997). However, these materials are not easy to work with. So far, they have shown favorable properties only when used as the spacer layer in tunnel junctions (Viret *et al.* 1997).

One should be careful about the meaning of spin polarization of conduction electrons. The diffusive phenomenon that gives rise to spin polarization by forcing a current through a magnetic nanostructure must be distinguished from the spin polarization observed when electrons tunnel out of a ferromagnet into, say, a superconductor. The clarification of Mazin (1999) appeared in *Physical Review Letters*. The notion that spins can be injected in a solid-state device dates back to the seminal paper of Dyakonov and Perel (1971). However, it remains one of the great challenges of spintronics using semiconductors to inject and modulate spin-polarized carriers electrically. In particular, spin injection from a metal into a semiconductor is not straightforward (Fert and Jaffrèes 2001; Gregg *et al.* 2003).

The spin polarization obtained by driving a current through a magnetic nanostructure is referred to as “*spin accumulation*”. It was accounted for simply in a Boltzmann description of spin-dependent transport, which extended the standard treatment by assuming spin-dependent distributions for “up” and “down” spins, and using distinct, spin-dependent electrochemical potentials (Valet and Fert 1993). The difference in electrochemical potential, in a first approximation, is proportional to the spin polarization. The polarization current is the difference between spin-up and spin-down currents (Fig. 3.5).

### 3.2.2 Spin-diffusion lengths

As can be expected, the spin polarization of the conduction electrons does not carry infinitely far. Spin-flip scattering processes take place that relax this spin polarization. This relaxation process was studied in the 1960s and 1970s by electron spin resonance (for a review, see Barnes 1981). One of the primary mechanism is spin-orbit scattering, either off impurities (Asik 1966) or directly on the host metal (Elliot 1953; Yafet 1963; Ball 1969). Conduction electron spin resonance (CESR) relies on the ability of conduction electron spins to carry the excitation generated by a microwave field from one side of a thin film to another (Edmonds *et al.* 1985). These measurements gave spintronics



**Fig. 3.5** Current polarization  $p_z(x) = \frac{j_+ - j_-}{j_+ + j_-}$  and normalized spin accumulation  $\delta m_z(x)$  for a single Co layer embedded between two semi-infinite Cu leads (top).  $\delta m_z(x)$  is defined as a dimensionless density rather than a magnetization and scaled by the ratio of the current to the Fermi velocity. Stiles *et al.* (2002a), copyright American Institute of Physics.

a precious database of electron spin cross-sections (Monod and Schultz 1968). Some aspects of this CESR work has direct connection with spintronics, such as the finding of a magnetic coupling between a copper and a permalloy film due to conduction electrons (Janossy and Monod 1976). As spintronics brings a renewed and keen interest in spin-relaxation processes, detailed modelling of relaxation mechanisms have been undertaken (Fabian *et al.* 1998, 1999).

The study of multilayers and spin-valves provided values of spin-diffusion length. The groups of Pratt and Bass at Michigan State University for example, were able to probe how the spin-diffusion length of spin-valves changed upon doping the spacer layer with various impurities (Yang *et al.* 1994). They relied on rather large pillars and used superconducting contacts (Pratt *et al.* 1991). Their extensive and precise investigations of spin-diffusion lengths are now summarized in a review (Bass and Pratt 2007).

Alternatively, nanowires with over a hundred Co-Cu layers could be electrodeposited in templates, thus forming large stacks that could be easily measured. By varying the thickness of the layer, it was possible also in this case to estimate the spin-diffusion length of the spacer layers (Doudin *et al.* 1996; Piraux *et al.* 1994, 1996).

While it is possible to describe spin diffusion in the framework of a thermodynamic model (Johnson and Silsbee 1985, 1987, 2003; Ansermet 2008), making connection with micromechanisms of spin relaxation requires a more elaborate description. For this purpose, a simple spin-dependent Boltzmann description of transport suffices (Valet and Fert 1993). Hence, we define a spin-flip mean-free path  $l_s$  ( $s = +, -$ ), with:

$$l_s^2 = \frac{1}{3} \tau_{sf} \left( \tau_s^{-1} + \tau_{sf}^{-1} \right)^{-1} v_F^2,$$

where  $\tau_s = \lambda_e/v_F$  and  $\lambda_e$  is the spin-dependent electron mean-free path (a momentum relaxation time),  $v_F$  is the Fermi velocity. The spin-dependent Boltzmann approach yields a spin-diffusion equation for the electrochemical potential difference  $\Delta\mu = \frac{\mu_+ - \mu_-}{2}$ :

$$\frac{\Delta\mu}{l_{sf}^2} = \nabla^2 \Delta\mu,$$

where the spin-diffusion length  $l_{\text{sf}}$  is found to be defined by the average:  $\frac{1}{l_{\text{sf}}^2} = \frac{1}{l_+^2} + \frac{1}{l_-^2}$ . It is usually so that the electron momentum relaxation rate  $\tau_s^{-1}$  is much larger than the spin-flip relaxation rate  $\tau_{\text{sf}}^{-1}$ . Thus, neglecting the spin dependence in  $\tau_+^{-1} \approx \tau_-^{-1} = \tau_e^{-1}$  for the estimate of the spin-diffusion length, yields:

$$l_{\text{sf}} = \frac{1}{\sqrt{6}} \sqrt{\tau_{\text{sf}} \tau_e} v_F.$$

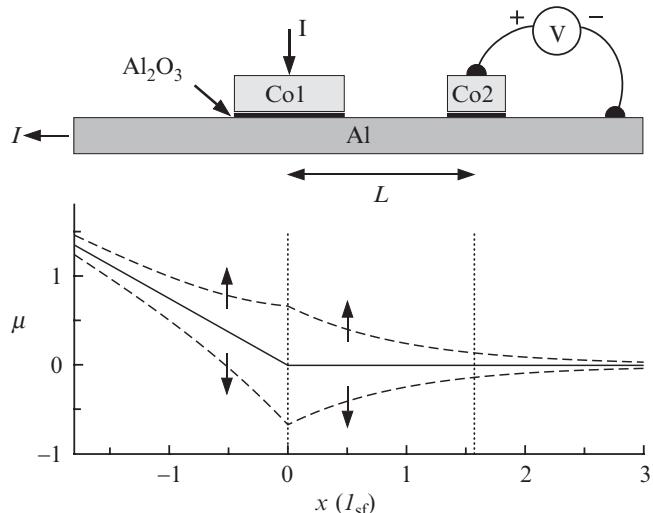
In the non-magnetic spacer, the spin-diffusion length depends strongly on the impurity content, as scattering at these impurities may determine  $\tau_{\text{sf}}$ . But for the sake of getting an order of magnitude estimate, let us assume a reasonable value of 100 nm for  $l_{\text{sf}}$ , and of 1 nm for the electron mean-free path  $\lambda_e \tau_e v_F$ . This implies a spin-flip time  $\tau_{\text{sf}}$  of about  $6 \times 10^{-11}$  s. Such a result can be compared with conduction electron spin resonance data (Monod and Schultz 1982). For example, it was found (Monod and Schultz 1963) that the spin lattice relaxation of conduction electrons of copper with traces of Cr relax with a time inversely proportional to the Cr concentration, with  $T_{\text{sl}} = 2 \times 10^{-8}$  s for just 1 ppm of Cr. In the making of a magnetic multilayer, one has to expect impurity contents in the per cent range rather than the ppm! Hence the spin-diffusion lengths end up in the tens of nm typically. It all depends on the method and the quality of the deposits. For example, Jedema reported  $l_{\text{sf}}$  of as much as 500 nm (Jedema *et al.* 2002a). While carbon nanotubes are beyond the scope of the present review, it is noteworthy to mention here that in them, the spin-diffusion length is quite long, owing to the very low spin-orbit coupling of their conduction electrons (low  $Z$  of carbon), and to the typical purity of the carbon nanotubes.

In the ferromagnet itself, the spin-diffusion length is quite short. At low temperatures, spin-orbit effects dominate, and at higher temperatures, additional relaxation by magnons shorten  $l_{\text{sf}}$  even further. For example, Piraux estimated  $l_{\text{sf}}$  from GMR data on Co-Cu multilayers and found a value of about 60 nm at low temperature, getting down to about 40 nm at room temperature (Piraux *et al.* 1998). Values of  $l_{\text{sf}}$  are reviewed by Bass and Pratt (2007, Table 3). In permalloy,  $l_{\text{sf}}$  appears to be of a few nm only at low temperature.

Spin diffusion has been reviewed by many authors, in particular Zutic *et al.* (2004, 2006), Silsbee (2004), Maekawa (2006).

### 3.2.3 Spin accumulation

One essential consequence of the spin dependence of the conductivity is the spin accumulation obtained in the metal spacer separating two ferromagnetic layers. Whereas such a spin accumulation is conceptually evident, its experimental verification by direct measurement turns out to be a challenge. Local probes such as MFM or spin-polarized STM have not been able to evidence spin accumulation, to the knowledge of this author (Wulfhekel *et al.* 2007). Muon spin resonance, which probes the hyperfine field (Luetkens *et al.* 2003, 2004), ought to be able to see its modulation under a dc current. XMCD depends on the full conduction electron band, so it seems not to be able to



**Fig. 3.6** Top: non-local measurement configuration, bottom: electrochemical potential for up and down spins, distances in units of the spin-diffusion length. Solid line: electrochemical potential in the absence of spin effects. Jedema *et al.* (2002b), copyright Nature Publishing Group.

detect slight changes at the Fermi level. There is one instance in which NMR was able to see a so-called Knight field: the modulation of the hyperfine field, caused by a current injected through a Schottky tunnel barrier into a quantum well at the Larmor frequency of the observed nuclei (Strand *et al.* 2005).

Direct evidence for spin accumulation comes from transport measurements. Johnson and Silsbee (1985, 1987), in a ground-breaking study for the spintronics to come, relied on a three-terminal device in order to distinguish spin currents from charge currents. Nowadays, such measurements are carried out in configurations such as that of Fig. 3.6, referred to as a non-local measurements (Jedema *et al.* 2002; Garzon *et al.* 2005; Tombros *et al.* 2006; Ji *et al.* 2007).

A current  $I$  is driven into one ferromagnetic electrode (Co1) while the potential  $V$  of a second nearby electrode (Co2) is monitored at zero current. The current  $I$  is collected in a third electrode away from the voltage electrode. This measurement is analogous to a three-electrode configuration in electrochemical measurements. If a non-zero voltage is detected, it must come from the electrochemical potential arising from spin-accumulation effects.

A stringent criterion for the detection of a spin accumulation is the so-called Hanle effect, in which the precession of spins as the electrons travel from one layer of a spin-valve to the other modulates the signal expected for a give the configuration of the magnetizations of the spin-valve (Jedema *et al.* 2002b; Lou *et al.* 2007).

A magnetic demonstration of spin accumulation was achieved with a non-local spin valve (Kimura *et al.* 2006). The accumulation of spins is thought to have caused the flipping of the magnetization of a nanomagnet positioned at the voltage-probing electrode. In this type of measurement of course, experimental arguments are needed to show that the current itself is not the cause of the switch.

## 3.3 Models for spin-polarized currents acting on magnetization

### 3.3.1 Early thoughts

Conduction electron spin resonance conducted in the 1960s and 1970s required modelling of s-d interactions (Herring 1963; Harrison 1979). Today, most efforts of modelling the action of spin-polarized current on magnetization are done in the parlance of the s-d model. The interaction of conduction electron spins with magnetic moments (not necessarily a ferromagnet) was of interest for modelling experimental results on magnetic resonance. For example, Hirst (1966) asked the question of the effect of the random diffusion of the spins of conduction electrons on the relaxation of the ferromagnetic resonance. He introduced a dyadic spin current  $J(\mathbf{r}, t)$  and added the term  $-\nabla \cdot J(\mathbf{r}, t)$  to the Laudau–Lifshitz equation for the magnetization.

In order to give a description of transmission conduction electron spin resonance, Winter (1971) combines the ideas of Hasegawa (1959) and Dyson (1955) to write coupled equations for the magnetizations  $\mathbf{M}_s$  and  $\mathbf{M}_d$  (see Wegrowe *et al.* 2007 for a modern version). In a thermodynamic approach to the dynamics of s-d coupling, Hasegawa points out that there are 3 systems that need to be considered,  $\mathbf{M}_s$ ,  $\mathbf{M}_d$  and  $\mathbf{L}$ , the translational motion.  $\mathbf{L}$  is said to be a fast system. The issue is whether we can assume that  $\mathbf{M}_s$  is strongly coupled to  $\mathbf{L}$  or whether it is  $\mathbf{M}_s$  and  $\mathbf{M}_d$  that are most strongly coupled. He concludes that in most cases, the situation is mid-way between these two limits. So, he proposes Bloch equations to account for all cases. Winter takes these and adds a diffusion term  $D\nabla^2\mathbf{M}_s$  for the conduction electron spins. There is no such term for  $\mathbf{M}_d$  because localized spins are in view. Hence, Winter writes:

$$\begin{aligned}\frac{d\mathbf{M}_s}{dt} &= \gamma_s \mathbf{M}_s \wedge \mathbf{H}_0 + \lambda \gamma_s \mathbf{M}_s \wedge \mathbf{M}_d - \frac{1}{T_{sd}} \mathbf{M}_s + \frac{1}{T_{ds}} \mathbf{M}_d \\ &\quad - \frac{1}{T_{sl}} (\mathbf{M}_s - \mathbf{M}_s^0) + D \nabla^2 \mathbf{M}_s \\ \frac{d\mathbf{M}_d}{dt} &= \gamma_d \mathbf{M}_d \wedge \mathbf{H}_0 + \lambda \gamma_d \mathbf{M}_d \wedge \mathbf{M}_s + \frac{1}{T_{sd}} \mathbf{M}_s - \frac{1}{T_{ds}} \mathbf{M}_d - \frac{1}{T_{dl}} (\mathbf{M}_d - \mathbf{M}_d^0).\end{aligned}$$

The phenomenological parameter  $\lambda$  is proportional to the exchange coupling  $J_{sd}$ . Following Heitler and Teller (1936), and Overhauser (1953), Hasegawa provides expressions for the spin-spin relaxation rates  $T_{sd}^{-1}$  and  $T_{ds}^{-1}$ , by analogy with relaxation of nuclear spins through hyperfine coupling. The terms with  $T_{sl}^{-1}$  and  $T_{dl}^{-1}$  are spin-lattice relaxation processes. Hasegawa concludes with a linearization and solves for the normal mode problem. These ideas were further expounded, and the experiments that illustrate them were reviewed, by Barnes (1981).

In connection with spintronics, it is instructive to see the case treated by Winter (1971, Chapter 10, eqn 10.2) for the response to an excitation by an rf field. He adds to the equation of evolution for the magnetization, the Maxwell equations and Ohm's law,  $\mathbf{j} = \sigma \mathbf{E}$ . He states as the boundary condition that there is no magnetization flow across the surface. This is where

the modern work departs radically from these older experiments: because of spin-accumulation effects, there is a magnetization flow at the boundaries!

### 3.3.2 The spin-transfer torque

In Section 3.2, we introduced the notion of spin accumulation. We focused on situations where the magnetizations of two adjacent layers were collinear. One of the major consequence of the existence of spin accumulation is the possibility to develop a torque on magnetization by passing a current through magnetic nanostructures defined on a length scale smaller than the spin-diffusion length. This was the notion first introduced in the seminal papers of Slonczewski (1996, 1999) and of Berger (1996, 2001). We can well imagine that spin accumulation is no longer parallel to the magnetization in the case where the adjacent magnetizations are no longer collinear (i.e. either parallel or antiparallel). A tensorial description of spin accumulation can be found in particular in Stiles and Zangwill (2002b) and in the review article of Brataas *et al.* (2006). The key of the concept of spin torque is that, when a spin current  $\mathbf{j}_m$  enters a ferromagnetic layer, its spin component that is perpendicular to the magnetization relaxes over very short distances.

Using a “sd” model for the interaction of conduction electron to the magnetization, Shapiro *et al.* (2003) write an equation of evolution for the spin accumulation  $\mathbf{m}$  of the form:

$$\frac{\partial \mathbf{m}}{\partial t} + \frac{\partial \mathbf{j}_m}{\partial t} + \frac{J_{sd}}{\hbar} \mathbf{m} \times \mathbf{M}_d = -\frac{\mathbf{m}}{\tau_{sf}}$$

where  $J_{sd}$  is the exchange constant of the sd model. As it should, this equation bears some resemblance with the phenomenology developed for electron spin resonance in alloys (Section 3.3.1). These authors show, then, that the component of the spin accumulation parallel to the magnetization,  $\mathbf{m}_{\parallel}$ , decays in the ferromagnet over a distance of the order of the spin-diffusion length  $l_{sf} = \sqrt{2D_0\tau_{sf}}$ . The diffusion constant may be written  $D_0 = (1/3)v_F\lambda_{mfp}$  where  $\lambda_{mfp}$  is the electron mean-free path. The component of the spin accumulation perpendicular to the magnetization decays according to:

$$\frac{\partial^2 \mathbf{m}_{\perp}}{\partial x^2} - \frac{\mathbf{m}_{\perp}}{l_{sf}^2} - \frac{\mathbf{m}_{\perp} \times \mathbf{M}_d}{\lambda_J^2} = 0$$

Here, the length scale is given by  $\lambda_J = \sqrt{2\hbar D_0/J}$ . It depends on the exchange constant  $J_{sd}$ . Hence,  $\lambda_J$  is very much shorter than the spin-diffusion length. Shapiro *et al.* argue that experimental values for  $J_{sd}$  are scarce and quote a value deduced from ferromagnetic resonance,  $J_{sd} \approx 0.1$  eV (Cooper and Uehling 1967). With a typical value for  $\lambda_{mfp}$ , this gives an  $\lambda_J$  of about 1 to 3 nm. An estimate for  $J_{sd}$  can be obtained from transport measurements through domain walls. Wegrowe *et al.* (2003) measured the resistance of a domain wall as a function of its thickness. The resistance  $R$ , starting from a value close to the expected “spin-coupled interface resistance” (Valet and Fert 1993), decreased with increasing wall thickness  $\delta$ , with  $\partial R/\partial \delta \approx 0.1\Omega/\text{nm}$ . Using the theoretical model of Levy and Zhang (1997), this implies for  $J_{sd}$  a value

of 0.14 eV. That this  $J_{sd}$  is small compared to an atomic exchange was already made clear by Mitchell (1957) and Kittel and Mitchell (1956) who used a value of 0.01 eV only, to account for their resonance observations, arguing that  $J_{sd}$  is not the atomic value, but a screened value, about a factor of 30 less than the atomic value. Such a small value is consistent with the notion (Herring 1963) that the sd exchange coupling is very small compared to the overall exchange energy.

Experiments measuring the angle of precession of the spin of an electron crossing a thin ferromagnetic layer can provide further information (Filipe *et al.* 1998; Weber *et al.* 2001; Cacho *et al.* 2002). For example, Weber *et al.* (2001) measured the precession angle as a function of thickness. Their data imply exchange energies of the order of 0.3, 0.2 and 0.08 eV for Fe, Co and Ni, respectively.

The description of Shapiro *et al.* (2003) is distinct from the arguments in the seminal work of Berger and Slonczewski. There, the precession of the spin of conduction electrons is considered in the exchange field of the interatomic exchange interaction (Slonczewski 2002). As the latter is widely dispersed over the Fermi sphere, an extremely rapid transverse spin relaxation is expected. Hence, decay over not much more than one atomic layer is expected in this view.

No matter what approach is used to infer the decay of the transverse spin accumulation, the practical conclusion is that it occurs over a distance of about 1 nm upon entering a ferromagnet. Using the notation of Shapiro *et al.* (2003), we write an equation of motion for the magnetization  $\mathbf{M}$ , staying with the “sd” picture for the effect of the spin accumulation:

$$\frac{d\mathbf{M}}{dt} = -\gamma \mathbf{M} \times (\mathbf{H}_e + J\mathbf{m}) + \alpha \mathbf{M} \times \frac{d\mathbf{M}}{dt},$$

where  $\gamma_0$  is the gyromagnetic ratio,  $\mathbf{H}_e$  the effective magnetic field, containing all contributions such as applied field, anisotropy field, etc... The added term  $J\mathbf{m}$  is meant to represent the coupling between the localized moments  $\mathbf{M}$  and the incoming spin current. The last term is the usual Gilbert damping. Evidently, the longitudinal part of  $J\mathbf{m}$ , that is, the part that is parallel to  $\mathbf{M}$ , has no effect on  $\mathbf{M}$ . We can replace  $\mathbf{m}$  by  $\mathbf{m}_\perp$  in this equation without changing it. As suggested by Zhang *et al.* (2002), we can write in all generality that  $\mathbf{m}_\perp$  contains two terms:

$$J\mathbf{m}_\perp = a\mathbf{M}^f \times \mathbf{M} + b\mathbf{M}^f,$$

where  $\mathbf{M}^f$  is the magnetization of the other layer, typically, it is the fixed one, whereas  $\mathbf{M}$  is the magnetization of the free layer of which one studies the magnetization dynamics. This expression for the incoming spin current inserted in the equation of evolution for  $\mathbf{M}$  allows for the possibility of having two terms. One term,  $b\mathbf{M} \times \mathbf{M}^f$ , represents a field  $b\mathbf{M}^f$  acting on  $\mathbf{M}$ . The other term,  $a\mathbf{M} \times (\mathbf{M}^f \times \mathbf{M})$ , is the term called spin-transfer torque (STT). It is the term that was first announced by Berger (1996, 2001) and Slonczewski (1996, 1999).

Thus, the Landau–Lifshitz equation in the approximation of a free layer, of volume  $V_m$ , being a single domain of magnetization, has the form (Kovalev

*et al.* 2007):

$$\frac{d\mathbf{M}}{dt} = -\gamma \mathbf{M} \times \mathbf{H}_e + \frac{\alpha}{M_s} \mathbf{M} \times \frac{d\mathbf{M}}{dt} + \gamma \frac{\hbar}{2e} \frac{I}{V_m} \times \left[ \eta_1 \frac{\mathbf{M}}{M_s} \times \left( \frac{\mathbf{M}^f}{M_s} \times \frac{\mathbf{M}}{M_s} \right) + \eta_2 \left( \frac{\mathbf{M}^f}{M_s} \times \frac{\mathbf{M}}{M_s} \right) \right]. \quad (3.1)$$

From the standpoint of an experimentalist contemplating a test of this model, it may be important to note that Kovalev *et al.* (2007) showed that the same equation can, algebraically, be recast in the form of the Slonczewski-type spin-transfer torque:

$$\frac{d\mathbf{M}}{dt} = -\gamma \mathbf{M} \times \mathbf{H}_e + \frac{\alpha}{M_s} \mathbf{M} \times \frac{d\mathbf{M}}{dt} - \gamma \frac{\hbar}{2e} \frac{I}{V_m} \frac{\mathbf{M}}{M_s} \times \left( [\mathbf{I}_{s1} + \mathbf{I}_{s2}] \times \frac{\mathbf{M}}{M_s} \right),$$

provided the coefficients  $\eta_1$  and  $\eta_2$  are angle dependent. Here,  $\mathbf{I}_{s1}$  and  $\mathbf{I}_{s2}$  are functions of the spin-dependent conductivities and the spin-mixing interface conductance. The latter is defined by the same workers in the framework of their so-called ‘‘magnetoelectronics’’ in the non-collinear configuration of the magnetizations (Brataas *et al.* 2006).

A description of spin-dependent transport with non-collinear magnetization can be found in various other places in the recent literature. The transport description is based on the assumption of semi-classical transport in bulk materials that is valid for diffuse and chaotic systems, as well as a separation of timescales of the electronic and magnetic degrees of freedom. Interfaces, because they are sharp, like collisions, must be treated as quantum-mechanical boundary conditions for electron-distribution functions and cast into a description of non-equilibrium transport. The crucial material parameter is the spin-mixing conductance  $G_{\uparrow\downarrow}$  that can be computed from *ab initio* electronic band structures. The real part of the mixing conductance (Brataas *et al.* 2000; Hernando *et al.* 2000; Piéchon and Thiaville 2007), a material parameter that describes the transport of spins that are non-collinear to the magnetization direction at the interface, is proportional to the torque acting on the ferromagnet in the presence of a non-collinear spin accumulation in the normal metal (Stiles and Zangwill 2002a; Xia *et al.* 2002).

The relative magnitude of the field term and the torque term has been discussed (Elliot *et al.* 2006; Edwards and Mathon 2007). These terms, being proportional to  $\eta_1$ , respectively  $\eta_2$ , depend on the respective quantities (Barnas *et al.* 2005, 2006; Gmitra *et al.* 2006):

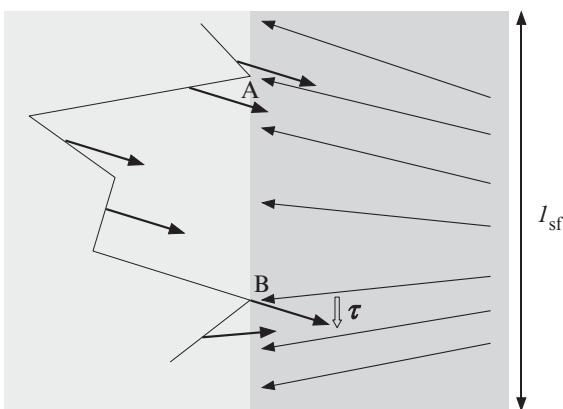
$$\begin{aligned} a &= \frac{\hbar}{e} \left[ \operatorname{Re} \{G_{\uparrow\downarrow}\} (\cot \theta (g_x \cos \varphi + g_y \sin \varphi) - g_z) \right. \\ &\quad \left. + \operatorname{Im} \{G_{\uparrow\downarrow}\} \frac{(g_x \cos \varphi + g_y \sin \varphi)}{\sin \theta} \right] \\ b &= \frac{\hbar}{e} \left[ -\operatorname{Im} \{G_{\uparrow\downarrow}\} (\cot \theta (g_x \cos \varphi + g_y \sin \varphi) - g_z) \right. \\ &\quad \left. + \operatorname{Re} \{G_{\uparrow\downarrow}\} \frac{(g_x \cos \varphi + g_y \sin \varphi)}{\sin \theta} \right]. \end{aligned}$$

Refer to Gmitra *et al.* (2006) for the definition of the angles that specify the orientation of  $\mathbf{M}$  with respect to  $\mathbf{M}_s$ . Again, the key ingredient of this effect is the spin accumulation occurring in the spacer layer, here denoted as  $\mathbf{g}$ .

There is an intuitive way to think of the spin-transfer torque. It is based on the notion of conservation of the total angular momentum. Under this view, if the conduction electron spin polarization that is perpendicular to the magnetization is cancelled, we must conclude that a torque is applied to the magnetization. This spin conservation is made clearer if one consider the form of the interaction of conduction electrons with magnons, as presented for example, by Mitchell (1957).

In order to get the dynamics of the interplay between the spin-polarized current and magnetization, one uses the very large difference in timescales of evolution for the magnetization and the electron spins. Thus, one works out the transport equation for a set magnetic configuration and deduces a spin accumulation  $\mathbf{g}$ . From it, a spin-transfer torque is deduced and a new magnetic configuration is deduced. This calculation cycle is repeated until the full time evolution of magnetization is obtained. See, for example, Stiles *et al.* (2004). A solution in terms of transfer matrices has been worked out by the group of Dieny (Manchon *et al.* 2007) for any arbitrary stacking of magnetic layer and metallic spacers.

Thus far, we have discussed spin torque as if the magnetization was uniform in each layer of a pillar. A spin polarization arose because of the spin-dependent conductivities, causing a spin accumulation as the current crossed at least two ferromagnetic layers. It has been pointed out that magnetic excitation may also be possible with an unpolarized current crossing a single ferromagnetic layer (Stiles 2004). The basic physics involved is described in the review article of Stiles and Miltat (2006). A graphical representation of the basic idea for the mechanism they have in mind is illustrated in Fig. 3.7. The same point was developed by Polianski (Polianski and Brouwer 2004; Shaffique *et al.* 2006). For Yanson *et al.* (2005), it is impurity scattering at the interface that is the driving mechanism for magnetic excitations. The dynamics of magnetization in the micromagnetics approximation is reviewed by Miltat *et al.* (2002).



**Fig. 3.7** Sketch suggestive of the principle by which a spin-polarized current may excite the magnetization of a single layer: at A, a total reflection is assumed, then diffusion takes place without spin flip before the electron hits the interface at B, where the orientation of the magnetization differs, and the transverse moment is absorbed.

### 3.3.3 Modified Landau–Lifshitz for domain walls

So far, we have addressed the interplay of spin polarization of the conduction electrons with the magnetization in geometries where the electrons crossed an interface between a non-magnetic metal and a ferromagnet. In the reference frame of the electron, this corresponds to an abrupt change of the exchange field. This situation is known as “sudden passage” in the terminology of magnetic resonance: the field is switched on or off in a time shorter than the spin-lattice relaxation time. When an electron crosses a domain wall, the spin dynamics turns out to be in a very different regime called an “adiabatic passage” in magnetic resonance: the field perceived in the frame of the electron is rotated at an angular velocity that is far less than the precession rate of the spin about the field. Thus, the spin tends to stay parallel to the field.

In the following, we leave aside the physics of ferromagnetic nanocontacts (Garcia *et al.* 1999; Tatara *et al.* 1999). It has been shown that a domain wall forms at the nanoconstriction (Bruno 1999; van Gorkom 1999a). Thus, nanocontacts may be small enough that quantum transport becomes relevant, which is far from the diffusive regime of nanowires (Sokolov *et al.* 2007). Likewise, we leave aside the transport regime of the seminal paper of Cabrera and Falicov (1974), in which the electron mean-free path is longer than the domain-wall width, though it has been pointed out that this special case may be realized in SrRuO<sub>3</sub>, an itinerant ferromagnet (Klein *et al.* 2000). The possibility of a localization effect in domain walls was considered (Tatara and Fukuyama 1997; Lyanda-Geller *et al.* 1998), and experimental data do not exclude this as a possibility (Rudiger *et al.* 1998).

The first insight into the proper description of spin dynamics in domain walls arose from trying to model domain-wall resistance. Positive and negative magnetoresistance, meaning an increase or a decrease in resistance due to a wall, appear to be possible in the diffusive regime (van Gorkom *et al.* 1999a,b). However, most data available show an increase of resistance in the presence of a wall (Ebels *et al.* 2000; Viret *et al.* 1996). In thin cobalt film, it was possible to generate a set of parallel domains (“stripe domains”) and decrease the number of walls by applying a field, which caused a marked decrease in resistance (Gregg *et al.* 1996a). By connecting a cobalt nanowire to a Gd film, it was possible to detect the change of resistance as a function of the thickness of the wall (Wegrowe *et al.* 2000). A resistance increase was also detected in individual walls in FePd (Danneau *et al.* 2002).

The basic physical concept invoked for this negative magnetoresistance is the idea that the spins nearly perfectly track the exchange field, but not quite. This concept had been already invoked for the spin transport in granular materials (Gregg *et al.* 1996b; Gehring 1995). Analytical expressions for the resistance contribution in terms of the mistracking of the spin of conduction electrons precessing about the exchange field were developed (Levy and Zhang 1997) and tested against experiments (Wegrowe *et al.* 2000).

While magnetoresistance is the effect of the presence of a domain wall on the conductivity of ferromagnetic metals, we wish to address here the question as to how to describe the effect of a spin-polarized current on the magnetization in a domain wall. The static and dynamical properties of walls in nanowires

have been reviewed by Thiaville and Nakatani (2006). The primary effect of spin-polarized current on magnetization was identified earlier by Bazaliy *et al.* (1998).

This can be understood in an sd-interaction picture. As the spins undergo a fast precession about the exchange field, they follow it, provided the change in field orientation is slow, the same argument as for transport consideration. So, if we assume a current density  $j_e$  flowing in the  $x$ -direction, in the presence of a magnetization  $\mathbf{M}(x)$ , we expect a spin current given by:

$$\mathbf{j}_x = P\mu_B \left( \frac{j_e}{e} \right) \frac{\mathbf{M}(x)}{M_s},$$

where  $P$  is the spin polarization. In ferromagnetic metals like Ni, Co and Fe,  $P$  has been measured using tunnelling into superconductors. So, if there is a slowly varying magnetization, the conduction electron spin contribution to the total angular momentum changes, and there must be a torque proportion to  $\mathbf{j}_s \cdot \frac{\partial \mathbf{M}}{\partial x}$ . Considering that the modulus of the magnetization is constant, this can be rewritten as:  $\tau_B \propto -\frac{1}{M_s^2} \mathbf{M} \times (\mathbf{M} \times \frac{\partial \mathbf{M}}{\partial x})$ . Various approaches for describing this phenomenon lead to a torque proportional to  $d\mathbf{M}/dx$  (Tatara and Kohno 2004; Ansermet 2004; Waintal 2004).

If the wall is thin, the spin dynamics of the conduction electrons approaches that experienced in magnetic layers separated by non-magnetic metals, and the spin-transfer torque contribution can be expected (Tatara and Kohno 2004). From pure geometrical considerations also, we can say that all possible orientations of the torque can be accounted for if a term normal to the magnetization and normal to the first term is added. Proper derivations (Tatara and Kohno 2004; Li and Zhang 2004; Zhang and Li 2004) lead effectively to two terms:

$$\begin{aligned} \frac{\partial \mathbf{M}}{\partial t} = & -\gamma \mathbf{M} \times \mathbf{H}_{\text{eff}} + \frac{\alpha}{M_s} \mathbf{M} \times \frac{\partial \mathbf{M}}{\partial t} - \frac{b_J}{M_s^2} \mathbf{M} \\ & \times \left( \mathbf{M} \times \frac{\partial \mathbf{M}}{\partial x} \right) - \frac{c_J}{M_s} \mathbf{M} \times \frac{\partial \mathbf{M}}{\partial x}. \end{aligned} \quad (3.2)$$

In this modified Landau–Lishitz equation, the familiar precession term and damping terms are augmented by what is referred to, in the order they appear, as the adiabatic and the non-adiabatic terms. Their respective coefficients are:

$$b_J = P\mu_B \left( \frac{j_e}{e} \right) \frac{1}{1 + \xi^2} \quad c_J = P\mu_B \left( \frac{j_e}{e} \right) \frac{\xi}{1 + \xi^2},$$

where  $\xi = \frac{\tau_{\text{ex}}}{\tau_{\text{sf}}}$ . For a typical ferromagnet, the exchange energy  $J_{\text{ex}}$  is of the order of 1 eV, the spin-flip time is in the picosecond range, so the parameter  $\xi$  is about 0.01. The non-adiabatic term is a correction term that accounts for the partial failure of the spins to track the magnetization as they go through the wall, quantified by the parameter  $\xi$  (Levy and Zhang 1997). The authors warn us against discarding this much smaller term. They show that the terminal velocity of a domain wall is independent of the strength of the parameter  $b_J$ .

but depends on this  $c_J$  term. The importance of the non-adiabatic term was also discussed in a phenomenological framework (Vanhaverbeke and Viret 2007). The case of a vortex wall has been analyzed theoretically by He *et al.* (2006).

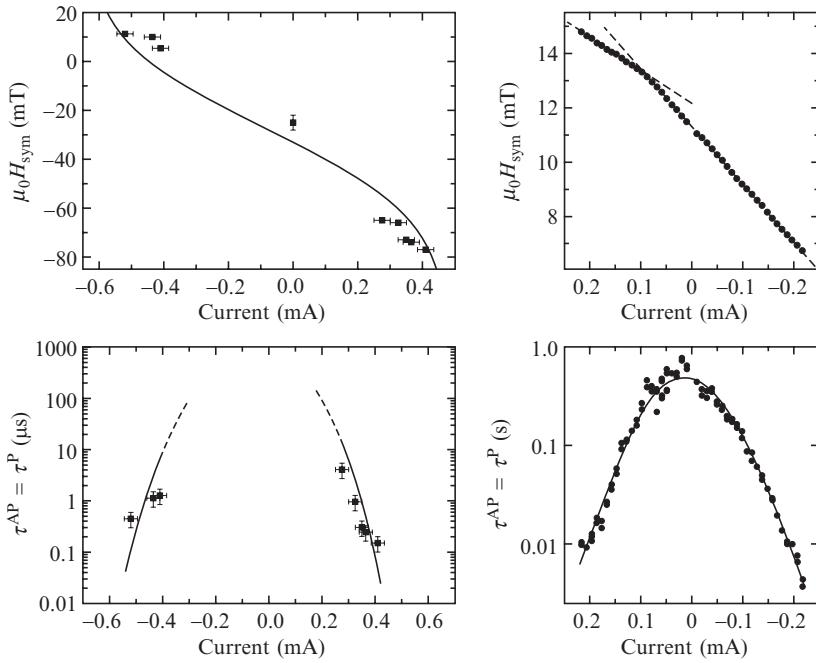
## 3.4 Current-induced magnetization switching

### 3.4.1 Time-to-switch experiments

The most frequent approach for studying the spin-transfer torque effect has been to observe a so-called “current-induced magnetization switching” (CIMS). This seems the most natural test since the seminal predictions of Berger (1996) and Slonczewski (1996) referred to switching the free layer of a spin valve. However, the initial evidence for spin-torque effects was different. The experimental observation recognized by most as the earliest verification of a spin-torque effect in spin-valves under a point contact, was the work of the collaboration between groups of Grenoble and Michigan State University (Tsoi *et al.* 1998a). The main observation was a peak in  $dV/dI$  measurements (this, and oscillations under dc current, are discussed in a section below), but mention is made also of observations of random switching (Fig. 3.2) as described in this section. Next came the work of the Cornell group (Myers 1999), also detecting  $dV/dI$  peaks with spin-valves. Nearly at the same time, the group of Radelaar published data (Theeuwen *et al.* 1999) that showed a change in switching field with current, that they interpret as evidence of CIMS. This author’s group initially looked at the effect of a current on homogeneous nanowires (Wegrowe *et al.* 1999a) (see Section 4.3 on domain walls). In the period of 2003 to 2007, the literature on the switching behavior of spin-valves became abundant, to the point that it is a challenge to review it fully and fairly.

Wernsdorfer *et al.* (1997) had shown how much can be learned about magnetic nanostructures from switching-time measurements. It is only natural to extend this type of study and see how spin injection affects these measurements. This author’s group found two-level fluctuations in spin-valves (Fabian *et al.* 2003). Repeated measurements allowed them to make histograms of the residence times in the parallel and antiparallel states of a single spin-valve. Average times were deduced from fits of these histograms to Poisson distributions. These average times were then measured as a function of field and current. As the current is high and the full measurement requires several days, it is important to have good samples or/and good luck! However, these measurements are worth the trouble, because they yield an in-depth characterization of current-induced magnetization switching. The Cornell group collected exceedingly good data sets, covering several decades in time (Emley *et al.* 2006). They examined random switching as a function of the temperature, thus gaining insight into the temperature dependence of the Gilbert damping parameter.

Tunnel junctions are known to be prone to a hierarchy of telegraph-type fluctuations, presumed at the time to be highly sensitive to the magnetic state of impurities in the barrier (Doudin *et al.* 1997; Sokolov *et al.* 2007). The Cornell



**Fig. 3.8** Value of applied field at which the magnetic state spends equal times in the P and AP configurations (top left and right) and value of this time as a function of current (bottom left and right), for metallic pseudospin-valves (left) and for MgO tunnel junctions (right). Lines are fits to eqn (3.3). Fabian *et al.* (2006), copyright IOP Publishing, Fuchs *et al.* (2006), copyright American Physical Society.

group found that under spin torque due to current through MgO tunnel barriers, random fluctuations occurred near the main jumps of the magnetization (Fuchs *et al.* 2006). Their analysis of the statistics of residence times follows the earlier works done with metallic spin-valves (Fig. 3.8). It is based on the work of Li and Zhang (2003) according to which, the residence times in the parallel (P) or antiparallel configurations (AP) can be written as:

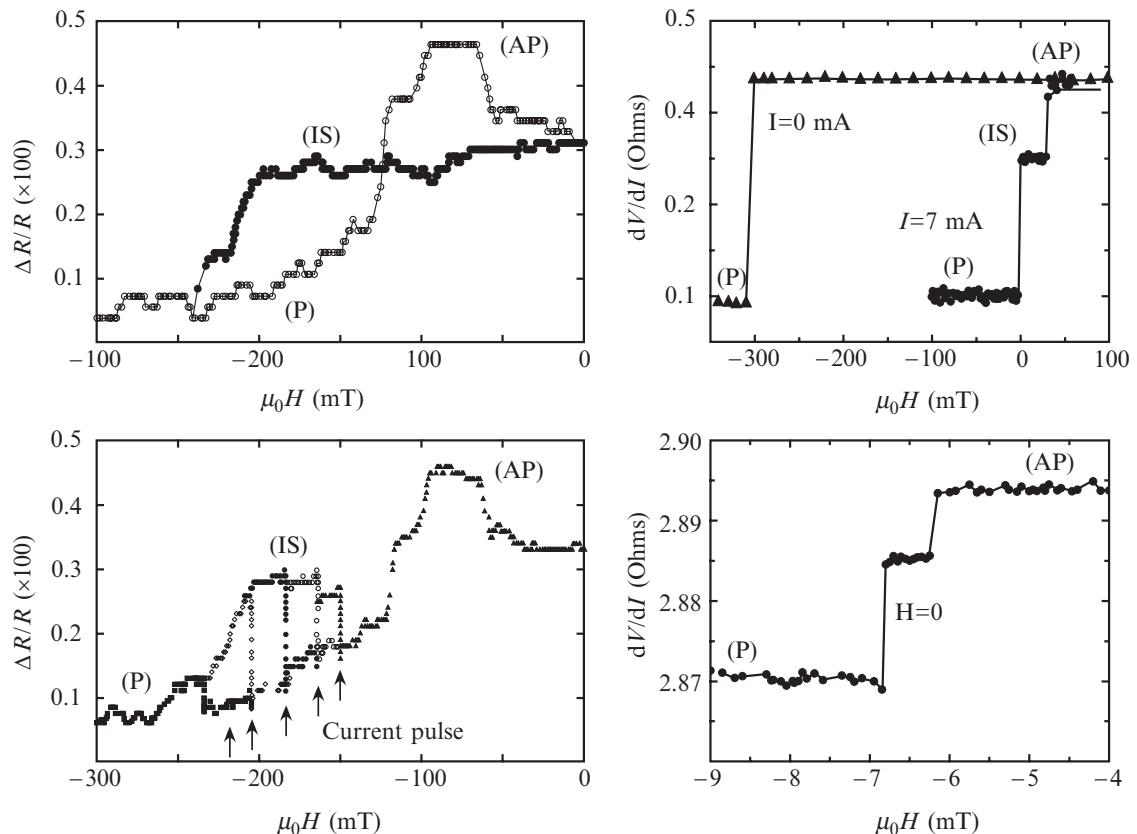
$$\tau_{\text{P(AP)}} = \tau_0 \exp \left[ \frac{E_a}{kT} \left( 1 \pm \frac{H - H_{\text{dip}}}{H_c} \right)^\eta \left( 1 \mp \frac{I}{I_c^{\text{P(AP)}}} \right) \right] \quad (3.3)$$

The pre-exponential factor  $\tau_0$  is expected to be of the order of 1 ns,  $E_a$  is a barrier energy for switching,  $\eta$  is known to be 3/2 unless the applied field  $H$  is exactly along the anisotropy axis (in which case  $\eta$  is 2),  $H_{\text{dip}}$  quantifies the magnetostatic coupling of the free layer to the fixed layer,  $I_c^{\text{P}}$  and  $I_c^{\text{AP}}$  are the critical currents at zero temperature.

Tunnel junctions, except for their spin-torque physics, are beyond the scope of this chapter. For MgO barriers, see, e.g. the work at Grandis (Diao *et al.* 2006; Huai *et al.* 2006).

### 3.4.2 Current-hysteresis measurements

The Cornell group (Myers *et al.* 1999) showed that a current-driven hysteresis of the magnetization can be obtained, owing to the spin-transfer torque. Their follow-up paper (Katine *et al.* 2000) presents clear current hysteresis data at low field (Fig. 3.2) and goes on to characterize peaks in  $dV/dI$  response as a function of current and applied field (see Section 5).



**Fig. 3.9** Reaching intermediate magnetic configurations with current-induced magnetization switching. Ravelosona *et al.* (2006), copyright American Institute of Physics.

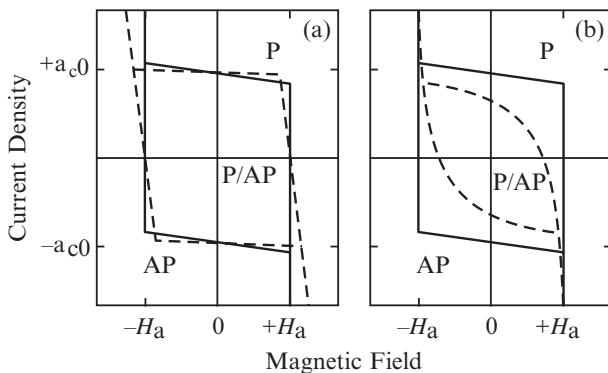
The author's group had a first look at switching with a current pulse (Wegrowe *et al.* 2001a) by making nanowires containing 5 layers of Co. The point about making a stack of 5 Co layers was to insure a high yield in sample production when seeking magnetic nanostructure with clean magnetic switching. In that study, we found that we could reach, by a pulse of current, intermediate states as compared to the states reached by a full magnetic field sweep that starts from saturation. This intermediate state could be accessed by a minor loop in the field (Fig. 3.9 top left) or by a pulse in current applied at the end of a field sweep coming from saturation (Fig. 3.9 bottom left).

Intermediate states were also found in genuine spin-valves (Yang 2005a). In a study of CIMS using a perpendicular polarizer, Ravelosona *et al.* (2006) found superb evidence for intermediate states also reached by a field (Fig. 3.9 top right) or by a current pulse (Fig. 3.9 bottom right). The author's group produced  $R(I)$  hysteresis of spin-valves (Wegrowe *et al.* 2002). However, these measurements were difficult because the nanowires were quite sensitive to high current densities for extended times.

In the more recent studies on switching by STT, a striking development has been the use of *perpendicular polarization*. One instance was just mentioned (Ravelosona *et al.* 2006). They used a combination of Co/Pt and Co/Ni multilayers. Others used CoFe/Pt multilayers (Hao and Wang, 2004, 2006) or FePt with Au as a spacer (Seki *et al.* 2006a,b,c). At Seagate, an orthogonal state between the magnetization of two adjacent layer was obtained in a synthetic antiferromagnet composed of two CoFe layers separated by a Ru layer about 1 nm thick (Covington *et al.* 2004).

In the brief review of the effect of spin-polarized current on magnetization, the possibility of a contribution in the form of a *spin current-induced field* was evoked. The group of Kent (NYU) with the spintronics team at IBM Yorktown Heights carried out a very detailed study that allowed them to quantify this field contribution (Zimmler *et al.* 2004). This field is estimated from their measurements at 4 K to be of  $1.5 \times 10^{-7}$  Oe(cm<sup>2</sup>/A). This is about a factor of 5 smaller than the spin-transfer torque, i.e. the original Slonczewski type (Slonczewski 1996) (first additional term in eqn (3.1)). They point out that the spin-transfer field effect tilts the boundaries of the so-called phase diagram (Fig. 3.10) showing which magnetic configurations, parallel (P) or antiparallel (AP) are stable at a given current and applied field. Unfortunately, as pointed out by these authors, thermal agitation has a similar effect on the pure Slonczewski spin-torque dynamics in the quasi-static regime (Fig 10(b)). Further experimental precautions are needed because, given the size of the field effect, the measurement protocol must distinguish this spin-transfer field from the field induced by the current, often called the “Oersted field”. The main point of the method is to observe the magnetization hysteresis of the free layer of a spin-valve with two opposite directions of the magnetization of the pinned layer. The magnetic coupling between the two layers is quantified by the observation of the shift of the midpoint of the field-driven hysteresis of the free layer.

Another group addressed the question of the spin current-induced field (Yang *et al.* 2005b, 2006). They conclude, contrary to all others, that this is effect dominates over the torque due to the fast relaxation of the transverse spin accumulation. However, it must be said that this group managed to make a nanoscopic lead to the pillar, on the free-layer side. They argue, on the basis



**Fig. 3.10** Solid lines: Form of the zero-temperature phase diagram of magnetic states in the Slonczewski torque model for a single-domain magnet with uniaxial anisotropy. Dashed lines: (a) Phase diagram with an additional effective field term of the form  $\eta_2 \hat{\mathbf{m}}_{\text{fixed}} \times \hat{\mathbf{m}}$ , assuming  $\eta_2 > 0$ . (b) Finite-temperature phase diagram in the torque model,  $\eta_1 \hat{\mathbf{m}} \times (\hat{\mathbf{m}}_{\text{fixed}} \times \hat{\mathbf{m}})$ . Regions in which there are precessional magnetic states are omitted. Zimmler *et al.* (2004) copyright American Physical Society.

of their simulations (Hamrle *et al.* 2005) that, by decreasing the width of the top electrode down to that of the nanopillar, the spin accumulation is increased, while the spin current is decreased. Hence, they measured the critical current for switching as a function of the size of the top electrode. They insured that the Oersted field was negligible by monitoring the effect of flipping the bottom layer, as in the work of Kent *et al.*

We mentioned in Section 3.2 the prediction of a spin torque effect in a *single ferromagnetic layer*. The experimental demonstration of switching was made in the form of current hysteresis of the resistance at a point contact to a Co layer, about 400 nm thick, covered by a natural CoO antiferromagnetic layer (Chen *et al.* 2004). As stated by the authors: “The exchange bias on the top surface of the Co layer appears to be essential for the observed STT effects in a single Co layer.” Notably, no effect could be seen if the CoO layer was replaced by a Au layer. The magnetoresistance detected in this single layer is a domain-wall resistance. The presence of a domain wall, or the nucleation of one, seems to be essential. Thus, the authors point to the earlier prediction by Berger of a “domain drag” effect (Berger 1978).

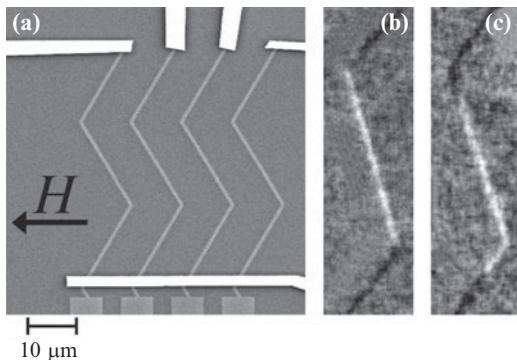
Contrary to this work, the group of Kent (Oozyilmaza *et al.* 2006) observed switching in a pure single layer, contacted by two Cu electrodes. They found current hysteresis with marked jumps of the resistance, five times larger than the AMR. This suggested to these authors that the current had generated a different magnetization state, presumed to be a vortex.

The earliest evidence of a spin-torque effect in a single ferromagnetic layer is given incidentally in one of the earliest experimental papers on spin-transfer torque (Myers *et al.* 1999). Indeed, in that paper, a peak in  $dV/dI$  measurement was observed (see their Fig. 1E), indicative of a sustained precession of the magnetization (see section below).

### 3.4.3 Domain-wall motion

The precursor to all attempts to move a wall with a current, to the exclusion in principle of the action of the field induced by the current, goes back to the experiments of Freitas and Berger (1985) on permalloy films, followed by those of Sahli and Berger (1994). Berger had considered the effect of the motion of electrons through a wall when the electron spins track the magnetization because of the s-d exchange interaction (Berger 1984, 1986).

As this author’s group was studying in the 1990s the magnetization-reversal properties of homogeneous nanowires, it was natural to try and send a pulse of current of large current density in our nanowires, prompted by the seminal papers of Berger (1996) and Slonczewski (1996). Thus, we first observed magnetization switching in Co nanowires (Wegrowe *et al.* 1999a). There, the magnetoresistance data show clearly that we had a wall trapped in the wire. Next, we tried nickel nanowires, the magnetoresistance of which indicated a single-domain behavior (Wegrowe *et al.* 2001b). While the part of the sample that dominated the magnetoresistance was clearly single domain, one end of the nanowire was in some instances ferromagnetic and possibly multidomain.



**Fig. 3.11** Permalloy zigzag structures (a) are subjected to a field pulse in the direction of the white arrow (b) forming a wall at the bend in the permalloy strip, that is pushed by an electron current. Klaeui *et al.* (2005b) copyright American Physical Society.

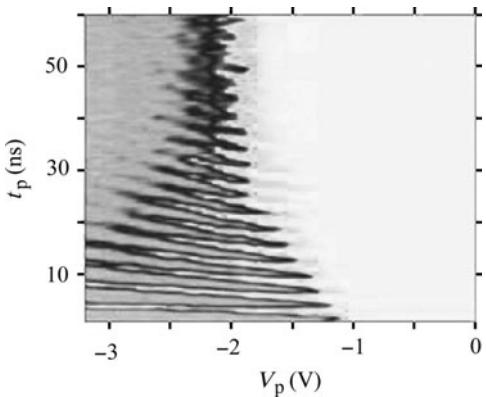
Its contribution to the resistance was vanishingly small because it was so much bigger than the nanowire. It served, however, to nucleate a wall under the effect of the current, that would then propagate through the nanowire. It was shown how the presence or not of such a ferromagnetic contact changed the field at which switching was triggered by a given current density (Wegrowe *et al.* 2004).

Demonstrating whether or not the so-called Oersted field matters, is a delicate question and detailed experimental investigations are needed to quantify this contribution to the overall effect of the current driven through the ferromagnet itself. Recent results of the NIST group show the role of the Oersted field on the magnetic configuration in their permalloy bars (Yuan *et al.* 2007).

Klaeui *et al.* (2004, 2005a), in a first set of experiments, confirmed the ability to move a wall with an injected current, using mesoscopic ring structures. They went on to demonstrate the motion of a domain in a zigzag structure of permalloy (Fig. 3.11), using spin-polarized scanning electron microscopy (Klaeui *et al.* 2005b).

One special feature of the research on the effect of spin-polarized current on walls is that various imaging techniques apply, as the structures can be made planar. A stunning example was provided by Yamaguchi *et al.* (2004a,b). Tsoi *et al.* (2003) used magnetic force microscopy. Besides the scientific interest in uncovering this new effect, there might be applications to be considered. The group of Cowburn pointed out that logic gates can be made with ferromagnetic loop structures (Allwood *et al.* 2002). Later, they showed also that a spin-polarized injected current could drive the wall and that walls could propagate quite fast, despite their lateral confinement (Atkinson *et al.* 2003). The group of S.S.P. Parkin at IBM Almaden (Thomas *et al.* 2006) showed that the probability to move a pinned wall with a current pulse depended critically on the duration of the pulse (Fig. 3.12). The timescale is that of the period of the free oscillation of the wall in its pinning potential.

The current density for all metals, as reported thus far, is of the order of  $10^7 \text{ A/cm}^2$ . However, the group of Ohno demonstrated wall motion induced by injected currents at the level of  $10^5 \text{ A/cm}^2$ , using the magnetic semiconductor GaMnAs. The lower saturation magnetization accounts for this lower current density (Chiba 2003; Yamanouchi 2004).



**Fig. 3.12** Probability of domain-wall motion (white for zero, gray for 1), as a function of pulse length and intensity. Thomas *et al.* (2006), copyright Nature Publishing Group.

## 3.5 Current-driven magnetic excitations

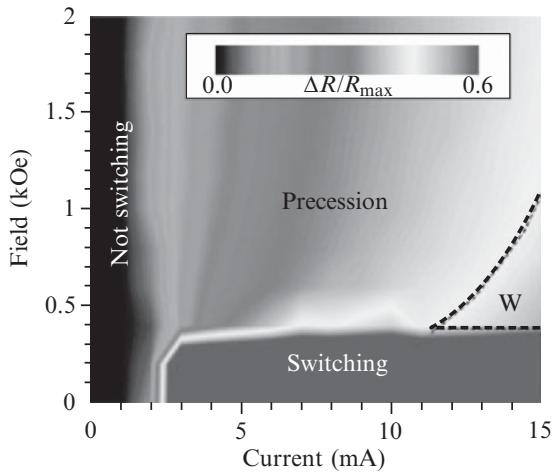
### 3.5.1 Highlights

Around the year 2000, it became apparent that a dc current may not only trigger a magnetization reversal, but under certain conditions of current and field, a precession of the magnetization can be maintained by the dc current. As pointed out with much fervour by Silva (2007), this is the first time since Oersted that we have a new means of tilting a magnetization.

Many simulations discuss various aspects of the phenomenon. The review of Stiles and Miltat (2006) provides much insight into the general ideas about spin torque as well as details on the full range of phenomena, from transport to micromagnetics. Our purpose here is to focus on experimental findings. Before reviewing them, we note in passing that, in parallel with the exploration of the use of spin-polarized currents to flip magnetization or to make it precess, there have been important results achieved in the so-called “precessional switching”. Although publications appeared over a spread of a few years, it appears that the results presented at conferences in 2001 on precessional switching were experimentally implemented and observed by several groups (Gerrits *et al.* 2002; Kaka *et al.* 2002; Schumacher *et al.* 2003).

Making point contacts to unbound Co-Cu multilayers, Tsoi *et al.* (1998a,b) observed spikes in the  $dV/dI$  measurements. The Cornell group formed lithographed point contacts to large films (Myers *et al.* 1999) and noticed also that at fields beyond saturation, in other words when switching is no longer expected, non-hysteretic peaks in  $dV/dI$ , like those of Tsoi *et al.* occurred. The same group went on to form Co/Cu/Co pillars, and found once again that if the field is raised enough, instead of switching, spikes in the  $dV/dI$  appear. Katine *et al.* declare them to be a “precessing spin wave state” (Katine *et al.* 2000).

While this phenomenon is interesting in itself as a further means of analysis of the spin-transfer torque effects, it has potential as an efficient means of flipping magnetization in random access memories. Efficient refers here both to the energy cost and the speed of writing. Thus, many groups observed and studied the transition from the regime of switching between the P and AP states



**Fig. 3.13** Micromagnetics predictions for the various regimes of the magnetization dynamics of a spin-valve subjected to spin-polarized current. The zone W marks the onset of a chaotic regime. Lee *et al.* (2004), copyright Nature Publishing Group.

of the magnetization of spin-valves, to the emergence of the precessional mode (Oezyilmaz *et al.* 2004, 2005; Piraux *et al.* 2007).

So-called “phase diagrams” can be worked out using the Landau–Lifshitz equation with spin-transfer torque under the assumption of uniform magnetization. The SPINTEC group, among others, has examined the various regimes of magnetization excitation beyond the single domain (Fig. 3.13). Such varied regimes are amply confirmed by experimental data (e.g. Chen *et al.* 2006).

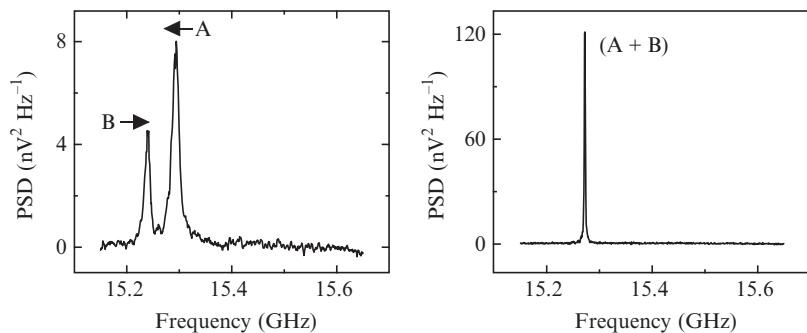
One conspicuous feature of the peaks in  $dV/dI$  was that they occur for one current sense only, as expected if a spin torque arises for electrons polarized in the fixed layer and entering into the other one. Tsoi *et al.* (2004) showed that if a spin-valve is made perfectly symmetrical, then both senses of current give rise to peaks in  $dV/dI$  curves, but one layer only gets excited.

While it was reasonable to assign these peaks in the  $dV/dI$  curves to a steady precession of the magnetization, it was necessary to verify it experimentally. In a pioneering study, Tsoi *et al.* (2000) used the joint effect of the injection of a spin-polarized current together with microwave radiation in order to identify the  $dV/dI$  peaks with long-wavelength magnon excitations. This assignment was ultimately confirmed by the work of the Cornell group, who achieved the first detection of the microwave signal produced by the oscillation of the resistance due to the precession of the magnetization at large angles (Kiselev *et al.* 2003, 2004). Since then, many groups have reported the electrical detection of the magnetization precession sustained by dc currents (e.g. Cros *et al.* 2005; Mancoff *et al.* 2006). SPINTEC, in collaboration with Headway, examined spin-transfer torque effects in spin-valves designed as giant magnetoresistive heads with current perpendicular to the interfaces (Deac *et al.* 2005). They found clear current-hysteretic behavior at low field and they could excite a precession at high fields. Likewise, Seagate detected microwave oscillations in tunnel junctions of MgO (Nazarov 2006). A group at the Mayo Clinic studied in extended datasets how the microwave oscillation frequency depends, in particular, on the applied field orientation (Cameron 2007).

It is rare to find a situation where a dc current excites a signal at GHz frequencies. It was observed by current injection in BiSb alloys (Nanney and George 1969). In this case, the exact mechanism remains uncertain. The authors suspected that the spin of the electrons had an effective temperature different from the lattice temperature, a potential mechanism for dynamic nuclear polarization (Feher 1958; Clark and Feher 1963).

This excitation of the precession of magnetization by a dc current is of great interest in our understanding spin-torque effects and magnetization dynamics in nanostructures. It may find application in writing magnetoresistive bits in MRAMs. The proposition has been made also to use this phenomenon as a means of generating microwave signals, possibly in a very compact and efficient manner. However, the power of the microwave signal out of a single nanopillar is very weak. Thus, the issue arises of putting several such nano-oscillators together and locking them together in phase. Phase locking has been demonstrated by Freescale (Mancoff *et al.* 2005), for oscillations at point contacts set 200 nm apart, in the frequency range of 10 to 24 GHz. In the same issue of Nature, the NIST group together with Hitachi (Kaka *et al.* 2005) published very sharp results (Fig. 14). Both used point contacts to CoFe/Cu/NiFe layers. The latter group found that locking of the individual oscillators occurred (Rippard *et al.* 2005) at frequencies several hundreds of MHz away from the oscillation frequency they had when they are isolated! They demonstrated that the phase locking arises by coupling through spin-waves in the base layer (Pufall *et al.* 2006, 2007).

Sankey *et al.* (2005) studied and discussed the issue of the coherence time and the determining mechanisms that drive it. This coherence time is directly related to the width of the spectrum of the microwave oscillations. The results point to thermal effects. At high temperature, the thermal fluctuations are sufficient to trigger switching between various non-uniform modes, giving rise to a sort of “motional narrowing”, an effect that is well known in magnetic resonance spectroscopy. The record result claimed in this publication is a line width  $\Delta f$  of about 5 MHz, corresponding to a coherence time of 190 ns. This result is achieved by using an antiferromagnetic layer to pin the magnetization near 45 degrees from the easy axis of the free layer, and applying the field along the exchange bias field. The line is then nearly homogeneously broadened,



**Fig. 3.14** Power spectra collected at point contacts A and B to a bilayer at a dc current just below (left) and in the range (right) of values when phase locking occurs. The arrows (left) indicate the frequency shift with increasing current. Kaka *et al.* (2005), copyright Nature Publishing Group.

having a width approaching the estimate:

$$\Delta f = \frac{4\pi\gamma\alpha n^2}{M_s V D^2} kT, \quad (3.4)$$

where  $\alpha$  is the Gilbert damping coefficient,  $\gamma$  the gyromagnetic ratio,  $M_s$  the saturation magnetization,  $n = 1$  or  $2$  for first and second harmonics,  $kT$  the thermal energy,  $V$  the volume of the free layer,  $D$  the length of the precession trajectory on the unit sphere. The estimate is based on a fluctuating field with a Gaussian distribution.

### 3.5.2 Zero-field oscillations

At one point, the spintronics community was set for the following simplifying picture: at low field, a spin-polarized current can switch the magnetization of a nanostructure that produces it, at high field, this current induces excitations of the magnetization. In this framework, it was a challenge to try and obtain oscillation at zero field. Two routes were found. Barnas *et al.* (2005, 2006) showed that asymmetric spin-valves, such as Py-Cu-Co, have a so-called “wavy torque” that allows oscillation induced by a dc current. The other approach is simply to add a third layer that provides a local field on the spin-valve.

Fert and coworkers used the theoretical ideas developed with J. Barnaas and observed precession at zero field (Bouille *et al.* 2007). Using two magnetic layers with different spin-diffusion lengths, they found that the angular dependence of the spin-transfer torque has the “wavy” form that insures instability under dc current (Barnas *et al.* 2005, 2006). Dieny and coworkers obtained oscillation at zero field simply by adding to a spin-valve a layer that produces the necessary magnetostatic coupling (Houssameddine *et al.* 2007). Katine, in collaboration with Chappert’s group used a spin-valve structure, making sure that the layer pinned with an antiferromagnet has its magnetization perpendicular to the easy axis of the free layer, both magnetizations being inplane (Devolder *et al.* 2007). This group suggests that the field-like term in eqn (3.1) is responsible for triggering the instability leading to the precession of the magnetization maintained by the current.

### 3.5.3 Vortices

The Cornell group (Pribiag *et al.* 2007) reported current-driven oscillations of vortices with a linewidth of as little as 300 kHz at 1.1 GHz in spin-valves composed of a 5-nm thick layer, a relatively thick Cu spacer (40 nm) and a thick permalloy base allowing for the formation of a vortex. Micromagnetics simulations showed that the resonance corresponds to a large amplitude precession of the core of the vortex at the interface to the Cu, that is, on the side of the incoming spin current.

In view of the existing literature on current-driven oscillation, quite extraordinary results have been obtained by the NIST group (Pufall 2007): oscillations occur under dc currents at zero or low applied field, their frequency

is very low, below 1 GHz, their bandwidth relatively narrow, the oscillation is detected as usual electrically, and here the signal reaches 1 mV! The authors suggest that the inhomogeneous Oersted field present in their configuration can nucleate a vortex state and the spin-transfer torque then in turn excites it.

### 3.5.4 Magnetic inhomogeneities

It has been argued that the onset of magnetic oscillation driven by dc current is made possibly by slight inhomogeneities, presumably caused by the Oersted field (Miltat 2003). A nanotechnology feat was achieved by the Cornell group, injecting the current in the free layer at a hole. Their micromagnetics simulation confirm the idea that magnetization switching by nucleation starts at this point (Ozatay *et al.* 2006). The final state turns out to be dependent on the current. Oscillations were observed in the form of peaks in  $dV/dI$  measurements. The same group showed that non-uniform magnetization states showed up in regular nanopillars at high currents (Kiselev *et al.* 2005).

With the present ongoing development of X-ray imaging, a direct insight into the inhomogeneous state of the magnetization in a nanoscopic layer is becoming possible. Acreman *et al.* (2006) paved the way for a completely new look at the details of magnetization switching dynamics under intense spin-polarized currents, with a resolution of about 30 nm!

## 3.6 Resonant-current excitation

### 3.6.1 Spin-valves and tunnel junctions

There has been remarkably little investigation of the ability of spin torque to excite a resonance. Looking at eqn (3.1), it appears that an alternating current may excite the self-resonance of a ferromagnet, just like an alternating microwave field is normally used for that. Since the resonance tilts the magnetization out of equilibrium, we may obtain an oscillating resistance because of the magnetoresistance of the material. Thus, not only is the volume where the *excitation* takes place reduced to the actual volume of the nanowire or nanopillar itself, but the electrical *detection* also is confined to the sample itself.

There are several instances of electrically detected magnetic resonance in the literature, outside of magnetic systems. It is based on the spin dependence of the rate of electron–hole recombination (Haberkorn and Dietz 1980; Wohlgenannt *et al.* 2001, 2002). One conspicuous feature of electrically detected magnetic resonance is, indeed, its great sensitivity (Rong *et al.* 1991, 1992; Maier *et al.* 1996; Stahis 1996; Eickelkamp *et al.* 1998; Wimbauer *et al.* 1998; Sato *et al.* 2001).

Juretschke (1960) predicted that at resonance, a dc voltage would develop at the edge of a ferromagnetic metallic strip, cross-ways with respect to the applied static field and the equilibrium magnetization. This galvanomagnetic effect (Jan 1957) arises because Ohm's law needs to be modified for ferromagnetic metals, to include anisotropic magnetoresistance and the extraordinary

Hall effect. This non-linearity leads to products of oscillating terms  $\mathbf{m}$  and  $\mathbf{j}$ , respectively the current and magnetization responses to the microwave irradiation. Thus, in a thin ferromagnet strip with a magnetoresistance  $\Delta\rho$ , a dc electric field arises of the form:

$$\mathbf{e}_0 = -(\Delta\rho/M^2) (\langle \mathbf{j} \times \mathbf{m} \rangle \times \mathbf{M}_0 + \langle \mathbf{j} \cdot \mathbf{m} \rangle \mathbf{M}_0).$$

The brackets here mean an average over time and over the sample cross-section normal to the equilibrium magnetization  $\mathbf{M}_0$ . These considerations have been revisited recently, in a paper that suggests thinking in terms of “microwave photoconductivity” (Gui *et al.* 2005). Juretschke and coworkers verified this prediction, using nickel (Egan *et al.* 1963) and permalloy (Moller and Juretschke 1970). Voltaic detection has been investigated in nanostructures also (Park *et al.* 2002) and in CoNb thin films (Oh *et al.* 2005). These are experiments where the *excitation* is a field, typically generated in a microwave cavity (Goennenwein *et al.* 2007), but it can also be the field produced in a microscopic current loop (Grollier *et al.* 2006; Bizières and Fremon 2007). Regarding resistive measurements, Gui *et al.* (2007) raise the question whether a bolometric effect is at play: the resonance occurs with dissipation and the temperature rise may cause a resistance change.

While electrical *detection* is of interest for those involved in using magnetic resonance, the key issue in the framework of this chapter is whether the spin torque can be used to *excite* a resonance, as opposed to using a magnetic field. The Cornell group (Sankey *et al.* 2006) concluded from their measurements that indeed, the spin-transfer torque could excite the ferromagnetic resonance of the free layer of a spin-valve. The sensitivity of electrical detection of resonance was once again confirmed. The size of the free layer in typical spin-valves for a spin-transfer torque experiment is of the order of  $10^6$  atoms, whereas a typical ESR experiment at the X band, using a cavity, requires about  $10^{12}$  spins or more.

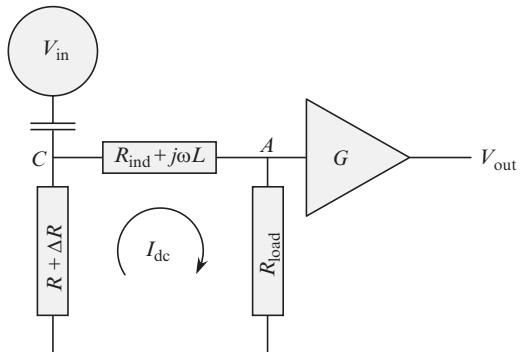
Evidence for resonant spin torque was also given in the case of tunnel junctions (Tulapurkar *et al.* 2005). The detection scheme can be thought of as follows, in a first approximation (Fig. 3.15). Under the effect of the current, we expect a change of the resistance  $\Delta R$  of the form  $\Delta R = \alpha_1 I_{\text{rf}} e^{i\phi}$ . The supplemental material of Tulapurkar *et al.* (2005), for example, gives the details of the calculation. Then, the current through the sample contains two contributions:

$$I_C = \frac{V_{\text{in}}}{R + \alpha_1 I_{\text{rf}}} \approx \frac{V_{\text{in}}}{R} \left( 1 - \frac{\alpha_1 I_{\text{rf}}}{R} \right).$$

To first order, we take  $I_{\text{rf}}$  to be  $\frac{V_{\text{in}}}{R}$ , neglecting dephasing between current and magnetic response, for the sake of this cursive presentation of the principle. We have then:

$$I_C \approx \frac{V_{\text{in}}}{R} \left( 1 - \alpha_1 \frac{V_{\text{in}}}{R^2} \right).$$

Clearly there is a quadratic contribution that has a non-vanishing dc contribution. This means a dc current  $I_{\text{dc}}$  runs into the loop of impedances at the amplifier input (Fig. 3.15). The results of this detection method must be analyzed in detail in order to establish the origin of the excitation:



**Fig. 3.15** Equivalent circuit for the dc detection of a resonance.

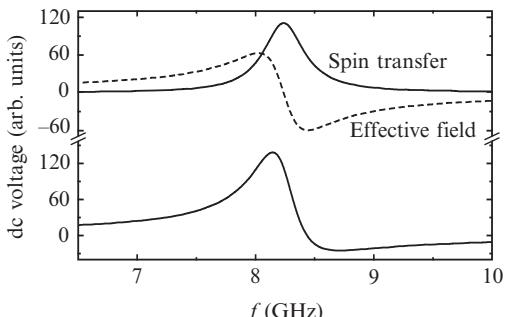
- How much of the resonance excitation is due to the spin torque?
- How big is the contribution of the effective exchange field term?
- How big is the contribution of the field induced by the current?

In the Yuasa paper (Tulapurkar *et al.* 2005), the authors point out that the spin-transfer term of eqn (3.1) (Section 3) gives rise to a Lorentzian line, whereas an effective field gives a dispersion-like resonance and the combined effect (Fig. 3.16) has the distortion that they have observed.

### 3.6.2 Resonant excitation of spin-waves

A natural extension to the above is the excitation of spin-waves by currents. First was the work of Rezende *et al.* (2000). A microwave current was injected at a point contact into an antiferromagnetically coupled Fe/Cr/Fe trilayers. The presence of spin-waves was detected by Brillouin scattering.

The group at Kaiserslautern (Demidov *et al.* 2004) managed the “tour de force” of detecting by Brillouin scattering off spin-waves excited by a microwave current driven through a pseudospin-valve nanopillar composed of a NiFe top layer, a Cu spacer and a film of CoFe. The laser spot on the film was focused down to a size of about 500 nm. A surprising result that they received was the presence of distinct peaks in the scattering intensity, indicative of discrete modes. Those are thought to be the quantized modes (Demokritov *et al.* 2001) confined to the pillar that radiate further into the film. These modes



**Fig. 3.16** Expected contributions to the resonance excitation of the spin torque and the exchange field. Tulapurkar *et al.* (2005), copyright Nature Publishing Group.

are selected out of the many possible because they presumably have the best coupling to the microwave current.

### 3.6.3 Resonant excitation of domain walls

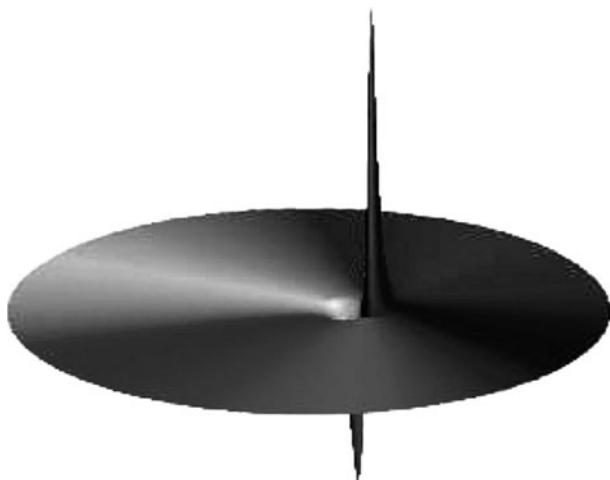
If we consider eqn (3.2), we can well imagine that the current term can be made resonant with the eigenmode of the magnetization that comprises a pinned wall. A Japanese group demonstrated the concept (Saitoh *et al.* 2004). They formed a trap for a wall, the pinning potential was such that the frequency of a small oscillation of the wall in its trap was in the tens of MHz range. At this frequency, there are no experimental complications that one has to address necessarily at microwave frequencies.

### 3.6.4 Resonant excitation of vortices

A ferromagnetic disc may present a vortex state. For example, Yamada *et al.* (2007) determined the resonance of a vortex to be in the hundreds of MHz range, and then simulated the effect of a resonant spin torque on the evolution of the vortex, finding that a reversal could occur (see Fig. 3.17). That such a reversal had occurred was then experimentally verified using a magnetic force microscope in order to monitor the orientation of the vortex of permalloy discs, 50 to 500 nm in diameter.

### 3.6.5 Spin pumping

A spin-polarized charge current can excite the magnetization through the effect of spin-transfer torque. The converse effect is referred to as spin pumping. The magnetization of a ferromagnetic layer is excited at resonance. This produces a spin current, as has been identified theoretically (Brataas *et al.* 2002; Tserkovnyak *et al.* 2002). The basic principles are not easily grasped, as they involve a non-local effect on the spin dynamics. However, the end results can



**Fig. 3.17** Simulation of a vortex about to switch. Yamada *et al.* (2007), copyright Nature Publishing Group.

be expressed in compact form (Tserkovnyak *et al.* 2005). The key ingredient is the spin-mixing conductance  $g_{\uparrow\downarrow}$  (see Section 2) at the interface between a ferromagnetic layer and a metal layer in metallic contact with it. Hence, as for the spin torque, the spin transport at the interface is treated quantum mechanically, while the transport in the layers is carried out semi-classically under the assumption of a diffusive regime. When the ferromagnet is thicker than the spin coherence length  $\frac{\pi}{|\mathbf{k}_{F\uparrow} - \mathbf{k}_{F\downarrow}|}$ , the spin current  $\mathbf{I}_s^{\text{pump}}$  emitted into the normal metal is:

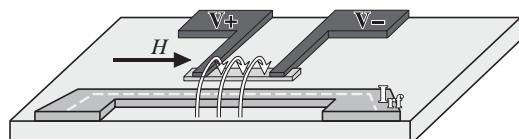
$$\mathbf{I}_s^{\text{pump}} = \frac{\hbar}{4\pi} g_{\uparrow\downarrow} \mathbf{m} \wedge \frac{d\mathbf{m}}{dt}.$$

Notice that this spin current does not involve charge current. It is an important prediction, as it opens the possibility of injecting spins (strictly, without charges) using the precession of the magnetization of an adjacent layer! This may solve some issues that arise when spin injection is carried out with electronic transport. In particular, spin injection does not work when injecting into semiconductors from a metallic ferromagnet because of the conductivity mismatch of the materials, which results in a screening of the spin polarization ahead of the semiconductor (Schmidt *et al.* 2000; Fert and Jaffr  s 2001).

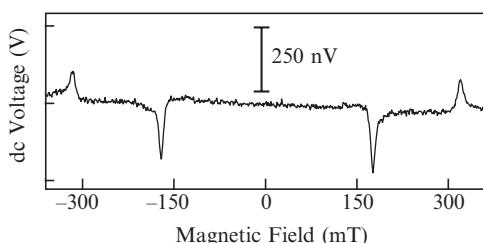
One consequence of spin pumping is the enhancement of the Gilbert damping. If the adjacent metal presents an efficient spin relaxation to the pumped spins, this metal acts as a sink, and hence increases the damping of the resonance in the ferromagnetic layer. This has been observed experimentally. (Urban *et al.* 2001; Mizukami *et al.* 2002; Lenz *et al.* 2004; Gerrits *et al.* 2006).

Tserkovnyak *et al.* (2005), in their review article, point out that spin pumping can lead to a form of dynamical exchange among metallic nanomagnets (Heinrich *et al.* 2003). Likewise, the simulations of Kim and Chappert (2005) of large-angle precession show that spin pumping can affect the effective switching field. If instead, the material is unable to relax these incoming spins, a spin accumulation arises. This may occur in a multilayer structure with a non-magnetic metal spacer between two ferromagnets.

The electrical detection of spin pumping has been demonstrated (Costache *et al.* 2006). A permalloy strip (Fig. 3.18) is exposed to a static field and a



**Fig. 3.18** top) Layout, right) dc voltage  $\Delta V = V_+ - V_-$  generated by a Pt/Py/Al device in response to the rf magnetic field generated by  $I_{rf}$  plotted as a function of the static magnetic field  $H$ . In order to remove some baseline distortion, the data were plotted as the difference in the voltage obtained at two frequencies 14.5 and 19.5 GHz. Thus, peaks and dips correspond to resonances at these frequencies. Costache *et al.* (2006) copyright American Physical Society.



microwave field. The strip is contacted by a Pt pad on one end, and to a Al pad on its other end. The authors exclude a rectifying effect, which would result from the compounded effect of the spurious rf current induced in the permalloy strip and the modulation of the resistance due to AMR. To do so, they changed the coupling of the detection arm so as to decrease drastically the induced current and found, nonetheless, about the same dc voltage signal.

### 3.7 Conclusion

The discovery of giant magnetoresistance (GMR) sparked many developments in materials physics. It revived the interest in tunnel junctions and brought them to outstanding performance, thanks to the introduction of MgO. It incited some to revisit the manganates, with extensive research worldwide (Coey *et al.* 1999). It sparked extended investigations of the physics of half-metallic ferromagnets (Dowben and Jenkins 2005). Spectacular magnetoconductance can be obtained (Viret *et al.* 1997), though more material science is needed to bring these performances to room temperature. As reviewed here, GMR led its aficionados to the discovery of a novel means of acting on the magnetization of a ferromagnet. A current of high enough density can cause switching, induce precessional motion, push walls or excite resonance.

Where else is spintronics going to take us? The giant Hall effect is one of the novel episodes of the hunt for remarkable responses of a nanostructured systems to magnetic fields. Systems based on super-paramagnetic particles find here a new life (Denardin *et al.* 2003) and new effects may occur in ferromagnetic semiconductors (Tang *et al.* 2003).

The spin Hall effect may some day become the method of generating spin accumulation, free of charge accumulation, by sorting spins out by spin-orbit scattering on impurities or by an intrinsic property of materials with complex electronic structure. It was first predicted by Dyakonov and Perel (1971). The theoretical underpinning was examined further (Froehlich *et al.* 2001; Meier 2003; Engel 2005). The group of Awschalom has detected, using Kerr microscopy, the polarization of electron spins due to current flowing down a strip of GaAs (Sih *et al.* 2005a). This group admits that their observation probably corresponds to a spin-orbit effect at impurities (Sih *et al.* 2005b). In this case, the spin Hall effect is referred to as an extrinsic effect. Others claim that they have managed to detect the novel, intrinsic, form of the spin Hall effect (Wunderlich *et al.* 2005). This intrinsic effect is a pure quantum-mechanical phenomenon (Sinova *et al.* 2004). A strong effective field is necessary, such that the spins precess more than once during each scattering event (Kato *et al.* 2004). A 2D hole gas is expected to give rise to this intrinsic effect (Inoue *et al.* 2005).

New routes for spintronics may arise from using organic materials for spintronics (Sanvito 2007). Carbon nanotubes have the fascinating property of being able to carry spins over very long distances (Cottet *et al.* 2006). Organic films have been known for their potential in making great tunnel barriers. Alq<sub>3</sub> is the newest case in view (Awschalom and Flatté 2007).

There is a major effort worldwide aimed at developing magnetic semiconductors. Ga(Mn)As has by now well-established physical properties, but its

Curie temperature is too low. The spin-torque effect has been observed with walls (Yamaguchi *et al.* 2005) and in spin-valves (Elsen *et al.* 2006), with far smaller current densities than in 3d metals, as expected from the smaller saturation magnetization. Doped ZnO, and similar materials, may come into use in spintronics some day, but they remain controversial at the moment. One issue with semiconductors is the so-called conductivity mismatch. It is not feasible to inject spins from a metal to a semiconductor; the spin polarization dies out, in other words, relaxes, before the current reaches the semiconductor (Schmidt *et al.* 2000; Fert and Jaffr  s 2001).

The industry is gearing up to make magnetoresistive memories. Spintronics enters in several aspects of design. First, the bit itself is a magnetoresistive device, typically a tunnel junction. Second, the issue of writing a bit without supplying too much current and without cross-talk makes the idea quite appealing of using the spin-transfer torque effect as a means of flipping the magnetic bit. There is an abundance of review articles on spintronics that give further details on this.

Among the wildest dream that one can think of in spintronics, there is the SWASER, a sort of laser of magnons. This is the contention of L. Berger that a stimulated emission of magnons could be produced by spin injection (Berger 2002). There are indeed analogies between spin-valves and semiconductor junctions, each having two charge carriers, threshold conditions on the chemical potential differences, and a damping that is a function of the number of excitations. It appears that a ferromagnet may not be necessary. Watts and van Wees (2006) suggest that by combining an external spin-injection source with microwave field spin pumping in a paramagnetic medium, the system can be driven from one that absorbs microwave energy to one that emits microwave energy. Hence, when the medium is placed within a resonant circuit, these authors expect microwave amplification by stimulated emission of radiation (maser) driven by spin injection. Alternatively, Nogaret *et al.* (2007) suggest a radiating cascade of spin oscillators, which could be implemented with 2-dimensional electron systems. Either method would allow spintronics to open the way to microwave sources in the range of up to hundreds of gigahertz, tuneable by an electrostatic potential.

## References

- Abid, M., Abid, J.-P., Jannin, S., Serrano-Guisan, S., Palaci, I., Ansermet, J.-Ph. *J. Phys. Condens. Matter* **18**, 6085 (2006a).
- Abid, M., Abid, J.-P., Ansermet, J.-Ph. *J. Electrochem. Soc.* **153**(8), D138 (2006b).
- Acremann, Y., Strachan, J.P., Chembrolu, V., Andrews, S.D., Tyliszczak, T., Katine, J.A., Carey, M.J., Clemens, B.M., Siegmann, H.C., St  hr, *J. Phys. Rev. Lett.* **96**, 217202 (2006).
- Albert, F.J., Emley, N.C., Myers, E.B., Ralph, D.C., Buhrman, R.A. *Phys. Rev. Lett.* **89**, 226802 (2002).
- Allwood, D.A., Xiong Gang, Cooke, M.D., Faulkner, C.C., Atkinson, D., Vernier, N., Cowburn, R.P. *Science* **296**, 2003 (2002).
- AlMawlawi, D., Coombs, N., Moskovitz, M. *J. Appl. Phys.* **70**, 4421 (1991).

- Ansermet, J.-Ph. *IEEE Trans. Magn.* **40**(2), 358 (2004),  
 Ansermet, J.-Ph. *IEEE Trans. Magn.* **44**, 329 (2008).
- Appell, D. *Nature* **419**, 553–555 (2002).
- Asik, T., Ball, M.A., Slichter, C.P. *Phys. Rev. Lett.* **16**, 740 (1966).
- Atkinson, D., Allwood, D.A., Xiong, G., Cooke, M.D., Faulkner, C.C., Cowburn, R.P. *Nature Mater.* **2**, 85–87 (2003).
- Awschalom, D.D., Flatté, M.E. *Nature Phys.* **3**, 153 (2007).
- Baibich, M.N., Broto, J.M., Fert A., Nguyen Van Dau, F., Petroff, F., Etienne P., Creuzet, G., Friederich, A., Chazelas, J. *Phys. Rev. Lett.* **61**, 2472 (1988)
- Ball, M.A., Asik, J.R., Slichter, C.P. *Phys. Rev.* **181**, 662 (1969).
- Barnas, J., Fert, A., Gmitra, M., Weymann, I., Dugaev, V.K. *Phys. Rev. B* **72**, 024426 (2005).
- Barnas, J., Fert, A., Gmitra, M., Weymann, I., Dugaev, V.K. *Mater. Sci. Eng. B* **126**, 271 (2006).
- Barnes S.E. *Adv. Phys.* **30**, 801–938 (1981).
- Bass, J., W. Pratt, Jr. *P. J. Phys.: Condens. Matter* **19**, 183201 (2007).
- Bazaliy, Y.B., Jones, B.A., Zhang Shou-Cheng, *Phys. Rev. B* **57**(6), R3213 (1998).
- Berger, L. *J. Appl. Phys.* **49**, 2156 (1978).
- Berger, L. *J. Appl. Phys.* **55**, 1954 (1984).
- Berger, L. *Phys. Rev. B* **33**, 1572 (1986).
- Berger, L. *Phys. Rev. B* **54**, 9353 (1996).
- Berger, L. *J. Appl. Phys.* **89**, 5521 (2001).
- Berger, L. *J. Appl. Phys.* **91**, 6795 (2002).
- Berkowitz, A.E., Mitchell, J.R., Carey, M.J., Young, A.P., Zhang, S., Spada, F.E., Parker, F.T., Hjutten, A., Thomas, G. *Phys. Rev. Lett.* **68**(25), 3745 (1992).
- Binusch, G., Grünberg, P., Saurenbach, F., Zinn W. *Phys. Rev. B* **39**, 4828–4830 (1989).
- Bizières, N., Fremon, C. *Appl. Phys. Lett.* **92**, 092503 2008.
- Blondel, A., Meier, J.P., Doudin, B., Ansermet, J.-Ph. *Appl. Phys. Lett.* **65**, 3019 (1994)
- Boulle, O., Cros, V., Grollier, J., Pereira, L.G., Deranlot, C., Petroff, F., Faini, G., Barnbas, J., Fert, A. *Nature Phys.* **482**, (2007).
- Brataas, A., Bauer, F.E.W., Kelly, P.J. *Phys. Rep.* **427**, 157 (2006),
- Brataas, A., Tserkovnyak, Y., Bauer, G.E.W., Halperin, B.I. *Phys. Rev. B* **66**, 060404 (2002).
- Brataas, A., Nazarov, Y.V., Bauer, G.E.W. *Phys. Rev. Lett.* **84**, 2481 (2000).
- Bruno, P. *Phys. Rev. Lett.* **83**(12), 2425 (1999).
- Butler, W.H., Zhang, X.-G., Schulthess, T.C., Maclarens, J.M. *Phys. Rev. B* **63**, 054416 (2001).
- Cabrera, G.G., Falicov, L.M. *Phys. Status Solidi. B* **61**, 539 (1974).
- Cacho, C., Lassailly, Y., Drouhin, H.-J., Lampel, G., Peretti, J. *Phys. Rev. Lett.* **88**, 066601 (2002).
- Cameron, G.P., Schraufnagel, D.J., Fjerstad, W.H., Sokolov, V., Daniel, E.S., Gilbert, B.K. *IEEE Trans. Magn.* **43**(5), 1934 (2007).
- Carlier, D., Ansermet, J.-Ph. *J. Electrochem. Soc.* **153**(8), D128 (2006).

- Chen, T.Y., Ji, Y., Chien, C.L., Stiles, M.D. *Phys. Rev. Lett.* **93**, 026601 (2004).
- Chen, W., Rooks, M.J., Ruiz, N., Sun, J.Z., Kent, A.D. *Phys. Rev. B* **74**, 144408 (2006).
- Chiba, D., Yamanouchi, M., Matsukura, F., Ohno, H. *Science* **301**, 943–945 (2003).
- Clark, W.G., Feher, G. *Phys. Rev. Lett.* **10**(4), 134 (1963).
- Coey, J.M.D., Berkowitz, A.E., Balcells, L., Putris, F.F., Barry, A. *Phys. Rev. Lett.* **80**, 3815–3818 (1998).
- Coey, J.M.D., Viret, M., von Molnar, S. *Adv. Phys.* **48**, 167–293 (1999).
- Cooper, R.L., Uehling, E.A. *Phys. Rev.* **164**, 662 (1967).
- Costache, M.V., Sladkov, M., Watts, S.M., van der Wal, C.H., van Wees, B.J. *Phys. Rev. Lett.* **216603** (2006).
- Cottet, A., Kontos, T., Sahoo, S., Man, H.T., Choi, M.-S., Belzig, W., Bruder, C., Morpurgo, A.F., Schönenberger, C. *Semicond. Sci. Technol.* **21**, S78–S95 (2006).
- Covington, M., AlHajDarwish, M., Ding, Y., Gokemeijer, N.J., Seigler, M.A. *Phys. Rev. B* **69**, 184406 (2004).
- Cros, V., Boulle, O., Grollier, J., Hamzi, A., Muñoz, M., Pereira L.G., Petroff, F. *Comp. Rend. Phys.* **6**, 956 (2005).
- Danneau, R., Warin, P., Attané J.P., Petej, I., Beigné, C., Fermon, C., Klein, O., Marty, A., Ott, F., Samson, Y., Viret, M. *Phys. Rev. Lett.* **88**(15), 157201 (2002).
- Dauguet, P., Gandit, P., Chaussy, J. *J. Appl. Phys.* **79**, 5823 (1996).
- Deac, A., Lee, K.J., Liu, Y., Redon, O., Li, M., Wang, P., Nozières J.-P., Dieny, B. *J. Magn. Magn. Mater.* **290–291**, 42 (2005).
- Demidov, V.E., Demokritov, S.O., Hillebrands, B., Laufenberg, M., Freitas, P.P. *Appl. Phys. Lett.* **85**, 2866 (2004).
- Demokritov, S.O., Hillebrands, B., Slavin, A.N. *Phys. Rep.* **348**, 441 (2001).
- Denardin, J.C., Knobel, M., Zhang, X.X., Pakhomov, A.B. *J. Magn. Magn. Mater.* **262**, 15 (2003).
- Dennis, C.L., Siristhatikul, C., Ensell, G.J., Gregg, J.F., Thompson S.M. *J. Phys. D* **36**, 81 (2003).
- Devolder, T., Meftah, A., Ito, K., Katine, J.A., Crozat, P., Chappert, C. *J. Appl. Phys.* **101**, 063916 (2007).
- Diao, Z., Pakala, M., Panchula, A., Ding, Y., Apalkov, D., Wang, L.-C., Chen, E., Huai, Y. *J. Appl. Phys.* **99**, 08G510 (2006).
- Doudin, B., Blondel, A., Ansermet, J.-Ph. *J. Appl. Phys.* **79**, 6090 (1996).
- Doudin, B., Redmond, G., Gilbert, S.E., Ansermet, J.-Ph. *Phys. Rev. Lett.* **79**, 933 (1997).
- Dowben, P.A., Skomski, R. *J. Appl. Phys.* **95**, 7453 (2004).
- Dowben, P.A., Jenkins, S.J. in *Frontiers in Magnetic Materials*, (ed.) A.V. Narlikar (Springer, 2005).
- Dyakonov, M.I., Perel, V.I. *ThETF Pis. Red.* **13**, 657 (1971); *JETP Lett.* **13**, 467 (1971).
- Dyson, F.J. *Phys. Rev.* **98**(2), 349 (1955).
- Ebels, U., Radulescu, A., Henry, Y., Piraux, L., Ounadjela, K. *Phys. Rev. Lett.* **84**(5), 983 (2000).
- Edmonds, R.N., Harrison, M.R., Edwards, P.P. *An. Rep. Prog. Chem.* **82**, Chapter 9, Conduction Electron Spin Resonance in Metallic Systems (1985).

- Edwards, D.M., Mathon, J. *J. Phys.: Condens. Matter* **19**, 165210 (2007).
- Egan, W.G., Juretschke, H.J. *J. Appl. Phys.* **34**(5), 1477 (1963).
- Eickelkamp, T., Roth, S., Mehring, M., *Molec. Phys.* **95**(5), 967–972 (1998).
- Elliot, R.J. *Phys. Rev.* **89**, 689 (1953).
- Elliott, R.J., Epshtain, E.M., Gulyaev, Y.V., Zilberman, P.E. *J. Magn. Magn. Mater.* **300**, 122–126 (2006).
- Emley, N.C., Krivorotov, I.N., Ozatay, O., Garcia, A.G.F., Sankey, J.C., Ralph, D.C., Buhrman, R.A. *Phys. Rev. Lett.* **96**, 247204 (2006).
- Engel, H.A., Halperin, B.I., Rashba, E.I. *Phys. Rev. Lett.* **95**, 166605 (2005).
- Elsen, M., Boulle, O., George, J.-M., Jaffrè, H., Mattana, R., Cros, V., Fert, A., Lemaitre, A., Giraud, R., Faini, G., *Phys. Rev. B* **73**, 035303 (2006).
- Fabian, A., Terrier, C., Serrano Guisan, S., Hoffer, X., Dubey, M., Gravier, L., Ansermet, J.-Ph. *Phys. Rev. Lett.* **91**, 257209 (2003).
- Fábián, A., Terrier, Ch., Serrano-Guisan, S., Guitienne, Ph., Gravier, L., Ansermet, J.-Ph. *J. Phys. Condens. Matter* **18**, 1569 (2006).
- Fabian, J., Das Sarma, S. *Phys. Rev. Lett.* **81**, 5624 (1998).
- Fabian, J., Das Sarma, S. *Phys. Rev. Lett.* **83**, 1211 (1999).
- Feher, G. *Phys. Rev. Lett.* **3**(3), 135 (1959).
- Fert, A. *J. Phys. C (Solid State Phys.)* **2**(2), 1784 (1969).
- Fert, A., Jaffrè, H. *Phys. Rev. B* **64**, 184420 (2001).
- Fert, A., Campbell, I.A. *Phys. Rev. Lett.* **21**(16), 1190 (1968).
- Filipe, A., Drouhin, H.-J., Lampel, G., Lassailly, Y., Nagle, J., Peretti, J., Safarov, V.I., Schuhl, A. *Phys. Rev. Lett.* **80**, 2425 (1998).
- Freitas, P.P., Berger, L. *J. Appl. Phys.* **57**, 1266 (1985).
- Fröhlich, J., Pedrini, B., Schweigert, C., Walcher, J. *J. Stat. Phys.* **103**, 527 (2001).
- Fuchs, G.D., Katine, J.A., Kiselev, S.I., Mauri, D., Wooley, K.S., Ralph, D.C., Buhrman, R.A. *Phys. Rev. Lett.* **96**, 186603 (2006).
- Garcia, N., Munoz, M., Zhao, Y.W. *Phys. Rev. Lett.* **82**(14), 2923 (1999).
- Garzon, S., Zutic, I., Webb, R.A. *Phys. Rev. Lett.* **94**, 176601 (2005).
- Gehring, G.A., Gregg, J.F., Thompson, S.M., Watson, M.L. *J. Magn. Magn. Mater.* **140–144**, 501–502 (1995).
- Gerrits, Th., van den Berg, H.A.M., Hohlfeld, J., Baer, L., Rasing, Th. *Nature* **418**, 509 (2002).
- Gerrits, Th., Schneider, M.L., Silva, T.J. *J. Appl. Phys.* **99**, 023901 (2006).
- Gijs, M.A.M., Giesbers, J.B., Lenczowski, S.K.J. *Phys. Rev. Lett.* **70**, 3343 (1993).
- Gijs, M.A.M., Johnson, M.T., Reinders, A., Huisman, P.E., van de Veerdonk, R.J.M., Lenczowski, S.K.J., van Gansewinkel, R.M.J. *Appl. Phys. Lett.* **66**, 1839 (1995).
- Gijs, M.A.M., Bauer, G.E.W. *Adv. Phys.* **46**, 285 (1997).
- Gmitra, M., Barnas, J. *Phys. Rev. Lett.* **96**, 207205 (2006).
- Goennenwein, S.T.B., Schink, S.W., Brandlmaier, A., Boger, A., Opel M., Gross, R., Keizer, R.S., Klapwijk, T.M., Gupta, A., Huebl, H., Bihler, C., Brandt, M.S. *Appl. Phys. Lett.* **90**, 162507 (2007).
- Gregg, J.F., Borges, R.P., Jouguelet, E., Dennis, C.L., Petej, I., Thompson, S.M., Ounadjela, K. *J. Magn. Magn. Mater.* **265**, 274–289 (2003).

- Gregg, J.F., Allen, W., Ounadjela, K., Viret, M., Hehn, M., Thompson S.M., Coey, J.M.D. *Phys. Rev. Lett.* **77**(8), 1580 (1996).
- Gregg, J.F., Allen, W., Thompson, S.M., Watson, M.L., Gehring, G.A. *J. Appl. Phys.* **79**(8), 5593 (1996).
- Grollier, J., Costache, M.V., van der Wal, C.H., van Wees, B.J. *J. Appl. Phys.* **100**, 024316 (2006).
- Gui, Y.S., Holland, S., Mecking, N., Hu, C.-M. *Phys. Rev. Lett.* **95**, 056807 (2005).
- Gui, Y.S., Mecking, N., Wirthmann, A., Bai, L.H., Hu, C.-M. *Appl. Phys. Lett.* **91**, 082503 (2007).
- Haberkorn, R., Dietz, W. *Solid State Commun.* **35**, 505–508 (1980).
- Hamrle, J., Kimura, T., Yang, T., Otani, Y. *Phys. Rev. B* **71**, 094434 (2005).
- Hao, M., Wang, J.-P. *Appl. Phys. Lett.* **84**, 3103 (2004).
- Hao, M., Wang, J.-P. *Appl. Phys. Lett.* **88**, 172506 (2006).
- Harrison, W.A. *Solid State Theory* (Dover, New York, 1979).
- Hasegawa, H. *Prog. Theor. Phys.* **21**(4), 483 (1959).
- He, J., Li, Z., Zhang, S. *Phys. Rev. B* **73**, 184408 (2006).
- Heinrich, B., Tserkovnyak, Y., Woltersdorf, G., Brataas, A., Urban, R., Bauer, G.E.W. *Phys. Rev. Lett.* **90**, 187601 (2003).
- Heitler, W., Teller, E. *Proc. Roy. Soc. (London)* **A155**, 637 (1936).
- Hernando, D., Huertas, Y., Rado, G.T., Suhl, H., Nazarov, V., Brataas, A., Bauer, G.E.W. *Phys. Rev. B* **62**, 5700 (2000).
- Herring, C., in *Magnetism*, Vol. **V**, (eds) G.T. Rado, H. Suhl (Academic Press, 1963).
- Hirst, L.L. *Phys. Rev.* **141**(2), 503 (1966).
- Hofer, W.A., Palotás, K., Rusponi, S., Cren, T., Brune, H. *Phys. Rev. Lett.* **100**, 026806 (2008).
- Holody, P., Steren, L.B., Morel, R., Fert, A., Loloei, R., Schroeder, P.A. *Phys. Rev. B* **50**, 12999 (1994).
- Houssameddine, D., Ebels, U., Deelaet, D., Rodmacq, B., Firastrau, I., Ponthenier, F., Brunet, M., Thirion, C., Michel, J.-P., Prejbeaunu-Buda, L., Cyrille, M.-C., Redon, O., Dieny, B. *Nature Mater.* **6**, 447 (2007).
- Huai, Y., Pakala, M., Diao, Z., Apalkov, D., Ding, Y., Panchula, A. *J. Magn. Magn. Mater.* **304**, 88 (2006).
- Huang, Y., Duan, X., Cui, Y., Lieber, C.M. *Nano. Lett.* **2**, 101–104 (2002).
- Inoue, J., Ohno, H. *Science* **309**, 2004 (2005).
- Jánossy, A., Monod, P. *Solid State Commun.* **18**(2), 203 (1976).
- Jedema, F.J., Heersche, H.B., Filip, A.T., Baselmans, J.J.A., van Wees, B.J. *Nature* **416**, 713 (2002b).
- Jedema, F.J., Costache, M.V., Heersche, H.B., Baselmans, J.J., van Wees, B.J. *Appl. Phys. Lett.* **81**, 5162 (2002a).
- Ji, Y., Hoffmann, A., Jiang, J.S., Pearson, J.E., Bader, S.D. *J. Phys. Appl. Phys. D* **40**, 1280–1284 (2007).
- Johnson, M., Byers, J. *Phys. Rev. B* **67**, 125112 (2003).
- Johnson, M., Silsbee, R.H. *Phys. Rev. Lett.* **55**(17), 1790 (1985).
- Johnson, M., Silsbee, R.H. *Phys. Rev B* **35**(10), 4959 (1987).

- Julliere, M. *Phys. Lett.* **54A**, 225 (1975)
- Juretschke, J.J. *J. Appl. Phys.* **31**(8), 1401 (1960).
- Kaka, S., Russek, S.E. *Appl. Phys. Lett.* **80**, 2958 (2002).
- Kaka, S., Pufall, M.R., Rippard, W.H., Silval, T.J., Russek, S.E., Katine, J.A. *Nature* **437**, 389 (2005).
- Katine, J.A., Albert, F.J., Buhrman, R.A., Myers, E.B., Ralph, D.C. *Phys. Rev. Lett.* **84**, 3149 (2000).
- Kato, Y.K., Myers, R.C., Gossard, A.C., Awschalom, D.D. *Science* **306**, 1910 (2004).
- Kelly, D., Wegrowe, J.-E., Truong, T.-K., Hoffer, X., Ansermet, J.-Ph. *Phys. Rev. B* **68**, 134425 (2003).
- Kim, J.-V., Chappert, C. *J. Magn. Magn. Mater.* **286**, 56 (2005).
- Kimura, T., Otani, Y., Hamrle, J. *Phys. Rev. Lett.* **96**, 037201 (2006).
- Kiselev, S.I., Sankey, J.C., Krivorotov, I.N., Emley, N.C., Garcia, A.G.F., Buhrman, R.A., Ralph, D.C. *Phys. Rev. B* **72**, 064430 (2005).
- Kiselev, S.I., Sankey, J.C., Krivorotov, I.N., Emley, N.C., Rinkoski, M., Perez, C., Buhrman, R.A., Ralph, D.C. *Phys. Rev. Lett.* **93**(3), 036601 (2004).
- Kiselev, S.I., Sankey, J.C., Krivorotov, I.N., Emley, N.C., Schoelkopf, R.J., Buhrman, R.A., Ralph, D.C. *Nature* **425**, 380 (2003).
- Kittel, C., Mitchell, A.H. *Phys. Rev.* **101**, 1611 (1956).
- Klaeui, M., Vaz, C.A.F., Bland, J.A.C., Heyderman, L.J., Nolting, F., Pavlovska, A., Bauer, E., Cherifi, S., Heun, S., Locatelli, A. *Appl. Phys. Lett.* **85**, 5637–5639 (2004).
- Klaeui, M., Vaz, C.A.F., Bland, J.A.C., Wernsdorfer, W., Faini, G., Cambril, E., Heyderman, L.J., Nolting, F., Ruediger, U. *Phys. Rev. Lett.* **94**, 106601 (2005a).
- Klaeui, M., Jubert, P.-O., Allenspach, R., Bischof, A., Bland, J.A.C., Faini, G., Ruediger, U., Vaz, C.A.F., Vila, L., Vouille, C. *Phys. Rev. Lett.* **95**, 026601 (2005b).
- Klein, L., Kats, Y., Marshall, A.F., Reiner, J.W., Geballe, T.H., Beasley M.R., Kapitulinik, A. *Phys. Rev. Lett.* **84**(26), 6090 (2000).
- Kovalev, A.A., Bauer, G.E.W., Brataas, A. *Phys. Rev. B* **75**, 014430 (2007).
- Krivorotov, I.N., Emley, N.C., Garcia, A.G.F., Sankey, J.C., Kiselev, S.I., Ralph, D.C., Buhrman, R.A. *Phys. Rev. Lett.* **93**, 166603 (2004).
- Krivorotov, I.N., Emley, N.C., Sankey, J.C., Kiselev, S.I., Ralph D.C., Buhrman, R.A. *Science* **307**, 228 (2005).
- Lee, K.-J., Deac, A., Redon, O., Nozières, J.-P., Dieny, B. *Nature Mater.* **3**, 877 (2004).
- Lenz, K., Tolinski, T., Lindner, J., Kosubek, E., Baberschke, K. *Phys. Rev. B* **69**, 144422 (2004).
- Levy, P., Zhang, S. *Phys. Rev. Lett.* **79**(25), 5110 (1997).
- Li, Z., Zhang, S. *Phys. Rev. B* **68**, 024404 (2003).
- Li, Z., Zhang, S. *Phys. Rev. Lett.* **92**(20), 207203 (2004).
- Liu, K., Nagodawithana, K., Searson, P.C., Chien, C.L. *Phys. Rev. B* **51**, 7381 (1995).
- Lou, X., Adelmann, C., Crooker, S.A., Garlid, E.S., Zhang, J., Reddy, K.S.M., Flexner, S.D., Palmstrøm, C.J., Crowell, P.A. *Nature Phys.* **3**, 197 (2007).
- Luetkens, H., Korecki, J., Morenzoni, E., Prokscha, T., Suter, A., Birke, M., Garifianov, N., Khasanov, R., Slezak, T., Litterst, F.J. *J. Magn. Magn. Mater.* **272–276**, 1128–1129 (2004).

- Luetkens, H., Korecki, J., Morenzoni, E., Prokscha, T., Birke, M., Glückler, H., Khasanov, R., Klauss, H.-H., Slezak, T., Suter, A., Forgan, E.M., Niedermayer, Ch., Litterst, F.J. *Phys. Rev. Lett.* **91**, 017204 (2003).
- Lyanda-Geller, Y., Aleiner, I.L., Golbart, P.M. *Phys. Rev. Lett.* **81**(15), 3215 (1998).
- Maekawa, S. (ed.), *Concepts in Spin Electronics* (Oxford University Press, Ser. *Semi-con. Sci. Technol.*, 2006)
- Maier, A., Gruipp, A., Mehring, M. *Solid State Commun.* **99**(9), 623–626 (1996).
- Man Leo, C.T., Apel, P., Cheung, T., Westerberg, L., K.N. Yu, Zet, C., Spohr, R. *Nucl. Instrum. Methods Phys. Res. B* **265**, 621–625 (2007).
- Manchon, A., Ryzhanoval, N., Strelkov, N., Vedyayev, A., Dieny, B. *J. Phys.: Condens. Matter* **19**, 165212 (2007).
- Mancoff, F.B., Rizzo, N.D., Engel, B.N., Tehrani, S. *Nature* **437**, 393 (2005).
- Mancoff, F.B., Rizzo, N.D., Engel, B.N., Tehrani, S. *Appl. Phys. Lett.* **88**, 112507 (2006).
- Martin, C.R. *Science* **296**, 2198 (2002).
- Mazin, I.I. *Phys. Rev. Lett.* **83**, 1427 (1999).
- Meier, F., Loss, D. *Phys. Rev. Lett.* **90**, 167204 (2003).
- Meservey, R., Tedrow, P.M. *Phys. Rep.* **238**, 173 (1994).
- Miltat, J. School of magnetism, Brasov, Rumania, Sept. 2003 (slides on the WEB) (2003).
- Miltat, J., Albuquerque, G., Thiaville, A. in *Spin Dynamics in Confined Magnetic Structures I*, (eds) B. Hillebrands, K. Ounadjela: Topics *Appl. Phys.* **83**, 1–34 (2002) (Springer-Verlag, Berlin, Heidelberg, 2002).
- Mitchell, A.H. *Phys. Rev.* **105**, 1439 (1957).
- Mizukami, S., Ando, Y., Miyazaki, T. *Phys. Rev. B* **66**, 104413 (2002).
- Miyazaki, T., Tezuka, N. *J. Magn. Magn. Mater.* **139**, L231–L234 (1995).
- Moller, W.M., Juretschke, H.J. *Phys. Rev. B* **2**(7), 2651 (1970).
- Monod, P., Schultz, S. *Phys. Rev.* **173**, 645 (1968).
- Monod, P., Schultz, S. *Phys. Rev.* **173**, 645 (1963).
- Monod, P., Schultz, S. *J. Phys.* **43**, 393 (1982).
- Monsma, D.J., Vlutters, R., Lodder, J.C. *Science* **281**, 5375 (1998).
- Moodera, J.S., Kinder, L.R. *J. Appl. Phys.* **79**, 4724 (1996).
- Moodera, J.S., Kinder, L.R., Wong, T.M., Meservey, R. *Phys. Rev. Lett.* **74**, 3273 (1995).
- Mooller, W.M., Juretschke, H.J. *Phys. Rev. B* **2**(7), 2651 (1970).
- Mott, N.F. *Adv. Phys.* **13**, 325–422 (1964).
- Mott, N.F. *Proc. Roy. Soc. (London) A* **153**, 699–717 (1935).
- Myers, E.B., Ralph, D.C., Katine, J.A., Louie, R.N., Buhrman, R.A. *Science* **285**, 867 (1999).
- Nanney, C.A., George, E.V. *Phys. Rev. Lett.* **22**(20), 1062 (1969).
- Nazarov, A.V., Olson, H.M., Cho, H., Nikolaev, K., Gao, Z., Stokes, S., Pant, B.B. *Appl. Phys. Lett.* **88** 162504 (2006).
- Nogaret, A., Lambert, N.J., Peeters, F.M. *Phys. Rev. B* **76**, 075312 (2007).

- Oezyilmaz, B., Kent, A.D., Sun, J.Z., Rooks, M.J., Koch, R.H. *Phys. Rev. Lett.* **93**(17), 176604 (2004).
- Oezyilmaz, B., Kent, A.D., Rooks, M.J., Sun, J.Z. *Phys. Rev. B* **71**, 140403(R) (2005).
- Oezyilmaz, B., Kent, A.D. *Appl. Phys. Lett.* **88**, 162506 (2006).
- Oh, D.K., Lee, C.E., Lee, J.-H., Rhie, K. *J. Magn. Magn. Mater.* **293**, (2005) 880 (2005)
- Ohgai, T., Gravier, L., Hoffer, X., Ansermet, J.-Ph. *J. Appl. Electrochem.* **35**, 479 (2005).
- Ohno, H., Shen, A., Matsukura, F., Oiwa, A., Endo, A., Katsumoto, S., Iye, Y. *Appl. Phys. Lett.* **69**, 363 (1996).
- Ono, T., Shinjo, T. *J. Phys. Soc. Jpn.* **64**, 363 (1995).
- Overhauser, A. *Phys. Rev.* **89**, 689 (1953).
- Ozatay, O., Emley, N.C., Braganca, P.M., Garcia, A.G.F., Fuchs, G.D., Krivorotov, I.N., Buhrman, R.A., Ralph, D.C. *Appl. Phys. Lett.* **88**, 202502 (2006).
- Park, E.R., Oh, D.K., Lee, C.E., Kim, S.H., Lee, S.R., Kim, Y.K. *J. Kor. Phys. Soc.* **41**, L1–L3 (2002).
- Piéchon, F., Thiaville, A. *Phys. Rev. B* **75**, 174414 (2007).
- Piraux, L., George, J.M., Despres, J.F., Leroy, C., Ferain, E., Legras R., Ounadjela, K., Fert, A. *Appl. Phys. Lett.* **65**, 2488 (1994).
- Piraux, L., Dubois, S., Fert, A., Beliard, L. *Eur. Phys. J. B* **4**, 413 (1998).
- Piraux, L., Dubois, S., Marchal, C., Beuken, J.M., Filipozzi, L., Despres, J.F., Ounadjela, K., Fert, A. *J. Magn. Magn. Mater.* **156**, 317–320 (1996).
- Piraux, L., Renard, K., Guillemet, R., Matefi-Tempfli, S., Matefi-Tempfli, M., Antohe, V.A., Fusil, S., Bouzehouane, K., Cros, V. *Nano Lett.* **7**(9), 2563 (2007).
- Polianski, M.L., Brouwer, P.W. *Phys. Rev. Lett.* **92**(2), 026602 (2004).
- Pratt, Jr., W.P., Lee, S.F., Slaughter, J.M., Loloei, R., Scroeder, P.A., Bass, J. *Phys. Rev. Lett.* **66**, 3060 (1991).
- Pribiag, V.S., Krivorotov, I.N., Fuchs, G.D., Braganca, P.M., Ozatay, O., Sankey, J.C., Ralph, D.C., Buhrman, R.A. *Nature Phys.* **3**, 498 (2007).
- Pufall, M.R., Rippard, W.H., Russek, S.E., Kaka, S., Katine, J.A. *Phys. Rev. Lett.* **97**, 087206 (2006).
- Pufall, M.R., Rippard, W.H., Schneider, M.L., Russek, S.E. *Phys. Rev. B* **75**, 140404(R) (2007).
- Pufall, M.R., Rippard, W.H., Kaka, S., Russek, S.E., Silva, T.J., Katine, J.A., Carey, M. *Phys. Rev. B* **69**, 214409 (2004).
- Ravelosona, R., Mangin, S., Lemaho, Y., Katine, J.A., Terris, B.D., Fullerton, E. *Phys. Rev. Lett.* **96**, 186604 (2006).
- Rezende, S.M., de Aguiar, F.M., Lucena, M.A., Azevedo, A. *Phys. Rev. Lett.* **84**, 4212 (2000).
- Rippard, W.H., Pufall, M.R., Kaka, S., Silva, T.J., Russek, S.E., Katine, J.A. *Phys. Rev. Lett.* **95**(6), 067203 (2005).
- Rong, F.C., Buchwald, W.R., Poindexter, E.H., Warren, W.L., Keeble, D.J. *Solid State Electron.* **34**(8), 835–841 (1991).
- Rong, F.C., Gerardi, G.J., Buchwald, W.R., Poindexter, E.H., Umior, M.T., Keeble, D.J., Warren, W.L. *Appl. Phys. Lett.* **60**(5), 610 (1992).

- Rudiger, U., Yu, J., Zhang, S., Kent, A.D. *Phys. Rev. Lett.* **80**(25), 5639 (1998)
- Rusponi, S., Weiss, N., Cren, T., Epple, M., Brune, H. *Appl. Phys. Lett.* **87**, 162514 (2005).
- Saitoh, E., Miyajima, H., Yamaoka, T., Tatara, G. *Nature* **432**, 203 (2004).
- Sahli, E., Berger, L. *J. Appl. Phys.* **76**, 4787–4792 (1994).
- Sankey, J.C., Krivorotov, I.N., Kiselev, S.I., Braganca, P.M., Emley, N.C., Buhrman, R.A., Ralph, D.C., *Phys. Rev. B* **72**, 224427 (2005).
- Sankey, J.C., Braganca, P.M., Garcia, A.G.F., Krivorotov, I.N., Buhrman, R.A., Ralph, D.C. *Phys. Rev. Lett.* **96**, 227601 (2006).
- Sanvito, S. *J. Mater. Chem.* **17**, 4455 (2007).
- Sato, T., Yokoyama, H., Ohya, H., Kamada, H. *J. Magn. Res.* **153**, 113–116 (2001).
- Schmid, G.T., Ferrand, D., Molenkamp, L.W., Filip, A.T., van Wees, B.J. *Phys. Rev. B* **62**, R4790 (2000).
- Schumacher, H.W., Chappert, C., Crozat, P., Sousa, R.C., Freitas, P.P., Miltat, J., Fassbender, J., Hillebrands, B. *Rev. Lett.* **90**, 017201 (2003).
- Seki, T., Mitani, S., Yakushiji, K., Takanashi, K. *Appl. Phys. Lett.* **88**, 172504 (2006b).
- Seki, T., Mitani, S., Yakushiji, K., Takanashi, K. *J. Appl. Phys.* **99**, 08G521 (2006c).
- Seki, T., Mitani, S., Yakushiji, K., Takanashi, K. *Appl. Phys. Lett.* **89**, 172504 (2006a).
- Shaffique, A., Polianski, M.L., Brouwer, P.W. *Phys. Rev. B* **73**, 024425 (2006).
- Shapiro, A., Levy, P.M., Zhang, S. *Phys. Rev. B* **67**, 104430 (2003)
- Sih, V., Yuichiro, K., Awschalom, D.D. *Phys. World* **18**(11), 33 (2005b).
- Sih, V., Myers, R.C., Kato, Y.K., Lau, W.H., Gossard, A.C., Awschalom D.D. *Nature Phys.* **1**, 31–35 (2005a).
- Silsbee, R.H. *J. Phys.: Condens. Matter* **16**, R179–R207 (2004).
- Silva, T.J. *Nature Phys.* **3**, 447 (2007).
- Sinova, J., Culcer, D., Niu, Q., Sinitsyn, N.A., Junqwirth, T., MacDonald, A.H. *Phys. Rev. Lett.* **92**, 126603 (2004).
- Slonczewski, J.C. *Phys. B* **39**(10), 6995 (1989).
- Slonczewski, J.C. *J. Magn. Magn. Mater.* **159**, L1 (1996).
- Slonczewski, J.C.U.S. Patent No. 5695864 (1997).
- Slonczewski, J.C. *J. Magn. Magn. Mater.* **195**, L261 (1999).
- Slonczewski, J.C. *J. Magn. Magn. Mater.* **247**, 324 (2002).
- Sokolov, A., Zhang, C., Tsymbal, E.Y., Redepenning, J., Doudin, B. *Nature Nanotechnol.* **2**, 171 (2007).
- Spoehr, R. *Ion Tracks and Microtechnology, Basic Principles and Applications* (Vieweg Publ. Co., 1990).
- Stahis, J.H. *Appl. Phys. Lett.* **68**(12), 1669 (1996).
- Stiles, M.D., Zangwill, A. *J. Appl. Phys.* **91**(10), 6812 (2002a).
- Stiles, M.D., Zangwill, A. *Phys. Rev. B* **66**, 014407 (2002b).
- Stiles, M.D., Miltat, J. in *Spin Dynamics in Confined Magnetic Structures III*, (eds) B. Hillebrands, A. Thiaville, *Topics in Appl. Phys.* 101, 225–308 (Springer-Verlag Berlin Heidelberg, 2006).

- Stiles, M.D., Xiao, J., Zangwill, A. *Phys. Rev. B* **69**, 055408 (2004).
- Strand, J., Lou, X., Adelmann, C., Schultz, B.D., Isakovic, A.F., Palmstrøm, C.J., Crowell, P.A. *Phys. Rev. B* **72**, 155308 (2005).
- Sun, J.Z. US Patent No. 6130814 (1998).
- Tang, H.X., Kawakami, R.K., Awschalom, D.D., Roukes, M.L. *Phys. Rev. Lett.* **90**, 107201 (2003).
- Tatara, G., Fukuyama, H. *Phys. Rev. Lett.* **78**(19), 3773 (1997).
- Tatara, G., Kohno, H. *Phys. Rev. Lett.* **92**(8), 86601 (2004).
- Tatara, G., Zhao, Y.-W., Munoz, M., Garcia, N. *Phys. Rev. Lett.* **83**(10), 2030 (1999).
- Tehrani, S., Slaughter, J.M., Chen, E., Durlam, M., Shi, J., De-Herrera, M. *IEEE Trans. Magn.* **35**, 2814 (2000).
- Theeuwen, S.J.C.H., Caro, J., Wellock, K.P., Radelaar, S., Marrows, C.H., Hickey, B.J., Kozub, V.I. *Appl. Phys. Lett.* **75**, 3677 (1999).
- Thiaville, A., Nakatani, K. in *Spin Dynamics in Confined Magnetic Structures III*, (eds) B. Hillebrands, A. Thiaville, *Topics in Appl. Phys.* 101, 161–205. (Springer-Verlag Berlin Heidelberg, 2006).
- Thomas, L., Hayashi, M., Jiang, X., Moriya, R., Rettner, C., Parkin, S.S.P. *Nature* **443**, 197 (2006).
- Tombros, N., van der Molen, S.J., van Wees, B.J. *Phys. Rev. B* **73**, 233403 (2006).
- Tserkovnyak, Y., Brataas, A., Gerrit, E.W., Bauer, Bertrand, I. Halperin, *Rev. Mod. Phys.* **77**, 1375 (2005).
- Tserkovnyak, Y., Brataas, A., Bauer, G.E.W. *Phys. Rev. Lett.* **88**, 117601 (2002).
- Tsoi, M., Jansen, A.G.M., Bass, J., Chiang, W.-C., Seck, M., Tsoi, V., Wyder, P. *Phys. Rev. Lett.* **80**, 4281 (1998a).
- Tsoi, M., Jansen, A.G.M., Bass, J., Chiang, W.-C., Seck, M., Tsoi, V., Wyder, P. *Phys. Rev. Lett.* **81**, 493 (1998b).
- Tsoi, M., Jansen, A.G.M., Bass, J., Chiang, W.-C., Tsoi, V., Wyder, P. *Nature* **406**, 46–48 (2000).
- Tsoi, M., Fontana, R.E., Parkin, S.S.P. *Appl. Phys. Lett.* **83**, 2617–2619 (2003).
- Tsoi, M., Sun, J.Z., Parkin, S.S.P. *Phys. Rev. Lett.* **93**, 036602 (2004).
- Tulapurkar, A.A., Suzuki, Y., Fukushima, A., Kubota, H., Maehara, H., Tsunekawa, K., Djayaprawira, D.D., Watanabe, N., Yuasa, S. *Nature* **438**, 339–342 (2005).
- Urazhdin, S., Birge, N.O., Pratt Jr., W.P., Bass, J. *Phys. Rev. Lett.* **91**, 146803 (2003).
- Urban, R., Woltersdorf, G., Heinrich, B. *Phys. Rev. Lett.* **87**, 217204 (2001).
- Valet, T., Fert, A. *Phys. Rev. B* **48**, 7099 (1993)
- van Gorkom, R.P., Brataas, A., Bauer, G.E.W. *Phys. Rev. Lett.* **83**(21), 4401 (1999b).
- van Gorkom, R.P., Caro, J., Theeuwen, S.J.C.H., Wellock, K.-P., Gribov N.N., Radelaar, S. *Appl. Phys. Lett.* **74**(3), 422 (1999a).
- van Son, P.C., van Kempen, H., Wyder, P. *Phys. Rev. Lett.* **58**, 2271 (1987).
- Vanhaverbeke, A., Viret, M. *Phys. Rev. B* **75**, 024411 (2007).
- Vavra, W., Cheng, S.F., Fink, A., Krebs, J.J., Prinz, G.A. *Appl. Phys. Lett.* **66**, 2579 (1995).
- Viret, M., Vignoles, D., Cole, D., Coey, J.M.D., Allen, W., Daniel, D.S., Gregg, J.F. *Phys. Rev. B* **53**(13), 8464 (1996)

- Viret, M., Drouet, M., Nassar, J., Contour, J.P., Fermon, C., Fert, A. *Europhys. Lett.* **39**(5), 545 (1997).
- Waïntal, X., Viret, M. *Europhys. Lett.* **65**(3), 427–433 (2004).
- Watts, S.M., van Wees, B.J. *Phys. Rev. Lett.* **97**, 116601 (2006).
- Weber, W., Riesen, S., Siegmann, H.C. *Science* **291**, 1015 (2001).
- Wegrowe, J.-E., Fabian, A., Guittienne, Ph., Hoffer, X., Kelly, D., Ansermet, J.-Ph., Olive, E. *Appl. Phys. Lett.* **80**, 3775 (2002).
- Wegrowe, J.-E., Comment, A., Jaccard, A., Ansermet, J.-Ph., Dempsey, N.M., Nozières *Phys. Rev. B* **61**, 12216 (2000).
- Wegrowe, J.-E., Kelly, D., Hoffer, X., Guittienne Ph., Ansermet, J.-Ph. *J. Appl. Phys.* **89**(11), 7127 (2001a).
- Wegrowe, J.-E., Kelly, D., Truong, T., Guittienne, Ph., Ansermet, J.-Ph. *Europhys. Lett.* **56**, 748–754 (2001b).
- Wegrowe, J.-E., Kelly, D., Jaccard, Y., Guittienne, Ph., Ansermet, J.-Ph. *Europhys. Lett.* **45**(5), 626 (1999a).
- Wegrowe, J.-E., Ansermet, J.-Ph., Gilbert, S.E., US Patent No. 6172902 (1999b).
- Wegrowe, J.-E., Ciornei, M.C., Drouhin, H.-J. *J. Phys.: Condens. Matter* **19**, 165213 (2007).
- Wegrowe, J.-E., Dubey, M., Wade, T., Drouhin, H.-J., Konczykowski, M. *J. Appl. Phys.* **96**, 4490 (2004).
- Wegrowe, J.-E., Wade, T., Hoffer, X., Gravier, L., Bonard, J.-M., Ansermet, J.-Ph. *Phys. Rev. B* **67**, 104418 (2003).
- Wernsdorfer, W., Bonet Orozco, E., Barbara, B., Hasselbach, K., Benoit, A., Mailly, D., Doudin, B., Meier, J., Wegrowe, J.E., Ansermet, J.-Ph., Demony, N., Pascard, H., Demony, N., Loiseau, A., Francois, L., Duxin N., Pilani, M.P. *J. Appl. Phys.* **81**, 5543, and references therein (1997).
- Wimbauer, T., Brandt, M.S., Bayerl, M.W., Reinaheer, N.M., Stutzmann, M. *Phys. Rev. B* **58**(8), 4892 (1998).
- Winter, J. *Magnetic Resonance in Metals* (Clarendon Press, Oxford, 1971).
- Wohlgemann, M., Jiang, X.M., Vardeny, Z.V., Janssen, R.A.J. *Phys. Rev. Lett.* **88**, 197401 (2002).
- Wohlgemann, M., Tandon, K., Mazumdar, S., Ramasesha, S., Vardeny, Z.V. *Nature* **409**, 494 (2001).
- Wulfhekel, W., Schlickum, U., Kirschner, J. in *Scanning Probe Microscopy Electrical*, part 2 (Springer, New York, 2007).
- Wunderlich, J., Kaestner, B., Sinova, J., Jungwirth, T. *Phys. Rev. Lett.* **94**, 047204 (2005).
- Xia, K., Kelly, P.J., Bauer, G.E.W., Brataas, A., Turek, I. *Phys. Rev. B* **65**, 220401 (2002).
- Xiao, J.X., Jiang, J.S., Chien, C.L. *Phys. Rev. Lett.* **68**(5), 3749 (1992).
- Yafet, Y. in *Solid State Physics* (Academic Press, New York, 1963), **14**.
- Yamada, K., Kasai, S., Nakatani, Y., Kobayashi, K., Kohno, H., Thiaville A., Ono, T., *Nature Mater.* **6**, 269 (2007).
- Yamaguchi, A., Ono, T., Nasu, S., Miyake, K., Mibu, K., Shinjo, T. *Phys. Rev. Lett.* **92**, 077205 (2004).

- Yamanouchi, M., Chiba, D., Matsukura, F., Ohno, H. *Nature* **428**, 539 (2004).
- Yang, T., Hamrle, J., Kimura, T., Otani, Y. *Appl. Phys. Lett.* **87**, 162502 (2005b).
- Yang, T., Kimura, T., Otani, Y. *J. Appl. Phys.* **97**, 064304 (2005a).
- Yang, T., Hamrle, J., Kimura, T., Otani, Y. *J. Magn. Magn. Mater.* **301**, 389 (2006).
- Yang, Q., Holody, P., Lee, S.F., Henry, L.L., Loloei, R., Schroeder, P.A., Pratt, Jr. W.P., Bass, J. *Phys. Rev. Lett.* **72**, 3274 (1994).
- Yanson, I.K., Naidyuk, Y.G., Bashlakov, D.L., Fisun, V.V., Balkashin O.P., Korenivski, V., Konovalenko, A., Shekhter, R.I. *Phys. Rev. Lett.* **185502** (2005).
- Yuan, L., da Silva, F.C.S., Halloran, S.T., Fardi, H.Z., Pappas, D.P., found on the web with the title “Oersted Field Induced Magnetization Scissoring in Permalloy Bars” (2007).
- Yuasa, S., Nagahama, T., Fukushima, A., Suzuki, Y., Ando, K. *Nature Mater.* **3**, 868–871 (2004).
- Zhang, S., Li, Z. *Phys. Rev. Lett.* **93**, 127204 (2004).
- Zhang, S., Levy, P.M., Fert, A. *Phys. Rev. Lett.* **88**, 236601 (2002).
- Zimmerler, M.A., Oezilimaz, B., Chen, W., Kent, A.D., Sun, J.Z., Rooks M.J., Koch, R.H. *Phys. Rev. B* **70**, 184438 (2004).
- Zutic, I., Fabian, J., Erwin, S.C. *IBM, J. Res. Dev.* **50**, 121 (2006).
- Zutic, I., Fabian, J., Das Sarma, S. *Rev. Mod. Phys.* **76**, 323 (2004).

# 4

# Molecular nanomagnets: Towards molecular spintronics

4.1 Introduction	136
4.2 Overview of molecular nanomagnets	139
4.3 Giant spin model for nanomagnets	141
4.4 Quantum dynamics of a dimer of nanomagnets	152
4.5 Resonant photon absorption in Cr <sub>7</sub> Ni antiferromagnetic rings	155
4.6 Photon-assisted tunnelling in single-molecule magnet	160
4.7 Environmental decoherence effects in nanomagnets	161
4.8 Molecular spintronics using single-molecule magnets	166
4.9 Conclusion	173
References	174

*W. Wernsdorfer*

## 4.1 Introduction

A revolution in electronics is in view, with the contemporary evolution of two novel disciplines, spintronics and molecular electronics. A fundamental link between these two fields can be established using molecular magnetic materials and, in particular, single-molecule magnets, which combine the classic macroscale properties of a magnet with the quantum properties of a nanoscale entity. The resulting field, molecular spintronics aims at manipulating spins and charges in electronic devices containing one or more molecules (Bogani and Wernsdorfer 2008; Roch *et al.* 2008).

The contemporary exploitation of electronic charge and spin degrees of freedom is a particularly promising field both at fundamental and applied levels. This discipline, called spintronics, has already seen some of its fundamental results turned into actual devices in a record time of 10 years and it holds great promises for the future (Wolf *et al.* 2001; Awschalom and Flatt 2007). Spintronic systems exploit the fact that the electron current is composed of spin-up and spin-down carriers that carry information encoded in their spin state and interact with magnetic materials differently. Information encoded in spins persists when the device is switched off; it can be manipulated with and without using magnetic fields and can be written using little energy, to cite just a few advantages of this approach.

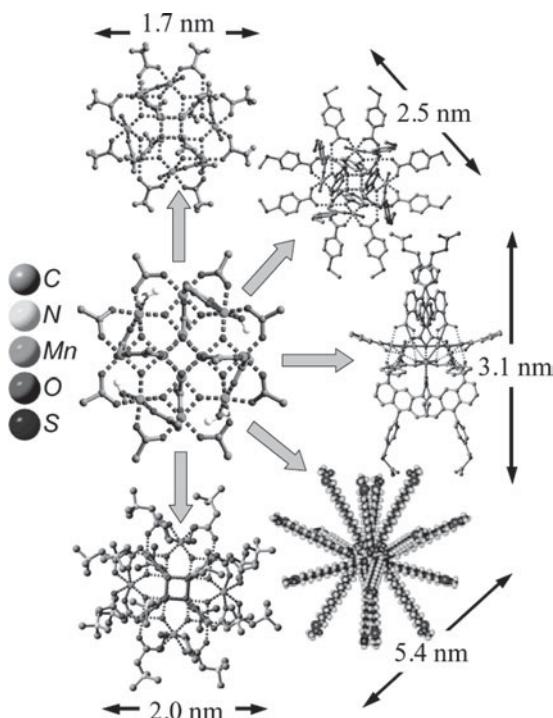
New efforts are now directed towards spintronic devices that preserve and exploit quantum coherence, so that fundamental investigations are shifting from metals to semiconducting (Wolf *et al.* 2001; Awschalom and Flatt 2007), and organic materials (Xiong *et al.* 2004), which potentially offer the best promises for cost, integration and versatility. For example, organic materials are already used in applications such as organic light-emitting diodes (OLED), displays and organic transistors. The concomitant trend towards

ever-smaller electronic devices (having already reached the nanoscale), and the tailoring of new molecules possessing increased conductance and functionalities are driving electronics to its ultimate molecular-scale limit (Tao 2006), and the so-called molecular electronics is now being intensively investigated.

In experiments of molecular electronics, the measuring devices are usually constituted by two nanoelectrodes and a bridging molecule in between, allowing the measurement of electron transport through a single molecule. As the measurement is performed at the molecular level, the observables are connected to molecular orbitals and not to Bloch waves as in bulk materials. Hence, new rules are found for these systems and it becomes possible to probe the quantum properties of the molecule directly. The electron-tunnelling processes in the electrode–molecule–electrode system can show the presence of Kondo or Coulomb-blockade effects, depending on the binding strength between the molecule and the electrodes, which can be tuned by selecting the appropriate chemical functional groups.

In this context, a new field of molecular spintronics is emerging that combines the concepts and the advantages of spintronics and molecular electronics (Sanvito and Rocha 2006; Bogani and Wernsdorfer 2008) which requires the creation of molecular devices using one or few magnetic molecules. Compounds of the single-molecule magnets (SMMs) class seem particularly attractive: their magnetization relaxation time is extremely long at low temperature, reaching years below 2 K, with record anisotropy barriers approaching 100 K (Milios *et al.* 2007). These systems, combining the advantages of the molecular scale with the properties of bulk magnetic materials, look attractive for high-density information storage and also, owing to their long coherence times (Ardavan *et al.* 2007; Carretta *et al.* 2007a,b; Bertaina *et al.* 2008), for quantum computing (Leuenberger and Loss 2001; Troiani *et al.* 2005). Moreover, their molecular nature leads to appealing quantum effects of the static and dynamic magnetic properties. The rich physics behind the magnetic behavior produces interesting effects like negative differential conductance and complete current suppression (Heersche *et al.* 2006; Jo *et al.* 2006), which could be used in electronics. Another advantage is that the weak spin-orbit and hyperfine interactions in organic molecules is likely to preserve spin-coherence over time and distance much longer than in conventional metals or semiconductors. Last but not least, specific functions (e.g. switchability with light, electric field, etc.) could be directly integrated into the molecule.

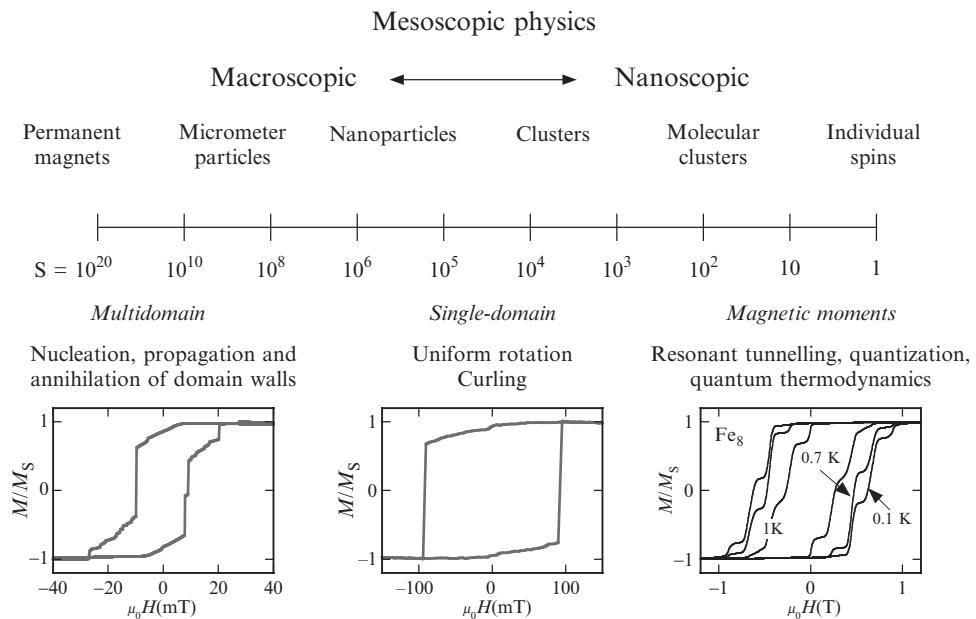
SMMs possess the right chemical characteristics to overcome several problems associated with molecular junctions. They are constituted by an inner magnetic core with a surrounding shell of organic ligands (Christou *et al.* 2000) that can be tailored to bind them on surfaces or into junctions (Coronado *et al.* 2004; Fleury *et al.* 2005; Naitabdi *et al.* 2005; Cornia *et al.* 2006) (Fig. 4.1). In order to strengthen magnetic interactions between the magnetic core ions, SMMs often have delocalized bonds, which can enhance their conducting properties. SMMs come in a variety of shapes and sizes and permit selective substitutions of the ligands in order to alter the coupling to the environment (Christou *et al.* 2000; Fleury *et al.* 2005; Cornia *et al.* 2006;



**Fig. 4.1** Representative examples of the peripheral functionalization of the outer organic shell of the  $\text{Mn}_{12}$  SMM. Different functionalizations used to graft the SMM to surfaces are displayed. Solvent molecules have been omitted. The diameter of the clusters is reported in the figure (reprinted from Bogani, L., Wernsdorfer, W. (2008) *Nature Mater.* **7**, 179, © 2008 Nature Publishing Group).

Bogani *et al.* 2007). It is also possible to exchange the magnetic ions, thus changing the magnetic properties without modifying the structure and the coupling to the environment (Ishikawa *et al.* 2005a,b). While grafting SMMs on surfaces has already led to important results, even more spectacular results will emerge from the rational design and tuning of single SMM-based junctions.

From a physics viewpoint, SMMs are the final point in the series of smaller and smaller units from bulk matter to atoms (Fig. 4.2). They combine the classic macroscale properties of a magnet with the quantum properties of a nanoscale entity. They have crucial advantages over magnetic nanoparticles in that they are perfectly monodispersed and can be studied in molecular crystals. They display an impressive array of quantum effects (that are observable up to higher and higher temperatures due to progress in molecular designs), ranging from quantum tunnelling of magnetization (Friedman *et al.* 1996; Thomas *et al.* 1996; Sangregorio *et al.* 1997; Wernsdorfer *et al.* 2006) to Berry phase interference (Wernsdorfer and Sessoli 1999; Wernsdorfer *et al.* 2005a) and quantum coherence (Ardavan *et al.* 2007; Carretta *et al.* 2007; Bertaina *et al.* 2008) with important consequences on the physics of spintronic devices. Although the magnetic properties of SMMs can be affected when they are deposited on surfaces or between leads (Bogani *et al.* 2007), these systems remain a step ahead of non-molecular nanoparticles, which show large size and anisotropy distributions, for a low structure versatility.



**Fig. 4.2** Scale of size that goes from macroscopic down to nanoscopic sizes. The unit of this scale is the number of magnetic moments in a magnetic system (roughly corresponding to the number of magnetic atoms). At macroscopic sizes, a magnetic system is described by magnetic domains that are separated by domain-walls. Magnetization reversal occurs via nucleation, propagation, and annihilation of domain walls (hysteresis loop on the left). When the system size is of the order of magnitude of the domain-wall width or the exchange length, the formation of domain walls requires too much energy. Therefore, the magnetization remains in the so-called single-domain state, and the magnetization reverses by uniform rotation or non-uniform modes (middle). SMMs are the final point in the series of smaller and smaller units from bulk matter to atoms and magnetization reverses via quantum tunnelling (right) (reprinted from Wernsdorfer, W. (2001) *Adv. Chem. Phys.* **118**, 99, © 2008 John Wiley & Sons Limited, reproduced with permission).

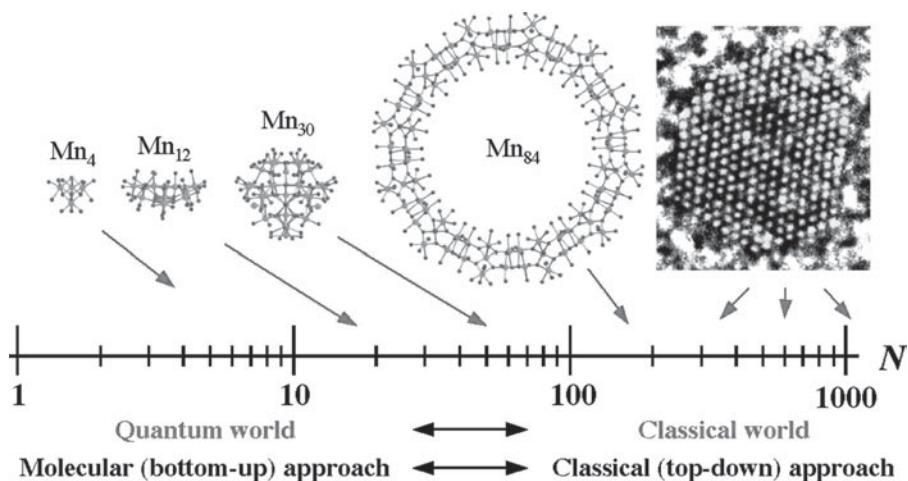
This chapter introduces the basic concepts that are needed to understand the quantum phenomena observed in molecular nanomagnets. Most tunnelling studies presented here were performed by magnetization measurements on single crystals using an array of micro-SQUIDs (Wernsdorfer 2001). This chapter concludes by showing the new trends towards molecular spintronics using junctions (Bogani and Wernsdorfer 2008) and nano-SQUIDs (Cleuziou *et al.* 2006).

## 4.2 Overview of molecular nanomagnets

Molecular nanomagnets or single-molecule magnets (SMMs) are mainly organic molecules that have one or several metal centers with unpaired electrons. These polynuclear metal complexes are surrounded by bulky ligands (often organic carboxylate ligands). The most prominent examples are a dodecanuclear mixed-valence manganese-oxo cluster with acetate ligands, short  $\text{Mn}_{12}$  acetate (Lis 1980), and an octanuclear iron(III) oxo-hydroxo cluster of formula  $[\text{Fe}_8\text{O}_2(\text{OH})_{12}(\text{tacn})_6]^{8+}$  where tacn is a macrocyclic ligand, short  $\text{Fe}_8$  (Wieghardt *et al.* 1984). Both systems have a spin ground state of  $S = 10$  and an Ising-type magnetic anisotropy, which stabilizes the spin states with  $m = \pm 10$  and generates an energy barrier of about 67 K for the reversal of

magnetization for  $\text{Mn}_{12}$  acetate (Caneschi *et al.* 1991; Sessoli *et al.* 1993a,b) and 25 K for  $\text{Fe}_8$  (Barra *et al.* 1996).

Thermally activated quantum tunnelling of the magnetization was first evidenced in both systems (Novak and Sessoli 1995; Paulsen and Park 1995; Friedman *et al.* 1996; Thomas *et al.* 1996; Sangregorio *et al.* 1997). Theoretical discussion of this assumes that thermal processes (principally phonons) promote the molecules up to high levels with small quantum numbers  $|m|$ , not far below the top of the energy barrier, and the molecules then tunnel inelastically to the other (Abragam and Bleaney 1970; Villain *et al.* 1994, 1997; Politi *et al.* 1995; Hartmann-Boutron *et al.* 1996; Garanin and Chudnovsky 1997; Fort *et al.* 1998; Leuenberger and Loss 2000b). Thus, the transition is almost entirely accomplished via thermal transitions and the characteristic relaxation time is strongly temperature dependent. For  $\text{Fe}_8$ , however, the relaxation time becomes temperature independent below 0.36 K (Sangregorio *et al.* 1997; Ohm *et al.* 1998a) showing that a pure tunnelling mechanism between the only populated ground states  $m = \pm S = \pm 10$  is responsible for the relaxation of the magnetization. On the other hand, in the  $\text{Mn}_{12}$  acetate system one sees temperature-independent relaxation only for strong applied fields and below about 0.6 K (Perenboom *et al.* 1998; Kent *et al.* 2000). During the last few years, many new molecular nanomagnets were presented (Aubin *et al.* 1998; Caneschi *et al.* 1999; Price *et al.* 1999; Yoo *et al.* 2000) that also show tunnelling at low temperatures. The largest molecular nanomagnet is currently a  $\text{Mn}_{84}$  molecule (Tasiopoulos *et al.* 2004) that has a size of a magnetic nanoparticle (Fig. 4.3). The record anisotropy barrier of 89 K is currently found for a  $\text{Mn}_6$  SMM (Milios *et al.* 2007).



**Fig. 4.3** Size scale spanning atomic to nanoscale dimensions. On the far right is shown a high-resolution transmission electron microscopy view along a [110] direction of a typical 3-nm diameter cobalt nanoparticle exhibiting a face-centered cubic structure and containing about 1000 Co atoms. The  $\text{Mn}_{84}$  molecule is a 4.2-nm diameter particle (Tasiopoulos *et al.* 2004). Also shown for comparison are the indicated smaller Mn nanomagnets, which are drawn to scale. An alternative means of comparison is the Néel vector ( $N$ ), which is the scale shown. The arrows indicate the magnitude of the Néel vectors for the indicated SMMs, which are 7.5, 22, 61, and 168 for  $\text{Mn}_4$ ,  $\text{Mn}_{12}$ ,  $\text{Mn}_{30}$  and  $\text{Mn}_{84}$ , respectively (reprinted from Tasiopoulos, A., Vinslava, A., Wernsdorfer, W., Abboud, K., Christou, G. (2004) *Angew. Chem. Int. Ed. Engl.* **43**, 2117, © 2008 John Wiley & Sons Limited, reproduced with permission).

### 4.3 Giant spin model for nanomagnets

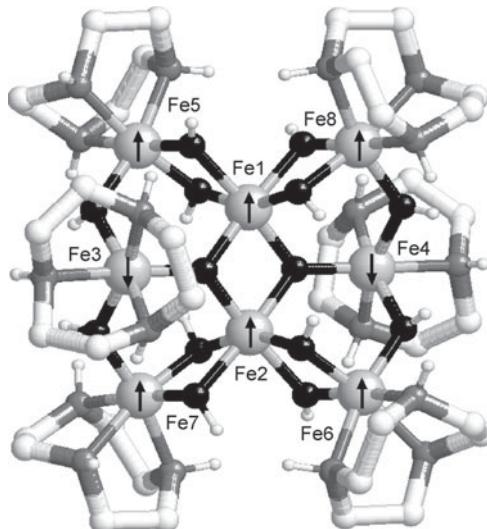
A magnetic molecule that behaves like a small nanomagnet must have a large uniaxial easy-axis-type magnetic anisotropy and a large ground-state spin. A typical example is the octanuclear iron(III) oxo-hydroxo cluster of formula  $[Fe_8O_2(OH)_{12}(tacn)_6]^{8+}$  where tacn is a macrocyclic ligand (1,4,7-triazaazacyclononane), short Fe<sub>8</sub> (Fig. 4.4) (Wieghardt *et al.* 1984).

The internal iron(III) ions are octahedrally coordinated to the two oxides and to four hydroxo bridges. The outer iron(III) ions coordinate three nitrogens and three hydroxyls. Spin-polarized neutron scattering showed that all Fe ions have a spin 5/2, six spins up and two down (Pontillon *et al.* 1999). This rationalizes the  $S = 10$  spin ground state that is in agreement with magnetization measurements.

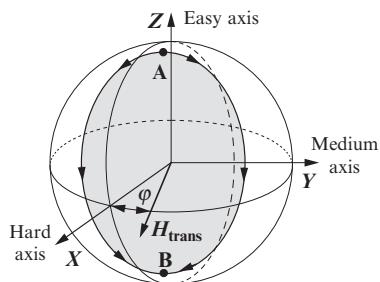
In principle, a multispin Hamiltonian can be derived taking into account all of the exchange interactions and the single-ion magnetic anisotropies. However, the Hilbert space is very large ( $6^8 \approx 10^6$ ) and the exchange coupling constants are not well known. A giant spin model is therefore often used that describes in an effective way the ground spin state multiplet. For a nanomagnet like the Fe<sub>8</sub> molecular cluster it has the following Hamiltonian

$$\mathcal{H} = -DS_z^2 + E(S_x^2 - S_y^2) + g\mu_B\mu_0\vec{S} \cdot \vec{H}, \quad (4.1)$$

$S_x$ ,  $S_y$ , and  $S_z$  are the three components of the spin operator,  $D$  and  $E$  are the anisotropy constants that were determined via high-frequency electron paramagnetic resonance (HF-EPR) ( $D/k_B \approx 0.275$  K and  $E/k_B \approx 0.046$  K (Barra *et al.* 1996)), and the last term of the Hamiltonian describes the Zeeman energy associated with an applied field  $\vec{H}$ . This Hamiltonian defines the hard, medium, and easy axes of magnetization in the  $x$ -,  $y$ -, and  $z$ -directions, respectively (Fig. 4.5). It has an energy-level spectrum with  $(2S + 1) = 21$  values that, to a first approximation, can be labelled by the quantum numbers  $m =$



**Fig. 4.4** Schematic view of the magnetic core of the Fe<sub>8</sub> cluster. The oxygen atoms are black, the nitrogen atoms are gray, and carbon atoms are white. The arrows represent the spin structure of the ground state  $S = 10$  (reprinted from Wernsdorfer, W. (2001) *Adv. Chem. Phys.* **118**, 99, © 2008 John Wiley & Sons Limited, reproduced with permission).



**Fig. 4.5** Unit sphere showing degenerate minima A and B that are joined by two tunnel paths (heavy lines). The hard, medium, and easy axes are taken in the x-, y-, and z-directions, respectively. The constant transverse field  $H_{\text{trans}}$  for tunnel-splitting measurements is applied in the  $xy$ -plane at an azimuth angle  $\varphi$ . At zero applied field  $\vec{H} = 0$ , the giant spin reversal results from the interference of two quantum spin paths of opposite direction in the easy anisotropy  $yz$ -plane. For transverse fields in the direction of the hard axis, the two quantum spin paths are in a plane that is parallel to the  $yz$ -plane, as indicated in the figure. By using Stokes' theorem it has been shown that the path integrals can be converted into an area integral, yielding that destructive interference—that is a quench of the tunnelling rate—occurs whenever the shaded area is  $k\pi/S$ , where  $k$  is an odd integer. The interference effects disappear quickly when the transverse field has a component in the  $y$ -direction because the tunnelling is then dominated by only one quantum spin path (reprinted from W. Wernsdorfer, R. Sessoli, (1999) *Science* **284**, 133, © 1999 American Association for the Advancement of Science).

–10, –9, . . . , 10 choosing the  $z$ -axis as the quantization axis. The energy spectrum, shown in Fig. 4.6, can be obtained by using standard diagonalization techniques of the  $[21 \times 21]$  matrix describing the spin Hamiltonian  $S = 10$ . At  $\vec{H} = 0$ , the levels  $m = \pm 10$  have the lowest energy. When a field  $H_z$  is applied, the energy levels with  $m < -2$  increase, while those with  $m > 2$  decrease (Fig. 4.6). Therefore, the energy levels of positive and negative quantum numbers cross at certain fields  $H_z$ . It turns out that for  $\text{Fe}_8$  the levels cross at fields given by  $\mu_0 H_z \approx n \times 0.22 \text{ T}$ , with  $n = 1, 2, 3, \dots$ . The inset of Fig. 4.6 displays the details at a level crossing where transverse terms containing  $S_x$  or  $S_y$  spin operators turn the crossing into an “avoided level crossing”. The spin  $S$  is “in resonance” between two states when the local longitudinal field is close to an avoided level crossing. The energy gap, the so-called “tunnel splitting”  $\Delta$ , can be tuned by an applied field in the  $xy$ -plane (Fig. 4.5) via the  $S_x H_x$  and  $S_y H_y$  Zeeman terms (Section 4.3.2).

The effect of these avoided level crossings can be seen in hysteresis-loop measurements (Fig. 4.7). When the applied field is near an avoided level crossing, the magnetization relaxes faster, yielding steps separated by plateaus. As the temperature is lowered, there is a decrease in the transition rate due to reduced thermally assisted tunnelling.

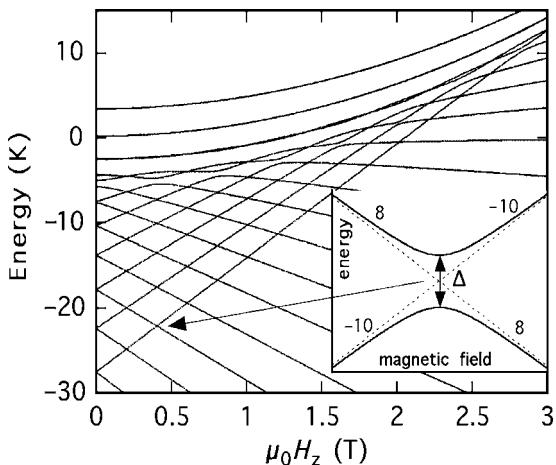
### 4.3.1 Landau–Zener tunnelling in $\text{Fe}_8$

The non-adiabatic transition between the two states in a two-level system was first discussed by Landau (1932); Stückelberg (1932), and Zener (1932). The original work by Zener concentrates on the electronic states of a diatomic molecule, while Landau and Stückelberg considered two atoms that undergo a scattering process. Their solution of the time-dependent Schrödinger equation of a two-level system could be applied to many physical systems and it became an important tool for studying tunnelling transitions. The Landau–Zener model has also been applied to spin tunnelling in nanoparticles and clusters (Miyashita 1995, 1996; Rose and Stamp 1998; Leuenberger and Loss 2000a; Thorwart *et al.* 2000). The tunnelling probability  $P$  when sweeping the longitudinal field  $H_z$  at a constant rate over an avoided energy-level crossing (Fig. 4.8) is given by

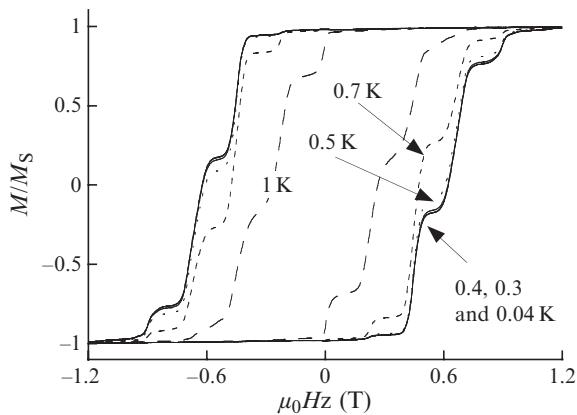
$$P_{m,m'} = 1 - \exp \left[ -\frac{\pi \Delta_{m,m'}^2}{2\hbar g \mu_B |m - m'| \mu_0 dH_z/dt} \right]. \quad (4.2)$$

Here,  $m$  and  $m'$  are the quantum numbers of the avoided level crossing,  $dH_z/dt$  is the constant field sweeping rates,  $g \approx 2$ ,  $\mu_B$  the Bohr magneton, and  $\hbar$  is Planck's constant.

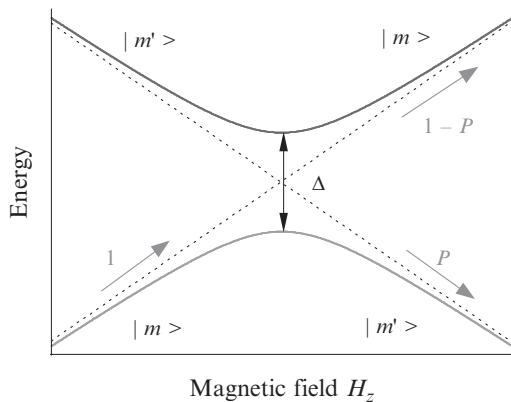
With the Landau–Zener model in mind, we can now start to understand qualitatively the hysteresis loops (Fig. 4.7). Let us start at a large negative magnetic field  $H_z$ . At very low temperature, all molecules are in the  $m = -10$  ground state (Fig. 4.6). When the applied field  $H_z$  is ramped down to zero, all molecules will stay in the  $m = -10$  ground state. When ramping the field over the  $\Delta_{-10,10}$  region at  $H_z \approx 0$ , there is a Landau–Zener tunnel probability  $P_{-10,10}$  to tunnel from the  $m = -10$  to the  $m = 10$  state.  $P_{-10,10}$  depends



**Fig. 4.6** Zeeman diagram of the 21 levels of the  $S = 10$  manifold of  $\text{Fe}_8$  as a function of the field applied along the easy axis (eqn (4.1)). From bottom to top, the levels are labelled with quantum numbers  $m = \pm 10, \pm 9, \dots, 0$ . The levels cross at fields given by  $\mu_0 H_z \approx n \times 0.22 \text{ T}$ , with  $n = 1, 2, 3, \dots$ . The inset displays the detail at a level crossing where the transverse terms (terms containing  $S_x$  or/and  $S_y$  spin operators) turn the crossing into an avoided level crossing. The greater the tunnel splitting  $\Delta$ , the higher the tunnel rate.



**Fig. 4.7** Hysteresis loops of a single crystal of  $\text{Fe}_8$  molecular clusters at different temperatures. The longitudinal field ( $z$ -direction) was swept at a constant sweeping rate of  $0.014 \text{ T/s}$ . The loops display a series of steps, separated by plateaux. As the temperature is lowered, there is a decrease in the transition rate due to reduced thermal-assisted tunnelling. The hysteresis loops become temperature independent below  $0.35 \text{ K}$ , demonstrating quantum tunnelling at the lowest energy levels (reprinted from W. Wernsdorfer, R. Sessoli, (1999) *Science* **284**, 133, © 1999 American Association for the Advancement of Science).



**Fig. 4.8** Detail of the energy-level diagram near an avoided level crossing.  $m$  and  $m'$  are the quantum numbers of the energy level.  $P_{m,m'}$  is the Landau-Zener tunnel probability when sweeping the applied field from the left to the right over the anticrossing. The greater the gap  $\Delta$  and the lower the sweeping rate, the higher is the tunnel rate (eqn (4.2)).

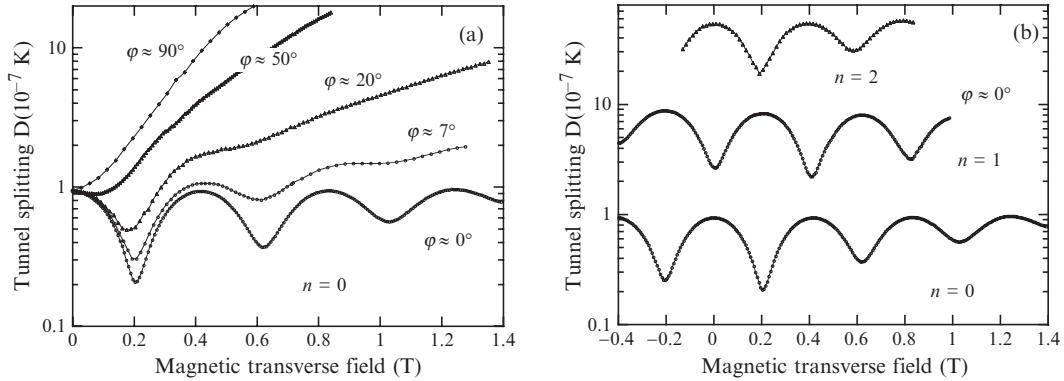
on the sweeping rate (eqn (4.2)); that is, the lower the sweeping rate, the larger the value of  $P_{-10,10}$ . This is clearly demonstrated in the hysteresis-loop measurements showing larger steps for slower sweeping rates (Wernsdorfer and Sessoli 1999; Wernsdorfer *et al.* 2005a). When the field  $H_z$  is now further increased, there is a remaining fraction of molecules in the  $m = -10$  state that forms a metastable state. The next chance to escape from this state is when the field reaches the  $\Delta_{-10,9}$  region. There is a Landau-Zener tunnel probability  $P_{-10,9}$  to tunnel from the  $m = -10$  to the  $m = 9$  state. As  $m = 9$  is an excited state, the molecules in this state de-excite quickly to the  $m = 10$  state by emitting a phonon. An analogous procedure happens when the applied field reaches the  $\Delta_{-10,10-n}$  regions ( $n = 2, 3, \dots$ ) until all molecules are in the  $m = 10$  ground state; that is, the magnetization of all molecules is reversed. As phonon emission can only change the molecule state by  $\Delta m = 1$  or 2, there is a phonon cascade for higher applied fields.

In order to apply quantitatively the Landau-Zener formula (eqn (4.2)), we first saturated the crystal of Fe<sub>8</sub> clusters in a field of  $H_z = -1.4$  T, yielding an initial magnetization  $M_{\text{in}} = -M_s$ . Then, we swept the applied field at a constant rate over one of the resonance transitions and measured the fraction of molecules that reversed their spin. This procedure yields the tunnelling rate  $P_{-10,10-n}$  and thus the tunnel splitting  $\Delta_{-10,10-n}$  (eqn (4.2)) with  $n = 0, 1, 2, \dots$

We first checked the predicted Landau-Zener sweeping-field dependence of the tunnelling rate. We found a good agreement for sweeping rates between 10 and 0.001 T/s (Wernsdorfer and Sessoli 1999). The deviations at lower sweeping rates are mainly due to the *hole-digging mechanism* (Wernsdorfer *et al.* 1999) that slows down the relaxation (Section 4.7.2). Our measurements showed for the first time that the Landau-Zener method is particularly adapted for molecular clusters because it works even in the presence of dipolar fields that spread the resonance transition provided that the field-sweeping rate is not too small.

### 4.3.2 Oscillations of tunnel splitting

An applied field in the  $xy$ -plane can tune the tunnel splittings  $\Delta_{m,m'}$  via the  $S_x$  and  $S_y$  spin operators of the Zeeman terms that do not commute with the spin Hamiltonian. This effect can be demonstrated by using the Landau-Zener method (Section 4.3.1). Figure 4.9 presents a detailed study of the tunnel splitting  $\Delta_{\pm 10}$  at the tunnel transition between  $m = \pm 10$ , as a function of transverse fields applied at different angles  $\varphi$ , defined as the azimuth angle between the anisotropy hard axis and the transverse field (Fig. 4.5). For small angles  $\varphi$  the tunnelling rate oscillates with a period of  $\sim 0.4$  T, whereas no oscillations showed up for large angles  $\varphi$  (Wernsdorfer and Sessoli 1999). In the latter case, a much stronger increase of  $\Delta_{\pm 10}$  with transverse field is observed. The transverse field dependence of the tunnelling rate for different resonance conditions between the state  $m = -10$  and  $(10 - n)$  can be observed by sweeping the longitudinal field around  $\mu_0 H_z = n \times 0.22$  T with  $n = 0, 1, 2, \dots$ . The corresponding tunnel splittings  $\Delta_{-10,10-n}$  oscillate with



**Fig. 4.9** Measured tunnel splitting  $\Delta$  as a function of transverse field for (a) several azimuth angles  $\varphi$  at  $m = \pm 10$  and (b)  $\varphi \approx 0^\circ$ , as well as for a quantum transition between  $m = -10$  and  $(10 - n)$ . Note the parity effect that is analogous to the suppression of tunnelling predicted for half-integer spins. It should also be mentioned that internal dipolar and hyperfine fields hinder a quench of  $\Delta$ , which is predicted for an isolated spin (reprinted from W. Wernsdorfer, R. Sessoli, (1999) *Science* **284**, 133, © 1999 American Association for the Advancement of Science).

almost the same period of  $\sim 0.4$  T (Fig. 4.9). In addition, comparing quantum transitions between  $m = -10$  and  $(10 - n)$ , with  $n$  even or odd, revealed a parity (or symmetry) effect that is analogous to the Kramers' suppression of tunnelling predicted for half-integer spins (Loss *et al.* 1992; von Delft and Henley 1992). A similar strong dependence on the azimuth angle  $\varphi$  was observed for all studied resonances.

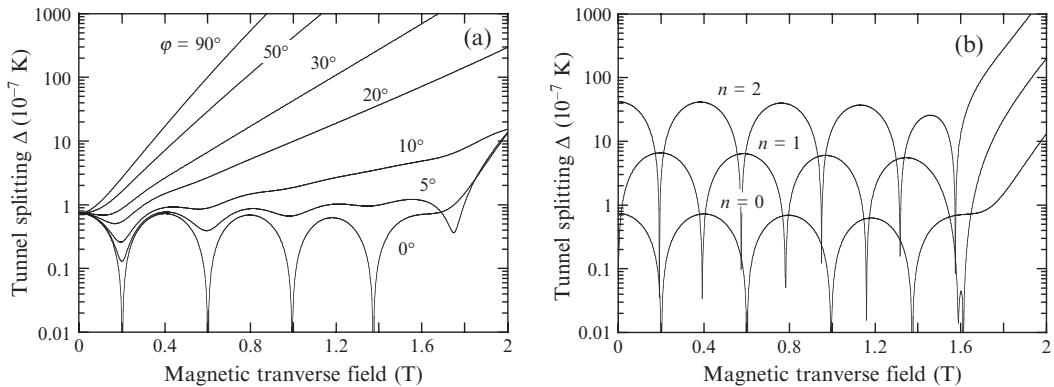
#### 4.3.2.1 Semi-classical descriptions

Before showing that the above results can be derived by an exact numerical calculation using the quantum operator formalism, it is useful to discuss semi-classical models. The original prediction of oscillation of the tunnel splitting was done by using the path-integral formalism (Feynman *et al.* 1970). Here (Garg 1993), the oscillations are explained by constructive or destructive interference of quantum spin phases (Berry phases) of two tunnel paths (instanton trajectories) (Fig. 4.5). Since our experiments were reported, the Wentzel–Kramers–Brillouin theory has been used independently by Garg (1999) and Villain and Fort (2000). The surprising fact is that although these models are derived semi-classically, and should have higher-order corrections in  $1/S$ , they appear to be exact as written! This was first noted by Garg (1999) and Villain and Fort (2000) and then established by (Kececioglu and Garg 2001).

The period of oscillation is given by (Garg 1993):

$$\Delta H = \frac{2k_B}{g\mu_B} \sqrt{2E(E + D)}, \quad (4.3)$$

where  $D$  and  $E$  are as defined in eqn (4.1). We find a period of oscillation of  $\Delta H = 0.26$  T for  $D = 0.275$  K and  $E = 0.046$  K as found by Barra *et al.* (1996). This is somewhat smaller than the experimental value of  $\sim 0.4$  T. We believe that this is due to higher-order terms of the spin Hamiltonian that are neglected in Garg's calculation. These terms can easily be included in the operator formalism as shown in the next section.



**Fig. 4.10** Calculated tunnel splitting  $\Delta$  as a function of transverse field for (a) quantum transition between  $m = \pm 10$  at several azimuth angles  $\varphi$  and (b) quantum transition between  $m = -10$  and  $(10 - n)$  at  $\varphi = 0^\circ$  (Section 4.3.2.2). The fourth-order terms suppress the oscillations of  $\Delta$  for large transverse fields  $|H_x|$  (reprinted from W. Wernsdorfer, R. Sessoli, (1999) *Science* **284**, 133, © 1999 American Association for the Advancement of Science).

#### 4.3.2.2 Exact numerical diagonalization

In order to quantitatively reproduce the observed periodicity we included fourth-order terms in the spin Hamiltonian (eqn (4.1)) as employed in the simulation of inelastic neutron scattering measurements (Caciuffo *et al.* 1998; Amoretti *et al.* 2000) and performed a diagonalization of the  $[21 \times 21]$  matrix describing the  $S = 10$  system. For the calculation of the tunnel splitting we used  $D = 0.289$  K,  $E = 0.055$  K (eqn (4.1)) and the fourth-order terms as defined by Amoretti *et al.* (2000) with  $B_4^0 = 0.72 \times 10^{-6}$  K,  $B_4^2 = 1.01 \times 10^{-5}$  K,  $B_4^4 = -0.43 \times 10^{-4}$  K, which are close to the values obtained by EPR (Barra *et al.* 2000) and neutron-scattering studies (Amoretti *et al.* 2000).

The calculated tunnel splittings for the states involved in the tunnelling process at the resonances  $n = 0, 1$ , and  $2$  are reported in Fig. 4.10, showing the oscillations as well as the parity effect for odd resonances.

#### 4.3.2.3 Spin-parity effect

The spin-parity effect is among the most interesting quantum phenomena that can be studied at the mesoscopic level in SMMs. It predicts that quantum tunnelling is suppressed at zero applied field if the total spin of the magnetic system is half-integer but is allowed in integer-spin systems. Enz and Schilling (1986) and Van Hemmen and Süto (1986) were the first to suggest the absence of tunnelling as a consequence of Kramers degeneracy. The Kramers theorem asserts that no matter how unsymmetric the crystal field, an ion possessing an odd number of electrons must have a ground state that is at least doubly degenerate, even in the presence of crystal fields and spin-orbit interactions (Kramers 1930).

The predicted spin-parity effect can be observed by measuring the tunnel splitting as a function of transverse field (Wernsdorfer *et al.* 2002b). An integer-spin system is rather insensitive to small transverse fields, whereas a half-integer spin system is much more sensitive. However, a half-integer spin system will also undergo tunnelling at zero external field as a result of

environmental degrees of freedom such as hyperfine and dipolar couplings or small intermolecular exchange interaction.

The nicest observation of the spin-parity effect has been seen for two molecular  $Mn_{12}$  clusters with a spin ground state of  $S = 10$  and  $S = 19/2$  showing oscillations of the tunnel probability as a function of a transverse field being due to topological quantum phase interference of two tunnel paths of opposite windings (Section 4.3.2.1). Spin-parity-dependent tunnelling was established for the first time in these compounds by comparing the quantum phase interference of integer and half-integer spin systems (Wernsdorfer *et al.* 2005a).

### 4.3.3 A classical approach with applications to the quantum regime

Recently, the molecular (or bottom-up) approach has reached the size regime from the classical (or top-down) approach to nanoscale magnetic materials (Tasiopoulos *et al.* 2004). Indeed, a giant  $Mn_{84}$  SMM was reported with a 4-nm diameter torus structure, exhibiting both magnetization hysteresis and quantum tunnelling. The study of such large systems is greatly complicated by the fact that the spin Hilbert space is huge and it is impossible to treat such systems with exact matrix diagonalization methods. However, since some SMMs are now as large as some classical nanoparticles, it raises the interesting possibility that classical models commonly employed to study the latter may be used to obtain a first-order understanding of large molecular systems (Wernsdorfer *et al.* 2005c). Indeed, we proposed the use of the classical Néel–Brown model (Néel 1949; Brown 1963; Coffey *et al.* 1995) of thermally activated magnetization reversal of a magnetic single-domain particle in order to study large SMMs (Wernsdorfer *et al.* 2005c). This method allows us to determine important parameters that characterize the magnetic properties of the SMM: the energy barrier, the magnetic anisotropy constant, the spin,  $\tau_0$ , and the cross-over temperature from the classical to the quantum regime. The method is particularly useful for SMMs having low-lying energy states and not showing quantum tunnelling steps in hysteresis loops. In such systems electron paramagnetic resonance (EPR) measurements often exhibit only very broad absorption peaks that do not allow the determination of the magnetic anisotropy.

#### 4.3.3.1 The Néel–Brown model of thermally activated magnetization reversal

The method is based on the Néel–Brown model of thermally activated magnetization reversal of a magnetic single-domain particle that has two equivalent ground states of opposite magnetization separated by an energy barrier due to magnetic anisotropy (Néel 1949; Brown 1963; Coffey *et al.* 1995). The system can escape from one state to the other either by thermal activation over the barrier at high temperatures or by quantum tunnelling at low temperatures. At sufficiently low temperatures and at zero field, the energy barrier between the two states of opposite magnetization is much too high to detect an escape process. However, the barrier can be lowered by applying a magnetic field in the opposite direction to that of the particle's magnetization. When the applied

field is close enough to the reversal field, thermal fluctuations are sufficient to allow the system to overcome the barrier, and the magnetization is reversed.

This stochastic escape process can be studied via the relaxation-time method consisting of the measurement of the probability that the magnetization has not reversed after a certain time. In the case of an assembly of identical and isolated particles, it corresponds to measurements of the relaxation of magnetization. According to the Néel–Brown model, the probability that the magnetization has not reversed after a time  $t$  is given by:

$$P(t) = e^{-t/\tau}, \quad (4.4)$$

and  $\tau$  (inverse of the reversal rate) can be expressed by an Arrhenius law of the form:

$$\tau(T, H) = \tau_0 e^{\Delta E(H)/k_B T}, \quad (4.5)$$

where  $\Delta E(H)$  is the field-dependent energy barrier height and  $\tau_0$  is the inverse of the attempt frequency. In most cases,  $\Delta E(H)$  can be approximated by

$$\Delta E(H) \approx E_0 \left(1 - H/H_c^0\right)^\alpha, \quad (4.6)$$

where  $H_c^0$  is the reversal field at zero temperature,  $E_0$  is the barrier height at zero applied field, and  $\alpha$  is a constant of the order of unity (for most cases  $1.5 \leq \alpha \leq 2$ ). In the case of a Stoner–Wohlfarth particle (Néel 1947; Stoner and Wohlfarth 1948) with uniaxial anisotropy and the field applied along the easy axis of magnetization, all constants can be determined analytically (Néel 1947, 1949):  $\alpha = 2$ ,  $E_0 = KV$ , and  $H_c^0 = 2K/M_s$ , where  $K$  is the uniaxial anisotropy constant,  $V$  is the particle volume, and  $M_s$  is the saturation magnetization. For SMMs with dominating uniaxial anisotropy:  $\alpha = 2$ ,  $E_0 = DS^2$ , and  $H_c^0 = 2DS/g\mu_0\mu_B$ . However, in general, all constants depend slightly on the fine details of the magnetic anisotropy and the direction of the applied field  $H$  (Thiaville 1998, 2000).

In order to study the field dependence of the relaxation time  $\tau(T, H)$  and to obtain the parameters of the model, the decay of magnetization has to be studied at many applied fields  $H$  and temperatures  $T$ . This is experimentally very time consuming and complicated by the fact that the equilibrium magnetization is temperature dependent and difficult to obtain for long relaxation times. In addition, for fast relaxation times the initial magnetization depends on the field sweep rates used for the application of the field. The number of exploitable orders of magnitude for  $\tau$  values is therefore limited for relaxation-time measurements.

A more convenient method of studying the magnetization decay is by ramping the applied field at a given rate (Wernsdorfer *et al.* 1997a) and measuring the coercive field  $H_c$  (the field value to obtain zero magnetization), which is then measured as a function of the field sweep rate and temperature.

The mathematical transformation from a reversal-time probability (eqns (4.4) and (4.5)) to a reversal-field probability was first given by Kurkijärvi (1972) for the critical current in SQUIDs. Later, Gunther and Barbara (1994) calculated similar expressions for magnetic domain wall junctions. A more general calculation was evaluated by Garg (1995). Here, we

use a simplified version (Wernsdorfer *et al.* 1997a): The probability density of reversal of a stochastic process is  $-dP/dt = P/\tau$  and the maximum of the probability density can be derived from  $d^2P/dt^2 = P(1 + d\tau/dt)/\tau^2 = 0$ . This gives  $d\tau/dt = -1$ . The application to eqn (4.5) leads to  $\Delta E(H) = k_B T \ln[k_B T / (\tau_0 \frac{dE}{dH} \frac{dH}{dt})]$ . Using eqn (4.6) we find eqn (4.7). The mean reversal field of an assembly of identical particles or SMMs is approximated by the coercive field  $H_c$ :

$$H_c(T, v) \approx H_c^0 \left( 1 - \left[ \frac{kT}{E_0} \ln \left( \frac{c}{v} \right) \right]^{1/\alpha} \right), \quad (4.7)$$

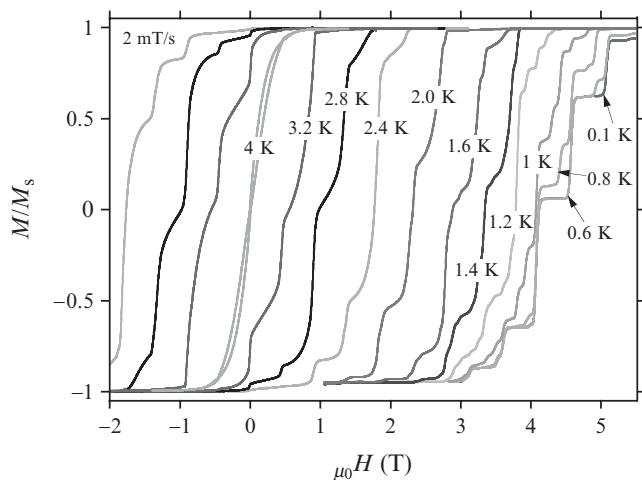
where the field-sweeping rate is given by  $v = dH/dt$ ;  $H_c^0$  is the coercive field at zero temperature, and  $c = H_c^0 k_B T / [\tau_0 \alpha E_0 (1 - H_c/H_c^0)^{\alpha-1}]$  (Wernsdorfer *et al.* 1997a).

#### 4.3.3.2 Application to Mn<sub>12</sub> SMMs

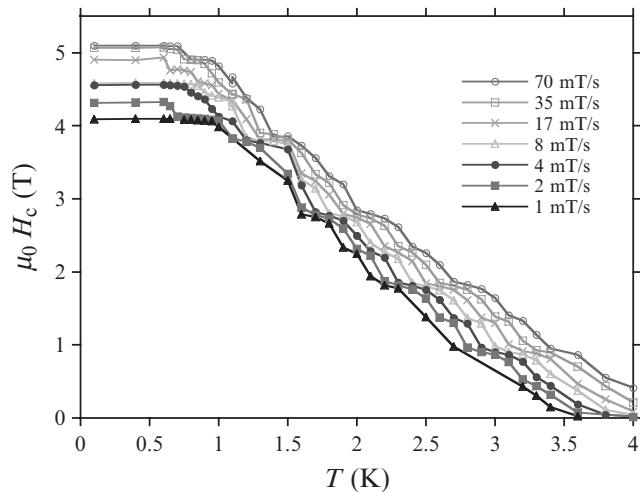
The above model is now applied to a Mn<sub>12</sub> SMMs (Wernsdorfer *et al.* 2005c) with a spin ground state of  $S = 10$ . Figure 4.11 shows typical hysteresis loops with a series of quantum steps separated by plateaus (Wernsdorfer *et al.* 2006). In order to apply the above approach, the temperature and field-sweep-rate dependences of the coercive fields  $H_c$  were measured and plotted in Fig. 4.12. As expected for a thermally activated process,  $H_c$  increases with decreasing temperature and with increasing field sweep rate. Furthermore, all measurements showed an almost logarithmic dependence of  $H_c$  on the field sweep rate.  $H_c$  becomes temperature independent below about 0.6 K.

The validity of eqn 7 was tested by plotting the set of  $H_c(T, v)$  values as a function of  $[T \ln(c/v)]^{1/2}$ , where  $c = H_c^0 k_B T / \tau_0 2 E_0 (1 - H_c/H_c^0)$ . If the underlying model is sufficient, all points should collapse onto one straight line by choosing the proper values for the constant  $\tau_0$ . We found that the data of  $H_c(T, v)$  fall on a master curve provided  $\tau_0 = 2.1 \times 10^{-7}$  s.

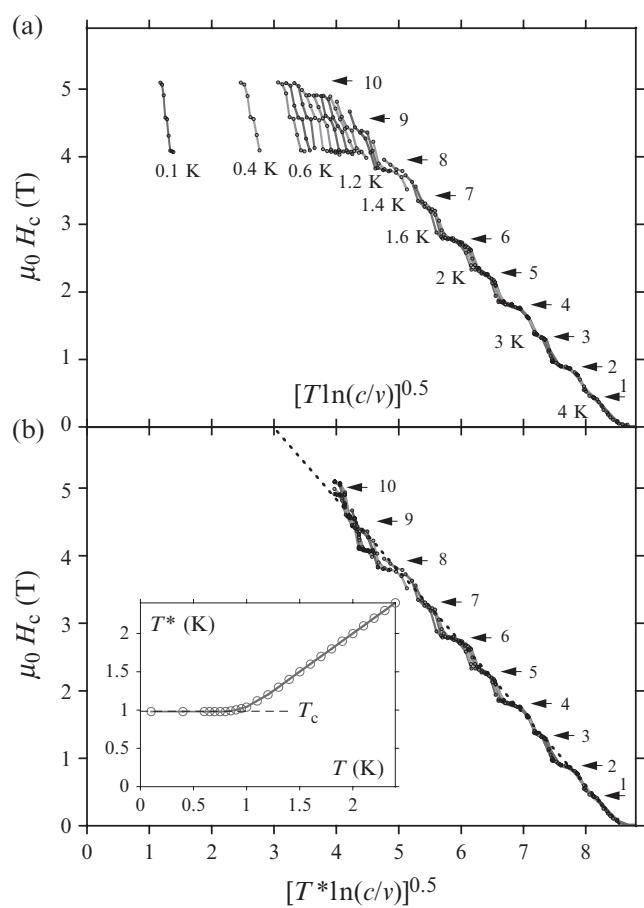
At low temperatures, strong deviations from the master curve are observed (Fig. 4.13(a)). In order to investigate the possibility that these low-temperature



**Fig. 4.11** Hysteresis loops of single crystals of a Mn<sub>12</sub> molecular cluster at different temperatures and a constant field sweep rate indicated in the figure. The loops display a series of steps, separated by plateaus. As the temperature is lowered, there is a decrease in the transition rate due to reduced thermal-assisted tunnelling. The hysteresis loops become temperature independent below 0.6 K, demonstrating quantum tunnelling at the lowest energy levels (reprinted from Wernsdorfer, W., Murugesu, M., Tasiopoulos, A.J., Christou, G. (2005) *Phys. Rev. B* **72**, 212406, © 2005 American Physical Society).



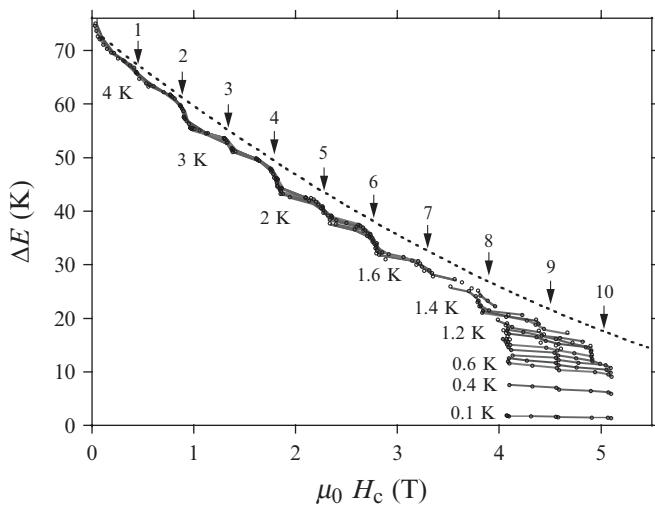
**Fig. 4.12** Coercive field  $H_c$  for  $\text{Mn}_{12}$  as a function of temperature. Note the steps of  $H_c$  coming from the resonant tunnelling steps in the hysteresis loops 11 (reprinted from Wernsdorfer, W., Murugesu, M., Tasiopoulos, A.J., Christou, G. (2005) *Phys. Rev. B* **72**, 212406, © 2005 American Physical Society).



**Fig. 4.13** (a) Scaling plot of the coercive field  $H_c(T, v)$  of  $\text{Mn}_{12}$  for field sweep rates between 0.0001 and 0.1 T/s and several temperatures: 0.1 K, 0.4 K, from 0.6 to 1 K in steps of 0.05 K, and from 1 to 4 K in steps of 0.1 K. The arrows indicate the step index  $n = -(m + m')$ , where  $m$  and  $m'$  are the quantum numbers of the corresponding level crossing. Note the parity effect of the steps: even  $n$  have larger steps than odd  $n$ . (b) Same data of  $H_c(T, v)$  and the same scales but the real temperature  $T$  is replaced by an effective temperature  $T^*$  (see inset), which restores the scaling below 1.1 K (reprinted from Wernsdorfer, W., Murugesu, M., Tasiopoulos, A.J., Christou, G. (2005) *Phys. Rev. B* **72**, 212406, © 2005 American Physical Society).

deviations are due to escape from the metastable potential well by tunnelling, a common method for classical models is to replace the real temperature  $T$  by an effective temperature  $T^*(T)$  in order to restore the scaling plot (Wernsdorfer *et al.* 1997b). In the case of tunnelling,  $T^*(T)$  should saturate at low temperatures. Indeed, the variation of  $T^*(T)$  as shown in the inset of Fig. 4.13(b), can restore unequivocally the scaling plot demonstrated by a straight master curve (Fig. 4.13(b)). The flattening of  $T^*$  corresponds to a saturation of the escape rate, which is a necessary signature of tunnelling. The cross-over temperature  $T_c$  can be defined as the temperature where the quantum rate equals the thermal one. The inset of Fig. 4.13(b) gives  $T_c = 0.97$  K. The slopes and the intercepts of the master curves give  $E_0 = 72.4$  K and  $H_c^0 = 9.1$  T. The  $E_0$  values are in good agreement with those extracted from Arrhenius plots (69 K). This result allows us to estimate the spin ground state using  $S = 2E_0/(g\mu_B\mu_0 H_c^0)$ :  $S = 11$ . This differs slightly from the  $S = 10$  determined via magnetization measurements. This deviation is due to quantum effects in the thermally activated regime and is considered further below.

Several points should be mentioned: (i) the classical regime of the model corresponds in most SMMs to the thermally activated tunnelling regime with tunnelling close to the top of the energy barrier. Because all parameters are deduced from this regime, small deviations from the exact values are expected; (ii) the field dependence of the energy barrier can be obtained directly using eqn (4.7) and is plotted in Fig. 4.14; (iii) eqn (4.7) is not valid for fields that are close to  $H = 0$  because the model only takes into account the transitions from the metastable to the stable well. However, close to  $H = 0$ , transitions between both wells are possible, leading to a rounding of the master curve at small fields; (iv) the method can be applied to powder samples with random orientations of the molecules. In this case,  $\alpha \approx 1.5$ ,  $vE_0 = DS^2$ , where  $v$  can be calculated (Thiaville 1998, 2000), and the intercept of the master curve gives  $H_c^0/2$ ; (v) in the case of a distribution of anisotropies, different parts of



**Fig. 4.14** Field dependence of the energy barrier of  $Mn_{12}$  obtained from eqn (4.7) and the set of  $H_c(T, v)$  data from Fig. 4.13. The arrows indicate the step index  $n = -(m + m')$ , where  $m$  and  $m'$  are the quantum numbers of the corresponding level crossing. Note the step-like reduction of the energy barrier due to resonant tunnelling and the parity effect of the steps: even  $n$  have larger steps than odd  $n$ . The dotted line gives the classical barrier  $\Delta E = E_0(1 - H/H_a)^2$  with  $E_0 = 74$  K and  $H_a = 9.8$  T (reprinted from Wernsdorfer, W., Murugesu, M., Tasiopoulos, A.J., Christou, G. (2005) *Phys. Rev. B* **72**, 212406, © 2005 American Physical Society).

the distribution can be probed by applying the method at different  $M$  values; (vi) this method is insensitive to small intermolecular interactions when  $H_c$  is larger than the typical interaction field; and (vii) the method can be generalized for 1D, 2D, and 3D networks of spins. In this case, eqn (4.6) describes a nucleation barrier.

#### 4.4 Quantum dynamics of a dimer of nanomagnets

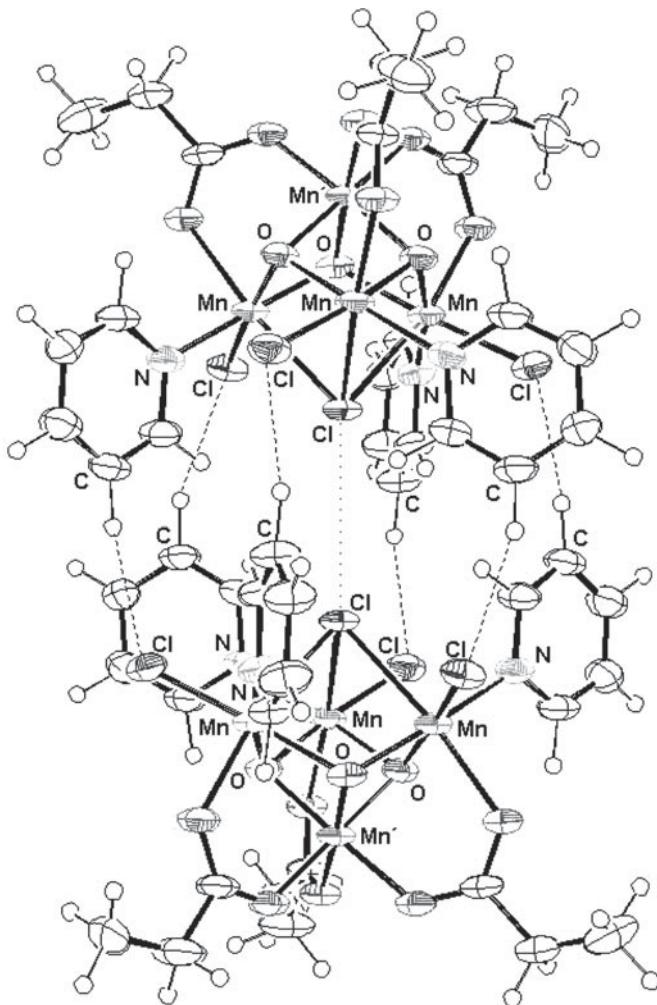
We present here a new family of dimers of nanomagnets (Wernsdorfer *et al.* 2002a) in which antiferromagnetic coupling between two single-molecule magnets (SMMs) results in quantum behavior different from that of the individual SMMs. Each SMM acts as a bias on its neighbor, shifting the quantum tunnelling resonances of the individual SMMs. Hysteresis-loop measurements on a single crystal of SMM dimers established quantum tunnelling of the magnetization via entangled states of the dimer. This shows that the dimer really does behave as a quantum-mechanically coupled dimer, and also allows the measurement of the longitudinal and transverse superexchange coupling constants (Tiron *et al.* 2003). The experimental evidence for entangled states was confirmed by an electron paramagnetic resonance (EPR) study (Hill *et al.* 2003).

The compound  $[\text{Mn}_4\text{O}_3\text{Cl}_4(\text{O}_2\text{CEt})_3(\text{py})_3]$  crystallizes in the hexagonal space group  $R\bar{3}(\text{bar})$  with two  $\text{Mn}_4$  molecules per unit cell lying head-to-head on a crystallographic  $S_6$  symmetry axis (Wernsdorfer *et al.* 2002a) (Fig. 4.15). Each  $\text{Mn}_4$  monomer has a ground-state spin of  $S = 9/2$ , well separated from the first excited state  $S = 7/2$  by a gap of about 300 K (Hendrickson *et al.* 1992). The Mn–Mn distances and the Mn–O–Mn angles are similar and the uniaxial anisotropy constant is expected to be the same for the two dimer systems. These dimers are held together via six C–H $\cdots$ Cl hydrogen bonds between the pyridine (py) rings on one molecule and the Cl ions on the other, and one Cl $\cdots$ Cl van der Waals interaction. These interactions lead to an antiferromagnetic superexchange interaction between the two  $\text{Mn}_4$  units of the  $[\text{Mn}_4]_2$  dimer (Wernsdorfer *et al.* 2002a). Dipolar couplings between  $\text{Mn}_4$  molecules can be easily calculated and are more than one order of magnitude smaller than the exchange interaction.

Before presenting the measurements, we summarize a simplified spin Hamiltonian describing the  $[\text{Mn}_4]_2$  dimer (Wernsdorfer *et al.* 2002a). Each  $\text{Mn}_4$  SMM can be modelled as a *giant spin* of  $S = 9/2$  with Ising-like anisotropy (eqn (4.1)). The corresponding Hamiltonian is given by

$$\mathcal{H}_i = -DS_{z,i}^2 + \mathcal{H}_{\text{trans},i} + g\mu_B\mu_0\vec{S}_i \cdot \vec{H}, \quad (4.8)$$

where  $i = 1$  or 2 (referring to the two  $\text{Mn}_4$  SMMs of the dimer),  $D$  is the uniaxial anisotropy constant, and the other symbols have their usual meaning. Tunnelling is allowed in these half-integer ( $S = 9/2$ ) spin systems because of a small transverse anisotropy  $\mathcal{H}_{\text{trans},i}$  containing  $S_{x,i}$  and  $S_{y,i}$  spin operators and transverse fields ( $H_x$  and  $H_y$ ). The exact form of  $\mathcal{H}_{\text{trans},i}$  is not important in this discussion. The last term in eqn (4.8) is the Zeeman energy



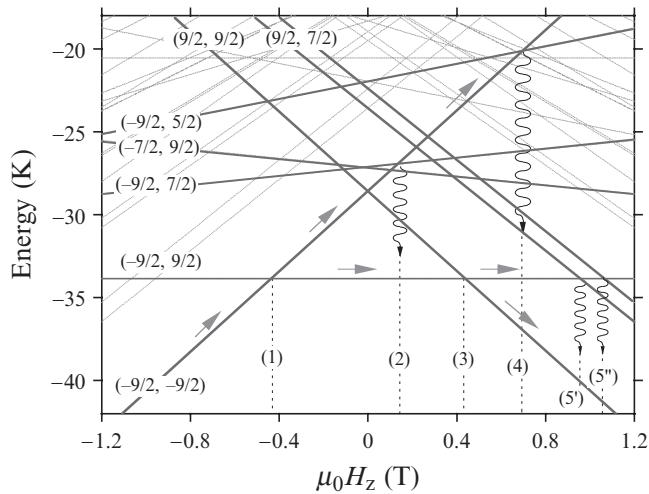
**Fig. 4.15** The structure of the  $[Mn_4]_2$  dimer of  $[Mn_4O_3Cl_4(O_2CEt)_3(py)_3]$ . The small circles are hydrogen atoms. The dashed lines are  $C-H \cdots Cl$  hydrogen bonds and the dotted line is the close  $Cl \cdots Cl$  approach. The labels Mn and Mn' refer to  $Mn^{III}$  and  $Mn^{IV}$  ions, respectively (reprinted from Wernsdorfer, W., Aliaga-Alcalde, N., Hendrickson, D., Christou, G. (2002) *Nature* **416**, 406, © 2002 Nature Publishing Group).

associated with an applied field. The  $Mn_4$  units within the  $[Mn_4]_2$  dimer are coupled by a weak superexchange interaction via both the six  $C-H \cdots Cl$  pathways and the  $Cl \cdots Cl$  approach. Thus, the Hamiltonian ( $\mathcal{H}$ ) for  $[Mn_4]_2$  is

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + J_z S_{z,1} S_{z,2} + J_{xy} (S_{x,1} S_{x,2} + S_{y,1} S_{y,2}), \quad (4.9)$$

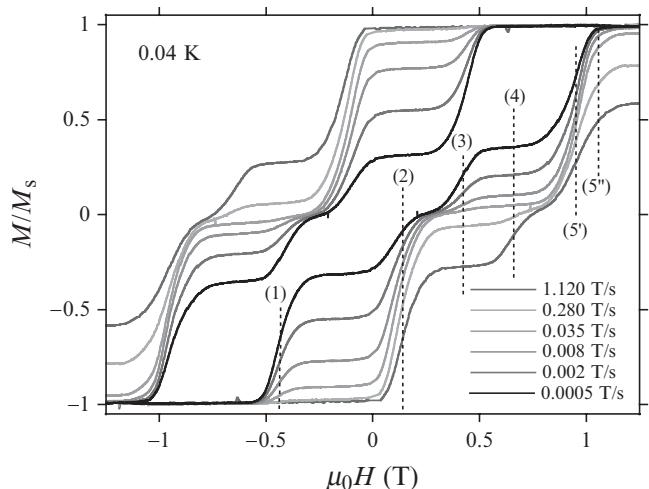
where  $J_z$  and  $J_{xy}$  are, respectively, the longitudinal and transverse superexchange interactions.  $J_z = J_{xy}$  is the case of isotropic superexchange. The  $(2S+1)^2 = 100$  energy states of the dimer can be calculated by exact numerical diagonalization and are plotted in Fig. 4.16 as a function of applied field along the easy axis. Each state of  $[Mn_4]_2$  can be labelled by two quantum numbers  $(M_1, M_2)$  for the two  $Mn_4$  SMMs, with  $M_1 = -9/2, -7/2, \dots, 9/2$  and  $M_2 = -9/2, -7/2, \dots, 9/2$ . The degeneracy of some of the  $(M_1, M_2)$  states is lifted by transverse anisotropy terms. For the sake of simplicity, we will discuss mainly the effect of the transverse superexchange interaction

**Fig. 4.16** Low-lying spin-state energies of the  $[\text{Mn}_4]_2$  dimer, calculated by exact numerical diagonalization using eqn (4.9) with  $D = 0.77 \text{ K}$  and  $J = 0.13 \text{ K}$ , as a function of applied magnetic field  $H_z$  (Zeeman diagram). The bold energy levels are labelled with two quantum numbers  $(M_1, M_2)$ . Dotted lines, labelled **1** to **5**, indicate the strongest tunnel resonances: **1**:  $(-9/2, -9/2)$  to  $(-9/2, 9/2)$ ; **2**:  $(-9/2, -9/2)$  to  $(-9/2, 7/2)$ , followed by relaxation to  $(-9/2, 9/2)$ ; **3**:  $(-9/2, 9/2)$  to  $(9/2, 9/2)$ ; **4**:  $(-9/2, -9/2)$  to  $(-9/2, 5/2)$ , followed by relaxation to  $(-9/2, 9/2)$ ; **5**:  $(-9/2, 9/2)$  to  $(7/2, 9/2)$ , followed by relaxation to  $(9/2, 9/2)$ . For clarity, degenerate states such as  $(M, M')$  and  $(M', M)$  and lifted degenerate states such as  $(M, M \pm 1)$ ,  $(M, M \pm 2)$ ... are not both listed. For example, the  $(9/2, 7/2)$  and  $(7/2, 9/2)$  states are strongly split into a symmetric (labelled  $5''$ ) and antisymmetric (labelled  $5'$ ) combination of  $(9/2, 7/2)$  and  $(7/2, 9/2)$  states. This splitting is used to measure the transverse superexchange interaction constant  $J_{xy}$ . Cotunneling and other two-body tunnel transitions have a lower probability of occurrence and are neglected (reprinted from Tiron, R., Wernsdorfer, W., Foguet-Albiol, D., Aliaga-Alcalde, N., Christou, G. (2003) *Phys. Rev. Lett.* **91**, 227203, © 2003 American Physical Society).



$\mathcal{J}_{\text{trans}} = J_{xy}(S_{x,1}S_{x,2} + S_{y,1}S_{y,2}) = J_{xy}(S_{+,1}S_{-,2} + S_{-,1}S_{+,2})/2$ , where  $S_{+,i}$  and  $S_{-,i}$  are the usual spin-raising and -lowering operators. Because  $\mathcal{J}_{\text{trans}}$  acts on  $(M, M \pm 1)$  states in the first order of perturbation theory, the degeneracy of those states is strongly lifted. For example, the  $(9/2, 7/2)$  and  $(7/2, 9/2)$  states are strongly split into a symmetric (labelled  $5''$ ) and antisymmetric (labelled  $5'$ ) combination of  $(9/2, 7/2)$  and  $(7/2, 9/2)$  states. Similarly for the  $(-9/2, -7/2)$  and  $(-7/2, -9/2)$  states. Measuring this energy splitting allows us to determine the transverse superexchange interaction constant  $J_{xy}$  because the latter is proportional to the former.

Figure 4.17 shows typical hysteresis loops (magnetization versus magnetic field scans) with the field applied along the easy axis of magnetization of  $[\text{Mn}_4]_2$ , that is, parallel to the  $S_6$  axis. These loops display step-like features separated by plateaus. The step heights are temperature independent below  $\sim 0.35 \text{ K}$  (Wernsdorfer *et al.* 2002a). The steps are due to resonant



**Fig. 4.17** Hysteresis loops for the  $[\text{Mn}_4]_2$  dimer at several field sweep rates and  $40 \text{ mK}$ . The tunnel transitions (manifested by steps) are labelled from **1** to **5**, see Fig. 4.1 (reprinted from Tiron, R., Wernsdorfer, W., Foguet-Albiol, D., Aliaga-Alcalde, N., Christou, G. (2003) *Phys. Rev. Lett.* **91**, 227203, © 2003 American Physical Society).

quantum tunnelling of the magnetization (QTM) between the energy states of the [Mn<sub>4</sub>]<sub>2</sub> dimer (see captions of Figs. 4.16 and 4.17 for a discussion of 5 tunnel transitions). QTM has been previously observed for most SMMs, but the novelty for [Mn<sub>4</sub>]<sub>2</sub> dimers is that the QTM is now the collective behavior of the complete  $S = 0$  dimer of exchange-coupled  $S = 9/2$  Mn<sub>4</sub> quantum systems. This coupling is manifested as an exchange bias of all tunnelling transitions, and the resulting hysteresis loop consequently displays unique features, such as the absence for the first time in a SMM of a QTM step at zero field (Wernsdorfer *et al.* 2002a).

Even though the five strongest tunnelling transitions are observed in Fig. 4.17, fine structure was not observed. For example, the hysteresis loops do not show the splitting of the (9/2, 7/2) states (labelled 5' and 5''), which we suspected might be due to line broadening. Usually, line broadening in SMMs is caused by dipolar and hyperfine interactions (Prokof'ev and Stamp 1998), and distributions of anisotropy and exchange parameters. In most SMMs, the zero-field resonance is mainly broadened by dipolar and hyperfine interactions because distributions of anisotropy parameters do not affect the zero-field resonance. For an antiferromagnetically coupled dimer, however, this resonance is shifted to negative fields. Therefore, a distribution of the exchange coupling parameter  $J_z$  can further broaden this resonance. In fact, we showed that the latter is the dominant source of broadening (Tiron *et al.* 2003). We used the “quantum hole-digging” method (see Section 4.7.2) (Prokof'ev and Stamp 1998; Wernsdorfer *et al.* 1999, 2000; Alonso and Fernandez 2001; Tupitsyn *et al.* 2004) to provide direct experimental evidence for the transitions 5' and 5'', which established tunnelling involving entangled dimer states and allowed us to determine  $J_{xy}$  (Tiron *et al.* 2003).

## 4.5 Resonant photon absorption in Cr<sub>7</sub>Ni antiferromagnetic rings

Magnetic molecules are currently considered among the most promising electron-spin-based quantum systems for the storing and processing of quantum information. For this purpose, ferromagnetic (Leuenberger and Loss 2001) and antiferromagnetic (Meier *et al.* 2003a,b) systems have attracted increasing interest (Troiani *et al.* 2005a,b). In the latter case the quantum hardware is thought of as a collection of coupled molecules, each corresponding to a different qubit. The main advantages would arise from the fact that they are extremely small and almost identical, allowing, in a single measurement, statistical averages of a large number of qubits to be obtained. The magnetic properties can be modelled with an outstanding degree of accuracy. And most importantly, the desired physical properties can be engineered chemically.

The suitability of Cr-based antiferromagnetic molecular rings for the qubit implementation has been proposed (Troiani *et al.* 2005a,b). The substitution of one metal ion in a Cr-based molecular ring with dominant antiferromagnetic couplings allows the engineering of the energy-level structure and the ground-state degeneracy (Overgaard *et al.* 2002; Larsen *et al.* 2003). A Cr<sub>7</sub>Ni

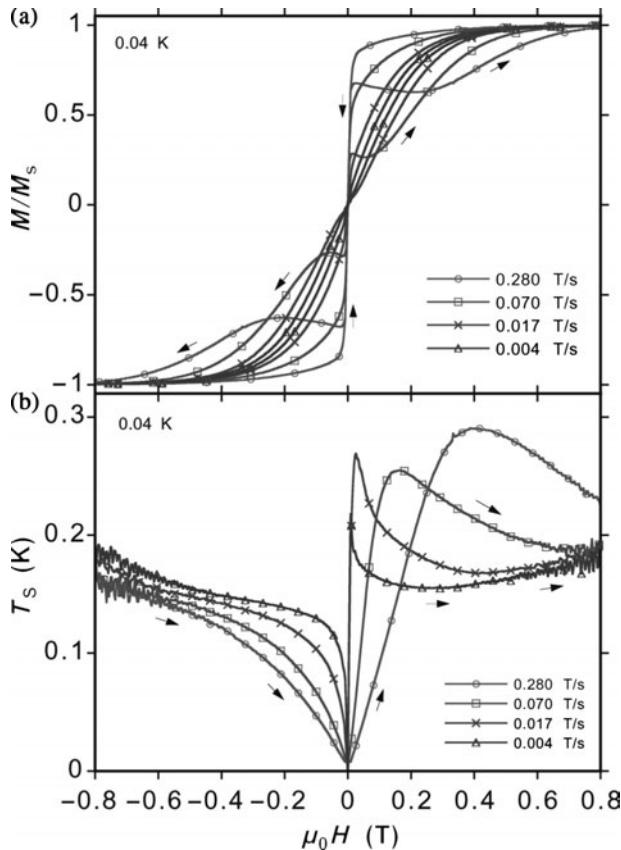
molecular ring was characterized by means of low-temperature specific-heat and torque-magnetometry measurements, thus determining the microscopic parameters of the corresponding spin Hamiltonian. The energy spectrum and the suppression of the leakage-inducing S-mixing render the Cr<sub>7</sub>Ni molecule a suitable candidate for the qubit implementation (Carretta *et al.* 2005; Troiani *et al.* 2005a,b).

In this section we report the first microsuperconducting quantum interference device (micro-SQUID) (Wernsdorfer *et al.* 2004) studies of the Cr<sub>7</sub>Ni molecular ring (Wernsdorfer *et al.* 2005b). Electron paramagnetic resonance (EPR) methods are combined with high-sensitivity magnetization measurements. We found very narrow resonant photon absorption lines that are mainly broadened by hyperfine interactions. Similar measurements were performed on Ni<sub>4</sub> molecules (del Barco *et al.* 2004). Quantum coherence has recently been observed (Ardavan *et al.* 2007).

The Cr<sub>7</sub>Ni molecular ring, is based on a ring with formula [Cr<sub>8</sub>F<sub>8</sub>(O<sub>2</sub>CCMe<sub>3</sub>)<sub>16</sub>]. The eight chromium(III) ions lie at the corners of a regular octagon (Overgaard *et al.* 2002). Each edge of the octagon is bridged by one fluoride ion and two pivalate ligands. There is a large cavity at the center of the ring. If a single chromium(III) ion is replaced by a metal(II) ion, for example nickel(II), this makes the ring anionic and a cation can be incorporated in the cavity. Thus, we can make [H<sub>2</sub>NMe<sub>2</sub>][Cr<sub>7</sub>NiF<sub>8</sub>(O<sub>2</sub>CCMe<sub>3</sub>)<sub>16</sub>] (Larsen *et al.* 2003). The Cr<sub>8</sub> and Cr<sub>7</sub>Ni compounds are found to be isostructural and crystallize in the tetragonal space group, *P4*.

The measurements were made in a dilution cryostat using a 20-μm-sized single crystal of Cr<sub>7</sub>Ni. The magnetic probe was a micro-SQUID array (Wernsdorfer 2001, 2004) equipped with three coils allowing application of a field in any direction and with sweep rates up to 10 T/s. The electromagnetic radiation was generated by a frequency synthesizer triggered with a nanosecond pulse generator. This setup allowed a continuous variation of the frequency from 0.1 Hz to 20 GHz, with pulse lengths ∼1 ns to continuous radiation (Thirion *et al.* 2003). Using a 50-μm-sized gold radio-frequency (RF) loop, the RF radiation field was directed in a plane perpendicular to the applied static field  $\mu_0 H$ . The microwave power of the generator could be varied from −80 to 20 dBm ( $10^{-11}$  to  $10^{-1}$  W). The sample absorbs only a small fraction of the generator power. This fraction is, however, proportional to the microwave power of the generator. The microwave amplitude  $B_{\text{RF}}$  can be estimated with the method described by Wernsdorfer *et al.* (2004). We found  $B_{\text{RF}} \approx 1$  mT at 4 GHz and 15 dBm, which is more than 1000 times larger than in our previous work on V<sub>15</sub> (Wernsdorfer *et al.* 2004).

Figure 4.18(a) shows magnetization versus applied field curves for several field sweep rates at a cryostat temperature of 0.04 K. The magnetization loops exhibit a clear hysteresis that is characteristic of the phonon-bottleneck regime with a spin-phonon relaxation time to the cryostat heat bath of a few seconds (Chiorescu *et al.* 2000b). Note that the degeneracy of the Kramers doublet is lifted due to internal transverse fields (mainly the transverse hyperfine fields). In order to quantify the out-of-equilibrium effect, Fig. 4.18(b) presents the same data as in Fig. 4.18(a) but the magnetization  $M$  is converted into a spin



**Fig. 4.18** (a) Magnetization ( $M$ ) hysteresis loops for several field-sweep rates at a cryostat temperature of 0.04 K. The loops are normalized by the saturation magnetization  $M_s$  at 1.5 T. (b) Spin temperature  $T_S$  for field sweeps from negative to positive fields, obtained by inversion of eqn (4.10), where  $M(T_S)$  are the data in (a) (reprinted from Wernsdorfer, W., Mailly, D., Timco, G.A., Winpenny, R.E.P. (2005) *Phys. Rev. B* **72**, 060409(R), © 2005 American Physical Society).

temperature  $T_S$  using the equation (Abragam and Bleaney 1970):

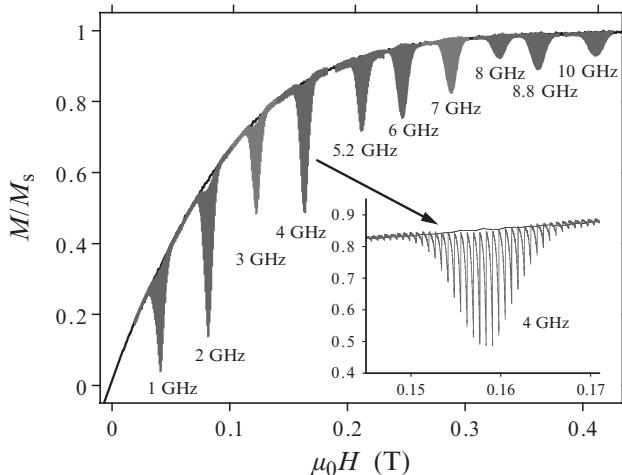
$$M(T_S)/M_s = \tanh(g\mu_B S\mu_0 H/k_B T_S), \quad (4.10)$$

with  $S = 1/2$  and  $g = 2.1$  (Larsen *et al.* 2003). Figure 4.18(b) shows clearly a strong adiabatic cooling when sweeping the field down to zero field. Note that this cooling mechanism might be used before qubit operations to reach extremely low temperatures even at relatively high cryostat temperatures. High-frequency noise from the RF loop around the sample leads to spin temperatures at 1 T being higher than the cryostat temperature.

Figure 4.19 shows magnetization curves  $M(H)$  in the quasi-static regime with a field-sweep rate slow enough (0.14 mT/s) to keep the system at equilibrium. During the field sweep, RF pulses were applied to the sample with a pulse length of 1  $\mu$ s and a period of 4 s between each pulse. Depending on the RF frequency, clear dips are observed that result from resonant absorptions of photons associated with spin transitions between the quantum numbers  $m_s = 1/2$  and  $-1/2$ . After each pulse, the magnetization relaxes back to the equilibrium magnetization (see the fine structure in the inset of Fig. 4.19).

Typical relaxation measurements at a constant applied field after RF pulses of different durations are shown in Fig. 4.20. The relaxation is exponential

**Fig. 4.19** Magnetization curves measured with and without irradiation. The cryostat temperature was 40 mK and the field-sweep rate of 0.14 mT/s was slow in order to keep the system at equilibrium. The electromagnetic radiation was pulsed with a period of 4 s and a pulse length of 1  $\mu$ s. The RF frequencies are indicated and the RF amplitude is slightly frequency dependent. Inset: Enlargement of the 4-GHz resonance. The fine structure is due to the RF pulses (reprinted from Wernsdorfer, W., Mailly, D., Timco, G.A., Winpenny, R.E.P. (2005) *Phys. Rev. B* **72**, 060409(R), © 2005 American Physical Society).

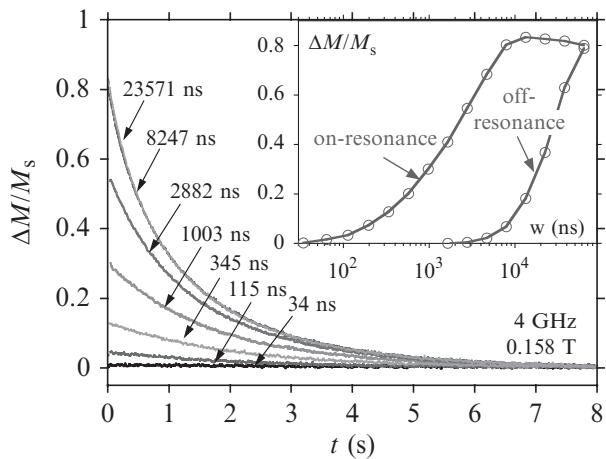


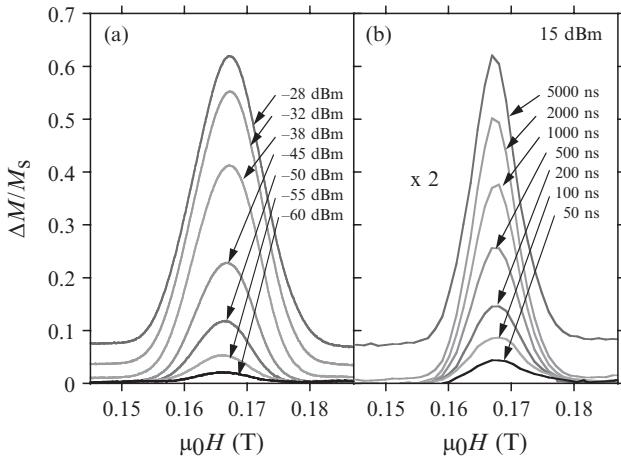
with the rate independent of the pulse length. Detailed studies showed that the relaxation rate is dominated by the phonon-bottleneck regime, that is the spin-phonon relaxation time to the cryostat heat bath.

The inset of Fig. 4.20 presents the change of magnetization  $\Delta M$  between the magnetization before and after the pulse as a function of the pulse length  $w$ .  $\Delta M$  increases linearly with  $w$  for short pulses of a few tens of ns. It saturates for  $w \approx 10 \mu$ s and decreases for very long pulses because of cryostat heating effects. Non-resonant photon absorption is also observed for very long pulses.

The resonant photon absorption lines are often used to estimate a lower bound of the decoherence time of a qubit. We therefore investigated in more detail the linewidth observed in Fig. 4.19. Figure 4.21(a) presents a typical power dependence of the linewidth for continuous irradiation at 4.2 GHz. Resonant photon absorption is clearly visible for a generator power

**Fig. 4.20** Relaxation of magnetization after a RF pulse of 4 GHz. The pulse lengths  $w$  are indicated. Inset: magnetization variation  $\Delta M$  after a RF pulse versus the pulse length  $w$  for an on-resonance field (0.1582 T) and off-resonance field (0.1722 T) (reprinted from Wernsdorfer, W., Mailly, D., Timco, G.A., Winpenny, R.E.P. (2005) *Phys. Rev. B* **72**, 060409(R), © 2005 American Physical Society).





**Fig. 4.21** (a) Magnetization variation  $\Delta M$  between the equilibrium curves measured without and with continuous irradiation. The microwave frequency was 4.2 GHz. The microwave powers of the generator are indicated. (b) Magnetization variation  $\Delta M$  after a RF pulse of 4.2 GHz and several pulse lengths. The cryostat temperature was 40 mK.  $\Delta M$  is multiplied by a factor two (reprinted from Wernsdorfer, W., Mailly, D., Timco, G.A., Winpenny, R.E.P. (2005) *Phys. Rev. B* **72**, 060409(R), © 2005 American Physical Society).

larger than  $-60$  dBm ( $1$  nW). The line saturated at about  $-20$  dBm ( $10$   $\mu$ W). Figure 4.21(b) presents the absorption line for the pulsed technique (see Fig. 4.19) for several pulse lengths and a generator power of  $15$  dBm ( $32$  mW,  $B_{RF} \approx 1$  mT). The resonant photon absorption is clearly visible for pulse lengths longer than  $10$  ns. Note that the linewidths in Fig. 4.21(a) are nearly twice as large as those in Fig. 4.21(b).

In our case of an assembly of identical spins, the line broadening is mainly due to dipolar and hyperfine interactions. The dipolar coupling energy can be estimated with  $E_{\text{dip}}/k_B \approx (g\mu_B S)^2/V \approx 0.1$  mK ( $S = 1/2$  and  $V = 6.3$  nm<sup>3</sup>) (Troiani *et al.* 2005b). The hyperfine coupling with the nuclear spins can be obtained by considering the dipolar interaction of one Cr ion ( $s = 3/2$ ) with the neighboring F nucleus having a nuclear spin  $I = 1/2$ . With  $g_F = +5.26$  and the distance of  $d = 0.2$  nm between F and Cr ions, the interaction energy is about  $0.4$  mK for each of the eight F nuclear spins (Troiani *et al.* 2005b). The hyperfine line broadening of all eight F nuclear spins is about  $3$  mK, which corresponds to  $5$  mT, in good agreement with the observed Gaussian linewidths of about  $\sigma = 4$  mT of Figs. 4.19 and 4.21.

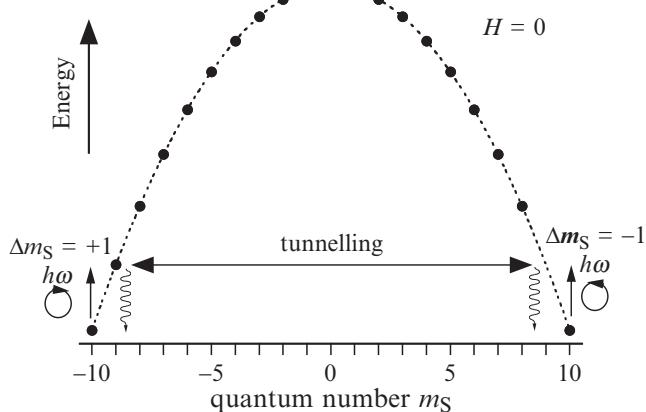
Finally, we discuss the possibility of observing Rabi oscillations with the present set-up. Due to inhomogeneous broadening only a lower bound of the coherence time  $\tau_c$  can be estimated from the resonance lines in Fig. 4.21:  $\tau_c \approx B_v/(\sigma\nu) \approx 10$  ns with  $B_v = 0.166$  T,  $\sigma = 4$  mT, and  $\nu = 4.2$  GHz. The corresponding number of coherent flips of the spin system is given by  $N = \tau_c/\tau_{\text{Rabi}}$  with  $\tau_{\text{Rabi}} = 2\pi/(\gamma B_{RF}) \approx 40$  ns for  $B_{RF} \approx 1$  mT. We obtain  $N \approx 0.25$  showing that there is no hope to see Rabi oscillations in the present conditions. In order to get  $N \gg 1$ , it will be necessary to increase further the radiation field  $B_{RF}$ , to reduce substantially the hyperfine broadening by substituting the F ions with OH groups, and to minimize the dipolar coupling by doping the crystal of Cr<sub>7</sub>Ni molecules with Cr<sub>8</sub> molecules.

## 4.6 Photon-assisted tunnelling in single-molecule magnet

It has also been proposed that molecular nanomagnets could be used as quantum computers by implementing Grover's algorithm (Leuenberger and Loss 2001). For this to occur it is necessary to be able to generate an arbitrary superposition of eigenstates of these systems. The suggested way to do this was through the use of multifrequency coherent magnetic radiation in the microwave and radio-frequency range. This would first introduce and amplify the desired phase for each  $m_S$  state and this information could be finally read-out by standard magnetic resonance techniques. In this approach advantage is taken of the non-equidistance of the  $m_S$  levels of the ground multiplet arising from the large axial anisotropy of these systems, which allows coherent populations of the different  $m_S$  levels. A theoretical study pointed out that a very accurate control of a pulse-shape technique, both in amplitude, duration and choice of frequency is needed to fulfill the condition to design quantum-computing devices in molecular nanomagnets (Zhou *et al.* 2002). In addition to such basic difficulties, we will see below that the total microwave power conveyed onto the samples cannot exceed a critical value above which nonlinear effects occur.

In order to investigate the feasibility of the proposed process, any preliminary experiment should aim to understand the effects of microwave absorption on the spin dynamics of these systems at low temperature. The measurements were performed using micro-Hall bars (Sorace *et al.* 2003). Continuous microwave radiation was generated by a couple of Gunn diodes equipped with calibrated attenuators. With this experimental setup, which has a very good stability of emitted power and a narrow bandwidth (100 kHz), only relative powers at the output of the diode can be known, which will be referred to as  $\rho$  in the following. We irradiated the sample using a 6-mm waveguide equipped with infra-red filters in order to reduce heating. The circular polarization was maximized to around 97%. The study was performed on a 0.1-mm Fe<sub>8</sub> single crystal.

As schematically depicted in Fig. 4.22, microwave radiation with a frequency of 115 GHz corresponds to the energy separation between the ground states  $m_S = \pm S$  and the first excited states  $m_S = \pm(S - 1)$  of Fe<sub>8</sub> in zero applied magnetic field (Barra *et al.* 1996; Hill *et al.* 2002). If the radiation is linearly polarized, the populations of the first excited states ( $m_S = \pm(S - 1)$ ) in both wells will be enhanced equally (equal transition probability for  $\Delta m_S = \pm 1$ ). On the contrary, the use of circular polarization has the advantage to distinguish between  $\Delta m_S = +1$  (left polarization,  $\sigma^-$  photons) or  $\Delta m_S = -1$  (right polarization,  $\sigma^+$  photons) (Abragam and Bleaney 1970), and the population of only one of the two excited states will be enhanced (Fig. 4.22). An excess of tunnelling from one well to the other is then expected. Therefore, circular polarization can help to distinguish between spin-phonon relaxation, and spin-phonon relaxation modified by the absorption of photons. The first equally affects the two sides of the barrier, i.e. the two branches of the hysteresis loop, while the second modifies the population of only one side of



**Fig. 4.22** Schematic representation of photon-assisted tunnelling. On irradiating a Fe<sub>8</sub> sample with a radiation of wavelength corresponding to the  $m_S = -10$  to  $-9$  splitting (vertical arrow), an enhancement of the fraction of molecules that tunnel from the first excited state is expected (horizontal arrow). The use of circularly polarized radiation allows selecting only one side of the well and distinguishing between spin-phonon and spin-photon transitions (reprinted from Sorace, L., Wernsdorfer, W., Thirion, C., Barra, A.-L., Pacchioni, M., Mailly, D., Barbara, B. (2003) *Phys. Rev. B* **68**, 220407(R), © 2003 American Physical Society).

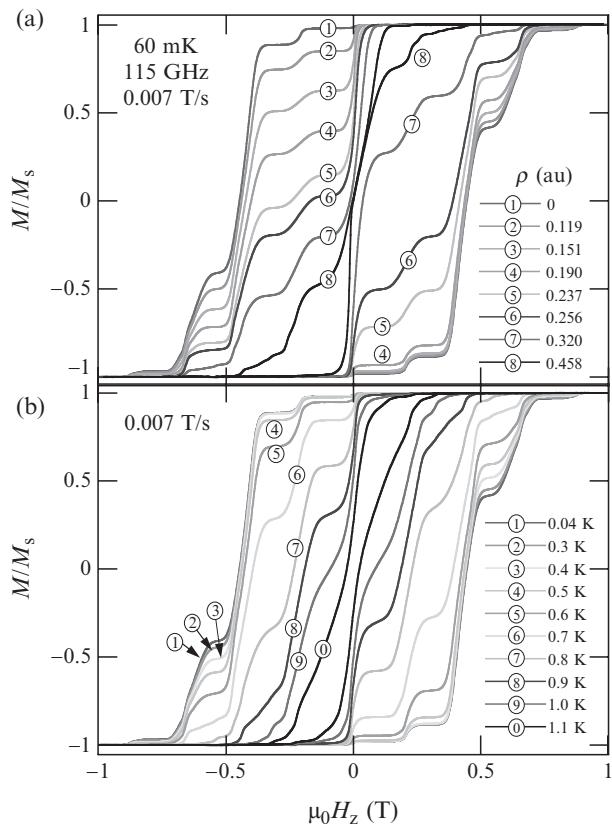
the barrier, i.e. one branch of the hysteresis loop. Any difference observed between the two branches of the hysteresis loop, has to be traced back to photon absorption.

Figure 4.23(a) shows the hysteresis loops of a Fe<sub>8</sub> single crystal with the easy axis parallel to the applied field, measured at 60 mK under irradiation. The tunnelling transition near zero field is strongly enhanced for a radiation at 115 GHz. This is in agreement with a photon-induced population transfer from  $m_S = 10$  to  $m_S = 9$ , and agrees with earlier HF-EPR studies showing strong zero-field absorption at about 116 GHz (Barra *et al.* 1996; Hill *et al.* 2002). Figure 4.23(a) also shows the expected asymmetry of the hysteresis loops in the presence of circularly polarized radiation. In particular, the height of the zero-field step (first tunnel resonance,  $n = 0$ ), obtained when sweeping the field from negative saturation, is much less affected than when sweeping from positive saturation. This behavior is completely different from phonon-assisted tunnelling (Fig. 4.23(b)) and clearly establishes that tunnelling is assisted by photons for the matching frequency of 115 GHz. The observation of a more symmetric shape of the hysteresis curve at high microwave power can be explained both by the incomplete microwave polarization and by phonon emission, leading to relaxation on both sides of the barrier.

A detailed study showed that at the lowest power, the tunnel probability increases linearly with power, whereas at higher powers a strongly non-linear regime is observed (Sorace *et al.* 2003). The latter might be due to multispin and coherent photon transitions.

## 4.7 Environmental decoherence effects in nanomagnets

At temperatures below 0.36 K, Fe<sub>8</sub> molecular clusters display a clear cross-over from thermally activated relaxation to a temperature-independent quantum regime, with a pronounced resonance structure of the relaxation time as a function of the external field (Section 4.4.3). It was surprising, however, that the observed relaxation of the magnetization in the quantum regime was



**Fig. 4.23** Magnetic hysteresis loops of Fe<sub>8</sub> at a field sweep rate of 0.007 T/s and at 60 mK under irradiation with microwaves at 115 GHz and for several microwave powers  $\rho$  (a). The easy axis of the crystal is oriented along the applied field and perpendicular to the radiation oscillating magnetic field. The observed increasing of the tunnelling rate at zero field, as a consequence of the absorption of photons induced by circularly polarized radiation, becomes evident by comparing the zero-field steps after positive or negative saturation. For comparison, the thermal behavior is presented in (b) (reprinted from Sorace, L., Wernsdorfer, W., Thirion, C., Barra, A.-L., Pacchioni, M., Mailly, D., Barbara, B. (2003) *Phys. Rev. B* **68**, 220407(R), © 2003 American Physical Society).

found to be non-exponential and the resonance width orders of magnitude too large (Sangregorio *et al.* 1997; Ohm *et al.* 1998a). The key to understanding this seemingly anomalous behavior involves the hyperfine fields as well as the evolving distribution of the weak dipole fields of the nanomagnets themselves (Prokof'ev and Stamp 1998). Both effects were shown to be the main source of decoherence at very low temperature. At higher temperatures, phonons are another source of decoherence.

In the following sections, we focus on the low-temperature and low-field limits, where phonon-mediated relaxation is astronomically long and can be neglected. In this limit, the  $m = \pm S$  spin states are coupled due to the tunnelling splitting  $\Delta_{\pm S}$ , which is about  $10^{-7}$  K for Fe<sub>8</sub> (Section 4.3.2) with  $S = 10$ . In order to tunnel between these states, the longitudinal magnetic energy bias  $\xi = g\mu_B S H_{\text{local}}$  due to the local magnetic field  $H_{\text{local}}$  on a molecule must be smaller than  $\Delta_{\pm S}$ , implying a local field smaller than  $10^{-8}$  T for Fe<sub>8</sub> clusters. Since the typical intermolecular dipole fields are of the order of 0.05 T, it seems at first that almost all molecules should be blocked from tunnelling by a very large energy bias. Prokof'ev and Stamp have proposed a solution to this dilemma by proposing that fast dynamic nuclear fluctuations broaden the resonance, and the gradual adjustment of the dipole fields in the sample caused by the tunnelling brings other molecules into resonance and

allows continuous relaxation (Prokof'ev and Stamp 1998). Some interesting predictions are briefly reviewed in the following section.

### 4.7.1 Prokof'ev–Stamp theory

Prokof'ev and Stamp were the first who realized that there are localized couplings of environmental modes with mesoscopic systems that cannot be modelled with an “oscillator bath” model (Feynman and Vernon 1963) describing delocalized environmental modes such as electrons, phonons, photons, and so on. They found that these localized modes such as nuclear and paramagnetic spins are often strong and described them with a spin-bath model (Prokof'ev and Stamp 1996). We do not review this theory (Prokof'ev and Stamp 2000) but focus on one particular application that is interesting for molecular clusters (Prokof'ev and Stamp 1998). Prokof'ev and Stamp showed that at a given longitudinal applied field  $H_z$ , the magnetization of a crystal of molecular clusters should relax at short times with a square-root time dependence that is due to a gradual modification of the dipole fields in the sample caused by the tunnelling

$$M(H_z, t) = M_{\text{in}} + (M_{\text{eq}}(H_z) - M_{\text{in}}) \sqrt{\Gamma_{\text{sqrt}}(H_z)t}. \quad (4.11)$$

Here,  $M_{\text{in}}$  is the initial magnetization at time  $t = 0$  (after a rapid field change), and  $M_{\text{eq}}(H_z)$  is the equilibrium magnetization at  $H_z$ . The rate function  $\Gamma_{\text{sqrt}}(H_z)$  is proportional to the normalized distribution  $P(H_z)$  of molecules that are in resonance at  $H_z$

$$\Gamma_{\text{sqrt}}(H_z) = c \frac{\xi_0}{E_D} \frac{\Delta_{\pm S}^2}{4\hbar} P(H_z), \quad (4.12)$$

where  $\xi_0$  is the linewidth coming from the nuclear spins,  $E_D$  is the Gaussian half-width of  $P(H_z)$ , and  $c$  is a constant of the order of unity that depends on the sample shape. If these simple relations are exact, then measurements of the short time relaxation as a function of the applied field  $H_z$  give directly the distribution  $P(H_z)$ , and they allow one to measure the tunnel splitting  $\Delta_{\pm S}$ , which is described in the next section.

### 4.7.2 Hole-digging method to study dipolar distributions and hyperfine couplings

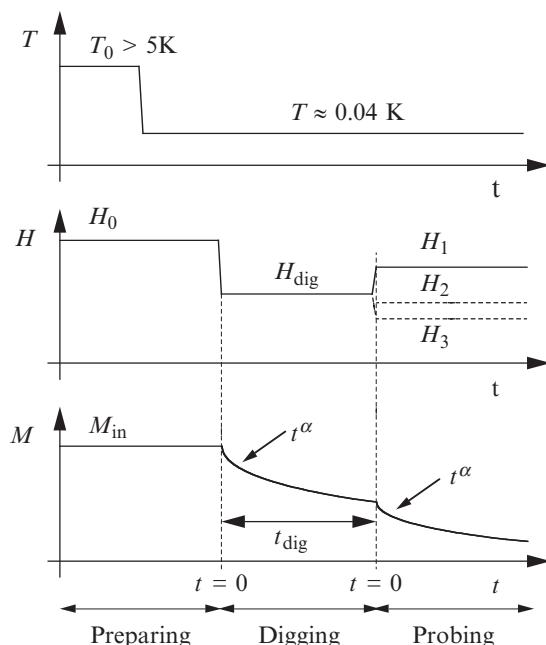
Motivated by the Prokof'ev–Stamp theory, we developed a new technique—which we call the *hole-digging method*—that can be used to observe the time evolution of molecular states in crystals of molecular clusters. It allowed us to measure the statistical distribution of magnetic bias fields in the Fe<sub>8</sub> system that arise from the weak dipole fields of the clusters themselves. A hole can be “dug” into the distribution by depleting the available spins at a given applied field. Our method is based on the simple idea that after a rapid field change, the resulting short time relaxation of the magnetization is directly related to the number of molecules that are in resonance at the given applied field. Prokof'ev

and Stamp have suggested that the short time relaxation should follow a  $\sqrt{t}$ -relaxation law (eqn (4.11)). However, the hole-digging method should work with any short time relaxation law—for example, a power law

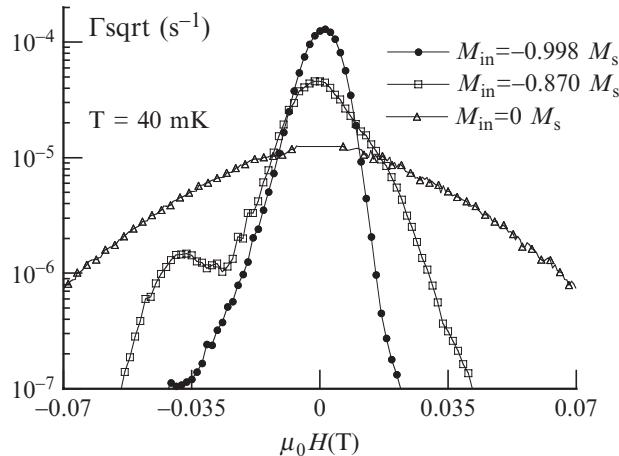
$$M(H_z, t) = M_{\text{in}} + (M_{\text{eq}}(H_z) - M_{\text{in}})(\Gamma_{\text{short}}(H_z)t)^{\alpha}, \quad (4.13)$$

where  $\Gamma_{\text{short}}$  is a characteristic short time relaxation rate that is directly related to the number of molecules that are in resonance at the applied field  $H_z$ , and  $0 < \alpha < 1$  in most cases.  $\alpha = 0.5$  in the Prokof'ev–Stamp theory (eqn (4.11)) and  $\Gamma_{\text{sqrt}}$  is directly proportional to  $P(H_z)$  (eqn (4.12)). The *hole-digging method* can be divided into three steps (Fig. 4.24):

- 1. Preparing the initial state.** A well-defined initial magnetization state of the crystal of molecular clusters can be achieved by rapidly cooling the sample from high down to low temperatures in a constant applied field  $H_z^0$ . For zero applied field ( $H_z = 0$ ) or rather large applied fields ( $H_z > 1$  T), one yields the demagnetized or saturated magnetization state of the entire crystal, respectively. One can also quench the sample in a small field of a few milliteslas yielding any possible initial magnetization  $M_{\text{in}}$ . When the quench is fast (< 1 s), the sample's magnetization does not have time to relax, either by thermal or by quantum transitions. This procedure yields a frozen thermal equilibrium distribution, whereas for slow cooling rates the molecule spin states in the crystal may tend to a partially ordered ground state. We found that for our fastest cooling rates of about 1 s, partial ordering occurs. However, we developed a Landau–Zener demagnetization method allowing us to reach a randomly disordered state (Wernsdorfer *et al.* 2005).



**Fig. 4.24** Schema of the hole-digging method presenting the time dependence of temperature, applied field, and magnetization of the sample (reprinted from Wernsdorfer, W. (2001) *Adv. Chem. Phys.* **118**, 99, © 2008 John Wiley & Sons Limited, reproduced with permission).



**Fig. 4.25** Field dependence of the short-time square-root relaxation rates  $\Gamma_{\text{sqrt}}(H_z)$  for three different values of the initial magnetization  $M_{\text{in}}$ . According to eqn (4.12), the curves are proportional to the distribution  $P(H_z)$  of magnetic energy bias due to local dipole field distributions in the sample. Note the logarithmic scale for  $\Gamma_{\text{sqrt}}$ . The peaked distribution labelled  $M_{\text{in}} = -0.998 M_s$  was obtained by saturating the sample, whereas the other distributions were obtained by thermal annealing.  $M_{\text{in}} = -0.870 M_s$  is distorted by nearest-neighbor lattice effects. The peak at 0.04 T as well as the shoulder at 0.02 T and 0.04 T are originated by the clusters that have one nearest-neighbor cluster with reversed magnetization: The peak at 0.04 T corresponds to the reversal of the neighboring cluster along the **a** crystallographic axis, which almost coincides with the easy axis of magnetization, while the shoulder at 0.02 T and 0.04 T are due to the clusters along **b** and **c** (reprinted from Wernsdorfer, W., Ohm, T., Sangregorio, C., Sessoli, R., Mailly, D., Paulsen, C. (1999) *Phys. Rev. Lett.* **82**, 3903, © 1999 American Physical Society).

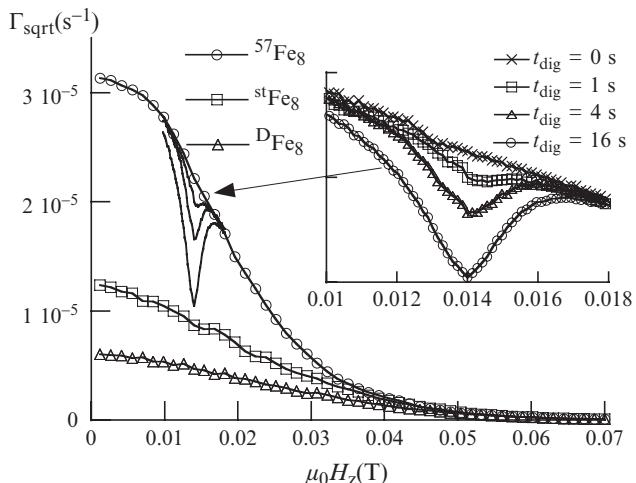
2. **Modifying the initial state—hole digging.** After preparing the initial state, a field  $H_{\text{dig}}$  is applied during a time  $t_{\text{dig}}$ , called the “digging field and digging time”, respectively. During the digging time and depending on  $H_{\text{dig}}$ , a fraction of the molecular spins tunnel (back and/or fourth); that is, they reverse the direction of magnetization. Note that the field-sweeping rate to apply  $H_{\text{dig}}$  should be fast enough to minimize the change of the initial state during the field sweep.
3. **Probing the final state.** Finally, a field  $H_z^{\text{probe}}$  is applied (Fig. 4.24) to measure the short time relaxation from which one yields  $\Gamma_{\text{short}}$ , (eqn (4.13)) which is related to the number of spins that are still free for tunnelling after step (2).

The entire procedure is then repeated many times but at other fields  $H_z^{\text{probe}}$  yielding  $\Gamma_{\text{short}}(H_z, H_{\text{dig}}, t_{\text{dig}})$ , which is related to the distribution of spins  $P(H_z, H_{\text{dig}}, t_{\text{dig}})$  that are still free for tunnelling after the hole digging. For  $t_{\text{dig}} = 0$ , this method maps out the initial distribution.

### 4.7.3 Intermolecular dipole interaction in Fe<sub>8</sub>

We applied the hole-digging method to several samples of molecular clusters and quantum spin glasses. The most detailed study has been done on the Fe<sub>8</sub> system. We found the predicted  $\sqrt{t}$  relaxation (eqn (4.11)) in experiments on fully saturated Fe<sub>8</sub> crystals (Ohm *et al.* 1998a,b) and on non-saturated samples (Wernsdorfer *et al.* 1999). Figure 4.25 displays a detailed study of the dipolar distributions, revealing a remarkable structure that is due to next-nearest-neighbor effects (Wernsdorfer *et al.* 1999). These results are in good agreement with simulations (Cuccoli *et al.* 1999; Tupitsyn *et al.* 2004).

For a saturated initial state, the Prokof'ev–Stamp theory allows one to estimate the tunnel splitting  $\Delta_{\pm S}$ . Using eqns (4.3), (4.9), and (4.12) of Prokof'ev and Stamp 1998, along with integration, we find  $\int \Gamma_{\text{sqrt}} d\xi = c \frac{\xi_0}{E_D} \frac{\Delta_{\pm S}^2}{4h}$ , where  $c$  is a constant of the order of unity that depends on the sample shape. With  $E_D = 15$  mT,  $\xi_0 = 0.8$  mT,  $c = 1$ , and  $\Gamma_{\text{sqrt}}$  (Wernsdorfer *et al.* 1999, 2000), we find  $\Delta_{\pm 10} = 1.2 \times 10^{-7}$  K that is close to the result of  $\Delta_{\pm 10} = 1.0 \times 10^{-7}$  K



**Fig. 4.26** Comparison of the short time relaxation rates of three different  $\text{Fe}_8$  samples at  $T = 40 \text{ mK}$  with  $H_{\text{trans}} = 0$  and  $M_{\text{init}} = 0$ . The inset displays a typical example of a hole that was dug into the distribution by allowing the sample to relax for the time  $t_{\text{dig}}$  at  $\mu_0 H_{\text{dig}} = 14 \text{ mT}$  (reprinted from Wernsdorfer, W., Caneschi, A., Sessoli, R., Gatteschi, D., Cornia, A., Villar, V., Paulsen, C. (2000) *Phys. Rev. Lett.* **84**, 2965, © 2000 American Physical Society).

obtained by using a Landau–Zener method (Section 4.3.1) (Wernsdorfer and Sessoli 1999; Wernsdorfer *et al.* 2005a).

Whereas the hole-digging method probes the longitudinal dipolar distribution ( $H_z$  direction), the Landau–Zener method can be used to probe the transverse dipolar distribution by measuring the tunnel splittings  $\Delta$  around a topological quench.

#### 4.7.4 Hyperfine interaction in $\text{Fe}_8$

The strong influence of nuclear spins on resonant quantum tunnelling in the molecular cluster  $\text{Fe}_8$  was demonstrated for the first time (Wernsdorfer *et al.* 2000) by comparing the relaxation rate of the standard  $\text{Fe}_8$  sample with two isotopic modified samples: (i)  $^{56}\text{Fe}$  is replaced by  $^{57}\text{Fe}$ , and (ii) a fraction of  $^1\text{H}$  is replaced by  $^2\text{H}$ . By using the hole-digging method, we measured an intrinsic broadening that is driven by the hyperfine fields (Fig. 4.26). Our measurements are in good agreement with numerical hyperfine calculations (Wernsdorfer *et al.* 2000; Tupitsyn *et al.* 2004). For  $T > 1.5 \text{ K}$ , the influence of nuclear spins on the relaxation rate is less important, suggesting that spin–phonon coupling dominates the relaxation rate.

### 4.8 Molecular spintronics using single-molecule magnets

Molecular spintronics combines the ideas of two novel disciplines, spintronics and molecular electronics. The resulting field aims at manipulating spins and charges in electronic devices containing one or more molecules (Bogani and Wernsdorfer 2008). The main advantage is that the weak spin-orbit and hyperfine interactions in organic molecules is likely to preserve spin-coherence over time and distance much longer than in conventional metals or semiconductors. In addition, specific functions (e.g.

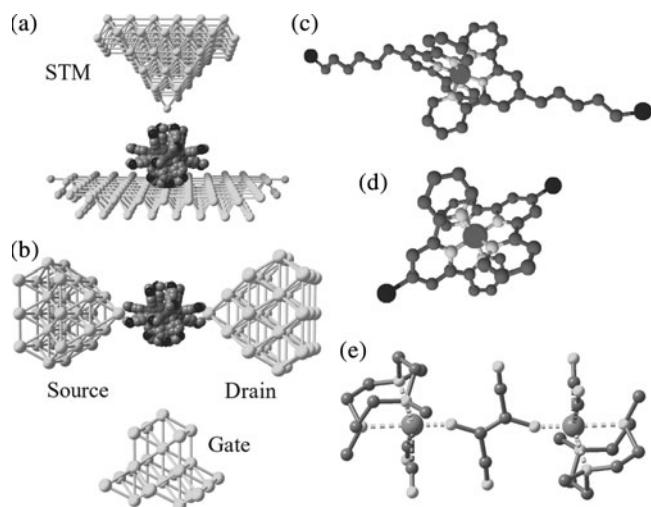
switchability with light, electric field, etc.) could be directly integrated into the molecule.

In order to lay the foundation of molecular spintronics, several molecular devices have been proposed (Bogani and Wernsdorfer 2008): molecular spin-transistor, molecular spin-valve and spin-filter, molecular double-dot devices, and carbon-nanotube-based nano-SQUIDs (Cleuziou *et al.* 2006). The main purpose is to fully control the initialization, the manipulation and the read-out of the spin states of the molecule and to perform basic quantum operations. The main targets for the coming years concern fundamental science, as many issues, experimental, technological and theoretical, must be addressed before applications, for instance in quantum electronics, can be realistically considered.

### 4.8.1 Molecular spin-transistor

The first scheme we consider is a magnetic molecule attached between two non-magnetic electrodes. One possibility is to use a scanning tunnelling microscope tip as the first electrode and the conducting substrate as the second one (Fig. 4.27(a)). So far, only a few atoms on surfaces have been probed in this way, revealing interesting Kondo effects (Wahl *et al.* 2007) and single-atom magnetic anisotropies (Hirjibehedin *et al.* 2007). The next scientific step is to pass from atoms to molecules in order to observe richer physics and to modify the properties of the magnetic objects. Although isolated SMMs on gold have been obtained (Coronado *et al.* 2004; Fleury *et al.* 2005; Naitabdi *et al.* 2005; Cornia *et al.* 2006), the rather drastic experimental requirements, i.e. very low temperatures and high magnetic fields, have not yet been achieved. The first theoretical work predicted that quantum tunnelling of the magnetization is detectable via the electric current flowing through the molecule (Kim and Kim 2004), allowing therefore the read-out of the quantum dynamics of a single molecule.

Another possibility concerns break-junction devices (Park *et al.* 1999), which integrate a gate electrode. Such a three-terminal transport device, called



**Fig. 4.27** Transport experiments on SMMs.  
 (a) Schematic using a STM tip to perform transport on surface grafted SMMs.  
 (b) Schematic of SMM-based molecular transistors, in which a gate voltage can modulate transport.  
 (c)  $[\text{Co}(\text{TerPy})_2]$  molecular magnet with alkyl spacers, permitting transport in the weakly coupled regime (Park *et al.* 2002).  
 (d)  $[\text{Co}(\text{TerPy})_2]$  molecular magnet with no spacers, showing strong coupling and the Kondo effect (Park *et al.* 2002).  
 (e) Divanadium  $[(\text{N}, \text{N}', \text{N}'\text{-trimethyl-1,4,7-triazacyclononane})_2\text{V}_2(\text{CN})_4(-\text{C}_4\text{N}_4)]$  molecular magnet showing the Kondo effect only in the charged state (Liang *et al.* 2002) (reprinted from Bogani, L., Wernsdorfer, W. (2008) *Nature Mater.* 7, 179, © 2008 Nature Publishing Group).

a molecular spin-transistor, is a single-electron transistor with non-magnetic electrodes and a single magnetic molecule as the island. The current passes through the magnetic molecule via the source and drain electrodes, and the electronic transport properties are tuned via a gate voltage  $V_g$  (Fig. 4.27(b)). Similarly to molecular electronics, two experimental regimes can be distinguished, depending on the coupling between molecule and electrodes.

#### 4.8.1.1 Weak-coupling limit

In the weak-coupling limit charging effects dominate the transport. Transport takes place when a molecular orbital is in resonance with the Fermi energy of the leads and electrons can then tunnel through the energy barrier into the molecular level and out into the drain electrode. The resonance condition is obtained by shifting the energy levels with  $V_g$  and the measurements show Coulomb-blockade diamonds (Hanson *et al.* 2007).

The experimental realization of this scheme has been achieved using  $\text{Mn}_{12}$  with thiol-containing ligands (Fig. 4.27(b)), which bind the SMM to the gold electrodes with strong and reliable covalent bonds (Heersche *et al.* 2006). An alternative route is to use short but weak-binding ligands (Jo *et al.* 2006): in both cases the peripheral groups act as tunnel barriers and help conserving the magnetic properties of the SMM in the junction. As the electron transfer involves the charging of the molecule, we must consider, in addition to the neutral state, the magnetic properties of the negatively and positively charged species. This introduces an important difference with respect to the homologous measurements on diamagnetic molecules, where the assumption is often made that charging of the molecule does not significantly alter the internal degrees of freedom (Kouwenhoven *et al.* 1997). Because crystals of the charged species can be obtained, SMMs permit direct comparison between spectroscopic transport measurements and more traditional characterization methods. In particular, magnetization measurements, electron paramagnetic resonance, and neutron spectroscopy can provide energy-level spacings and anisotropy parameters. In the case of  $\text{Mn}_{12}$ , positively charged clusters possess a lower anisotropy barrier (Chakov *et al.* 2005). As revealed by the first Coulomb-blockade measurements, the presence of these states is fundamental to explain transport through the clusters (Heersche *et al.* 2006; Jo *et al.* 2006). Negative differential conductance was found that might be due to the magnetic characteristics of SMMs.

Studies in a magnetic field showed a first evidence of the spin-transistor properties (Jo *et al.* 2006). Degeneracy at zero field and non-linear behavior of the excitations as a function of field are typical of tunnelling via a magnetic molecule. In these first studies, the lack of a hysteretic response can be due, besides environmental effects (Bogani *et al.* 2007), to the alternation of the molecules during the grafting procedure, to the population of excited states with lower energy barriers, or might also be induced by the source-drain voltage scan performed at each field value.

Theoretical investigations in the weak-coupling regime predict many interesting effects. For example, a direct link between shot-noise measurements and the detailed microscopic magnetic structure of SMMs has been proposed (Romeike *et al.* 2006c), allowing the connection of structural and magnetic

parameters to the transport features and therefore a characterization of SMMs using transport measurements. This opens the way to rational design of SMMs for spintronics and to test the physical properties of related compounds. The first step in this direction has already been made by comparing the expected response of chemically related SMMs (Romeike *et al.* 2007). Note that this direct link cannot be established for nanoparticles or quantum dots (QDs) because they do not possess a unique chemical structure.

A complete theoretical analysis as a function of the angle between the easy axis of magnetization and the magnetic field showed that the response persists whatever the orientation of the SMM in the junction and that even films of SMMs should retain many salient properties of single-molecule devices (Elste and Timm 2007; Timm 2007).

#### 4.8.1.2 Strong-coupling limit

For strong electronic coupling between the molecule and the leads, higher-order tunnel processes become important, leading to the Kondo effect (Cronenwett *et al.* 1998; Goldhaber-Gordon *et al.* 1998; Nygard *et al.* 2000; van der Wiel *et al.* 2000). This regime has been attained using paramagnetic molecules containing one (Park *et al.* 2002) or two magnetic centers (Liang *et al.* 2002), but remains elusive for SMMs.

The first mononuclear magnetic molecule investigated (Fig. 4.27c) is a  $\text{Co}^{2+}$  ion bound by two terpyridine ligands, TerPy, attached to the electrodes with chemical groups of variable length (Park *et al.* 2002). The system with the longer alkyl spacer, due to a lower transparency of the barrier, displays Coulomb-blockade diamonds, which are characteristic of the weak-coupling regime, but no Kondo peak. Experiments conducted as a function of magnetic field reveal the presence of excited states connected to spin excitations, in agreement with the effective  $S = 1/2$  state usually attributed to  $\text{Co}^{2+}$  ions at low temperatures but a Land factor  $g = 2.1$  is found. This is unexpected for  $\text{Co}^{2+}$  ions, characterized by high spin-orbit coupling and magnetic anisotropy, and this point needs further investigation. The same complex with the thiol directly connected to the TerPy ligand (Fig. 4.27(d)) shows strong coupling to the electrodes, with exceptionally high Kondo temperatures around 25 K (Park *et al.* 2002).

Additional physical effects of considerable interest were obtained using a simple molecule containing two magnetic centers (Liang *et al.* 2002). This molecule, the divanadium molecule (Fig. 4.27(e)), was again directly grafted to the electrodes, so as to have the highest possible transparency (Liang *et al.* 2002). The molecule can be tuned with the gate voltage  $V_g$  into two differently charged states. The neutral state, due to antiferromagnetic coupling between the two magnetic centers, has  $S = 0$ , while the positively charged state has  $S = 1/2$ . Kondo features are found, as expected (Cronenwett *et al.* 1998; Goldhaber-Gordon *et al.* 1998; Nygard *et al.* 2000; van der Wiel *et al.* 2000) only for the state in which the molecule has a non-zero spin moment. This nicely demonstrates that magnetic molecules with multiple centers and antiferromagnetic interactions permit switching of the Kondo effect on and off, depending on their charge state. The Kondo temperature is again exceptionally high, exceeding 30 K, and its characterization as a function of  $V_g$  indicates that

not only spin but also orbital degrees of freedom play an important role in the Kondo resonance of single molecules. Molecular magnets, in which spin-orbit interaction can be tuned without altering the structure (Ishikawa *et al.* 2005b), are appealing to investigate further this physics.

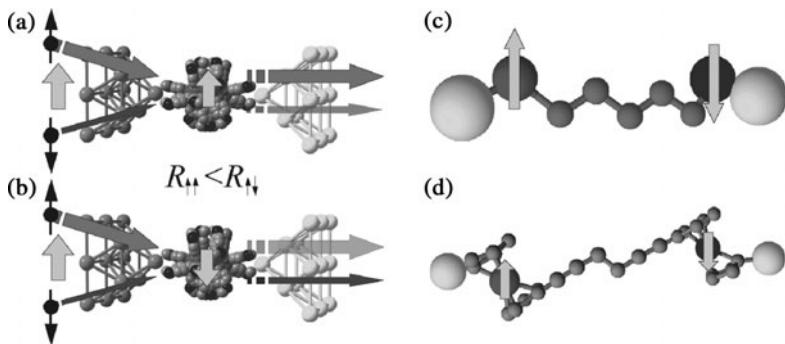
The Kondo temperatures observed in the two cases (Liang *et al.* 2002; Park *et al.* 2002) are much higher than those obtained for QDs and carbon nanotubes and are extremely encouraging. The study of the super-paramagnetic transition of SMMs while in the Kondo regime thus seems achievable, possibly leading to an interesting interplay of the two effects. In order to observe the Kondo regime one might start with small SMMs (Ishikawa *et al.* 2005b, Aromi and Brechin 2006), with core states more affected by the proximity of the leads and use short and strongly bridging ligands to connect SMMs to the electrodes (Park *et al.* 2002; Cornia *et al.* 2006).

Theoretical investigations have explored the rich physics of this regime (Romeike *et al.* 2006a,b,c), revealing that the Kondo effect should even be visible in SMMs with  $S > 1/2$  (Romeike *et al.* 2006b). This is in contrast to expectations for a system with an anisotropy barrier, where the blocked spin should hinder cotunnelling processes. However, in SMMs, the presence of a transverse anisotropy induces a Kondo resonance peak (Romeike *et al.* 2006b). The observation of this new physical phenomena should be possible because of the tunability of SMMs, allowing a rational choice of the physical parameters governing the tunnelling process: low-symmetry transverse terms are particularly useful, because selection rules apply for high-symmetry terms.

The first theoretical predictions argued that the Kondo effect should be present only for half-integer spin molecules. However, the particular quantum properties of SMMs allow for the Kondo effect even for integer spins. In addition, the presence of the so-called Berry-phase interference (Garg 1993; Wernsdorfer and Sessoli 1999; Wernsdorfer *et al.* 2005a), a geometrical quantum phase effect, can produce not only one Kondo resonance peak, but a series of peaks as a function of applied magnetic field (Leuenberger and Mucciolo 2006). These predictions demonstrate how the molecular nature of SMMs and the quantum effects they exhibit differentiate them from inorganic QDs and nanoparticles and should permit the observation of otherwise prohibited phenomena.

#### **4.8.2 Molecular spin-valve**

A molecular spin-valve (SV) (Sanvito and Rocha 2006) is similar to a spin-transistor but contains at least two magnetic elements (Figs. 4.28a and b). SVs change their electrical resistance for different mutual alignments of the magnetizations of the electrodes and of the molecule, analogous to a polarizer-analyzer setup. Non-molecular devices are already used in hard disc drives, owing to the giant- and tunnel-magnetoresistance effects. As good efficiency has already been demonstrated for organic materials (Xiong *et al.* 2004), molecular SVs are actively sought (Rocha *et al.* 2005; Waldron *et al.* 2006). As only a few examples of molecular SVs exist (Pasupathy *et al.* 2004; Hueso *et al.* 2007), the fundamental physics behind these devices remains largely

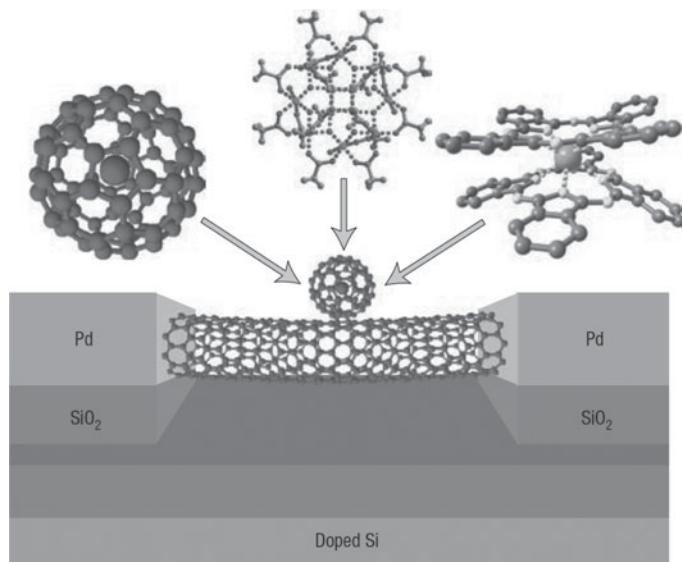


**Fig. 4.28** Spin-valves based on molecular magnets. Yellow arrows represent the magnetization. (a) Parallel configuration of the magnetic source electrode and molecular magnetization, with diamagnetic drain electrode. Spin-up majority carriers (thick arrow) are not affected by the molecular magnetization, while the spin-down minority carriers (thin arrow) are partially reflected back. (b) Antiparallel configuration: majority spin-up electrons are only partially transmitted by the differently polarized molecule, while the minority spin-down electrons pass unaffected. Assuming that the spin-up contribution to the current is larger in the magnetic contact, this configuration has higher resistance than that of the previous case. (c) Theoretical schematic of a spin-valve configuration with non-magnetic metal electrodes (Sanvitto and Rocha 2006) and (d) proposed molecular magnet between gold electrodes: a conjugated molecule bridges the cobaltocene and ferrocene moieties (Liu *et al.* 2006) (reprinted from Bogani, L., Wernsdorfer, W. (2008) *Nature Mater* **7**, 179, © 2008 Nature Publishing Group).

unexplored and will likely be the focus of considerable attention in the near future. The simplest SV consists of a diamagnetic molecule in between two magnetic leads, which can be metallic or semiconducting. The first experiments sandwiched a C<sub>60</sub> fullerene between Ni electrodes, showing a very large negative magnetoresistance effect (Pasupathy *et al.* 2004). Another interesting possibility is to use carbon nanotubes connected with magnetic half-metallic electrodes transforming spin information into large electrical signals (Liu *et al.* 2006).

A SMM-based SV can have one or two magnetic electrodes (Figs. 4.28(a) and (b)), or the molecule can possess two magnetic centers in between two non-magnetic leads (Figs. 4.28(c) and (d)), in a scheme reminiscent of early theoretical models of SVs (Sanvitto and Rocha 2006). Molecules with two magnetic centers connected by a molecular spacer are well known in molecular magnetism and a double metallocene junction has been theoretically studied (Liu *et al.* 2006). This seems a good choice, as the metallocenes leave the d-electrons of the metals largely unperturbed.

Theory indicates that, when using SMMs, the contemporary presence, at high bias, of large currents and slow relaxation will reveal a physically interesting regime (Elste and Timm 2006; Timm and Elste 2006). Only spins parallel to the molecular magnetization can flow through the SMM and the current will display, for a time equivalent to the relaxation time, a very high spin polarization. For large currents this process can lead to a selective drain of spins with one orientation from the source electrode, thus transferring a large amount of magnetic moment from one lead to the other. This phenomenon, due to a sole SMM, has been named giant spin amplification (Timm and Elste 2006) and offers a convenient way to read the magnetic state of the molecule. The switching of the device seems more complicated, at first sight, involving a two-step process that includes the application of a magnetic field and the variation of the bias voltage. However, it has recently been suggested that the spin-polarized current itself can be sufficient to switch the magnetization of a SMM (Misiorny and Barnas 2007). The switching can be detected in the current as a step if both leads are magnetic and have parallel magnetization, or as a sharp peak for the antiparallel configuration.



**Fig. 4.29** Molecular double-dot devices. Magnetic molecules proposed for grafting on suspended carbon nanotubes connected to Pd electrodes (form left to right): a C<sub>60</sub> fullerene including a rare-earth atom, the Mn<sub>12</sub> SMM and the rare-earth-based double-decker [Tb(phtalocyanine)<sub>2</sub>] SMM. The gate voltage of the double-dot device is obtained by a doped Si substrate covered by a SiO<sub>2</sub> insulating layer (reprinted from Bogani, L., Wernsdorfer, W. (2008) *Nature Mater* 7, 179, © 2008 Nature Publishing Group).

#### 4.8.3 Molecular multidot devices

A double-dot device (Fig. 4.29) is one possible way of molecular spintronics (Bogani and Wernsdorfer 2008). It is a three-terminal device, where the current passes through a non-magnetic quantum conductor (quantum wire, nanotube, molecule, or quantum dot (QD)). The magnetic molecule is only weakly coupled to the non-magnetic conductor but its spin can influence the transport properties, permitting read-out of the spin state with minimal backaction. Several mechanisms can be exploited to couple the two systems. One appealing way is to use a carbon nanotube as a detector of the magnetic flux variation, possibly using the nano-SQUID (Cleuziou *et al.* 2006). Other possibilities involve the indirect detection of the spin state through electrometry. Indeed, a non-magnetic quantum conductor at low temperatures behaves as a QD for which charging processes become quantized, giving rise to Coulomb-blockade and Kondo effects depending on the coupling to the leads. Any slight change in the electrostatic environment (controlled by the gate) can induce a shift of the Coulomb diamonds of the device, leading to a conductivity variation of the QD at constant gate voltage. QDs are therefore accurate electrometers. When the QD is coupled, even weakly, with a magnetic object, due to the Zeeman energy the spin flip at non-zero field induces a change of the electrostatic environment of the QD. This effect, called the magneto-Coulomb effect, therefore enables detection of the magnetization reversal of the molecule.

Another route is weak exchange or dipole coupling between the magnetic molecule and the QD. It is interesting to probe these effects as a function of the number of trapped electrons because an odd or even number of electrons should lead to different couplings. The main advantage of these schemes is that the coupling to the leads and the injected current does not alter the magnetic

properties of the molecule. Because coupling is small, these devices might allow a non-destructive read-out of the spin states.

## 4.9 Conclusion

In conclusion, we have presented detailed measurements that have demonstrated that molecular nanomagnets offer a unique opportunity to explore the quantum dynamics of a large but finite spin. We focused our discussion on the Fe<sub>8</sub> molecular nanomagnet because it is the first system where studies in the pure quantum regime were possible. The tunnelling in this system is remarkable because it does not show up at the lowest orders of perturbation theory.

A new family of supramolecular, antiferromagnetically exchange-coupled dimers of single-molecule magnets (SMMs) has recently been reported (Wernsdorfer *et al.* 2002a). Each SMM acts as a bias on its neighbor, shifting the quantum tunnelling resonances of the individual SMMs. Hysteresis-loop measurements on a single crystal of SMM dimers have established quantum tunnelling of the magnetization via entangled states of the dimer. This showed that the dimer really does behave as a quantum-mechanically coupled dimer. The transitions are well separated, suggesting long coherence times compared to the timescale of the energy splitting (Tiron *et al.* 2003). This result is of great importance if such systems are to be used for quantum computing.

Molecules with small spin have also been studied. For example, time-resolved magnetization measurements were performed on a spin 1/2 molecular complex, so-called V<sub>15</sub> (Chiorescu *et al.* 2000a). Despite the absence of a barrier, magnetic hysteresis is observed over a time scale of several seconds. A detailed analysis in terms of a dissipative two-level model has been given, in which fluctuations and splittings are of the same energy. Spin–phonon coupling leads to long relaxation times and to a particular “butterfly” hysteresis loop (Chiorescu *et al.* 2000b; Dobrovitski *et al.* 2000). We presented magnetization measurements on a crystal of Cr<sub>7</sub>Ni antiferromagnetic rings with a spin 1/2. Irradiation with microwaves at frequencies between 1 and 10 GHz leads to observation of very narrow resonant photon absorption lines that are broadened by hyperfine and spin–spin interactions.

The use of circularly polarized microwaves allowed us to show for the first time the phenomenon of photon-assisted tunnelling in magnetism, using a single-molecule magnet Fe<sub>8</sub> (Sorace *et al.* 2003). In accordance with the selection rules for EPR spectroscopy (Abragam and Bleaney 1970), circularly polarized radiation promotes the transition  $m_S = 10$  to 9 with  $\Delta m_S = -1$ , giving an effect of magnetic dichroism at millimeter wavelengths. At the lowest powers, the tunnel probability increases linearly with power, whereas at higher powers a strongly non-linear regime is observed. The latter might be due to multispin and coherent photon transitions.

Quantum coherence between states of opposite magnetization has been observed on molecules with small spin (Ardavan *et al.* 2007; Carretta *et al.* 2007; Bertaina *et al.* 2008) and without barrier. However, for most SMMs, dipole–dipole and hyperfine interactions are source of decoherence. In other

words, when a spin has tunnelled through the barrier, it experiences a huge modification of its environment (hyperfine and dipolar) which prohibits the back tunnelling. Prokof'ev and Stamp suggested three possible strategies to suppress the decoherence (Prokof'ev and Stamp 1995; Morello *et al.* 2003). (i) Choose a system where the NMR frequencies far exceed the tunnel frequencies making any coupling impossible. (ii) Isotopically purify the sample to remove all nuclear spins. (iii) Apply a transverse field to increase the tunnel rate to frequencies much larger than hyperfine-field fluctuations. Several groups are currently working on such proposals.

Concerning the perspectives of the field of single-molecule magnets, we expect that chemistry is going to play a major role through the synthesis of novel larger spin clusters with strong anisotropy (Milios *et al.* 2007). We want to stress that there are already many other molecular nanomagnets (the largest is currently a Mn<sub>84</sub>, Fig. 4.3) which are possible model systems. We believe that more sophisticated theories are needed that describe the dephasing effects of the environment onto the quantum system. These investigations are important for studying the quantum character of molecular clusters for applications like “quantum computers”. The first implementation of Grover’s algorithm with molecular nanomagnets has been proposed (Leuenberger and Loss 2001). Antiferromagnetic systems have attracted increasing interest. In this case the quantum hardware is thought of as a collection of coupled molecules, each corresponding to a different qubit (Meier *et al.* 2003a,b; Troiani *et al.* 2005a,b). In order to explore these possibilities, new and very precise setups are currently being built and new methods and strategies are being developed. The field of molecular nanomagnets evolves towards molecular electronics and spintronics, which are both rapidly emerging fields of nanoelectronics with a strong potential impact for the realization of new functions and devices helpful for information storage as well as quantum information. New projects aim at the merging of the two fields by the realization of molecular junctions that involve a molecular nanomagnet. In order to tackle the challenge of controlled connection at the single molecule level, molecular self-assembly on nanojunctions obtained by the technique of electromigration was used (Heersche *et al.* 2006; Jo *et al.* 2006). Furthermore, a new nano-SQUID with carbon nanotube Josephson junctions was developed (Cleuziou *et al.* 2006), which should be sensitive enough to study individual magnetic molecules that are attached to the carbon nanotube. Such techniques will lead to enormous progress in the understanding of the electronic and magnetic properties of isolated molecular systems and they will reveal intriguing new physics (Bogani and Wernsdorfer 2008).

## References

- Abragam, A., Bleaney, B. *Electron Paramagnetic Resonance of Transition Ions* (Clarendon Press, Oxford, 1970).
- Alonso, J.J., Fernandez, J.F. *Phys. Rev. Lett.* **87**, 097205 (2001).
- Amoretti, G., Caciuffo, R., Combet, J., Murani, A., Caneschi, A. *Phys. Rev. B* **62**, 3022 (2000).

- Ardavan, A., Rival, O., Morton, J.J., Blundell, S.J., Tyryshkin, A.M., Timco, G.A., Winpenny, R.E. *Phys. Rev. Lett.* **98**, 057201 (2007).
- Aromi, G., Brechin, E.K. *Struct. Bond.* **122**, 1 (2006).
- Aubin, S.M.J., Dilley, N.R., Wemple, M.B., Christou, G., Hendrickson, D.N. *J. Am. Chem. Soc.* **120**, 839 (1998).
- Awschalom, D.D., Flatt, M.M. *Nature Phys.* **3**, 153 (2007).
- Barra, A., Gatteschi, D., Sessoli, R. *Chem. Eur. J.* **6**, 1608 (2000).
- Barra, A.-L., Debrunner, P., Gatteschi, D., Schulz, C.E., Sessoli, R. *EuroPhys. Lett.* **35**, 133 (1996).
- Bertaina, S., Gambarelli, S., Mitra, T., Tsukerblat, B., Muller, A., Barbara, B. *Nature* **453**, 203 (2008).
- Bogani, L., Cavigli, L., Gurioli, M., Novak, R.L., Mannini, M., Caneschi, A., Pineider, F., Sessoli, R., Clemente-León, M., Coronado, E., Cornia, A., Gatteschi, D. *Adv. Mater.* **19**, 3906 (2007).
- Bogani, L., Wernsdorfer, W. *Nature Mater.* **7**, 179 (2008).
- Brown, W.F. *Phys. Rev.* **130**, 1677 (1963).
- Caciuffo, R., Amoretti, G., Murani, A., Sessoli, R., Caneschi, A., Gatteschi, D. *Phys. Rev. Lett.* **81**, 4744 (1998).
- Caneschi, A., Gatteschi, D., Laugier, J., Rey, P., Sessoli, R., Zanchini, C. *J. Am. Chem. Soc.* **113**, 5873 (1991).
- Caneschi, A., Gatteschi, D., Sangregorio, C., Sessoli, R., Sorace, L., Cornia, A.M., Novak, A., Paulsen, C., Wernsdorfer, W. *J. Magn. Magn. Mater.* **200**, 182 (1999).
- Carretta, S., Santini, P., Amoretti, G., Affronte, M., Ghirri, A., Sheikin, I., Piligkos, S., Timco, G., Winpenny, R.E.P. *Phys. Rev. B* **72**, 060403(R) (2005).
- Carretta, S., Santini, P., Amoretti, G., Guidi, T., Copley, J.R., Qiu, Y., Caciuffo, R., Timco, G.A., W.R.E. *Phys. Rev. Lett.* **98**, 167401 (2007).
- Chakov, N.E., Soler, M., Wernsdorfer, W., Asbboud, K.A., Christou, G. *Inorg. Chem.* **44**, 5304 (2005).
- Chiorescu, I., Wernsdorfer, W., Barbara, B., Muller, A., Bogge, H. *J. Appl. Phys.* **87**, 5496 (2000).
- Chiorescu, I., Wernsdorfer, W., Müller, A., Bögge, H., Barbara, B. *Phys. Rev. Lett.* **84**, 3454 (2000).
- Christou, G., Gatteschi, D., Hendrickson, D., Sessoli, R. *MRS Bull.* **25**, 66 (2000).
- Cleuziou, J.-P., Wernsdorfer, W., Bouchiat, V., Ondarcuhu, T., Monthoux, M. *Nature Nanotech.* **1**, 53 (2006).
- Coffey, W.T., Crothers, D.S.F., Dormann, J.L., Kalmykov, Y.P., Waldron, J.T. *Phys. Rev. B* **52**, 15951 (1995).
- Cornia, A., Fabretti, A.C., Zobbi, L., Caneschi, A., Gatteschi, D., Mannini, M., Sessoli, R. *Struct. Bond.* **122**, 133 (2006).
- Coronado, E., Forment-Aliaga, A., Gaita-Arino, A., Gimenez-Saiz, C., Romero, F., Wernsdorfer, W. *Angew. Chem. Int. Ed. Engl.* **43**, 6152 (2004).
- Cronenwett, S.M., Oosterkamp, T.H., Kouwenhoven, L.P. *Science* **281**, 540 (1998).
- Cuccoli, A., Fort, A., Rettori, A., Adam, E., Villain, J. *Euro. Phys. J. B* **12**, 39 (1999).
- del Barco, E., Kent, A.D., Yang, E.C., Hendrickson, D.N. *Phys. Rev. Lett.* **93**, 157202 (2004).

- Dobrovitski, V.V., Katsnelson, M.I., Harmon, B.N. *Phys. Rev. Lett.* **84**, 3458 (2000).
- Elste, F., Timm, C. *Phys. Rev. B* **73**, 235305 (2006).
- Elste, F., Timm, C. *Phys. Rev. B* **75**, 195341 (2007).
- Enz, M., Schilling, R. *J. Phys. C* **19**, L711 (1986).
- Feynman, R., Vernon, F. *Ann. Phys.* **24**, 118 (1963).
- Feynman, R.P., Leighton, R.B., Sand, M. *The Feynman lectures on Physics, vol. 3* (Addison-Wesley Publishing Company, London, 1970).
- Fleury, B., Catala, L., Huc, V., David, C., Zhong, W.Z., Jegou, P., Baraton, L., Pelacin, S., Albouy, P.A., Mallah, T. *Chem. Commun.* 2020 (2005).
- Fort, A., Rettori, A., Villain, J., Gatteschi, D., Sessoli, R. *Phys. Rev. Lett.* **80**, 612 (1998).
- Friedman, J.R., Sarachik, M.P., Tejada, J., Ziolo, R. *Phys. Rev. Lett.* **76**, 3830 (1996).
- Garanin, D., Chudnovsky, E. *Phys. Rev. B* **56**, 11102 (1997).
- Garg, A. *EuroPhys. Lett.* **22**, 205 (1993).
- Garg, A. *Phys. Rev. B* **51**, 15592 (1995).
- Garg, A. *Phys. Rev. Lett.* **83**, 4385 (1999).
- Goldhaber-Gordon, D., Gores, J., Kastner, M.A., Shtrikman, H., Mahalu, D., Meirav, U. *Phys. Rev. Lett.* **81**(23), 5225 (1998).
- Rose, G., Stamp, P. *Low-Temp. Phys.* **113**, 1153 (1998).
- Gunther, L., Barbara, B. *Phys. Rev. B* **49**, 3926 (1994).
- Hanson, R., Kouwenhoven, L., Petta, J., Tarucha, S., Vandersype, L. *Rev. Mod. Phys.* **79**, 1217 (2007).
- Hartmann-Boutron, F., Politi, P., Villain, J. *Int. J. Mod. Phys.* **10**, 2577 (1996).
- Heersche, H.B., de Groot, Z., Folk, J.A., van der Zant, H.S.J., Romeike, Wegewijs, Zobbi, C.M.R.L., Barreca, D., Tondello, E., Cornia, A. *Phys. Rev. Lett.* **96**, 206801 (2006).
- Hemmen, J.L.V., Suto, S. *EuroPhys. Lett.* **1**, 481 (1986).
- Hendrickson, D.N., et al. *J. Am. Chem. Soc.* **114**, 2455 (1992).
- Hill, S., Maccagnano, S., Park, K., Achey, R.M., North, J.M., Dalal, N.S. *Phys. Rev. B* **65**, 224410 (2002).
- Hirjibehedin, C.F., Lin, C.Y., Otte, A.F., Ternes, M., Lutz, C. P., Jones, B.A., Heinrich, A.J. *Science* **317**, 1199 (2007).
- Hueso, L.E., Pruneda, J.M., Ferrari, V., Burnell, G., Valds-Herrera, L.P., Simons, M.D., Littlewood, P.B., Artacho, E., Fert, A., Mathur, N.D. *Nature* **445**, 410 (2007).
- Ishikawa, N., Sugita, M., Wernsdorfer, W. *J. Am. Chem. Soc.* **127**, 3650 (2005a).
- Ishikawa, N., Sugita, M., Wernsdorfer, W. *Angew. Chem. Int. Ed.* **44**, 2 (2005b).
- Jo, M.H., Grose, J.E., Baheti, K., Deshmukh, M.M., Sokol, J.J., Rumberger, E.M., Hendrickson, D.N., Long, J.R., Park, H., Ralph, D.C. *Nano. Lett.* **6**, 2014 (2006).
- Kececioglu, E., Garg, A. *Phys. Rev. B* **63**, 064422 (2001).
- Kent, A., Zhong, Y., Bokacheva, L., Ruiz, D., Hendrickson, D., Sarachik, M. *EuroPhys. Lett.* **49**, 521 (2000).
- Kim, G., Kim, T. *Phys. Rev. Lett.* **92**, 137203 (2004).
- Kouwenhoven, L.P., Marcus, C.M., McEuen, P.L., Tarucha, S., Wingreen, R.M.W.N. *Mesoscopic Electron Transport, vol. 345 of Series E*, (eds) Sohn, L.L., Kouwenhoven, L.P. and Schoen, G. (Kluwer, London, 1997) p. 105.

- Kramers, H.A. *Proc. Acad. Sci. Amsterdam* **33**, 959 (1930).
- Kurkijarvi, J. *Phys. Rev. B* **6**, 832 (1972).
- Landau, L. *Phys. Z. Sowjetunion* **2**, 46 (1932).
- Larsen, F.K., McInnes, E.J.L., El., Mkami, H., Overgaard, J., Piligkos, S., Rajaraman, G., Rentschler, E., Smith, A.A., Smith, G.M.H., Boote, V., Jennings, M., Timco, G.A., Winpenny, R.E.P. *Angew. Chem. Int. Ed.* **42**, 101 (2003).
- Lü, R., Hu, H., Zhu, J.-L., Wang, X.-B., Chang, L., Gu, B.-L. *Phys. Rev. B* **61**, 14581 (2000a).
- Lü, R., Kou, S.-P., Zhu, Chang, J.-L., L., Gu, B.-L. *Phys. Rev. B* **62**, 3346 (2000b).
- Leuenberger, M.N., Loss, D. *Phys. Rev. B* **61**, 12200 (2000a).
- Leuenberger, M.N., Loss, D. *Phys. Rev. B* **61**, 1286 (2000b).
- Leuenberger, M.N., Loss, D. *Nature* **410**, 789 (2001).
- Leuenberger, M.N., Mucciolom, E.R. *Phys. Rev. Lett.* **97**, 126601 (2006).
- Liang, J.-Q., Mueller-Kirsten, H., Park, D., Pu, F.-C. *Phys. Rev. B* **61**, 8856 (2000).
- Liang, W., Bockrath, M. Park, H. *Phys. Rev. Lett.* **88**, 126801 (2002).
- Lis, T. *Acta Crystallogr. B* **36**, 2042 (1980).
- Liu, R., Ke, S.-H., Baranger, H.U., Yang, W. *J. Am. Chem. Soc.* **128**, 6274 (2006).
- Loss, D., DiVincenzo, D.P., Grinstein, G. *Phys. Rev. Lett.* **69**, 3232 (1992).
- Meier, F., Levy, J., Loss, D. *Phys. Rev. Lett.* **90**, 047901 (2003a).
- Meier, F., Levy, J., Loss, D. *Phys. Rev. B* **68**, 134417 (2003b).
- Miliotis, C.J., Vinslava, A., Wernsdorfer, W., Moggach, S., Parsons, S., Perlepes, S.P., Christou, G., Brechin, E.K. *J. Am. Chem. Soc.* **129**, 2754 (2007).
- Misiorny, M., Barnas, J. *Phys. Rev. B* **75**, 134425 (2007).
- Miyashita, S. *J. Phys. Soc. Jpn.* **64**, 3207 (1995).
- Miyashita, S. *J. Phys. Soc. Jpn.* **65**, 2734 (1996).
- Morello, A., Mettes, F.L., Luis, F., Fernández, J.F., Krzystek, J., Aromi, G., Christou, G., de Jongh, L.J. *Phys. Rev. Lett.* **90**, 017206 (2003).
- Naitabdi, A., Bucher, J., Gerbier, P., Rabu, P., Drillon, M. *Adv. Mater.* **17**, 1612 (2005).
- Neel, L. *C.R. Acad. Science* **224**, 1550 (1947).
- Neel, L. *Ann. Geophys.* **5**, 99 (1949).
- Novak, M., Sessoli, R. *Quantum Tunneling of Magnetization-QTM'94, L. vol. 301 of NATO ASI Series E Applied Sciences*, (eds) Gunther, L. and Barbara, B., Kluwer, (Academic Publishers, London, 1995) p. 171.
- Nygard, J., Cobden, D.H., Lindelof, P.E. *Nature* **408**, 342 (2000).
- Ohm, T., Sangregorio, C., Paulsen, C. *Euro. Phys. J. B* **6**, 195 (1998a).
- Ohm, T., Sangregorio, C., Paulsen, C. *J. Low Temp. Phys.* **13**, 1141 (1998b).
- Overgaard, J., Iversen, B.B., Palii, S.P., Timco, G.A., Gerbeleu, N.V., Larsen, F.K. *Chem. Eur. J.* **8**, 2775 (2002).
- Park, H., Kim, A.K.L., Alivisatos, A.P., Park, J., McEuen, P.L. *Appl. Phys. Lett.* **75**, 301 (1999).
- Park, J., Pasupathy, A.N., Goldsmith, J.I., Chang, C., Yaish, Y., Petta, J.R., Rinkoski, M., Sethna, J.P., Abruna, H.D., McEuen, P.L., Ralph, D.C. *Nature* **417**, 722 (2002).

- Pasupathy, A., Bialczak, R., Martinek, J., Grose, J., Donev, L., McEuen, P., Ralph, D. *Science* **306**, 86 (2004).
- Paulsen, C., Park, J.-G. *Quantum Tunneling of Magnetization-QTM'94, vol. 301 of NATO ASI Series E Applied Sciences*, (eds) Gunther, L. and Barbara, B., Kluwer, Academic Publishers, London, 1995 p. 189.
- Perenboom, J., Brooks, J., Hill, S., Hathaway, T., Dalal, N. *Phys. Rev. B* **58**, 330 (1998).
- Politi, P., Rettori, A., Hartmann-Boutron, F., Villain, J. *Phys. Rev. Lett.* **75**, 537 (1995).
- Pontillon, Y., Caneschi, A., Gatteschini, D., Sessoli, R., Ressouche, E., Schweizer, J., Lelievre-Berna, E. *J. Am. Chem. Soc.* **121**, 5342 (1999).
- Price, D.J., Lionti, F., Ballou, R., Wood, P., Powell, A.K. *Philos. Trans. R. Soc. Lond. A* **357**, 3099 (1999).
- Prokof'ev, N., Stamp, P. *Quantum Tunneling of Magnetization-QTM'94, L. vol. 301 of NATO ASI Series E Applied Sciences*, (eds) Gunther, L. and Barbara, B. (Kluwer, Academic Publishers, London, 1995) p. 369.
- Prokof'ev, N., Stamp, P. *J. Low-Temp. Phys.* **104**, 143 (1996).
- Prokof'ev, N., Stamp, P. *Phys. Rev. Lett.* **80**, 5794 (1998).
- Prokof'ev, N., Stamp, P. *Rep. Prog. Phys.* **63**, 669–726 (2000).
- Roch, N., Florens, S., Bouchiat, V., Wernsdorfer, W., Balestro, F. *Nature* **453**, 633 (2008).
- Rocha, A.R., Garcia-Suarez, V.M., Bailey, S.W., Lambert, C.J., Ferrer, J., Sanvito, S. *Nature Mater.* **4**, 335 (2005).
- Romeike, C., Wegewijs, M.R., Schoeller, H. *Phys. Rev. Lett.* **97**, 206601 (2006a).
- Romeike, C., Wegewijs, M.R., Schoeller, H. *Phys. Rev. Lett.* **96**, 196601 (2006b).
- Romeike, C., Wegewijs, M.R., Schoeller, H. *Phys. Rev. Lett.* **96**, 196805 (2006c).
- Romeike, C., Wegewijs, R., Park, M.R., Ruben, M., Wenzel, W., Schoeller, H. *Phys. Rev. B* **75**, 064404 (2007).
- Sangregorio, C., Ohm, T., Paulsen, C., Sessoli, R., Gatteschini, D. *Phys. Rev. Lett.* **78**, 4645 (1997).
- Sanvito, S., Rocha, A.R. *J. Comput. Theor. Nanosci.* **3**, 624642 (2006).
- Sessoli, R., Gatteschini, D., Caneschi, A., Novak, M.A. *Nature* **365**, 141 (1993a).
- Sessoli, R., Tsai, H.-L., Schake, A.R., Wang, S., Vincent, J.B., Folting, K., Gatteschini, D., Christou, G., Hendrickson, D.N. *J. Am. Chem. Soc.* **115**, 1804 (1993b).
- S. Hill, Edwards, R., Aliaga-Alcalde, N., Christou, G. *Science* **302**, 1015 (2003).
- Sorace, L., Wernsdorfer, W., Thirion, C., Barra, A.-L., Pacchioni, M., Mailly, D., Barbara, B. *Phys. Rev. B* **68**, 220407(R) (2003).
- Stuckelberg, E. *Helv. Phys. Acta* **5**, 369 (1932).
- Stoner, E.C., Wohlfarth, E.P. *Philos. Trans. London Ser. A* **240**, 599 (1948). Reprinted in *IEEE Trans. Magn. MAG-27* 3475 (1991).
- Tao, N.J. *Nature Nanotechnol.* **1**, 173 (2006).
- Tasiopoulos, A., Vinslava, A., Wernsdorfer, W., Abboud, K., Christou, G. *Angew. Chem. Int. Ed. Engl.* **43**, 2117 (2004).
- Thiaville, A. *J. Magn. Magn. Mater.* **182**, 5 (1998).
- Thiaville, A. *Phys. Rev. B* **61**, 12221 (2000).
- Thirion, C., Wernsdorfer, W., Mailly, D. *Nature Mater.* **2**, 524 (2003).

- Thomas, L., Lonti, F., Ballou, R., Gatteschi, D., Sessoli, R., Barbara, B. *Nature* **383**, 145 (1996).
- Thorwart, M., Grifoni, M., Hanggi, P. *Phys. Rev. Lett.* **85**, 860 (2000).
- Timm, C. *Phys. Rev. B* **76**, 014421 (2007).
- Timm, C., Elste, F. *Phys. Rev. B* **73**, 235304 (2006).
- Tiron, R., Wernsdorfer, W., Foguet-Albiol, D., Aliaga-Alcalde, N., Christou, G. *Phys. Rev. Lett.* **91**, 227203 (2003).
- Troiani, F., Affronte, M., Carretta, S., Santini, P., Amoretti, G. *Phys. Rev. Lett.* **94**, 190501 (2005a).
- Troiani, F., Ghirri, A., Affronte, M., Carretta, S., Santini, P., Amoretti, G., Piligkos, S., Timco, G., Winpenny, R. *Phys. Rev. Lett.* **94**, 207208 (2005b).
- Tupitsyn, I., Stamp, P., Prokof'ev, N.V. *Phys. Rev. B* **69**, 132406 (2004).
- van der Wiel, W.G., Franceschi, S.D., Fujisawa, T., Elzerman, J.M., Tarucha, S., Kouwenhoven, L.P. *Science* **289**, 2105 (2000).
- Villain, J., Fort, A. *Euro. Phys. J. B* **17**, 69 (2000).
- Villain, J., Hartmann-Boutron, F., Sessoli, R., Rettori, A. *EuroPhys. Lett.* **27**, 159 (1994).
- Villain, J., Wurger, A., Fort, A., Rettori, A. *J. de Phys. I* **7**, 1583 (1997).
- von Delft, J., Henley, C.L. *Phys. Rev. Lett.* **69**, 3236 (1992).
- Wahl, P., Simon, P., Diekhoner, L., Stepanyuk, V., Bruno, P., Schneider, M., Kern, K. *Phys. Rev. Lett.* **98**, 056601 (2007).
- Waldron, D., Haney, P., Larade, B., MacDonald, A., Guo, H. *Phys. Rev. Lett.* **96**, 166804 (2006).
- Wernsdorfer, W., Sessoli, R. *Science* **284**, 133 (1999).
- Wernsdorfer, W. *Adv. Chem. Phys.* **118**, 99 (2001).
- Wernsdorfer, W., Aliaga-Alcalde, N., Hendrickson, D., Christou, G. *Nature* **416**, 406 (2002a).
- Wernsdorfer, W., Bhaduri, S., Boskovic, C., Christou, G., Hendrickson, D. *Phys. Rev. B* **65**, 180403(R) (2002b).
- Wernsdorfer, W., Bhaduri, S., Vinslava, A., Christou, G. *Phys. Rev. B* **72**, 214429 (2005d).
- Wernsdorfer, W., Caneschi, A., Sessoli, R., Gatteschi, D., Cornia, A., Villar, V., Paulsen, C. *Phys. Rev. Lett.* **84**, 2965 (2000).
- Wernsdorfer, W., Chakov, N.E., Christou, G. *Phys. Rev. Lett.* **95**, 037203 (2005a).
- Wernsdorfer, W., Mailly, D., Timco, G.A., Winpenny, R.E.P. *Phys. Rev. B* **72**, 060409(R) (2005b).
- Wernsdorfer, W., Müller, A., Mailly, D., Barbara, B. *EuroPhys. Lett.* **66**, 861 (2004).
- Wernsdorfer, W., Murugesu, M., Christou, G. *Phys. Rev. Lett.* **96**, 057208 (2006).
- Wernsdorfer, W., Murugesu, M., Tasiopoulos, A.J., Christou, G. *Phys. Rev. B* **72**, 212406 (2005c).
- Wernsdorfer, W., Ohm, T., Sangregorio, C., Sessoli, R., Mailly, D., Paulsen, C. *Phys. Rev. Lett.* **82**, 3903 (1999).
- Wernsdorfer, W., Bonet Orozco, E., Hasselbach, K., Benoit, A., Barbara, B., Demony, N., Loiseau, A., Boivin, D., Pascard, H., Mailly. D. *Phys. Rev. Lett.* **78**, 1791 (1997a).

- Wernsdorfer, W., Orozco, E.B., Hasselbach, K., Benoit, A., Mailly, D., Kubo, O., Nakano, H., Barbara, B. *Phys. Rev. Lett.* **79**, 4014 (1997).
- Wieghardt, K., Pohl, K., Jibril, I., Huttner, G. *Angew. Chem. Int. Ed. Engl.* **23**, 77 (1984).
- Wolf, S.A., Awshalom, D.D., Buhrman, R.A., Daughton, J.M., von Molnr, S., Roukes, M.L., Chtchelkanova, A.Y., Treger, D.M. *Science* **294**, 449 (2001).
- Xiong, Z.H., Wu, D., Valy Vardeny, Z., Shi, J. *Nature* **427**, 821–824 (2004).
- Yoo, J., Brechin, E.K., Yamaguchi, A., Nakano, M., Huffman, J.C., Maniero, A.L., Brunel, L.C., Awaga, K., Ishimoto, H., Christou, G., Hendrickson, D.N. *Inorg. Chem.* **39**, 3615 (2000).
- Yoo, S.-K., Lee, S.-Y. *Phys. Rev. B* **62**, 3014 (2000).
- Zener, C. *Proc. R. Soc. London, Ser. A* **137**, 696 (1932).
- Zhou, B., Tao, R., Shen, S., Liang, J.Q. *Phys. Rev. A* **66**, 010301 (2002).

# Si/SiGe heterostructures in nanoelectronics

D.J. Paul

5

## 5.1 Introduction

The first transistor point contact was demonstrated by Bardeen and Brattain (1948) at Bell Laboratories using germanium and the first silicon transistor was demonstrated by Tannerbaum in 1954 also at Bell Laboratories. The first mention of the alloy silicon-germanium (SiGe) in electronic devices was actually in the original patent by Bell Laboratories for the bipolar transistor where the idea of a SiGe-base heterostructure bipolar transistor (HBT) was first discussed. The first paper in the literature using the alloy SiGe can be traced to Glickman (1955) with the work on magnetoresistance, but it was not until the pioneering work of Erich Kasper and coworkers at the then AEG Research Laboratories in Ulm, Germany in 1975 that Si/SiGe heterostructures could be realised. Since that time the field and use of SiGe alloys in electronics have expanded substantially. In the 1980s research was dominated by growth development, in the 1990s by the development of the HBT (which can be found in many present-day mobile phones) and in the 2000s by the incorporation of SiGe into microprocessors and mainstream complementary metal oxide semiconductor (CMOS) nanoelectronics. Most high-end microprocessors being sold in 2008 have SiGe alloys incorporated into them at some level to increase performance. There are many other examples of Si/SiGe heterostructures being researched in nanoelectronics for different applications and a number of them will be reviewed in this chapter.

## 5.2 Growth of silicon-germanium alloys

The phase diagram of SiGe alloys shows that Si and Ge are completely miscible, forming  $\text{Si}_{1-x}\text{Ge}_x$  solid solutions with  $x$  ranging from 0 to 1. The  $\text{Si}_{1-x}\text{Ge}_x$  system, however, has the solid and liquid phases separated by a large regime of coexistence, which is typical for systems with strong segregation. It is therefore extremely difficult, if not impossible, to grow

5.1 Introduction	181
5.2 Growth of silicon-germanium alloys	181
5.3 Strain	183
5.4 Band structure	186
5.5 Mainstream nanoelectronic applications	189
5.6 Resonant tunnelling diodes	194
5.7 SiGe quantum cascade emitters	197
5.8 Conclusions	202
References	202

high-quality single-crystal  $\text{Si}_{1-x}\text{Ge}_x$  alloys from a melt of the constituent elements for any arbitrary concentration of Ge. A few papers have discussed adding metals to the Si and Ge melt to produce  $\text{Si}_{1-x}\text{Ge}_x$  but many of the resulting materials have far too high concentrations of background impurities and high defect densities to be of use for electronic or optoelectronic applications.

With the advent of bandgap engineering in the 1970s and the lack of high-quality single  $\text{Si}_{1-x}\text{Ge}_x$  crystals from melts, the main techniques that have developed for the growth of SiGe alloys have been epitaxial techniques. Molecular beam epitaxy (MBE) and chemical vapor deposition (CVD) are the two main techniques being used for the production of SiGe material. As both these techniques are epitaxial, they deposit  $\text{Si}_{1-x}\text{Ge}_x$  alloys on top of a substrate, typically silicon (100), and can form multilayer or superlattice films with virtually any Ge content  $x$ . These techniques are therefore enabling for the formation of bandgap-engineered structures on a silicon wafer.

The growth of  $\text{Si}_{1-x}\text{Ge}_x$  alloys on top of a starting substrate can result in either a 2D or 3D growth mechanism dependent on growth temperature, flux of growth species (i.e. Si and Ge atoms or molecules containing Si or Ge), defects or step heights on the substrate surface, surface tension but also any strain that may result from the 4.2% lattice mismatch between bulk Si and Ge layers. If the strain energy is significant then growth undulations and surface roughening can develop, which in the extreme case can lead to the formation of quantum dots. For most heterolayer electronic devices, 2D growth through the Frank–van der Merwe growth mode is desirable (Kasper and Paul 2005). Full 3D growth can occur through the Volmer–Weber growth mode, which has demonstrated tensile-strained Si quantum dots on bulk Ge substrates, while many Ge quantum dots on Si substrates have been produced by the combined 2D/3D Stranski–Krastanov growth mode (Kasper and Paul 2005).

MBE was the original technique used for growth of SiGe heterostructures and relies on an ultrahigh-vacuum chamber with sources of Si, Ge and any dopants required for the doping of the heterostructure layers. Two main types exist: solid source and gas source. Both techniques can switch each source of material and dopant on or off using either shutters in front of the sources or in the case of gas sources, simply switching the gas lines on or off. The solid-source systems are very similar to those used for III–V growth of materials but require electron-beam evaporators for Si and Ge due to the high melting points of the materials. While B can easily be used in a radiatively heated effusion cell (Knudsen cell) for p-type doping, n-type doping frequently requires secondary ion implantation, where some Si atoms from the electron-beam evaporators are ionized and a small potential between the evaporator and the growth substrate is used to “implant” the dopant atoms into the epitaxial film. This is required for dopants such as P, Sb or As, which surface segregate during SiGe growth resulting in poor incorporation into the epitaxial film. Other techniques exist for doping including bolting a full ion implanter onto the MBE chamber to allow implanted doping *in situ*. Graphite meander heaters are used to heat the substrates with the wafer typically rotated to increase uniformity. The great

advantage of MBE over other growth techniques is that the Ge content of the epitaxially deposited film is dependent on the fluxes of the source material and not on the substrate temperature.

CVD is the main growth technique used by the silicon microelectronics industry and is also the main technique used for  $\text{Si}_{1-x}\text{Ge}_x$  epitaxial growth in research. The growth by CVD is from the pyrolysis at an elevated temperature of a precursor gas. Precursor gases for  $\text{Si}_{1-x}\text{Ge}_x$  CVD includes  $\text{SiH}_4$ ,  $\text{Si}_2\text{H}_6$ ,  $\text{GeH}_4$  and  $\text{SiH}_2\text{Cl}_2$ , while doping is achieved using  $\text{AsH}_3$ ,  $\text{PH}_3$  and  $\text{B}_2\text{H}_6$ . The use of halides as growth precursors requires the dissociation of a hydrogen atom, adsorption of the resultant molecule and then desorption of the remaining hydrogen for the incorporation of the source material into a heterolayer. A full review of CVD growth processes can be found in Caymax and Leong (1994). CVD can also be set up to allow selective or differential growth, which is heavily used by industry. Most SiGe HBTs have a graded  $\text{Si}_{1-x}\text{Ge}_x$  base grown selectively so that the base is a single-crystal  $\text{Si}_{1-x}\text{Ge}_x$  heterolayer and the contacts to the base heterolayer are on top of shallow trench isolation (STI) oxides to reduce the capacitance and increase the maximum oscillation frequency ( $f_{\max}$ ) of the transistors. For deep submicrometer MOSFETs, Ohmic contacts are grown using small amounts of Ge along with a dopant to produce uniaxial strain through the transistor, which improves the mobility significantly in the MOSFET. These applications will be discussed in later sections of the chapter.

## 5.3 Strain

Since Ge has a 4.2% larger lattice constant than Si, most  $\text{Si}_{1-x}\text{Ge}_x$  heterolayers are strained. If a  $\text{Si}_{1-x}\text{Ge}_x$  2D heterolayer is grown on top of a bulk Si (001) substrate then the inplane strain of the heterolayer is given by

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{\partial u_x}{\partial x} = \frac{a_{\text{Si}} - a_{\text{SiGe}}}{a_{\text{SiGe}}},$$

where  $u_x$  is the lattice displacement,  $a_{\text{Si}}$  is the lattice constant of the Si and  $a_{\text{SiGe}}$  is the lattice constant of the relaxed, bulk  $\text{Si}_{1-x}\text{Ge}_x$  layer. The lattice constants of bulk, relaxed  $\text{Si}_{1-x}\text{Ge}_x$  alloys have a small deviation from Vegard's law and can be calculated (Kasper and Lyutovich 2000) using

$$a_{\text{Si}_{1-x}\text{Ge}_x} = 0.5431 + 0.01992x + 0.002733x^2 \text{ (nm)}.$$

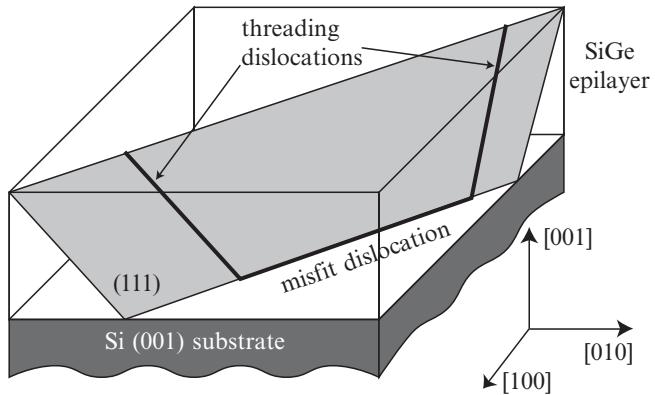
The lattice changes from cubic (diamond) to tetragonal with the applied compressive strain and the perpendicular strain is given by

$$\varepsilon_{xx} = \frac{-2\nu}{1-\nu} \varepsilon_{xx},$$

where  $\nu$  is Poisson's ratio (0.28 for Si). For such a pseudomorphic heterolayer, the stored elastic energy as a function of  $x$  for a layer thickness,  $h$  is given by

$$E \sim h\varepsilon_{xx}^2 = 0.0016xh.$$

Provided the misfit between the epitaxially grown heterolayer and the substrate is small, the epilayer will be coherent and lattice matched to the



**Fig. 5.1** A schematic diagram of a misfit dislocation with the two threading arms that propagate along the {111} planes.

substrate. As the thickness of the epitaxial layer increases, however, there exists a maximum thickness, called the critical thickness,  $h_c$  above which it costs too much energy to elastically strain addition heterolayers in coherence with the substrate. Above the critical thickness, defects will appear to relieve the elastic strain, in this case misfit dislocation with threads on the {111} planes. These threading dislocations are at  $60^\circ$  to the (001) growth plane and are shown in Fig. 5.1. The heterolayer relaxes and the resulting defects can degrade the optical and electrical performance of any devices fabricated on the material.

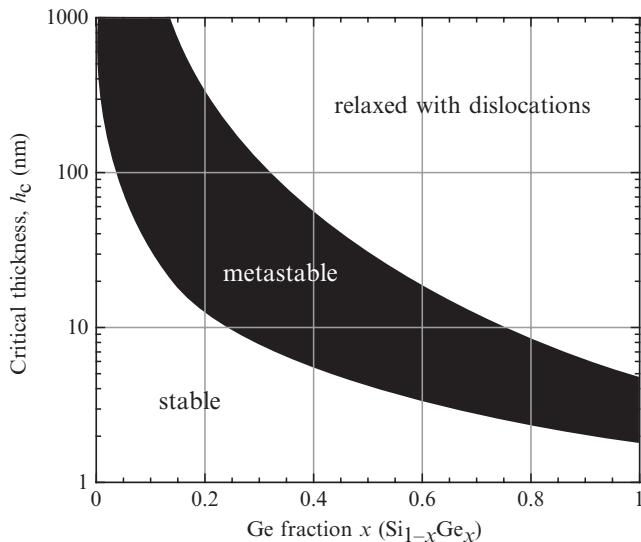
There are a number of different derivations of critical thickness in the literature but the most widely used is that by Matthews and Blakeslee (1976). The critical thickness was calculated by balancing the force of the threading dislocation arm (Fig. 5.1) in an uncapped epitaxial layer with the restoring line tension force from the self-energy of the extra interfacial dislocation created during relaxation, which when solved produces

$$h_c \simeq \frac{b}{2\pi f} \frac{(1 - \nu \cos^2 \theta)}{(1 + \nu) \cos \lambda} \left[ \ln \left( \frac{h_c}{b} \right) + 1 \right].$$

In this equation  $b$  is the Burger's vector ( $=0.4\text{nm}$ ),  $\theta$  is the angle between the dislocation line and its Burger's vector,  $\lambda$  is the angle between the Burger's vector and the direction in the interface normal to the dislocation direction and  $f$  is the misfit between the heterolayer and the substrate ( $= (a_{\text{SiGe}} - a_{\text{Si}})/a_{\text{Si}}$ ). This can be reduced to a more workable form by taking the misfit between pure Si and Ge as  $f = 0.0418$ ,  $\cos \lambda = \cos \theta = 0.5$  for  $60^\circ$  dislocations to produce

$$h_c \simeq \frac{0.55}{x} \ln (10h_c) \quad (h_c \simeq 1.7793x^{-1.2371} \text{ nm}).$$

The solution to this equation is plotted in Fig. 5.2. For some heterolayers grown at low temperatures, it has been found that this equilibrium critical thickness can be exceeded without significant relaxation of the heterolayer. These metastable epilayers have been investigated by People and Bean (1985a,b) and

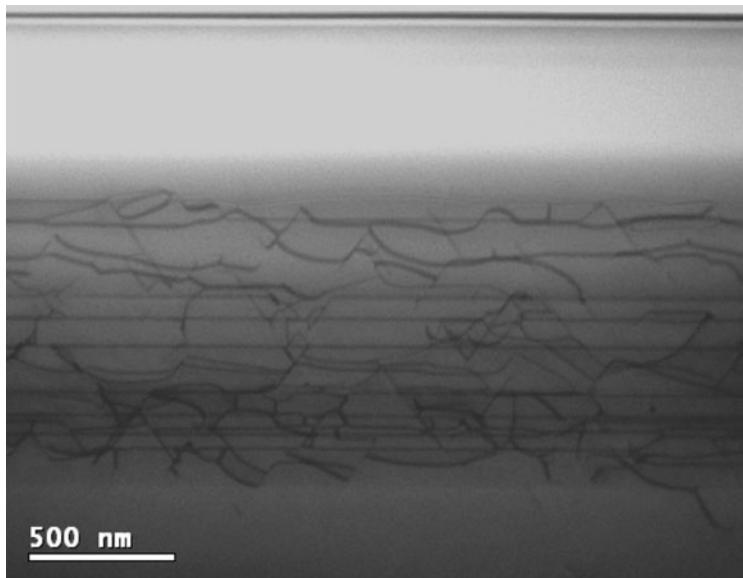


**Fig. 5.2** The Matthews and Blakeslee critical thickness as a function of Ge content,  $x$  for pseudomorphic growth of a  $\text{Si}_{1-x}\text{Ge}_x$  heterolayer on top of a bulk Si (100) substrate. Also included is a metastable curve for MBE material grown at 550 °C from People and Bean (1985a,b).

the typical critical thickness for heterolayers grown at 550 °C in MBE systems is plotted in Fig. 5.2. While such layers are frequently grown and used in research they are thermodynamically metastable and at some stage will relax. Therefore, all industrial designs of SiGe devices will always be designed with strained heterolayers below the equilibrium critical thickness.

As will be shown later in the chapter, there are also many applications for tensile-strained layers where a  $\text{Si}_{1-x}\text{Ge}_x$  layer is grown on a substrate with a larger lattice constant i.e.  $\text{Si}_{1-y}\text{Ge}_y$  where  $x < y$ . To achieve this, a strain-relaxed buffer layer must be grown to provide a relaxed  $\text{Si}_{1-y}\text{Ge}_y$  “virtual” substrate for a subsequent heterolayer to be grown onto. If a single  $\text{Si}_{1-y}\text{Ge}_y$  layer is grown well above the critical thickness and annealed, then a near-relaxed buffer (>90% relaxation) will result, but it will have threading dislocation densities through the surface up to  $10^{12} \text{ cm}^{-2}$ , which is far too high for any applications. The solution is to slowly grade the Ge content (Fig. 5.3), which results in significantly lower threading dislocation densities. Provided the growth is at high enough growth rates and temperatures, then the threading dislocations glide to the edges of the wafer provided the grading rate is low enough so that the misfits and threads do not pile up with each other. The best virtual substrates of composition  $\text{Si}_{0.8}\text{Ge}_{0.2}$  that are commercially available have threading dislocation densities as low as  $5000 \text{ cm}^{-2}$ . For higher Ge compositions or if an aggressive grading rate is used, e.g. >10% Ge per  $\mu\text{m}$  (see Paul 2004) then the dislocation density can be significantly higher. These graded buffers have been an enabling technology for many research devices and will be used in a number of the device sections later in the chapter.

There are also attempts at producing thin virtual substrates as the graded relaxed buffers layers require to be many  $\mu\text{m}$  thick for low defect densities and smooth surfaces. Techniques that are being investigated include

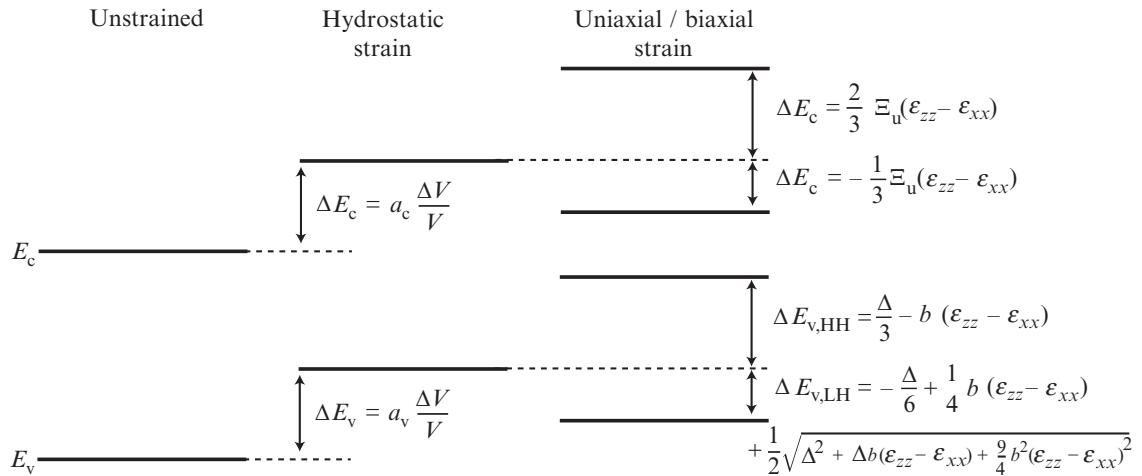


**Fig. 5.3** A transmission electron micrograph of a graded Si<sub>1-x</sub>Ge<sub>x</sub> buffer grown on top of a Si (001) substrate. The bottom of the buffer has  $x = 0$ , while the top of the buffer has a constant composition of Si<sub>0.8</sub>Ge<sub>0.2</sub>. It is clear from the image that very few threading dislocations thread through the surface of the wafer. A large number of misfit segments are visible in the graded part of the buffer.

He implantation and Ge condensation (see Paul 2004) and while some promising results are emerging, there is no clear winner yet. Also, bond-and-etchback techniques have been used to form relaxed Si<sub>1-y</sub>Ge<sub>y</sub> on SiO<sub>2</sub> with a silicon substrate. Silicon-on-insulator (SOI) has been used in microelectronics for years and now tensile-strained Si on insulator (SSOI), relaxed Si<sub>1-y</sub>Ge<sub>y</sub> on insulator (SiGeOI) and germanium-on-insulator (GeOI) are available at least at a research level as starting substrates. These substrates remove the highly defected graded part of the buffer and leave behind a high-quality device layer.

## 5.4 Band structure

The band structure of heterodevices is fundamental in the design, operation and performance of electronic devices. For Si<sub>1-x</sub>Ge<sub>x</sub> alloys, the strain can have a significant impact on the band structure and properties of devices. Strain can lift the degeneracy of degenerate bands and can also change the band alignments between two different materials at a heterointerface. There are two components of strain that are important. Hydrostatic strain (the change in the volume of the material) will change the position of a band while uniaxial or biaxial strain will split the degeneracy of bands (Fig. 5.4). Uniaxial strain is strain only along one direction, while biaxial strain is applied along two directions in a plane. A set of parameters called the deformation potentials define the changes to each band through hydrostatic and uniaxial or biaxial strain and these are defined in Fig. 5.4. Table 5.1 provides a summary of these parameters for the elements Si and Ge taken from predominantly experimental values used in **k.p**-modelling of Si<sub>1-x</sub>Ge<sub>x</sub> heterostructures (Paul 2008). For Si<sub>1-x</sub>Ge<sub>x</sub> alloys, interpolation schemes are used to find the



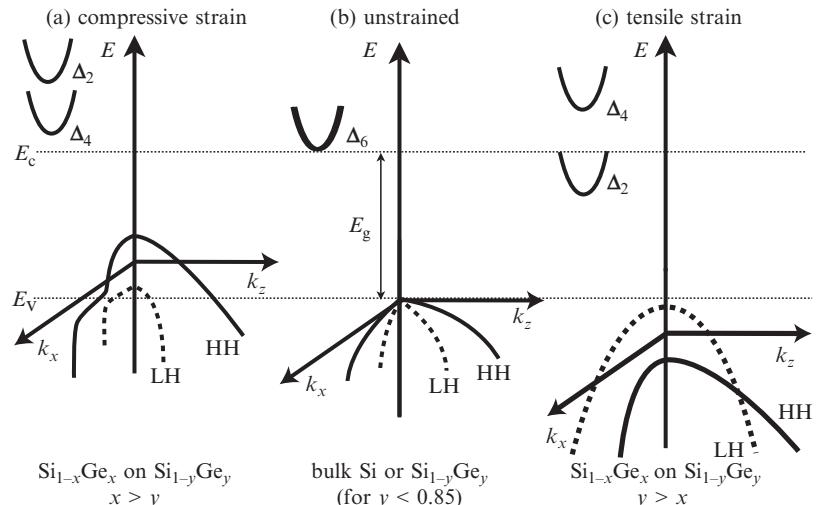
**Fig. 5.4** The changes to the conduction band and valence bands through the application of hydrostatic and uniaxial or biaxial strain applied along the <100> directions.

appropriate deformation potentials that range from linear to parabolic interpolation. For the conduction band, different values are required for the  $\Gamma$ -, L- and  $\Delta$ -valleys.

Figure 5.5 shows the bands at the valence and conduction band edges of  $\text{Si}_{1-x}\text{Ge}_x$  alloys under (a) compressive-strain (b) unstrained and (c) tensile-strain conditions. For the unstrained bands, up to a Ge content of 0.85, the band structure is Si-like with the conduction-band minima being 6 degenerate bands at the  $\Delta$  points of the Brillouin zone forming six  $\Delta$ -valleys. For compressive strain, corresponding to the active heterolayer having a larger Ge content than the substrate, these six degenerate valleys are split with 4 of the bands lower in energy and 2 of the bands higher in energy (Fig. 5.5(a)).

**Table 5.1** Lattice constants, elastic constants, energy gaps and deformation potentials required to calculate the strained- $\text{Si}_{1-x}\text{Ge}_x$  bandstructure (from Paul 2008).

Parameter	Silicon	Germanium
$a$ (nm)	0.54304	0.56579
$c_{11}$ (GPa)	165.77	128.53
$c_{12}$ (GPa)	63.93	48.28
$c_{44}$ (GPa)	79.62	66.80
$E_g$ at 291 K (eV)	1.1135	0.664
$\Delta$ (eV)	0.044	0.289
$a_v$ (eV)	1.80	1.24
$b$ (eV)	-2.10	-2.86
$a_c^\Gamma$ (eV)	-10.39	-10.41
$a_c^L$ (eV)	-0.66	-1.54
$a_c^X$ (eV)	3.3	2.55
$\Xi_u^\Gamma$ (eV)	0	0
$\Xi_u^L$ (eV)	16.14	16.2
$\Xi_u^\Delta$ (eV)	8.6	9.42

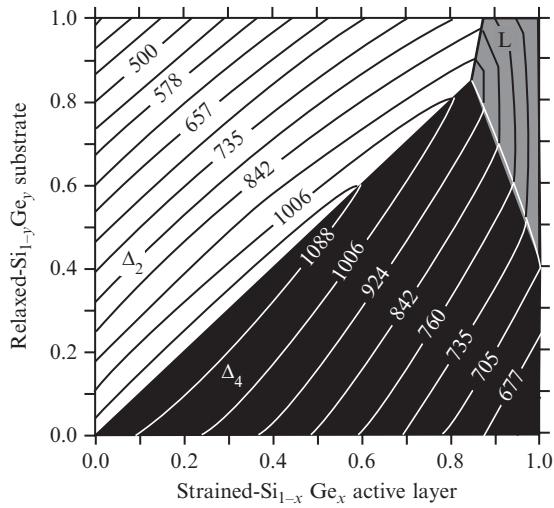


**Fig. 5.5** The bands at the valence and conduction band edges for (a) compressive strain, (b) unstrained and (c) tensile-strained  $\text{Si}_{1-x}\text{Ge}_x$  heterolayers grown on  $\text{Si}_{1-y}\text{Ge}_y$  substrates.

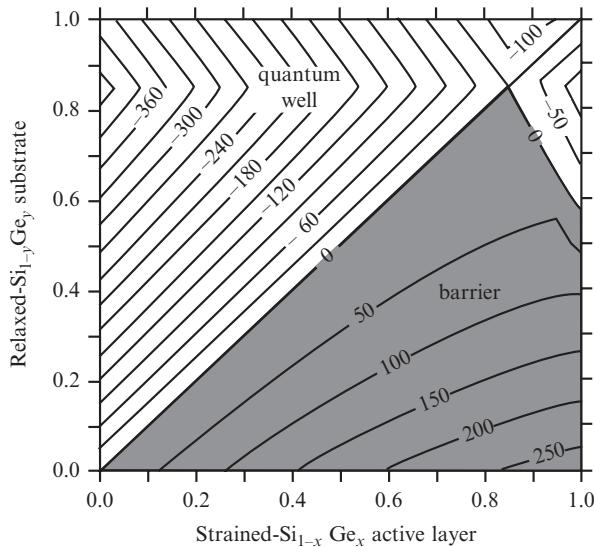
For tensile strain, corresponding to the active heterolayer having a lower Ge content than the substrate, the situation is reversed with the  $\Delta_2$  valleys lowest in energy (Fig. 5.5(c)). For electrons, the tensile strain results in significantly higher mobility at room temperature since electrons only occupy 2 rather than  $6\Delta$ -valleys and the reduced intervalley scattering significantly increases the mobility when there is sufficient strain that the valleys are split by  $>10 k_B T$  where  $k_B$  is Boltzmann's constant and  $T$  is the temperature. The valence band is also changed by the application of strain. For compressive strain the heavy hole (HH) state is the lowest in energy with the light-hole (LH) band pushed to higher energy. For tensile strain, the LH band becomes the lowest ground state.

The bandgap of  $\text{Si}_{1-x}\text{Ge}_x$  alloys is therefore dependent on both the Ge content of the heterolayer but also the amount of strain (i.e. dependent on the Ge content of the relaxed  $\text{Si}_{1-y}\text{Ge}_y$  substrate). Below Ge contents of  $\sim 0.85$ , the conduction-band minima are Si-like with  $\Delta$ -valleys along the  $\langle 100 \rangle$  directions being the conduction-band edge. Above  $\sim 0.85$  Ge content, the L-valleys along the  $\langle 111 \rangle$  directions become the minima, which is the case for bulk Ge. The band structure is therefore frequently described as Ge-like. This is shown in Fig. 5.6 for all possible levels of strain for heterolayers when grown on a relaxed  $\text{Si}_{1-y}\text{Ge}_y$  substrate (Rieger and Vogl 1993, 1994; Paul 2004). The bandgaps are reduced as the Ge content (and strain) are increased in the heterolayers.

It is not just the bandgaps that are important for heterolayers. The alignment of bands at a heterointerface between two different  $\text{Si}_{1-x}\text{Ge}_x$  alloys is also important for the design of devices. The alignment for the conduction-band edge is shown in Fig. 5.7, while the alignment for the valence-band edge is shown in Fig. 5.8 (Rieger and Vogl 1993, 1994; Paul 2004). For all these band structure graphs mapping out alignments and bandgaps, not all Ge content heterolayers can be practically used as the thickness of heterolayers is limited by the critical thickness.



**Fig. 5.6** The bandgap in meV for  $\text{Si}_{1-x}\text{Ge}_x$  alloys under different levels of strain. The white region has a Si-like band structure with the  $\Delta_2$  valleys along  $\langle 100 \rangle$  directions lowest in energy while the black region has a Si-like band structure with the  $\Delta_4$  valleys lowest in energy. The gray region has a Ge-like band structure where the L-valleys along the  $\langle 111 \rangle$  directions are the minima in energy.

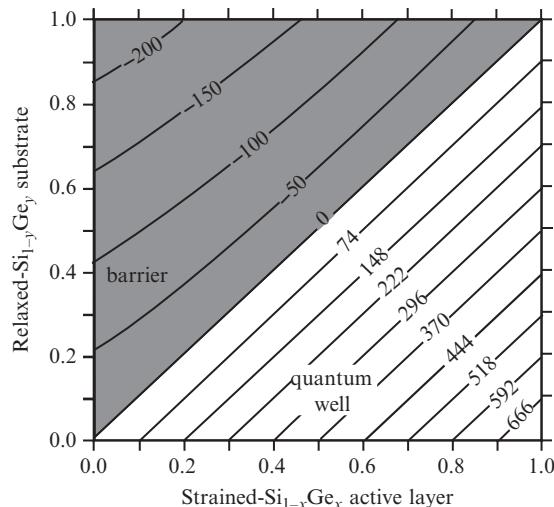


**Fig. 5.7** The conduction-band discontinuity in meV for strained  $\text{Si}_{1-x}\text{Ge}_x$  grown on relaxed  $\text{Si}_{1-y}\text{Ge}_y$ .

## 5.5 Mainstream nanoelectronic applications

### 5.5.1 Heterojunction bipolar transistors (HBTs)

The first major use of SiGe alloys in a microelectronics application was in the heterojunction bipolar transistor (HBT) that was first demonstrated by Patton *et al.* (1988) and entered production in late 1998. There are many review articles on the subject (Taur and Ning 1998; Meyerson 2000; Jain *et al.* 2001). The addition of a small amount of Ge to the base of Si n-p-n bipolar transistor will reduce the bandgap in the base of the transistor. This reduction of energy significantly improves the injection efficiency of electrons from the emitter into the base as most of the reduction of bandgap occurs in the conduction

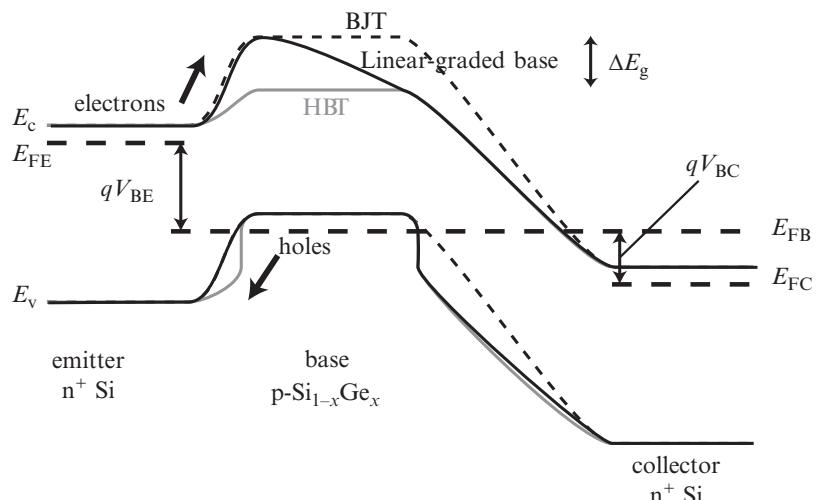


**Fig. 5.8** The valence-band discontinuity in meV for strained  $\text{Si}_{1-x}\text{Ge}_x$  grown on relaxed  $\text{Si}_{1-y}\text{Ge}_y$ .

band (Fig. 5.9). The full advantages of the addition of Ge to the base requires a close look at the equations governing the operation of a bipolar transistor. The collector current density for any n-p-n bipolar transistor in the common emitter configuration is given by

$$J_c = -\frac{q D_n n_i^2 B}{W_B N_A} \left[ \exp \left( \frac{q V_{BE}}{k_B T} \right) \right],$$

where  $q$  is the electron charge,  $D_n$  is the diffusion constant for minority electrons in the base,  $n_i$  is the intrinsic carrier density in the base,  $W_B$  is the base width,  $N_A$  is the acceptor doping density in the base and  $V_{BE}$  is the voltage applied across the base emitter interface (see Fig. 5.9). The collector current density for a HBT with a constant Ge composition base  $\text{Si}_{1-x}\text{Ge}_x$  can



**Fig. 5.9** A schematic diagram of the band structure of a silicon n-p-n bipolar transistor with a SiGe HBT and a linearly graded  $\text{Si}_{1-x}\text{Ge}_x$  base transistor drawn on top for comparison.

be found simply by replacing the intrinsic carrier density in the base,  $n_{iB}$  by an effective carrier density given by

$$n_{iB,\text{eff}}^2(\text{SiGe}) = n_{iB}^2(\text{Si}) \exp\left(\frac{\Delta E_{gB}}{k_B T}\right),$$

which includes a contribution from the reduction of the bandgap,  $\Delta E_g$ , due to the addition of Ge in the base. This change in bandgap has been shown to be experimentally (Paul 2004)

$$\Delta E_{gB} = 430x - 20.6x^2 \text{ (meV).}$$

Therefore, the addition of a small amount of Ge to the base of a bipolar transistor results in an exponential increase in the collector current density relative to the reduction of the bandgap of the base.

The collector current density is not the only important parameter for bipolar transistor design. The gain of the transistor,  $\beta$  is given by the collector current density divided by the base current density and for an HBT is

$$\beta = \frac{D_n N_D L_p}{D_p N_A W_B} \exp\left(\frac{\Delta E_{gB}}{k_B T}\right),$$

where  $L_p$  is the diffusion length of holes in the emitter,  $N_D$  is the emitter doping density and  $D_p$  is the diffusion constant for minority holes in the emitter. Again, the gain of the transistor depends exponentially on the change of bandgap with the addition of Ge.

While the constant composition  $\text{Si}_{1-x}\text{Ge}_x$  base transistor does not increase the base transit time, which is the most important parameter in determining the unity gain cutoff frequency ( $f_T$ ), the smaller bandgap in the base allows higher doping and narrower bases compared to a standard Si bipolar for the same parameters and this results in higher operational frequencies. The constant compositional base HBT is not the only possible design. A linearly graded base  $\text{Si}_{1-x}\text{Ge}_x$  allows a built-in electric field to be designed across the base region, which accelerates electrons, increasing the  $f_T$  of the transistor (see Fig. 5.9). Many companies use the linearly graded  $\text{Si}_{1-x}\text{Ge}_x$  base in production. The calculation of the performance improvements for the linearly graded base transistor are much more involved compared to the constant composition. Table 5.2 shows the enhancements for both the constant composition base (HBT) and also the linearly graded base  $\text{Si}_{1-x}\text{Ge}_x$  transistors compared to standard Si bipolar devices. While the enhancements in gain and collector current densities for the linearly graded base  $\text{Si}_{1-x}\text{Ge}_x$  transistor are not as high as for the constant composition base HBT, the lower values are actually more appropriate for many analog applications and result in transistor performance that is easier for circuit designers to use. The linearly graded base  $\text{Si}_{1-x}\text{Ge}_x$  transistor also has a significant improvement in the Early voltage ( $V_A$ ), which is extremely important for low-noise analog circuits.

SiGe HBTs in production now have  $f_T$  and  $f_{\max}$  values around 280 GHz and provide direct competition for many III-V HEMTs and HBTs especially for analog markets. In research,  $f_T$  values up to 350 GHz have been demonstrated (Khater *et al.* 2004) and the thickness of the  $\text{Si}_{1-x}\text{Ge}_x$  heterolayer in the base

**Table 5.2** A table summarizing the changes to the main bipolar parameters for a box-profile SiGe HBT and also a linearly graded base HBT.

Parameter	SiGe HBT	Linear-graded SiGe base
$\frac{n_{iB}^2(\text{SiGe})}{n_{iB}^2(\text{Si})}$	$\exp\left[\frac{\Delta E_g}{k_B T}\right]$	$\exp\left[\frac{\Delta E_g}{k_B T} \frac{z}{W_B}\right]$
$\frac{J_C(\text{SiGe})}{J_C(\text{Si})}$	$\exp\left[\frac{\Delta E_g}{k_B T}\right]$	$\frac{\Delta E_g}{k_B T} \left(1 - \exp\left[\frac{-\Delta E_g}{k_B T}\right]\right)^{-1}$
$\frac{\beta(\text{SiGe})}{\beta(\text{Si})}$	$\exp\left[\frac{\Delta E_g}{k_B T}\right]$	$\frac{\Delta E_g}{k_B T} \left(1 - \exp\left[\frac{-\Delta E_g}{k_B T}\right]\right)^{-1}$
$\frac{V_A(\text{SiGe})}{V_A(\text{Si})}$	1	$\frac{k_B T}{\Delta E_g} \left(\exp\left[\frac{\Delta E_g}{k_B T}\right] - 1\right)$
$\frac{\tau_B(\text{SiGe})}{\tau_B(\text{Si})}$	1	$\frac{2k_B T}{\Delta E_g} \left(1 - \frac{k_B T}{\Delta E_g} \left(1 - \exp\left[\frac{\Delta E_g}{k_B T}\right]\right)\right)$

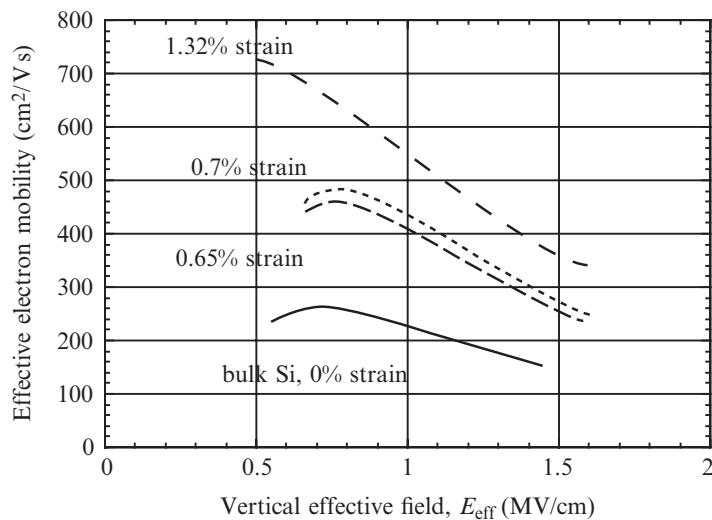
has now been scaled to below 10 nm. Further scaling of these devices is likely to further increase the performance.

### 5.5.2 CMOS

The largest section of the micro- and nanoelectronics market is in the production of CMOS circuits. For the last 40 years, the gate length on transistors,  $L_g$  has been scaled to smaller dimensions to improve the on-current of the transistor,  $I_{on}$  for a given gate width  $W$  as

$$I_{on} \approx \frac{W}{2L_g} \mu C_{ox} (V_g - V_T)^2.$$

Here,  $\mu$  is the mobility of the carriers in the channel,  $V_g$  is the gate voltage applied to the transistor and current flows or the transistor is switched on when  $V_g$  is above the threshold voltage,  $V_T$ . The performance of the transistor can also be improved by increasing the gate to channel capacitance,  $C_{ox}$ , which can be undertaken by scaling the gate oxide to smaller dimensions. This scaling was driven by Moore's law, in which in 1965, Gordon Moore pointed out that the size of transistors has to scale at approximately  $\times 0.7$  (i.e.  $\times 1/\sqrt{2}$ ) every year if companies were to make money. Moore's law is an economic law and has nothing to do with physics. Even today, 40 years on, the scaling of transistors is still being pursued and transistor gate lengths were 35 nm in production in 2008. The road to such gate lengths has not been completely smooth and new materials have been integrated into MOSFETs as the gate lengths have reduced. The present-day 35 nm gate length transistor uses process-induced uniaxial strain across the MOSFET channel to improve the  $I_{on}$  through the improvement of mobility with strain. For n-MOSFETs, tensile strain applied to Si results in the splitting of the  $\Delta$ -valleys with the electrons in the MOSFET channel only occupying the  $\Delta_2$  valleys. This results in reduced intervalley scattering that produces higher mobilities. Figure 5.10 shows experimental results from MOSFETs where tensile strain

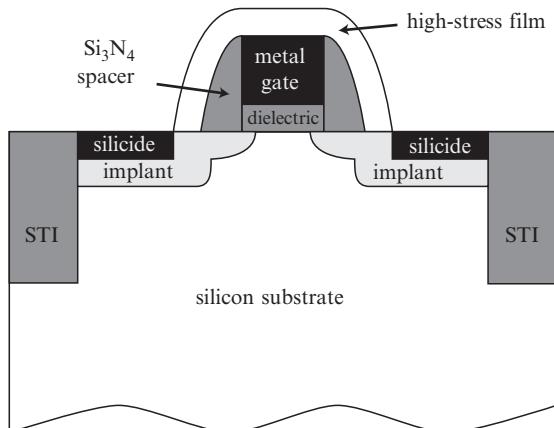


**Fig. 5.10** The effective electron mobility for n-MOSFETs as a function of vertical electric field for different levels of applied tensile strain at 300 K.

has been applied to the Si channel in MOSFETs and shows how the mobility can be increased by over 100% at room temperatures just by the addition of strain. For electrons the effective mass is unchanged by the application of strain.

The p-MOSFET is much more complicated due to the multiple HH and LH bands that are degenerate for unstrained devices at  $k_{\parallel} = 0$ . The application of strain not only splits the degeneracy at  $k_{\parallel} = 0$  but also changes the  $k_{\parallel}$  dispersion, resulting in different effective masses for holes. The changing effective mass can result in improved mobility. Typically, both tensile and compressive strain will increase the hole mobility but the changes are larger for compressive strain. p-MOSFETs are now made with  $\text{Si}_{1-x}\text{Ge}_x$  ohmic contacts, either through implanting a small amount of Ge with the boron p-type doping, or removing the Si at the sides of the transistor and regrowing a p- $\text{Si}_{1-x}\text{Ge}_x$  contact. Both techniques result in compressive uniaxial strain across the channel for transistors with gate lengths below 100 nm, which improves the mobility and  $I_{\text{on}}$  of the devices.

For transistors with gate lengths below 100 nm, process-induced strain can be invoked by many different techniques and technologies. Figure 5.11 shows the cross-section of a typical 35-nm gate length MOSFET with low-doped drain extensions at each side of the MOSFET channel to reduce the electric field across the device. The implants have already been discussed as being used to provide uniaxial strain across the channel through the addition of Ge for the p-MOSFET but the silicide that is used to reduce the contact and access resistance to the transistor can also invoke strain. Shallow trench isolation (STI) is required to provide electrical isolation between closely spaced transistors to prevent latch-up but it can be used to provide both tensile or compressive strain to the channel, depending on which crystallographic orientation is used for the trench. Even high-stress films such as silicon nitride and BPSG deposited over the top of the gate stack can provide process-induced strain to the channel to improve  $I_{\text{on}}$ .



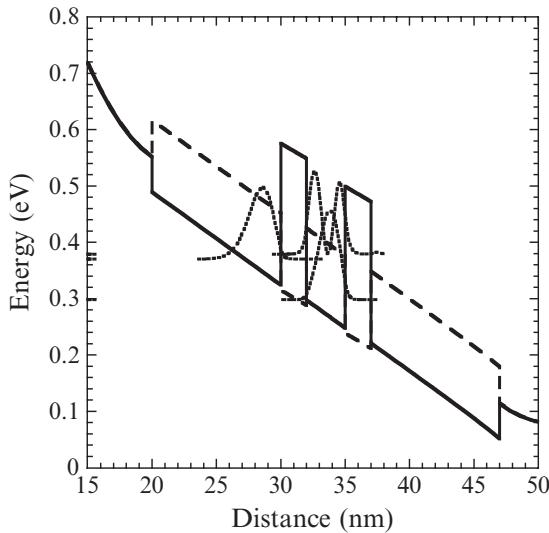
**Fig. 5.11** A schematic diagram of a modern deep submicrometer MOSFET. Shallow trench isolation (STI), Ohmic contact implants, silicides and high-stress films can all be used to engineer process-induced strain into the MOSFET channel.

While many people talking about Moore's law describe the scaling of the gate length in MOSFETs, the gate oxide has to be scaled as well. For the 65-nm technology node MOSFET devices, the gate oxide was only 1.2 nm thick (about 5  $\text{SiO}_2$  molecules thick) and such thin oxides are subject to large leakage currents as carriers can quantum-mechanically tunnel through this thin insulator. Therefore, Intel has introduce a high- $k$  dielectric to the 45-nm technology node devices that has an effective oxide thickness of less than 1 nm, while maintaining a physical thickness of 2 nm. This has the advantage of reducing the gate leakage current by up to 4 orders of magnitude, which has a major impact on the static power dissipation of MOSFET nanoelectronic circuits.

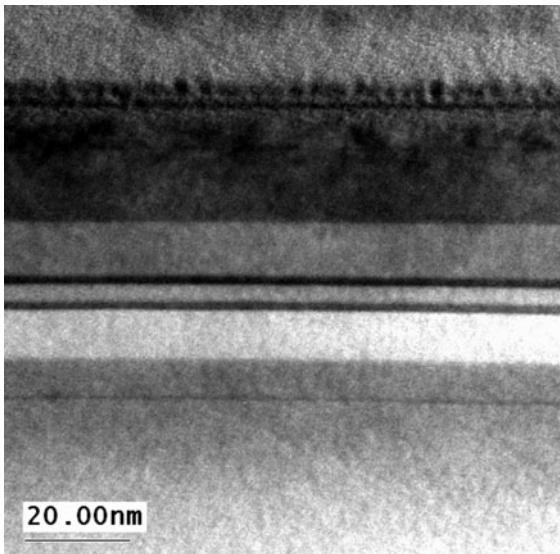
## 5.6 Resonant tunnelling diodes

Resonant tunnelling diodes (RTDs) are a true quantum nanoelectronic device and operate using quantum-mechanical tunnelling. The device is fabricated using two tunnel barriers with a quantum well sandwiched between the barriers. Electrons can only tunnel through the whole device when the chemical potential of the source contact is aligned to a subband state in the quantum well. Therefore, electrons can only tunnel from source to drain when the source contact is resonant with a subband state in the central quantum well.

RTDs are very common in the III-V material systems but as the diodes are only two terminal, most useful nanoelectronic circuit designs require the RTDs to be integrated with transistors to form tunnelling static random access memories (TSRAMs) or logic circuits (van der Wagt 1999 and Mazumder *et al.* 1998). The band structure in the SiGe system does not make it easy to design RTDs similar to III-V devices such as GaAs/AlGaAs RTDs. Figure 5.12 shows a Si/SiGe RTD structure grown on a relaxed  $\text{Si}_{0.8}\text{Ge}_{0.2}$  virtual substrate and consisting of a 3-nm tensile-strained Si quantum well sandwiched between two 2-nm  $\text{Si}_{0.4}\text{Ge}_{0.6}$  barriers. On each side of the barriers are 10-nm tensile-strained Si layers that provide a larger conduction-band discontinuity to produce higher barriers. The figure also shows the first two subband states in the



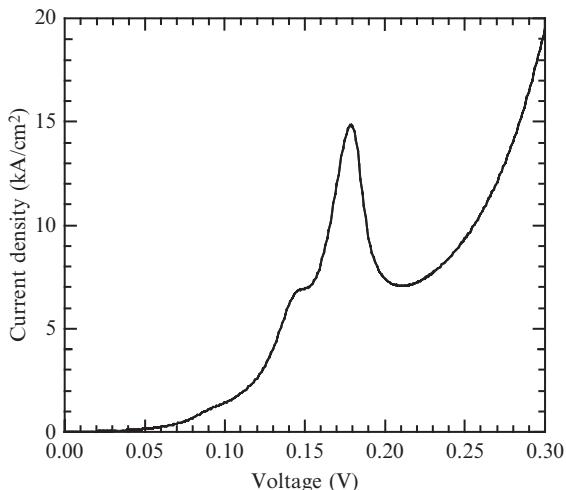
**Fig. 5.12** The conduction-band edge for a Si/SiGe RTD with 2-nm  $\text{Si}_{0.4}\text{Ge}_{0.6}$  barriers and a 3-nm Si quantum well.



**Fig. 5.13** A transmission electron micrograph of the 2-nm  $\text{Si}_{0.4}\text{Ge}_{0.6}$  barriers and a 3-nm Si quantum well in a Si/SiGe RTD. 10-nm Si cladding layers are used either side of the RTD structure to improve the barrier height.

quantum well and the subband state produced in the 10-nm tensile-strained Si spacer well. The tensile-strained Si layers have the  $\Delta_2$ -valleys as the ground state and the  $\Delta_4$ -valleys are raised in energy. A TEM image of the device grown by CVD is shown in Fig. 5.13.

Figure 5.14 shows the current–voltage characteristics for a  $20\text{-}\mu\text{m}$  diameter mesa device measured at 77 K. A strong resonance peak occurs when the lowest subband state in the 10-nm tensile-strained Si spacer well becomes resonant with the first subband state in the 3-nm central quantum well and electrons can quantum-mechanically tunnel through the device from source to drain. The drop in the current after the resonance results in a region of negative differential resistance (NDR) that is extremely useful for circuits. By biasing the RTD in



**Fig. 5.14** The current–voltage characteristics of the device in Fig. 5.13 showing clear NDR at 77 K.

this NDR region, the diode will oscillate, and for small diodes this oscillation frequency can be hundreds of GHz. RTDs to date are the fastest electron transit time device with oscillations recorded up to 712 GHz (Brown *et al.* 1991). The highest performance of Si/SiGe RTDs was demonstrated by See and Paul (2001) with a peak current density of 282 kA/cm<sup>2</sup> and peak to valley current ratio of 2.4 at room temperature. This performance is comparable to many of the best III–V demonstrated diodes designed with high peak current densities as required for high-frequency operation.

Perhaps the area where a SiGe RTD could have the most impact is in memory. TSRAMs require two RTDs to be integrated with a field effect transistor. The RTD is also one of the few proposed quantum devices that has achieved room-temperature and circuit operation and that now appears in the ITRS Roadmap for potential future device concepts. One of the major limitations in microprocessor performance is the static random access memory (SRAM) on the chip, which requires 6 CMOS transistors and consumes substantial amounts of static power. While 1 transistor/1 capacitor dynamic RAM (DRAM) is more compact, it is difficult to integrate the capacitor with a high-yield microprocessor fabrication technology without reducing the performance of the logic, and DRAM is also much slower than CMOS SRAM. In addition, the memory has to be refreshed, while SRAM has the advantage of not requiring refresh cycles that consume large amounts of power. The tunnelling SRAM (TSRAM) of 2 RTDs integrated with 1 transistor has previously demonstrated both fast and low-power memory (van der Wagt 1999) ideal for microprocessors and the transfer of this into CMOS technology has been modelled by Christian Pacha (Infineon) for the EC Technology Roadmap of European Nanoelectronics (Compano 2000) (see Table 5.3). Si-based tunnel diodes coupled with strained-Si MOSFETs (see CMOS section above) have the potential to operate at the same speed as CMOS SRAM but with integration densities closer to DRAM. This is coupled to a reduction in standby power by 7 orders of magnitude compared to CMOS SRAM and 5 orders of magnitude reduction compared to DRAM. As power dissipation is considered to be one

**Table 5.3** A table comparing TRSAM to other forms of semiconductor memory (Paul 2004).

Technology	Access time	Retention time	Speed (GHz) (1–2 μm pitch)	Density (Mbit/cm <sup>2</sup> )	Standby power (W/Mbit)
16 MB SRAM	1 ns	>10 years	0.03 to 0.1	5 to 10	0.01
256 MB DRAM	10 ns	<256 ms	0.01 to 0.02	60 to 150	$2 \times 10^{-4}$
256 MB flash	10 ns	>10 years	0.01 to 0.02	60 to 150	
256 MB 1T Si SRAM	10 ns	>10 years	0.01 to 0.04	50 to 100	$10^{-9}$
256 MB 1T strained-Si TSRAM	4 ns	>10 years	0.03 to 0.12	50 to 100	$10^{-9}$

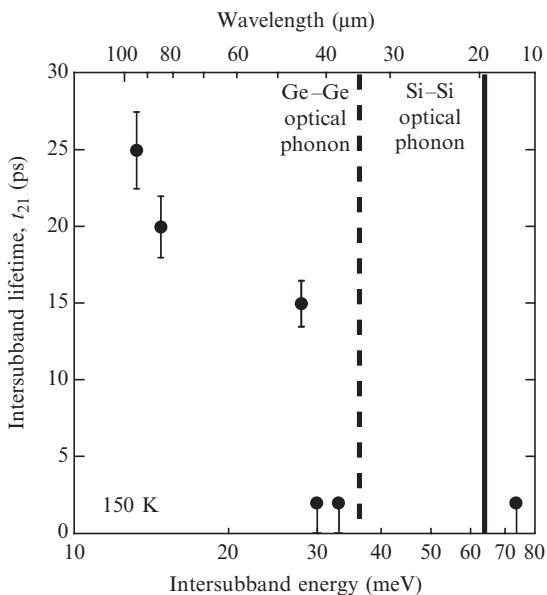
of the main problems as microprocessors are scaled down below 100-nm gate lengths, reduction of the standby power would have significant advantages for microprocessors.

## 5.7 SiGe quantum cascade emitters

The terahertz (THz) region of the electromagnetic spectrum potentially has a large number of applications including medical (Arnone *et al.* 2000) and security imaging, pollution monitoring, proteomics and bioweapons detection. The major limitation to the mainstream use of the technology has been the lack of cheap and practical THz sources. Most application demonstrations to date have used photoconductive antenna with pulsed femtosecond lasers but such systems are still far too expensive for many of the markets THz has the potential to address. The demonstration of GaAs quantum cascade lasers (QCLs) operating at terahertz frequencies (Köhler *et al.* 2002) potentially opens up much cheaper, high-power THz sources but to date these still typically operate with tens of mW power below 100 K (Williams *et al.* 2005). Higher-temperature operation has recently been demonstrated by the use of a double metal-reflector technology, but at the cost of reduced power.

The indirect bandgap of silicon, germanium and silicon-germanium alloys precludes any efficient interband emission of radiation as an appropriate phonon is required before an electron–hole pair can emit radiation. At lower frequencies, intersubband emission of radiation is an efficient method of photon emission without any phonons and the ability to tailor population inversion and gain is not dependent on a direct bandgap. Intersubband lasers in the form of the QCL were first proposed by Kazarinov and Suris (1971) but it was not until 1994 that Faist and colleagues demonstrated a working QCL in the InGaAs materials system. A number of other materials systems have been used to produce mid-infra-red QCLs (for a review see Capasso *et al.* 2002). Köhler *et al.* (2002) expanded the range of demonstrated lasing frequencies with the first demonstration of a QCL at terahertz (far-infra-red) frequencies but there is a gap between the acoustic and optical phonon energies called the restrahlen band where III–V QCLs cannot operate.

Si/SiGe QCLs are one of the technologies presently being investigated in an attempt to produce an electrically pumped Si-based laser. Due to the band structure, band discontinuities and intersubband lifetimes, it is easier to design



**Fig. 5.15** The intersubband lifetimes for LH/SO to HH intrawell transition in  $\text{Si}_{1-x}\text{Ge}_x$  quantum wells at 150 K as measured by far-infra-red pump-probe measurements. Also marked are the optical phonon energies for Ge–Ge and Si–Si bonds.

quantum cascade structures at low energies with emitted photons below the optical phonon energy for Si of 63 meV. Si/SiGe heterostructures have some enormous advantages over III–V materials for terahertz lasers. The non-polar nature of the Si–Si, Si–Ge and Ge–Ge bonds results in a lack of polar optical phonon scattering. Intersubband lifetime measurements of SiGe quantum wells below the optical phonon energy have demonstrated no reduction in the non-radiative lifetimes up to 225 K as expected for the Group IV materials (which have no polar optical phonon scattering (Kelsall *et al.* 2005; Calfano *et al.* 2007). As an example, the low-temperature lifetimes as measured by pulsed pump-probe spectroscopy are shown in Fig. 5.15 for mixed LH/SO to HH transitions inside  $\text{Si}_{1-x}\text{Ge}_x$  quantum wells as a function of subband spacing. As the optical phonon energy of Ge–Ge bonds is reached, the lifetimes are significantly reduced as optical phonon scattering becomes allowed in the Group IV system and the lifetimes quickly reduce to femtoseconds. These lifetimes are also typically longer than those measured in many III–V materials (Murdin *et al.* 1997). In GaAs quantum wells at similar intersubband energies the lifetimes are significantly reduced at temperatures above  $\sim 40$  K due to the polar optical phonon scattering (Murdin *et al.* 1997). This polar optical phonon scattering has also limited the operating temperature of GaAs THz QCLs to below 170 K. What is not so obvious is that the lack of the resonant, polar optical phonons, however, makes population inversion more difficult to design into Group IV systems. Other potential benefits of the Si devices include lower cost, mature processing technology and the integrability with Si microelectronics and foundries. Finally, III–V lasers cannot operate inside the forbidden restrahlen band, which for GaAs is about 8 to 9 THz. Group IV materials can emit radiation between these frequencies and have the potential to lase as well.

Si/SiGe quantum cascade emitters were first demonstrated at mid-infra-red frequencies using heavy-hole (HH) to HH transitions (Dehlinger *et al.* 2000). To date, only SiGe hole cascades have been demonstrated due to the high electron effective mass in the [001] tunnel direction typically around  $0.93 m_0$  ( $m_0$  is the free-electron mass) and n-type dopant segregation issues for gas-based growth systems. The first THz emission from a Si/SiGe cascade was demonstrated using an intrawell light-hole (LH) to HH transition at 2.9 THz and a HH<sub>2</sub> to HH<sub>1</sub> transition at 8.9 THz (Lynch *et al.* 2002). Interwell transitions between two different quantum wells were also demonstrated with the ability to tune the energy using electric fields (Bates *et al.* 2003).

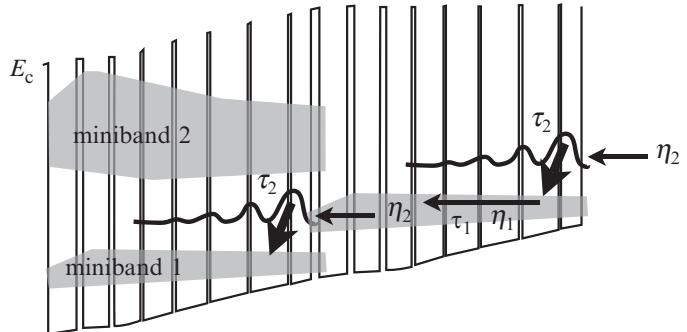
The valence-band intersubband transitions are significantly more complicated than conduction-band transitions especially under the influence of strain. The LH states are really mixed LH/split-off (SO) states and the LH and HH states can also have mixed character, especially when the states are close in energy and away from the Brillouin zone center. Modelling of such strained valence bands requires more complicated techniques than simple effective-mass Poisson–Schrödinger tools such as tight-binding, pseudopotential or  $\mathbf{k}\cdot\mathbf{p}$  theory. The calculated subband energies and in particular the mixed LH/SO states are strongly dependent on deformation and Dresselhaus parameters. Multiple values for these parameters can be found in the literature (see Table 5.1) and different interpolation schemes are also reported for  $\text{Si}_{1-x}\text{Ge}_x$  alloy values. While the strain and germanium content affect the LH-HH relative energies, the energy difference is also sensitive to non-uniformity, diffusion of Ge, segregation of Ge and strain fluctuations. The HH states are less dependent on Ge composition and strain and are more strongly dependent on the widths of quantum wells. For this reason HH to HH designs and particularly the bound-to-continuum design (Faist *et al.* 2001) are much more tolerant of the physical, chemical, material and technological limits to accurate growth of heterostructures.

One of the most successful designs for III–V QCLs has been the bound-to-continuum (Faist *et al.* 2001) active region that has also been demonstrated in the mid-infra-red in SiGe (Diehl *et al.* 2002). This design uses a miniband as the lower radiative state of the laser to produce fast depopulation through scattering in the miniband and increases the injection efficiency into the upper laser radiative states by having a single bound state. Figure 5.16 shows the typical structure of two periods of an eight quantum well bound-to-continuum quantum cascade active region for electrons. The gain in such a structure is defined in terms of the cross-section,  $\sigma$  and the change in carrier density,  $\Delta n$  as

$$\text{Gain} = \sigma \Delta n = \sigma \frac{J}{e} \left[ \eta_2 \tau_2 \left( 1 - \frac{\tau_1}{\tau_{21}} \right) - \eta_1 \tau_1 \right].$$

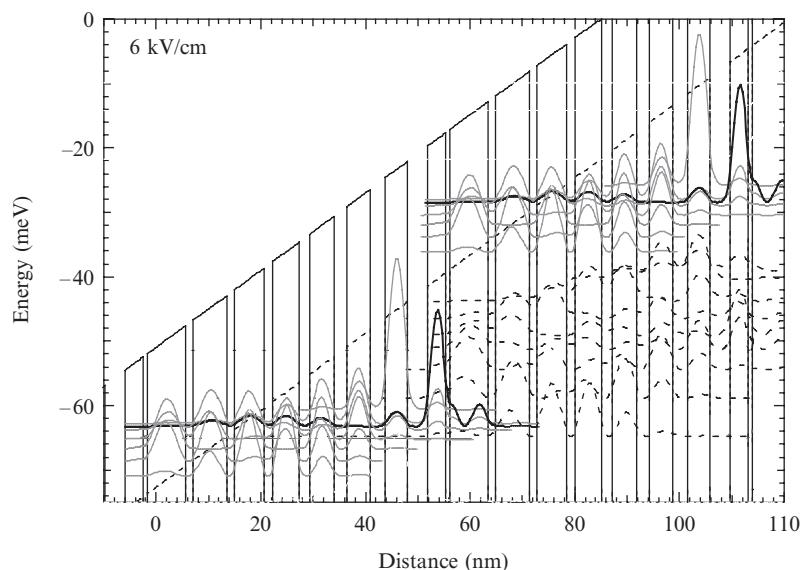
Here,  $J$  is the current in the cascade,  $\eta_2$  is the injection efficiency into the bound upper radiative transition state that has a lifetime,  $\tau_2$ ,  $\eta_1$  is the extraction efficiency in the miniband with lifetime  $\tau_1$  and  $\tau_{21}$  is the intersubband non-radiative lifetime for state 2. The bound-to-continuum design optimizes  $\eta_2 \tau_2$  and minimizes  $\eta_1 \tau_1$  to produce high gain.

An example of a Si/SiGe THz bound-to-continuum structure is shown in Fig. 5.17. The figure shows two periods of an eight quantum well HH to

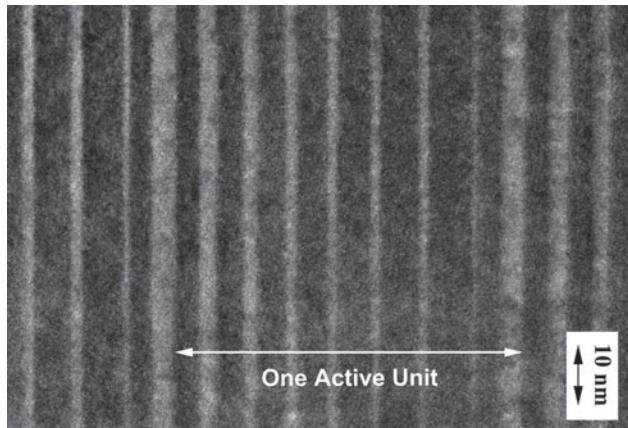


**Fig. 5.16** A schematic diagram of two periods of an eight quantum well bound-to-continuum design.  $\eta_2$  is the injection efficiency into the bound state with lifetime  $\tau_2$  and  $\eta_1$  is the extraction efficiency in the miniband with lifetime  $\tau_1$ .

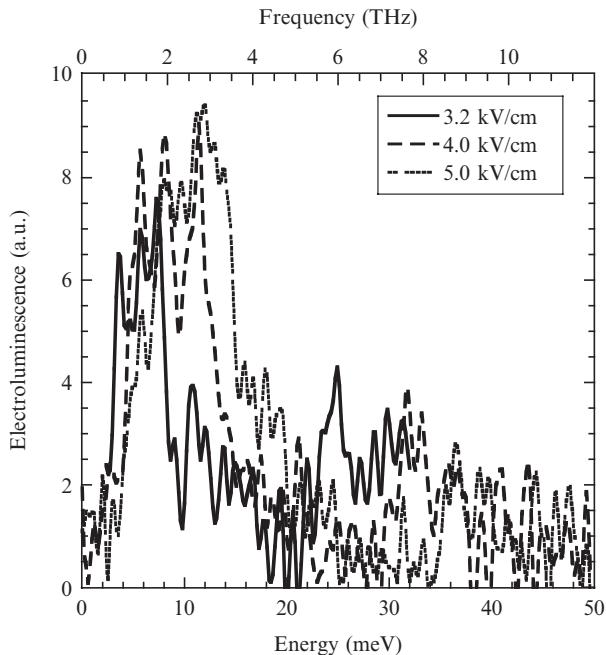
HH design with the holes flowing from the bottom left to the top right of the figure. The solid lines are the HH subband states and the dashed subband states in the figure are the mixed LH/SO states. The quantum wells are compressive  $\text{Si}_{0.723}\text{Ge}_{0.277}$  and the barriers are tensile-strained Si that have been designed so that they are strain symmetrized with an average lattice constant equivalent of  $\text{Si}_{0.8}\text{Ge}_{0.2}$ : therefore the structure is grown on a  $\text{Si}_{0.8}\text{Ge}_{0.2}$  virtual substrate. This balancing of the strain is extremely important as it allows an infinite superlattice to be grown well above the critical thicknesses for the compressive and tensile layers. The layer thicknesses in nm starting with the injection barriers (barriers in plain, quantum wells in bold) are 3.7/3.6/0.8/7.3/1.4/6.4/1.5/5.6/1.7/5.0/2.0/4.7/2.4/4.5/2.9/4.3. Figure 5.18 shows a transmission electron micrograph of the grown wafer. While devices from this wafer showed good current–voltage characteristics with a clear voltage threshold, no electroluminescence was observed. This is due to the LH/SO states between the HH radiative subband states (see Fig. 5.17). Holes can be



**Fig. 5.17** The band structure of a Si/SiGe cascade showing the HH states in solid gray and the LH states in dashed black. The bound state is shown in solid black, the HH band edge in solid black and the LH band edge in dashed black.



**Fig. 5.18** A transmission electron micrograph of an eight-period Si/SiGe bound-to-continuum quantum cascade structure where the thinnest barrier is only 0.9 nm.



**Fig. 5.19** The electroluminescence spectra as a function of electric field for a Si/SiGe bound-to-continuum cascade device at 4.2 K with 10% duty cycle. A clear Stark shift of the peak to higher energy with higher electric field is observed.

easily scattering into these parasitic LH/SO subband states that significantly reduces the number of radiative transitions to the level that no electroluminescence is observed experimentally.

The solution to the parasitic LH/SO subband states is to produce a design where these states are pushed much higher in energy than the radiative HH states. This requires far higher Ge contents and strain in the quantum wells to be achieved. Figure 5.19 shows the results from one such structure that was grown on  $\text{Si}_{0.65}\text{Ge}_{0.35}$  virtual substrate with compressive- $\text{Si}_{0.5}\text{Ge}_{0.5}$  quantum wells and tensile-strained Si barriers. A low emission energy has been chosen that also makes it easier to strain the LH-SO states to higher energy.

Clear electroluminescence is shown in Fig. 5.19 along with a clear Stark shift of the peak to higher energy as the electric field across the sample is increased.

At present, no lasing has been observed from Si/SiGe quantum cascade structures. The gain that has been achieved is around  $7\text{ cm}^{-1}$ , while the minimum waveguide losses that have been demonstrated are  $12.5\text{ cm}^{-1}$  from a stationary surface plasmon waveguide (De Rossi *et al.* 2006) when mirror losses are included. The gain needs to be improved before the waveguide losses can be overcome and a laser demonstrated. Recently, a number of pure Ge quantum wells designs have been designed by Driscoll and Paiella with gain up to  $250\text{ cm}^{-1}$ . These are far more difficult to grow but if successful then the possibility of a Si-based laser should be achievable.

## 5.8 Conclusions

Silicon-germanium is now a mature field with HBTs and CMOS products in the market place. In the research field there are many areas where Si/SiGe heterostructures are being used to bandgap engineer nanoelectronic devices resulting in significant improvements in device performance. A number of these areas have good potential for eventually reaching production, while there are also many that allow fundamental research on the physics of materials and devices. All Si/SiGe devices are heavily influenced by the strain in the devices and this can be used to advantage, but also provides limitations especially with regard to critical thickness. Developing areas for SiGe research that are not included in this chapter include Si photonics, where optoelectronic solutions are being integrated with CMOS chips to provide complete photonic and nanoelectronic integration, and also quantum computing where single-electron quantum dots have now been demonstrated using Si/SiGe heterostructures. It is clear that the mature and cheap silicon technology will enable SiGe alloys to be used for many decades to come in nanoelectronic research.

## References

- Arnone, D.D., Ciesla, C.M., Pepper, M. *Phys. World.* **13**(4), 35 (2000).
- Bardeen, J., Brattain, W.H. *Phys. Rev.* **74**, 230 (1948).
- Bates, R., Lynch, S.A., Paul, D.J., Ikonic, Z., Kelsall, R.W., Harrison, P., Liew, S.L., Norris, D.J., Cullis, A.G., Tribe, W.R., Arnone, D.D. *Appl. Phys. Lett.* **83**, 4092 (2003).
- Brown, E.R., Söderström, J.R., Parker, C.D., Mahoney, L.J., Molvar, K.M. *Appl. Phys. Lett.* **58**, 2291 (1991).
- Califano, M., Vinh, N.Q., Phillips, P.J., Ikonic, Z., Kelsall, R.W., Harrison, P., Pidgeon, C.R., Murdin, B.N., Paul, D.J., Townsend, P., Zhang, J., Ross, I.M., Cullis, A.G. *Phys. Rev. B* **75**, 045338 (2007).
- Capasso, F., Paiella, R., Martini, R., Colombelli, R., Gmachl, C., Myers, T.L., Taubman, M.S., Williams, R.M., Bethea, C.G., Unterrainer, K., Hwang, H.Y., Sivco, D.L., Cho, A.Y., Sergent, A.M., Liu, H.C., Whittaker, E.A. *IEEE J. Quantum Electron.* **38**, 511 (2002).

- Caymax, M.R., Leong, W.Y. *Advanced Silicon and Semiconducting Silicon-Alloy Based Materials and Devices*, (ed.) Nijs, J.F.A. (Institute of Physics Publishing, Bristol, 1994), p. 141.
- Compano, R. *Technology Roadmap for European Nanoelectronics*, 2nd edn, European Commission, Brussels (available at <http://www.cordis.lu/ist/fetnidqf.htm>) (2000).
- Dehlinger, G., Diehl, L., Gennser, U., Sigg, H., Faist, J., Ensslin, K., Grützmacher, D., Müller, E. *Science* **290**, 2277 (2000).
- De Rossi, A., Carras, M., Paul, D.J. *IEEE J. Quantum. Electron.* **42**, 1233 (2006).
- Diehl, L., Mentese, S., Müller, E., Grützmacher, D., Sigg, H., Gennser, U., Sagnes, I., Campidelli, Y., Kermarrec, O., Bensahel, D., Faist, J. *Appl. Phys. Lett.* **81**, 4700 (2002).
- Driscoll, K., Paiella, R. *J. Appl. Phys.* **102**, 093103 (2007).
- Faist, J., Capasso, F., Sivco, D.L., Sirtori, C., Hutchinson, A.L., Cho, A.Y. *Science* **264**, 553 (1994).
- Faist, J., Beck, M., Aellen, T., Gini, E. *Appl. Phys. Lett.* **78**, 147 (2001).
- Glickman, M. *Phys. Rev.* **100**, 1146 (1955).
- Jain, S.C., Decoutere, S., Willander, M., Maes, H.E. *Semicond. Sci. Technol.* **76**, R67 (2001).
- Kazarinov, R.F. *Sov. Phys. Semicond.* **5**, 707 (1971).
- Kasper, E., Lyutovich, K. (Eds.) *Properties of Silicon Germanium and SiGe: Carbon EMIS Datareview Series No. 24* INSPEC IEE: London, U.K. (2000).
- Kasper, E., Paul, D.J. *Silicon Quantum Integrated Circuits* (Springer-Verlag, Berlin, 2005).
- Khater, M., Rieh, J.-S., Adam, T., Chinthakindi, A., Johnson, J., Krishnasamy, R., Meghelli, M., Pagette, F., Sanderson, D., Schnabel, C., Schonenberg, K.T., Smith, P., Stein, K., Stricker, A., Jeng, S.-J., Ahlgren, D., Freeman, G. *IEEE Electr. Devi. Meet. Tech. Dig.* **2004**, 247 (2004).
- Kelsall, R.W., Ikonic, Z., Murzyn, P., Pidgeon, C.R., Phillips, P.J., Carder, D., Harrison, P., Lynch, S.A., Townsend, P., Paul, D.J., Liew, S.L., Norris, D.J., Cullis, A.G. *Phys. Rev. B* **71**, 115326 (2005).
- Köhler, R., Tredicucci, A., Beltram, F., Beere, H.E., Linfield, E.H., Davies, A.G., Ritchie, D.A., Iotti, R.C., Rossi, F. *Nature* **417**, 156 (2002).
- Lynch, S.A., Bates, R., Paul, D.J., Norris, D.J., Cullis, A.G., Ikonic, Z., Kelsall, R.W., Harrison, P., Arnone, D.D., Pidgeon, C.R. *Appl. Phys. Lett.* **81**, 1543 (2002).
- Matthews, J.W., Blakeslee, A.E. *J. Cryst. Growth* **32**, 265 (1976).
- Mazumder, P., Kulkarni, S., Bhattacharya, M., Jian Ping Sun Haddad, G.I. *Proc. IEEE* **86**, 664 (1998).
- Meyerson, B.S. *IBM J. Res. Dev.* **44**, 391 (2000).
- Murdin, B.N., Heiss, W., Langerak, C.J.G.M., Lee, S.C., Galbraith, I., Strasser, G., Gornik, E., Helm, M., Pidgeon, C.R. *Phys. Rev. B* **55**, 5171 (1997).
- Patton, G.L., Subramanian, S.S., Delage, S.L., Tiwari, S., Stork, J.M.C. *IEEE Electron. Device Lett.* **9**, 165 (1988).
- Paul, D.J., See, P., Zozoulenko, I.V., Berggren, K.-F., Kabius, B., Holländer, B., Mantl, S. *Appl. Phys. Lett.* **77**, 1653 (2000).
- Paul, D.J. *Semicond. Sci. Technol.* **19**, R75 (2004).
- Paul, D.J. *Phys. Rev. B* **77**, 155323 (2008).

- People, R., Bean, J.C. *Appl. Phys. Lett.* **47**, 322 (1985a).  
People, R., Bean, J.C. *Appl. Phys. Lett.* **49**, 229 (1985b).  
Rieger, M.M., Vogl, P. *Phys. Rev. B.* **48**, 14276 (1993).  
Rieger, M.M., Vogl, P. *Phys. Rev. B.* **50**, 8138 (1994).  
See, P., Paul, D.J., Holländer, B., Mantl, S., Zozoulenko, I.V., Berggren, K.-F. *IEEE Electron. Device Lett.* **22**, 182 (2001).  
See, P., Paul, D.J. *IEEE Electron. Device Lett.* **22**, 582 (2001).  
Taur, Y., Ning, T.H. *Fundamentals of Modern VLSI Devices* (Cambridge University Press, Cambridge, UK, 1998).  
van der Wagt, J.P.A. *Proc. IEEE* **87**, 571 (1999).  
Williams, B.S., Kumar, S., Hu, Q., Reno, J.L. *Opt. Exp.* **13**, 3331 (2005).

# Quantum dots: Self-organized and self-limiting assembly

Dimitri D. Vvedensky

6

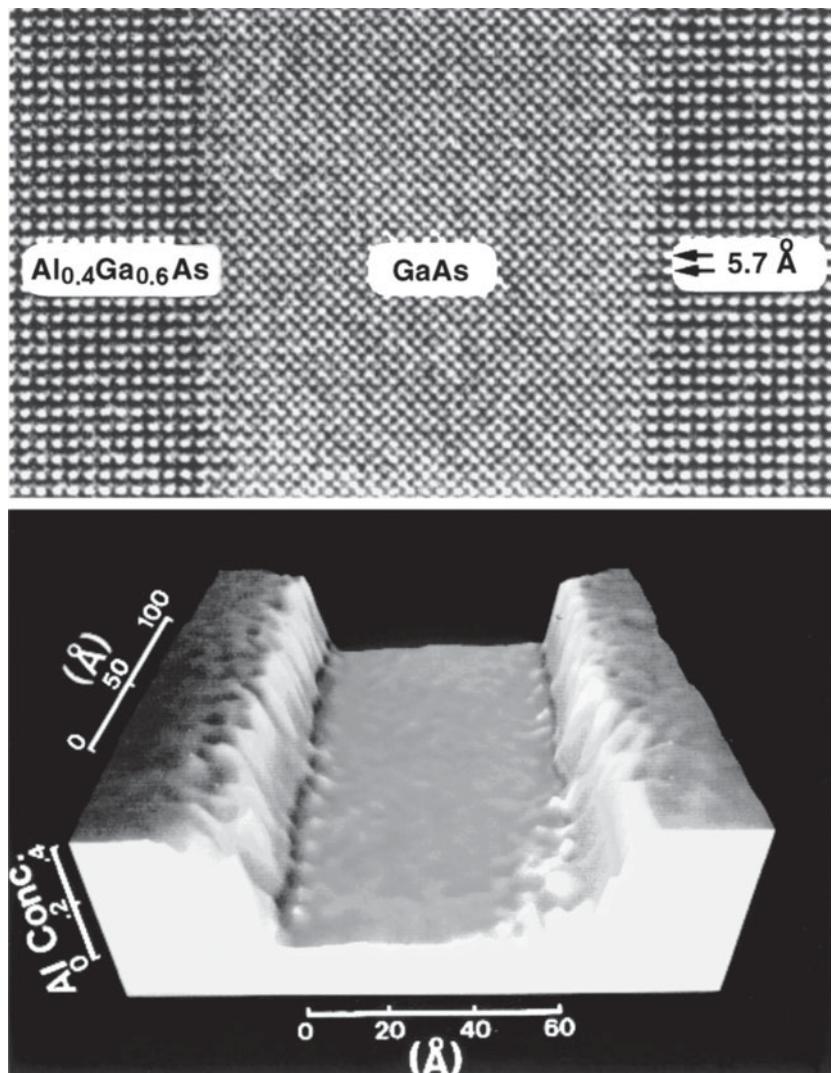
## 6.1 Introduction

### 6.1.1 The second industrial revolution

The invention of the transistor in the late 1940s sparked a revolution in information technology that persists unabated to this day. Smaller and faster computers, brighter lasers with lower power consumption, electronic memories with greater capacity and shorter access times, and information-transmission systems to meet the expanding demands of a connected information-intensive society have driven device design down to the atomic level. This has been accompanied by frontier concepts in quantum computing, encryption, and teleportation that represent an altogether different paradigm based on harnessing the intricacies of quantum mechanics.

The era of “quantum nanostructures” was born out of a suggestion in the late 1960s by Esaki and Tsu (1970). They proposed that structures made from alternate regions of different semiconductors (Fig. 6.1) would have a spatially varying potential energy that to an electron (or hole) would appear as a series of hills and valleys, or “wells.” The electrons would be repelled by the hills and attracted to the wells. If the well width could be made comparable to their de Broglie wavelength, then the electrons would exhibit “quantum confinement” as evidenced, for example, by the presence of discrete energy levels in their excitation spectra. The lateral electronic motion within the well would, however, remain unaffected. Such confinement would produce an effective *two-dimensional* (2D) electron system in what is now called a “quantum well” (Fig. 6.2(a)). Quantum wells find applications in low-noise high-frequency amplifiers that are used in satellite communications for improving the signal-to-noise ratio in mobile telephones. Quantum well lasers are used in fiber-optical communications, as reading heads in compact disk players, and as sensors for the detection of hazardous gases. The control over interface definition and doping profiles has also made planar nanostructures a popular testing ground for many fundamental ideas in condensed-matter

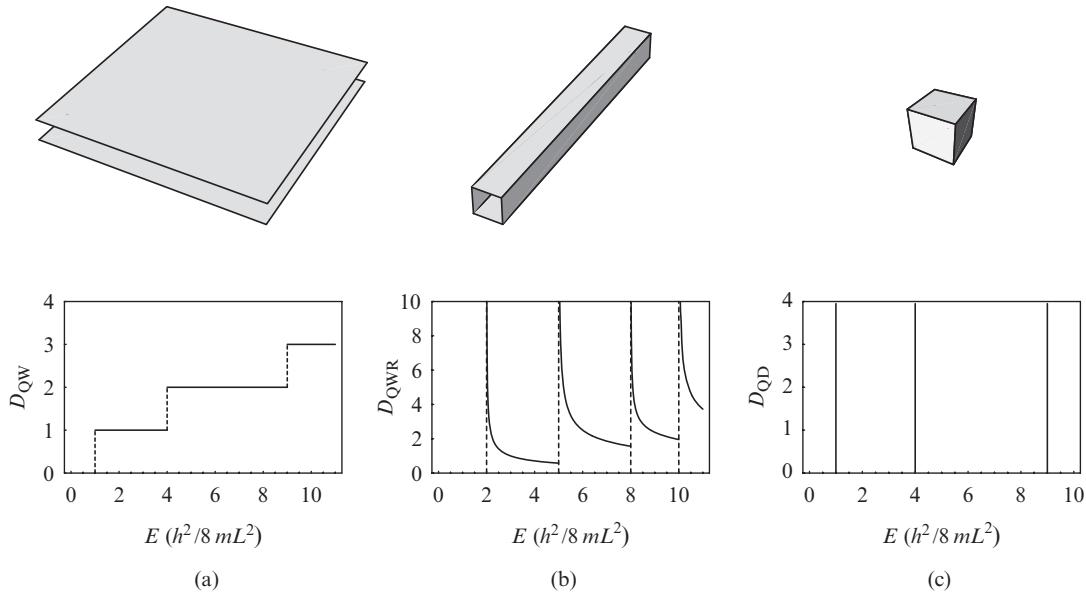
6.1 Introduction	205
6.2 Methods of epitaxial growth	210
6.3 Self-organization in Stranski–Krastanov systems	212
6.4 Site control of quantum dots on patterned substrates	220
6.5 Nanophotonics with quantum dots	227
6.6 Arrays of quantum dots	234
6.7 Summary and outlook	236
References	238



**Fig. 6.1** (Top) Chemical lattice image based on high-resolution transmission electron microscopy (TEM) of a GaAs quantum well between two Al<sub>0.4</sub>Ga<sub>0.6</sub>As barriers. (Bottom) Three-dimensional plot obtained from a vector pattern analysis of the image in the top panel. Height represents local composition, with changes in shading corresponding to statistically significant changes in composition. Reused with permission from P. Schwander, W.-D. Rau and A. Ourmazd, Composition mapping at high resolution, *J. Microscopy* **190**, 171 (1998). Published by Wiley InterScience.

physics based on confined electrons, which has led to the discovery of new physical phenomena such as the quantum Hall and fractional quantum Hall effects.

The observation of carrier confinement in quantum wells in the 1970s (Dingle *et al.* 1974) spawned a huge worldwide effort aimed at fabricating nanostructures with additional dimensional confinement. This was driven by the realization that the attractive optical and transport properties of such structures (Fig. 6.2(a)) stem from the fundamental changes in the densities of states and the Coulomb interactions between carriers as their effective dimensionality is reduced. Beginning with a quantum well, the imposition of an additional lateral potential barrier provides confinement in two directions, but free motion along the remaining direction in what is called a “quantum



**Fig. 6.2** Schematic illustration of (a) a quantum well, (b) a quantum wire, and (c) a quantum dot. The planar surfaces represent infinite potential barriers separated by a distance  $L$ . Electrons are thereby confined in one direction in (a), in two directions in (b), and in three directions in (c). The corresponding densities of states of free electrons within these structures are displayed in the bottom panels. The broken lines in (a) and (b) indicate the discrete levels of the confined states. Associated with each such level is a subband of continuous states corresponding to the unconfined motion.

“wire” (Fig. 6.2(b)). If electrons are confined in three dimensions to regions comparable to their de Broglie wavelength, the resulting structure is called a “quantum dot” (Fig. 6.2(c)).

Each quantum-mechanical state in the nanostructures in Fig. 6.2 is labelled by discrete quantum numbers for the confined degrees of freedom and continuous quantum numbers that are the wave numbers of the unconfined motion. Thus, for quantum wells and wires, there is a continuous subband associated with each discrete state. The energy separation  $\Delta E \sim 1/L^2$  between the discrete levels, which enables the subband separation to be engineered by varying the size of the confining region. The combination of discrete and continuous labels endows the density of states of each nanostructure in Fig. 6.2 with a characteristic profile. The density of states of a quantum well (Fig. 6.2(a)) has the appearance of a “staircase” because the constant density of states of a 2D free-electron system is associated with each discrete level resulting from the one-dimensional (1D) confinement. The density of states of a quantum wire (Fig. 6.2(b)) shows the discrete structure of 2D confinement and the square-root singularity of 1D free electrons, resulting in a series of sharp peaks. The density of states of a quantum dot is simply a series of discrete levels. This qualitative resemblance to an atomic energy spectrum is the reason that quantum dots are often referred to as “artificial atoms.”

The densities of states in Fig. 6.2 embody several assumptions about the structure and composition of nanostructures that are difficult, if not impossible, to realize in practice. The confining potential barriers, which are formed at interfaces between different semiconducting materials, are finite, so only the

first few subbands are bound. Additionally, these interfaces generally have a degree of structural and compositional disorder (Fig. 6.1) that can severely disrupt the density of states. For quantum wires, such non-uniformities along the wire translate into potential variations that lead to inhomogeneous broadening of the electronic spectra and to carrier localization at local potential minima (Taylor *et al.* 1991; Nikolić and MacKinnon 1994). This broadening can eventually smear out the features associated with quantum confinement, thus rendering the structure of little use as a quasi-1D system. In particular, localization effects make it difficult to study electrical transport through 1D channels, where the reduced dimensionality is expected to yield profound modifications of transport mechanisms.

A more fundamental limitation to the independent electron approximation in nanostructures is the effect of Coulomb interactions, which are responsible, for example, for the formation of excitons. For quantum wells, exciton formation is manifested by the appearance of strong, sharp features just below the band edges (Dingle *et al.* 1974). Theoretical studies (Ogawa and Takagahara 1991; Rossi and Molinari 1996) have shown that, in quantum wires, Coulomb correlations remove the singularity of the 1D joint density of states and greatly reduce the absorption above the band edge. The strong carrier confinement within quantum dots means that Coulomb effects play an even more prominent role than in quantum wells or wires. Coulomb interactions between carriers control recombination effects, which can be exploited to produce antibunched photons or entangled photon pairs, which will be discussed later in this chapter, the charging of quantum dots and the Coulomb blockade, which is the basis of the single-electron transistor (Kastner 2000) and the Kondo effect (Cronenwett *et al.* 1998).

### 6.1.2 Science and technology of quantum dots

The confinement of carriers to quantum wires or quantum dots is relatively easy to achieve when the lateral dimension of the confining region is  $\sim 500\text{--}1000\text{ \AA}$ . The techniques that enable the fabrication of such structures are based on lithography. However, the lithographic process often produces interfaces with a high defect density and even damage to the bulk material. Moreover, carrier confinement within regions of such lateral dimensions leads to a limited subband separation (typically a few meV), which in most cases is exceeded by thermal broadening. Most of the interesting physical behavior in such nanostructures is observed only at very low temperatures ( $T \sim 4\text{ K}$ ). This property distinguishes mesoscopic structures from true *quantum* nanostructures, in which subband separations are larger than the typical Coulomb interaction energy between carriers. For these and other reasons, fabricating quantum nanostructures directly with *in-situ* growth techniques provides an attractive alternative to *ex-situ* processing with lithography.

Quantum dots offer immense opportunities for next-generation semiconductor technology. When integrated into a solid-state matrix, quantum dots have atomic-like features such as discrete energy states and sharp optical transitions (Fig. 6.2(c)) that are determined by factors such as their size, shape, and

composition. This makes quantum dots potentially useful in extremely low-power consumption nanophotonic devices such as lasers and optical switches, providing a platform for information processing and transmission systems. Furthermore, like atoms or molecules, individual quantum dots can generate streams of single photons or polarization-entangled photon pairs, and may be used to store or process qubits in the solid state.

Much of the initial progress in quantum-dot science and technology has been achieved with self-assembled arrays formed by the Stranski–Krastanov growth mode during molecular beam epitaxy (MBE) and metalorganic vapor-phase epitaxy (MOVPE). Many fundamental features of three-dimensional (3D) confined charge carriers and their interactions with photons have been elucidated and new phenomena with potential applications to devices with novel functionality have emerged from these studies. However, the inherent randomness of the nucleation process on initially flat surfaces limits the utility of the resulting quantum-dot ensembles because of the absence of site control, size uniformity, and the inability to produce a single dot at a particular location. This has motivated the search for ways of improving the ordering in self-assembled arrays by modulating the lateral surface chemical potential with additional elements such as non-planar features or strained patterns to favor nucleation at desired sites. Another class of techniques utilizes growth-rate variations on different facets and nanocapillarity effects on non-planar surfaces to generate quantum-dot arrays that form at prescribed positions on a substrate. All of these techniques can produce dots without compromising their interfacial integrity and optical quality.

### 6.1.3 Chapter overview

In this chapter, we will review the main methods for fabricating quantum dots and discuss some of their applications. For the most part, our focus will be on III–V semiconductor quantum dots because of their attractive optical and electronic properties and the consequent wealth of theoretical and experimental studies of these systems. The primary fabrication methodologies of semiconductor quantum dots and other supported nanostructures are MBE and MOVPE, which are described in Section 6.2. Quantum dots are produced by invoking several strategies, including the Stranski–Krastanov growth mode on singular surfaces, which can be augmented by various modifications to the surface chemical potential to control nucleation, as outlined in Section 6.3, and directed self-assembly on patterned substrates, which is the subject of Section 6.4. We then examine the optical properties of single quantum dots and discuss applications of quantum dots and quantum-dot molecules to nanophotonics in Section 6.5. Some applications of quantum arrays are covered in Section 6.6. A summary and outlook are given in Section 6.7.

Numerous books, conference proceedings, and review articles have appeared on quantum dots in general (Bimberg *et al.* 1999; Chakraborty 1999; Masumoto and Takagahara 2002; Joyce *et al.* 2005), and specifically on their structural (Shchukin and Bimberg 1999; Goldman 2004; Joyce and Vvedensky 2004; Stangl *et al.* 2004; Schmidt 2007), transport (Beenakker 1997; van der

Wiel *et al.* 2003), electronic (Maksym *et al.* 2000; Hansen *et al.* 2007) and optical (Ustinov *et al.* 2003; Michler 2004; Skolnick and Mowbray 2004; Bryant and Solomon 2005) properties. These should be consulted for more detailed discussions than those provided here.

## 6.2 Methods of epitaxial growth

There are three principal ways of fabricating quantum dots: lithography, colloidal chemistry, and epitaxy. Electron-beam lithography and reactive-ion etching have been used to produce arrays of pillars by laterally patterning quantum wells (Scherer and Craighead 1986). While this yields nanostructures with 3D confinement (one direction from the original quantum well, the other two from the lateral etching) the etching process introduces contamination and produces large defect densities. The synthesis of colloidal quantum dots (Murray *et al.* 1993), typically based on II–VI materials, proceeds much like a standard chemical reaction in which the reactions are engineered to precipitate quantum dots from a solution or a host material, such as a polymer. Capping is often required to arrest the growth of the particles and to stabilize the resulting structures. Colloidal quantum dots have excellent fluorescence properties, which make them attractive for biological applications and large-area displays. The third method for producing quantum dots, and the main focus of this chapter, is epitaxy. Epitaxial techniques offer distinct advantages over lithography for producing quantum dots with controlled and reproducible properties that can be integrated into a solid-state system and placed at pre-determined positions on a substrate.

Epitaxial growth is a process during which a crystal is formed on an underlying crystalline surface as the result of deposition of new material onto that surface. The term “epitaxy”, which is a combination of the Greek words *epi*, meaning “upon”, and *taxis*, meaning “order,” was coined by Royer in the 1920s to convey the notion of growing a new crystal whose orientation is determined by a crystalline substrate and to distinguish epitaxial growth from polycrystalline and amorphous growth. The experimental methodologies that have been developed to achieve the required control over composition, doping, and interface definition in surface nanostructures are described in this section.

### 6.2.1 Molecular beam epitaxy

The simplest way conceptually of fabricating semiconductor nanostructures is with MBE (Joyce 1985; Cho 1994), which is essentially a two-step process carried out in an ultrahigh-vacuum (UHV) environment. In the first step, atoms or simple homoatomic molecules that are the constituents of the growing material (e.g. atomic Ga and either As<sub>2</sub> or As<sub>4</sub> for GaAs, atomic Si for Si) are evaporated from solid sources in heated cells, collimated into beams, and directed toward a heated substrate. The particles within these beams neither react nor collide with one another prior to their arrival onto the substrate, i.e. the particle flow to the substrate is *molecular*—hence the name *molecular beam epitaxy*.

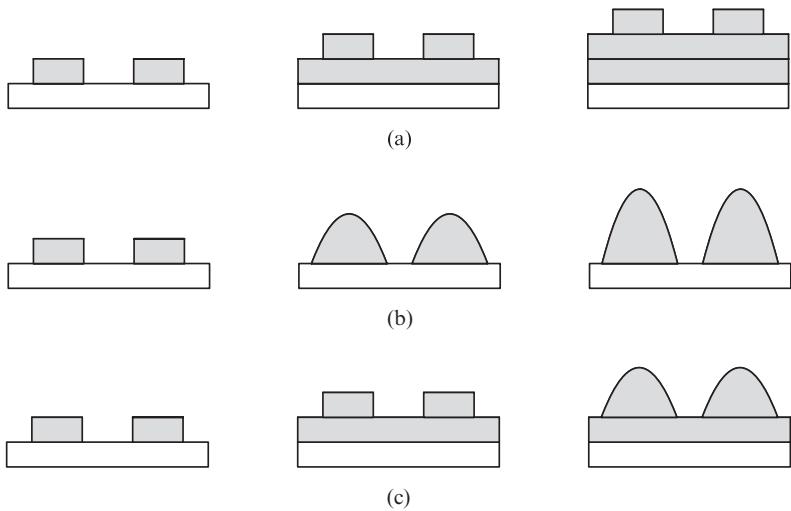
The second step of MBE is the migration of the deposited species on the surface prior to their incorporation into the growing material. The effectiveness of surface migration and the lifetime to incorporation depends on factors such as the deposition rates of the constituent species, the surface temperature, the decomposition rates of any molecular species, the substrate material and its crystallographic orientation, and the presence of any surface reconstruction. The dependence of the morphology on the deposition rate of material means that MBE (and other epitaxial growth techniques) are inherently non-equilibrium, or driven, processes. Thermodynamics provides the overall driving force for the morphological evolution of the surface, but the extent to which equilibrium is attained even locally is mediated by kinetics, i.e. the rates of processes that determine how effectively mass is transported across the surface (Madhukar 1983).

An important operational advantage of MBE is that the UHV environment enables the application of *in-situ* surface-analytical techniques to characterize the growing material at various levels of resolution. These include reflection high-energy electron-diffraction (RHEED) patterns to analyze the symmetry of surface reconstructions, the evolution of the specular beam as an indicator of the growth mode (Ichimiya and Cohen 2004), and scanning tunnelling and atomic force microscopies to probe static and time-resolved atomic arrangements (Voigtländer 2001). For MBE, therefore, a large body of theoretical and experimental work has established a conceptual and computational foundation for many fundamental aspects of the growth process (Pimpinelli and Villain 1998; Venables 2000).

### 6.2.2 Metalorganic vapor-phase epitaxy

An alternative to deposition by molecular beams is the hydrodynamic transport of material to the substrate from gas sources (Stringfellow 1989). In this scenario, which is called *vapor-phase epitaxy* (VPE), the constituents of the growing surface are delivered by heteroatomic molecules called *precursors*. For group-IV materials, these are hydrides, chlorides, or chloro-hydrides. For the growth of III–V materials, the Group-V elements are supplied as hydrides, e.g. AsH<sub>3</sub>, though for reasons of safety, organo-substituted hydrides are often used instead. The Group-III elements are delivered by precursors that contain carbon, i.e. they are metalorganic molecules, such as trimethylgallium (TMGa), Ga(CH<sub>3</sub>)<sub>3</sub>, and trimethylaluminum (TMAI), Al(CH<sub>3</sub>)<sub>3</sub>. For III–V materials, therefore, this technique is referred to as *metalorganic vapor-phase epitaxy* (MOVPE).

The pressures inside a vapor-phase reactor can vary from 10<sup>-2</sup> Torr up to atmospheric, so the flow of the gas is viscous and the chemicals reach the substrate by diffusion through a boundary layer. Thus, the delivery of material to the growing film relies on surface chemical reactions, as well as mass transport within the injected fluid as it flows through the reactor, the latter being highly dependent on the system pressure and the reactor design (Jensen 1993). This reactive hydrodynamic environment means that only optical techniques (Richter 2002) are appropriate for *in-situ* diagnostics. The growth morphology is typically analyzed *ex situ* with atomic force microscopy (AFM).



**Fig. 6.3** Schematic evolution of the (a) Frank–van der Merwe, (b) Volmer–Weber, and (c) Stranski–Krastanov heteroepitaxial growth morphologies. The unshaded regions represent the substrate and the shaded region the deposited material. In each panel, time increases from left to right as more material is deposited. In (a) growth proceeds by the formation of successive layers, while in (b), the 3D islands that appear at the onset of growth eventually merge into a complete layer. In (c), a wetting film, typically consisting of a few monolayers (ML), is formed initially, after which 3D islands appear.

## 6.3 Self-organization in Stranski–Krastanov systems

### 6.3.1 Thermodynamic considerations

Numerous experiments have revealed that, for small amounts of one material deposited onto the surface of a possibly different material, the epitaxial morphology falls into one of three distinct categories. By convention, these are referred to as: *Frank–van der Merwe* morphology, with flat single-crystal films consisting of successive largely complete layers, *Volmer–Weber* morphology, with 3D islands that leave part of the substrate exposed, and *Stranski–Krastanov* morphology, with 3D islands atop a thin flat “wetting” film that completely covers the substrate. These morphologies are illustrated in Fig. 6.3.

For lattice-matched systems, the Frank–van der Merwe and Volmer–Weber morphologies can be understood from thermodynamic wetting arguments based on interfacial free energies (Bauer 1958). We denote the free energy of the epilayer/vacuum interface by  $\gamma_e$ , that of the epilayer/substrate interface by  $\gamma_i$ , and that of the substrate/vacuum interface by  $\gamma_s$ . The Frank–van der Merwe growth mode is favored if

$$\gamma_e + \gamma_i < \gamma_s. \quad (6.1)$$

In this case, as the epilayers are formed, the free energy *decreases* initially before attaining a steady-state value for thicker films. This is the growth mode observed for homoepitaxial systems and for the AlAs/GaAs system, for which the lattice mismatch is 0.12%. Alternatively, if

$$\gamma_e + \gamma_i > \gamma_s, \quad (6.2)$$

then Volmer–Weber growth is favored. Here, the free energy *increases* if epilayers are formed on the substrate, rendering a uniform layer thermodynamically unstable against a break-up into regions where the substrate is covered

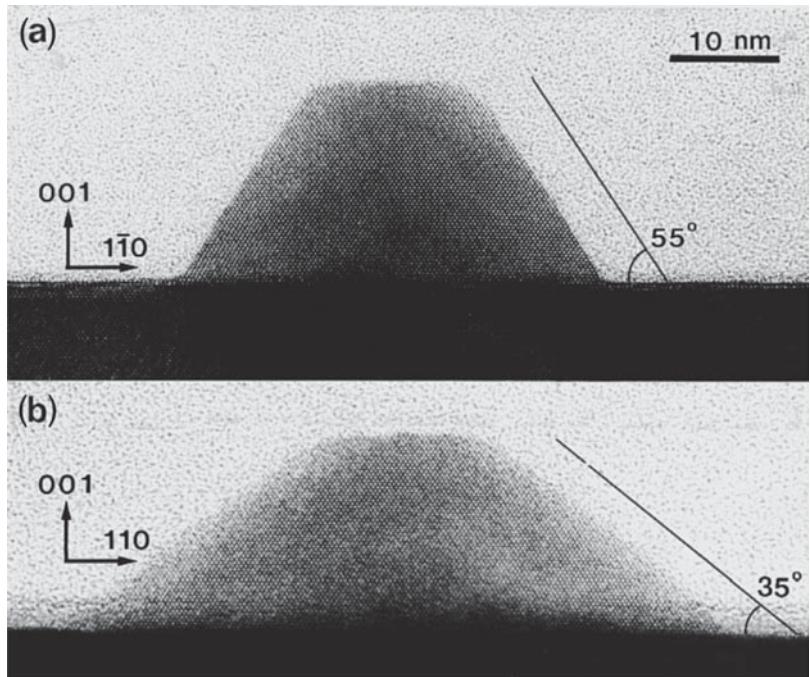
and those where it is uncovered. An example of a semiconductor system with this growth mode is Si on Ge(111) (Raviswaran *et al.* 2001).

The Stranski–Krastanov morphology is observed in systems where there is appreciable lattice mismatch between the deposited material and the substrate. The classical rationale for this growth mode is based on the accommodation of misfit strain, which changes the balance between the surface and interfacial free energies as the strain energy increases with the film thickness. Although the growth of wetting layers is favored initially, the build-up of strain energy eventually makes subsequent layer growth energetically unfavorable. The deposition of material beyond this point leads to the formation of 3D islands within which strain is relaxed by the formation of misfit dislocations. However, there is another scenario within the Stranski–Krastanov morphology: the formation of islands *without* dislocations—called *coherent* islands because their atomic structures are coherent with the substrate and the wetting layers—atop one or more wetting layers. This will be discussed in the next section.

### 6.3.2 General phenomenology

The prototypical cases for the Stranski–Krastanov growth of coherent 3D islands are InAs on GaAs(001) (Madhukar and Rajkumar 1990), which has a 7% lattice mismatch, and Ge on Si(001) (Eaglesham and Cerullo 1990), which has a 4% lattice mismatch. In both cases, the substrate exerts a compressive strain on the deposited material. When the 3D islands are embedded within epitaxial layers of a material with a wider bandgap, the carriers within the islands are confined by the potential barriers that surround each island, forming an array of quantum dots. Because these quantum dots are obtained directly by growth, with no additional processing, they are referred to as *self-organized* or *self-assembled* structures. Such quantum dots (Fig. 6.4) have several potentially useful characteristics (Leonard *et al.* 1993; Moison *et al.* 1994):

- (i) They are small enough to exhibit quantum effects in the confined carriers. The average diameter of the base of the dots is typically  $\sim 300 \text{ \AA}$  with an average height  $\sim 50 \text{ \AA}$ . To put these sizes into perspective, we consider the energy scale in Fig. 6.2(c). For a 150- $\text{\AA}$  square quantum dot, with an effective mass  $m^* \sim 0.1 m_e$  for InAs (Singh 1993) we obtain  $\Delta E \sim 80 \text{ meV}$ . This should be compared with the thermal energy corresponding to room temperature:  $k_B T \sim 25 \text{ meV}$ , which is smaller than  $\Delta E$ , but not overwhelmingly so.
- (ii) The dispersion about these averages is typically  $\pm 10\%$  for the base diameter and  $\pm 20\%$  for the height. Such dispersions in quantum-dot sizes are not ideal for the optical applications envisaged for these nanostructures (Section 6.3.3). This has motivated the development of several strategies for producing more uniform arrays. These will be described in later sections.
- (iii) The dot shapes are elongated truncated pyramids with faceted sidewalls along one direction. This suggests that the structures of the individual

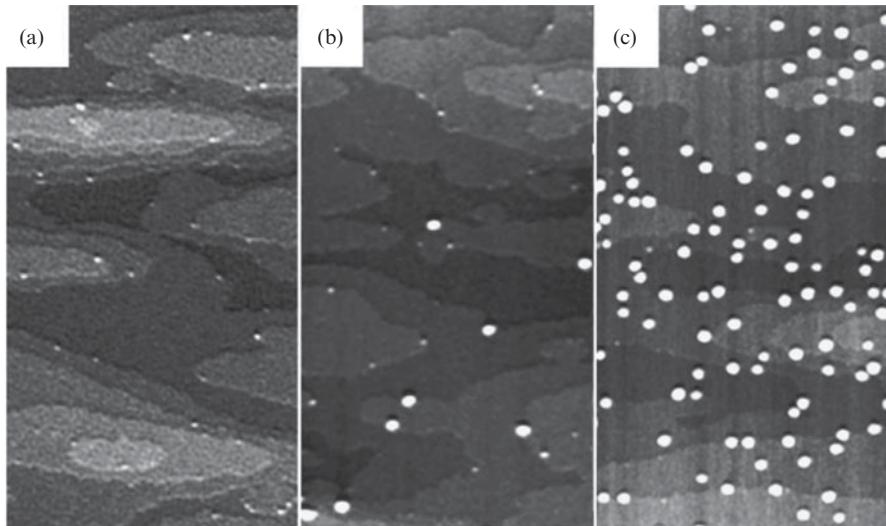


**Fig. 6.4** High-resolution cross-sectional TEM image of an uncapped InP island on GaInP grown by MOVPE at 580 °C along the (a) [110] and (b) [1-10] directions. The islands are elongated along [110] and the planes of atoms are noticeably curved toward the center of the island near the substrate, which is caused by the compressive strain of the substrate, but there is no evidence of any dislocations. Reused with permission from K. Georgsson, N. Carlsson, L. Samuelson, W. Seifert, and L.R. Wallenberg, *Appl. Phys. Lett.* **67**, 2981 (1995). Copyright 1995, American Institute of Physics.

dots are strongly influenced by thermodynamics, though reports of different shapes and a variety of multifaceted structures suggests that kinetic factors cannot be altogether neglected. Indeed, the quantum-dot morphology during growth (Costantini *et al.* 2003) and upon subsequent capping (Costantini *et al.* 2006) have been shown to result from a subtle interplay between kinetics and thermodynamics.

- (iv) One of the most important structural properties of self-organized quantum dots is that they are *coherent*, that is, the 3D islands are formed with no misfit dislocations at the interface with the substrate. This is especially desirable for designing lasers based on quantum dots because dislocations can act as centers for the non-radiative recombination of electron–hole pairs, i.e. recombination without the emission of a photon, with the excess energy of the recombined carriers transmitted to the vibrations of the crystal lattice. This degrades the performance of the laser by increasing the threshold current (the current at which the output of a laser is dominated by stimulated rather than spontaneous emission).

The growth of InAs on GaAs(001) proceeds first by the nucleation of 2D islands (Bressler-Hill *et al.* 1994; Bell *et al.* 2000) that coalesce to form coherently strained layers. These are the “wetting” layers in the conventional Stranski–Krastanov description (Fig. 6.3(c)). Coherent 3D islands are formed after the deposition of near 1.6 InAs bilayers (Fig. 6.5), though the precise point at which this transition occurs depends on the growth conditions. The transition from the growth of the wetting layers to the nucleation and growth



**Fig. 6.5** AFM image ( $0.9 \mu\text{m} \times 0.5 \mu\text{m}$ ) of InAs deposited in GaAs(001) at  $500^\circ\text{C}$  for coverages of (a) 1.54 ML, (b) 1.57 ML, (c) 1.64 ML. The 3D islands are shown as bright spots against the darker background of the wetting layers, with lighter shading corresponding to higher layers above the substrate. Several incomplete wetting layers are discernible and these images indicate that 3D islands form preferentially at step edges. The small 3D islands in (a) have a different distribution function from the 3D islands in (b) and (c), which suggests a different formation mechanism. Reused with permission from F. Arciprete, E. Placidi, V. Sessi, M. Fanfoni, F. Patella, and A. Balzarotti, *Appl. Phys. Lett.* **89**, 041904 (2006). Copyright 2006, American Institute of Physics.

of 3D islands is quite abrupt, occurring over less than 0.1 ML (Fig. 6.5). This transition can be followed *in situ* by RHEED, which shows a change from a streaky pattern, characteristic of layer-by-layer growth, to a spotty pattern that results from the transmission of the electrons through the 3D islands (Pashley *et al.* 2001).

What are the mechanisms that drive the transition from the 2D growth of wetting layers to the formation of 3D islands? Recent direct composition measurements (Walther *et al.* 2001) suggest that segregation is the controlling factor for determining the thickness of the wetting layer immediately prior to the appearance of 3D islands. For the InAs/GaAs system, kinetic Monte Carlo (KMC) simulations (Cullis *et al.* 2002; Migliorato *et al.* 2002) indicate that some of the Ga atoms in the topmost layer exchange positions with In atoms in the preceding layer to reduce the misfit strain with the substrate, which increases the In concentration in the top layer. When the next layer of In is deposited, this exchange process is repeated, but some of the top-layer In atoms will lie atop another In atom in the preceding layer, so the In concentration cannot be reduced as effectively as for earlier layers. As additional In is deposited, the concentration of the InAs layers increases, which increases the elastic energy between the growing film and the GaAs substrate. Eventually, the concentration in the topmost layer reaches a critical value for which the strain is relieved most efficiently not by further exchange, but by the formation of 3D islands. The point at which this occurs depends critically on the rate of the In exchange process and on the concentration of deposited atoms, both of which can be influenced by varying the growth conditions (temperature and fluxes). Accordingly, the critical wetting-layer thickness can range from

less than two layers, as for the InAs/GaAs, to over 10 layers. Intermixing was also invoked by Tu and Tersoff (2004) in their one-dimensional model for the critical thickness for island formation in heteroepitaxial systems.

This segregation mechanism also has implications for the concentration profiles within the 3D islands (Walther *et al.* 2001). The In-rich adatom population must be accommodated with the smaller Ga atoms. This results in a non-uniform concentration profile resulting from the preferential trapping of In atoms near the apex of the 3D islands, where there is maximum strain relief. This, in turn, induces a chemical-potential gradient along the sides of the island that favors Ga-atom incorporation near the base of the island.

### 6.3.3 Statistics of quantum-dot arrays

Interest in quantum dots produced by the Stranski–Krastanov growth mode was stimulated originally by the observation of intense photoluminescence (PL) (Section 6.5.1) from small InAs islands on GaAs(001). Optical studies of these structures found that *individual* dots exhibit high optical quality in terms of a narrow ( $\ll k_B T$ ) PL linewidth (Marzin *et al.* 1994). These sharp linewidths, together with their temperature dependence (Raymond *et al.* 1995), are consistent with a density of states consisting of a series of delta functions and an enhanced oscillator strength, both of which are expected for structures with 3D carrier confinement.

However, *arrays* of dots show a comparatively large inhomogeneous broadening ( $\gg k_B T$ ) of the PL emission due to the distributions of dot sizes and composition that are inherent to the self-organization process. This impedes the development of devices that rely on dot–field or dot–dot interactions. Although experiments carried out on single dots have demonstrated the basic principles of functionality of such devices, many of the desirable features of quantum dots and the scaling to meet manufacturing requirements depend on control over their location and properties, the latter being dependent on their sizes, shapes, and composition profiles. Thus, controlling the kinetics of 3D island nucleation and growth has become a central focus of current research.

Figure 6.5(c) shows an AFM scan of an array of 3D islands grown by deposition of InAs onto GaAs(001). Most apparent is the array of 3D islands, shown as bright spots against the darker region that represents the wetting layers on the GaAs(001) substrate. This image shows that quantum dots have a tendency to form near the upper steps of the wetting layer, presumably because of the better opportunity for strain relaxation than on the terraces. The absence of any appreciable local ordering gives rise to a distribution of local island environments, as evidenced by the strong variations of the interisland distances and numbers of neighboring islands. This produces a corresponding distribution of local growth rates because the catchment area of individual islands is determined by the number and proximity of neighboring islands that, in turn, leads to a distribution of 3D island sizes, as in the case of 2D homoepitaxial islands (Mulheran and Blackman 1996; Evans *et al.* 2006). The characterization of this distribution leads to some unexpected conclusions. We first provide some background.

One of the most important results to emerge from the theoretical description of homoepitaxial growth (Bartelt and Evans 1992, 1996; Ratsch *et al.* 1994b, 1995; Amar and Family 1995; Vvedensky 2000) is that, in the submonolayer regime, prior to an appreciable coalescence, the density  $n_s$  of  $s$ -atom ( $s \geq 2$ ) islands can be written as

$$n_s = \frac{\theta}{s_{av}^2} f\left(\frac{s}{s_{av}}\right), \quad (6.3)$$

where  $s_{av}$  is the average island size,  $\theta$  is the coverage, and  $f$  is a scaling function. The scaling function is “universal” in the sense that its dependence on the coverage, the growth rate and the substrate temperature enter only through  $s_{av} \sim \theta^z R^\chi$ , where  $R = D/F$  is the ratio of the adatom diffusion constant  $D$  to the deposition flux  $F$ , and  $z$  and  $\chi$  are (positive) scaling exponents. In other words,  $s_{av}$  is the only length scale necessary for describing island nucleation and growth. Thus, when island-size data are plotted as  $n_s s_{av}^2 / \theta$  versus  $s/s_{av}$ , the distributions for different  $\theta$  and  $R$  “collapse” onto a single curve, which is the scaling function  $f$ . There is substantial experimental, simulational, and theoretical evidence in support of this scaling hypothesis (Evans *et al.* 2006), which has also demonstrated the sensitivity of the scaling function to various nucleation and growth mechanisms (Avery *et al.* 1997 and references therein).

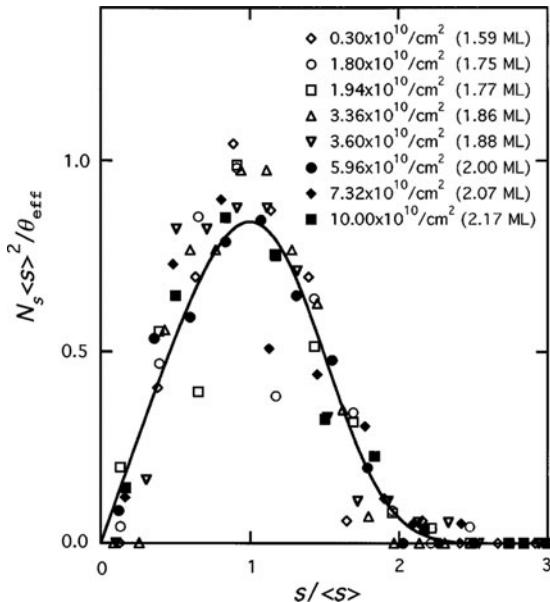
Despite their success in homoepitaxial systems, the applicability of scaling concepts to heteroepitaxial systems is not immediately apparent. Size-dependent effects induced by lattice misfit, such as adatom detachment, could pre-empt the validity of this scaling theory (Ratsch *et al.* 1994a) by introducing an elastic length scale that competes with  $s_{av}$ . Nevertheless, measurements (Ebiko *et al.* 1998, 1999; Krzyzewski *et al.* 2002; Arciprete *et al.* 2006; Fanfoni *et al.* 2007) of self-organized 3D islands (Fig. 6.6) have shown that the size (volume) distribution does indeed exhibit a data collapse and, moreover, that the scaling function corresponds to *irreversible aggregation*, that is, the 3D islands grow by accumulating material, with minimal subsequent detachment. Similar results were obtained for *dislocated* InN islands on GaN(0001) (Cao *et al.* 2003). That a data collapse in the presence of strain is obtained at all can be rationalized by a characteristic elastic length that is much larger than  $s_{av}$  (Aqua and Frisch 2008). But the irreversibility of the quantum-dot growth process merits closer examination.

The kinetics of island nucleation and growth are often described in terms of a “critical nucleus”  $i$  such that islands of size  $s > i^*$  do not dissociate (Venables 2000). An alternative formulation is based on the ratio  $\lambda$  of the net detachment rate to the net attachment rate of an island (Bartelt *et al.* 1995; Ratsch *et al.* 1995; Vvedensky 2000),

$$\lambda = \frac{R_d}{F\kappa + D\sigma\varrho}, \quad (6.4)$$

in which  $\kappa$  is the direct capture number from the incident flux,  $\sigma$  the diffusive capture number with an adatom density of  $\varrho$  and  $R_d$  is the escape rate of an adatom from a 3D island. KMC simulations of 2D homoepitaxy show that  $\lambda$  parametrizes a continuous family of scaling functions (Ratsch *et al.* 1995),

**Fig. 6.6** Scaling plot of the volume distribution of 3D islands measured during the Stranski-Krastanov growth of InAs on GaAs(001) at 490°C. The quantity  $\theta_{\text{eff}} = 1.5 \text{ ML}$  is the coverage *beyond* the critical coverage  $\theta_c$  for the 2D–3D transition, i.e.  $\theta_{\text{eff}} = \theta - \theta_c$ . The different symbols, which correspond to the size distributions for different coverages, fall near the solid curve, which is the scaling function for irreversible aggregation (Amar and Family 1995). Reprinted figure with permission from Y. Ebiko, S. Muto, D. Suzuki, S. Itoh, K. Shiramine, T. Haga, Y. Nakata and N. Yokoyama, *Phys. Rev. Lett.* **80**, 2650 (1998). Copyright 1998 by the American Physical Society.



but for  $\lambda \approx 1$ , the island-size distribution fits eqn (6.3), even when there is significant detachment (Koduvely and Zangwill 1999).

For homoepitaxial systems, the nucleation and growth of islands is determined by the balance between adatom attachment, which dominates at large fluxes and/or low temperatures, and detachment from islands, which becomes more important as the flux is decreased and/or the temperature increased. In the absence of detachment, i.e. irreversible growth, island-size distributions are broad because unfavorable island positions cannot be “corrected” as they can if there is significant detachment. An island that is nucleated near a large island has a diminished growth rate because of a smaller capture zone than in a region with few neighboring islands. The resulting spread of growth rates is reflected in the broad distribution of island sizes. But if detachment is allowed, a low growth rate could allow the island to dissolve. In a region of the surface with fewer neighboring islands, the growth rate is larger, so the nucleation of an island is followed by growth, rather than by dissolution. Thus, with increasing detachment, island-size distributions become narrower with an increasing regularity of island positions and sizes (Ratsch *et al.* 1995). Conversely, the broadest distribution is obtained for irreversible aggregation.

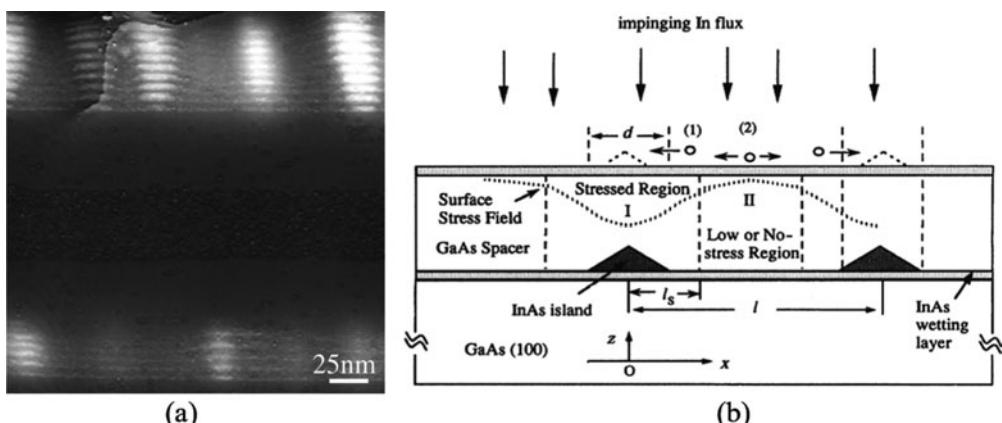
The formation of 3D islands during Stranski-Krastanov growth introduces new elements into this basic description. As the critical thickness is approached, the growth of the wetting layer becomes increasingly unfavorable. With ongoing deposition, the attachment of arriving atoms to the wetting layer is energetically unfavorable, which causes the adatom population to increase. This leads eventually to the formation of incipient 3D islands. Adatoms from the deposition flux, as well as those detached from the wetting layer, are readily incorporated into these islands (Joyce *et al.* 1998), which

grow rapidly (see Fig. 6.5). The large mass flow to the islands has the effect of a substantially increased deposition flux, which drives the system toward the limit of irreversible aggregation (small  $\lambda$ ). This picture of “sudden nucleation” is supported by recent experiments (Cullis *et al.* 2005; Fanfoni *et al.* 2007; Placidi *et al.* 2007). The broad 3D island-size distribution in Fig. 6.6 is therefore an inherent aspect of Stranski–Krastanov growth that cannot be alleviated simply by varying the growth conditions.

### 6.3.4 Multilayer quantum dots

The tendency of quantum dots to form near the steps of the wetting layer (Fig. 6.5) suggests that variations of the surface chemical potential could be engineered to improve the lateral ordering of quantum-dot ensembles. Substrate preparations that have been utilized toward this end include vicinality (Kitamura *et al.* 1995; Liao *et al.* 1999), strain modulation (Lee *et al.* 2000; Nakamura *et al.* 2002; Alchalabi *et al.* 2003; Leroy *et al.* 2005), and the etching of various patterns.

Strain-induced nucleation provides a natural way of ordering 3D islands in the vertical (growth) direction. If two or more layers of 3D islands are grown sufficiently close together, that is, with few enough spacer layers between them, then the islands on successive layers are observed to align (Fig. 6.7(a)), with the extent of the alignment decreasing with increasing interlayer separation (Xie *et al.* 1995). This behavior can be explained (Xie *et al.* 1995) in terms of the effect that the buried 3D islands have on the chemical potential of the migrating In adatoms in the next cycle (Fig. 6.7(b)). In the region above a buried island, the GaAs is under stress to accommodate the (larger) InAs lattice constant. The migrating In adatoms are attracted to these stressed regions, whereupon the large adatom concentration increases the nucleation



**Fig. 6.7** (a) Large-scale cross-sectional STM image showing low-temperature 5- and 10-period superlattices of InAs/GaAs quantum dots (bright regions). Reused with permission from B. Lita, R.S. Goldman, J.D. Phillips, and P.K. Bhattacharya, *Appl. Phys. Lett.*, **74**, 2824 (1999). Copyright 1999, American Institute of Physics. (b) Schematic representation of the two main migration pathways of In adatoms on a stressed InGaAs surface: (1) directional diffusion under the influence of a spatially varying chemical potential, and (2) isotropic diffusion that contributes to the nucleation of new islands not aligned below buried islands. Reprinted figure with permission from Q. Xie, A. Madhukar, P. Chen and N.P. Kobayashi, *Phys. Rev. Lett.* **75**, 2542 (1995). Copyright 1995 by the American Physical Society.

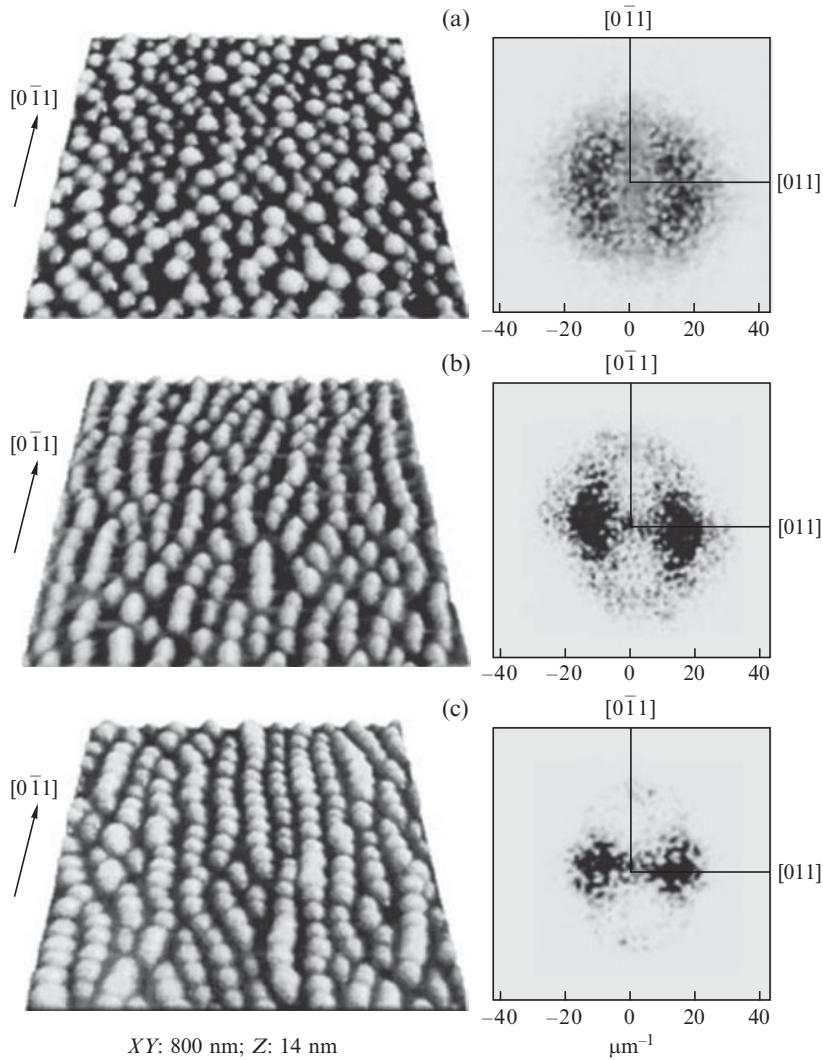
rate of 3D islands. In the region between the islands, where the chemical-potential gradient has a smaller magnitude, the possibility still remains of nucleating islands that do not lie above a buried island, but with a much reduced likelihood. The result is a striking vertical alignment of the quantum dots. A 1D model (Teichert *et al.* 1996; Tersoff *et al.* 1996) that incorporates these effects reproduces the observed tendency toward ordering.

An accompanying phenomenon to the *vertical* ordering in multilayer quantum dots that is not apparent in 1D models is an enhanced *lateral* ordering with a concomitant size regularization in successive layers (Tersoff *et al.* 1996; Lytvyn *et al.* 2007; Strel'chuk *et al.* 2007; Niu *et al.* 2008). The basic mechanism for this ordering is the same as that discussed above and depicted in Fig. 6.7(b). Because of the attractive chemical potential for migrating adatoms in regions above a buried island, nucleation is strongly favored there. The growth rate of islands is determined by their ability to capture migrating adatoms that, in turn, is determined by their proximity to neighboring islands (Mulheran and Blackman 1996; Evans *et al.* 2006). If, during one of the growth cycles, an island with an unfavorable local environment is formed by the natural fluctuations of nucleation, its growth rate would be lower than that of neighboring islands, so its influence on the elastic field will be correspondingly diminished in the next cycle. Hence, an island with such a local environment is unlikely to persist in successive cycles. Over repeated cycles, the local environments of islands, their growth rates, and, hence, their sizes tend to regularize.

An example of enhanced lateral ordering in a multilayer systems (Strel'chuk *et al.* 2007) is shown in Fig. 6.8 for  $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}/\text{GaAs}$  periodic structures that are comprised of 9.8 ML of the alloy interspersed with 60 ML of GaAs spacer layers. The critical thickness for the growth conditions used for this study is in the range 3.4–4.0 ML. The two-period sample (Fig. 6.8(a)) does not appear to exhibit any evidence of lateral ordering, but as the number of periods is increased to 7 (Fig. 6.8(b)) and then to 9 (Fig. 6.8(c)), ordering clearly develops along the [011] direction. This visual assessment of the surface morphologies is supported by 2D Fourier transforms of  $5\ \mu\text{m} \times 5\ \mu\text{m}$  samples, which are displayed adjacent to the corresponding morphologies in Fig. 6.8. There are two peaks along [011] that are indicative of the periodicity of the quantum dots along this direction. These are present even in the two-period sample, but become sharper for the 7- and 9-period samples.

## 6.4 Site control of quantum dots on patterned substrates

A primary goal of fabricating self-organized quantum dot arrays is minimizing the variations in size, shape, and composition so that, for example, as many dots as possible are optically active at a specified wavelength. But there is another aspect to fabricating quantum dots—site control. The precise positioning of quantum dots within larger systems or devices is an essential feature for designing and functionalizing semiconductor nanosystems. For example, the positioning of quantum wires or quantum dots with a precision of  $\sim 10\ \text{nm}$



**Fig. 6.8** (a) 3D AFM images of sections of the surface morphologies of  $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}/\text{GaAs}$  multilayer structures composed of (a) two periods, (b) 7 periods, and (c) 9 periods of 9.8 ML of the alloy interspersed with 60 ML of GaAs. The right panels show the corresponding 2D Fourier transforms of  $5 \mu\text{m} \times 5 \mu\text{m}$  samples. With kind permission from Springer Science + Business Media: V.V. Strel'chuk, P.M. Lytvyn, A.F. Kolomys, M. Ya. Valakh, Yu. I. Mazur, Zh. M. Wang and G.J. Salamo, "Lateral ordering of quantum dots and wires in the  $(\text{In},\text{Ga})\text{As}/\text{GaAs}(100)$  multilayer structures," *Semiconductors* **41**, 73 (2007), Fig. 1. Copyright 2007, MAIK Nauka/Interperiodica.

within an optical cavity is crucial for the coupling of these structures to the electromagnetic modes of the cavity, which, in turn, directly affects the performance of light sources or detectors based on these systems. In this section, we discuss a growth methodology that combines self-assembly and control over composition, size and positioning on patterned and non-planar substrates.

#### 6.4.1 Thermodynamic considerations

The formation of surface patterns during growth requires a lateral modulation of the surface mass current. The Stranski-Krastanov growth mode forms the basis of one method (Mui *et al.* 1995) of controlling the positions of quantum dots on a patterned substrate. A corrugated substrate, with alternate concave and convex regions (i.e. "hills" and "valleys," respectively), flattens during

growth to minimize its surface free energy. This flattening occurs by faster growth in the convex regions because there are more stable binding sites for migrating adatoms than in the concave regions. There results a net mass current into the convex regions, even in the absence of a deposition flux. Thus, during deposition of material onto the corrugated substrate, the thickness in the convex regions attains the point at which 3D islands appear *before* that in the concave regions. Hence, the 3D islands appear preferentially in the convex regions of the substrate.

There are other ways that growth on patterned substrates can influence the positioning of quantum dots. Lithography can be used to etch patterns, such as periodic arrays of holes, to produce a lateral modulation of the chemical potential. Under suitably chosen growth conditions, 3D islands nucleate only in the holes, which results in a regular array of quantum dots. Lithographic patterning can also be used to expose crystalline facets with different chemical, transport and structural properties. Growth on such surfaces by MOVPE introduces another design parameter—the variation of precursor decomposition rates on different facets. This has the effect of preferential deposition of the constituents of the growing material on particular facets. Both types of lithographic methodologies will be discussed below.

The chemical potential for each facet on an etched surface can be derived by a method due to Herring (1951), from which the chemical potential  $\mu_i$  of a facet surrounded by  $N$  other facets is obtained as (Ozdemir and Zangwill 1992; Biasiol *et al.* 2002)

$$\mu_i = \mu_0 + \frac{\Omega}{A_i} \sum_{j=1}^N [L_j (\gamma_j \operatorname{cosec} \theta_{ij} - \gamma_i \cot \theta_{ij})], \quad (6.5)$$

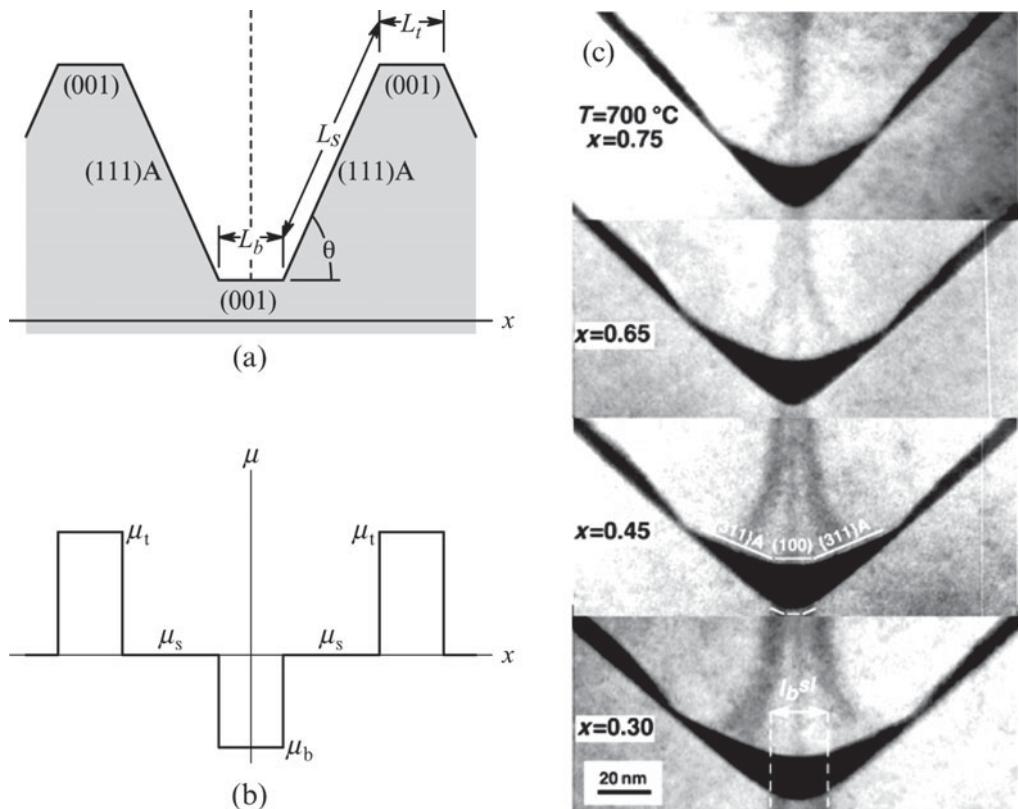
where  $\mu_0$  is the chemical potential of the bulk crystal,  $\Omega$  is the atomic volume,  $A_i$  is the area of the  $i$ th facet,  $L_j$  is the length between two adjacent facets,  $\gamma_i$  is the surface energy of the  $i$ th facet, and  $\theta_{ij}$  is angle between the  $i$ th and  $j$ th facets.

A surface that has been patterned with alternating (001) and (111)A facets is shown in Fig. 6.9(a). We suppose that the system extends indefinitely along the direction perpendicular to the plane of the page, but we will neglect any effects of growth along this direction. With this proviso, the lateral surface mass currents can be calculated with a 1D theory, and the chemical potentials  $\mu_t$ ,  $\mu_s$ ,  $\mu_b$  of the top, side, and bottom facets reduce to

$$\begin{aligned} \mu_t &= \mu_0 + \frac{2\gamma \Omega_0}{L_t}, \\ \mu_s &= \mu_0, \\ \mu_b &= \mu_0 - \frac{2\gamma \Omega_0}{L_b}, \end{aligned} \quad (6.6)$$

with

$$\gamma = \frac{\gamma_s}{\sin \theta} - \frac{\gamma_b}{\tan \theta}. \quad (6.7)$$



**Fig. 6.9** (a) Schematic 1D groove profile showing three facets: the (001) at the top and bottom, and the (111)A along the sidewalls. The lengths of the top, side, and bottom facets are  $L_t$ ,  $L_s$ , and  $L_b$ , respectively. (b) Surface chemical potential profile of the patterned substrate in the top panel. (c) Dark-field TEM cross-sections of self-limiting  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  grown by MOVPE at  $700^\circ\text{C}$  for the indicated values of  $x$ , where 5-nm thick GaAs markers have been inserted. Reprinted figure with permission from G. Biasiol, A. Gustafsson, K. Leifer and E. Kapon, *Physical Review B* **65**, 205306 (2002). Copyright 2002 by the American Physical Society.

The resulting chemical potential profile, which is depicted in Fig. 6.9(b), clearly shows that the faceted morphology in Fig. 6.9(a) is not an equilibrium configuration of the surface. The chemical potential differences across the surface induce a net mass flow from the top facet to the bottom facet until the surface becomes flat, just as in the case of the corrugated substrate discussed above.

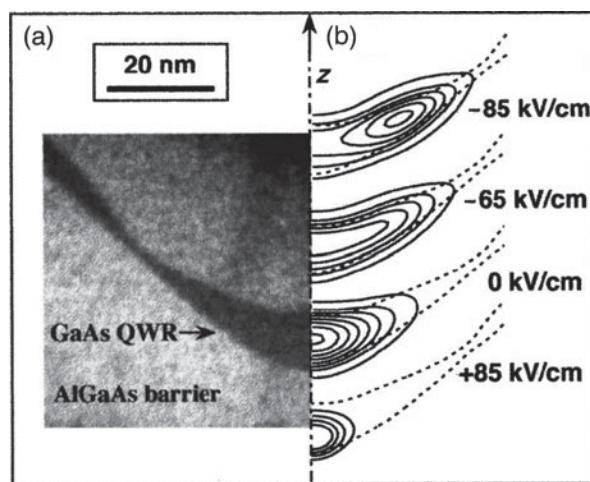
Figure 6.9(c) shows a series of dark-field TEM cross-sections of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  quantum wires grown by MOVPE for the indicated Al concentrations (Biasiol *et al.* 2002). The initial  $\text{GaAs}(001)$  substrate was etched with a series of V-grooves, as illustrated schematically in Fig. 6.9(a). During growth, the region near the apex of the V-groove evolves toward a stationary “self-limiting” profile (Fig. 6.9(c)). The term “self-limiting” is used because, as for self-organization, the system “selects” a kinetically limited characteristic profile that is determined by the growth conditions, which include the substrate temperature, the concentration of the alloy, the geometry of the V-groove, the growth rates of different facets, and the emergence of new facets, in this case the (311)A.

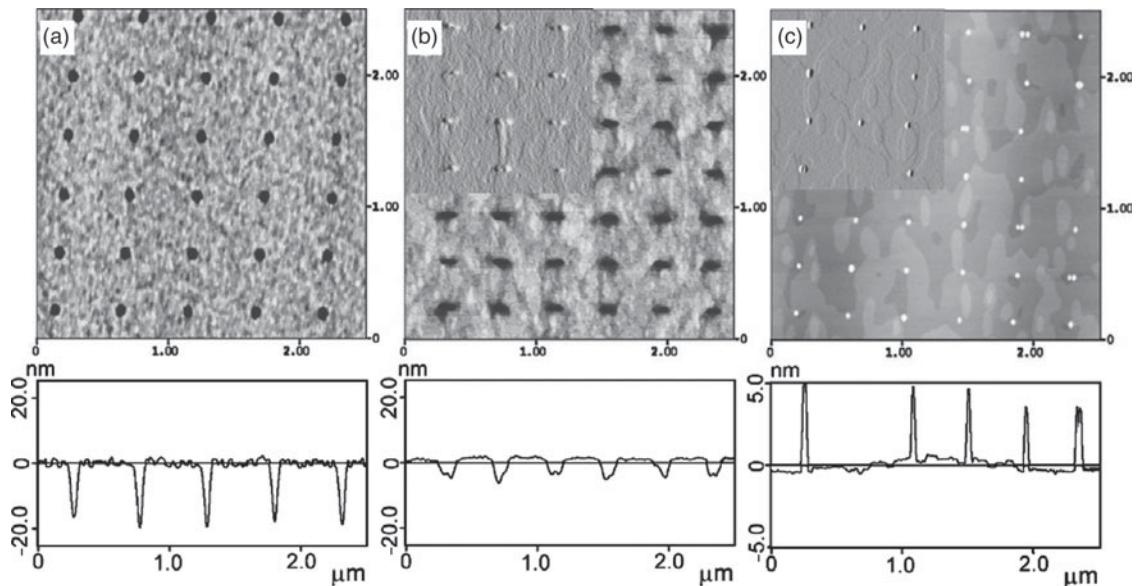
The images in Fig. 6.9(c) show that, near the apex of the V-groove, the profile becomes sharper and narrower with increasing Al concentration  $x$ . But the most important feature of these images is the segregation of the Ga near these apices. Although the surface chemical potential is attractive for both Ga and Al adatoms in this region (Fig. 6.9(b)), the lower mobility of the Al leads to the observed segregation. This is ultimately responsible for the formation of quantum wires near the apices. Carrier confinement in the vertical direction of such a nanostructure results in the usual manner from potential barriers provided by the valence- and conduction-band offsets at the atomically abrupt interfaces between the Ga-rich regions and the surrounding alloy. The confinement in the lateral direction, however, comes about because of the tapering of this confining potential well away from the apex of the V-groove. Lower-energy states in the quantum wire have probability densities that are greatest near the apices, where the vertical confining region is widest, and decrease away from the apex, as the confining region narrows. This leads to an effective lateral confinement (Fig. 6.10). With increasing energy (or external field), the wavefunctions spread out from the apex into the tapered region, but remain confined. A similar mechanism is operative for a serpentine superlattice of quantum wires (Miller *et al.* 1992).

#### 6.4.2 Ordered arrays of quantum dots on patterned substrates

The positioning of individual quantum dots at pre-determined locations on a substrate requires a pattern to promote nucleation at the desired locations. In a series of experiments using MBE on GaAs(001) (Atkinson *et al.* 2006), electron-beam lithography followed by conventional wet or dry etching was used to produce arrays of small holes 60–150 nm wide and 10–40 nm deep. The growth conditions were chosen to ensure that the amount of InAs deposited was less than the critical thickness for the appearance of 3D islands on the terraces (Section 6.3), so the islands formed solely in the holes.

**Fig. 6.10** (a) Left side of a (mirror-symmetric) V-groove quantum wire of the type shown in Fig. 6.9(c). (b) Contours corresponding to the right side of the (mirror-symmetric) ground-state wavefunction at the indicated external voltages in an electric field whose positive direction is along the positive  $z$ -axis. The confining effect of the tapered profile of the quantum wire is clearly evident. The calculational details are described by Vouilloz *et al.* (1998). Reused with permission from H. Weman, E. Martinet, M.-A. Dupertuis, A. Rudra, K. Leifer, and E. Kapon, *Appl. Phys. Lett.* **74**, 2334 (1999). Copyright 1999, American Institute of Physics.



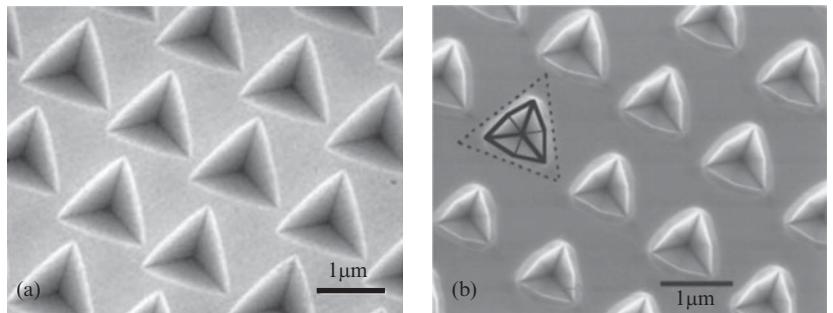


**Fig. 6.11**  $2.5 \mu\text{m} \times 2.5 \mu\text{m}$  AFM scans of a GaAs(001) wafer onto which an array of holes has been etched with *ex-situ* electron-beam lithography followed by conventional wet or dry etching and (a) *in-situ* hydrogen exposure, (b) the deposition of a 10-nm GaAs buffer film, and (c) after the deposition of 2 ML of InAs. The insets in (b) and (c) show the variation in cantilever amplitude, which is proportional to the surface gradient, for improved contrast. Representative line scans are shown below each surface scan. Note the missing island in the line scan in (c). Reused with permission from P. Atkinson, S.P. Bremner, D. Anderson, G.A.C. Jones, and D.A. Ritchie, *J. Vac. Sci. Technol. B* **24**, 1523 (2006). Copyright 2006, AVS The Science & Technology Society.

Figure 6.11 shows AFM scans of a dry-etched wafer after exposure to hydrogen (Fig. 6.11(a)), the deposition of 10 nm of GaAs buffer material (Fig. 6.11(b)), and the deposition of 2 ML of InAs (Fig. 6.11(c)). The line scans below each surface scan indicate the corresponding changes to the etched holes. Most apparent in Fig. 6.11(c) is the regular array of 3D islands within the etched holes. But notice also that, while a large proportion of the holes contain a single 3D island, some holes contain a pair of closely spaced islands, and some contain no island at all, having been completely filled by In(Ga)As. Incomplete subcritical wetting layers are discernible between the holes. Thus, there is expected to be substantial attachment and detachment of In and Ga adatoms from the edges of the complete layers and a large adatom population with a net mass current toward the attractive chemical potential of the holes. The distribution of occupancies of the holes is a sensitive function of the hole width, but in both the wet- and dry-etched cases, there is an optimal width for which a majority of holes are occupied by single quantum dots.

An alternative scenario uses MOVPE on substrates patterned to promote self-limiting growth of quantum dots that confine carriers in a manner analogous to that of V-groove quantum wires (Hartmann *et al.* 1997; 1999). Because this does not rely on the Stranski–Krastanov mechanism, a lattice mismatch between the constituents of the quantum dot and the substrate is not required. Our discussion here will focus on the lattice-matched AlAs/GaAs system (0.12% lattice mismatch). But there is another aspect to using MOVPE that

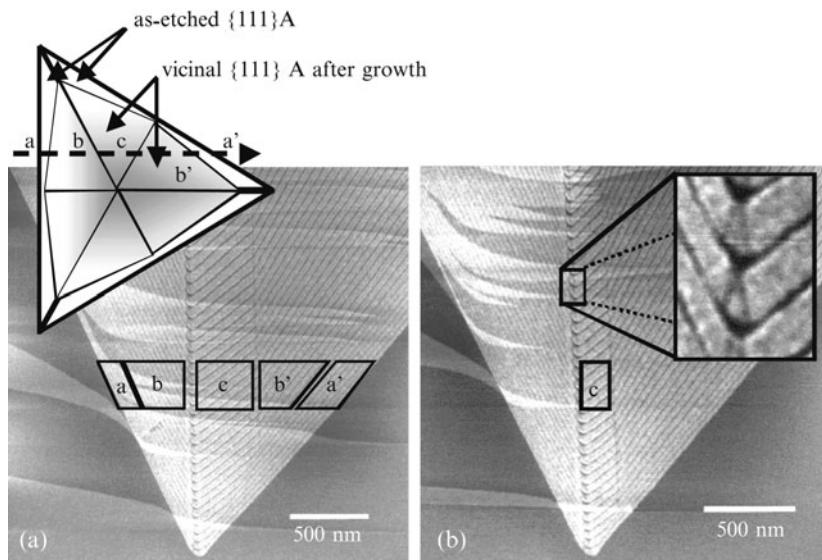
**Fig. 6.12** (a) Plan-view SEM image of a  $2\text{ }\mu\text{m}$  pitch pyramid array formed from (111)A bounding planes that have been wet etched onto a GaAs (111)B substrate. Reproduced with permission from A. Hartmann, Y. Ducommun, K. Leifer and E. Kapon, *J. Phys.: Condens. Matter* **11**, 5901 (1999). Copyright 1999 Institute of Physics. (b) Plan-view SEM image of the patterned surface in (a) after the growth of an AlGaAs/GaAs sequence, showing the development of vicinal (111)A facets during growth. The boundaries of the pyramids before growth are indicated by dashed lines and of the vicinal (111)A facets after growth by solid lines. Reused with permission from A. Hartmann, L. Loubies, F. Reinhardt and E. Kapon, *Appl. Phys. Lett.* **71**, 1314 (1997). Copyright 1997, American Institute of Physics.



merits some discussion—the facet dependence of the precursor decomposition rates. TMGa, or TMAI, and AsH<sub>3</sub> arrive at the substrate by diffusion through a boundary layer, after which these species and any fragments migrate along the terraces with minimal lateral interactions. The highly mobile precursors migrate across distances that are measured in micrometers. Decomposition reactions release Ga, Al, and As preferentially at step edges and other low coordination sites. Thus, facets on which decomposition rates are low can be used to transport the precursors into regions with higher decomposition rates and thereby engineer growth mainly in pre-determined regions of the substrate.

Figure 6.12(a) shows a scanning electron microscopy (SEM) image of the patterned substrate on which a regular array of quantum dots has been grown. The substrate is GaAs (111)B (As-terminated) with etched inverted pyramids bounded by three (111)A (Ga-terminated) planes. The facet dependence of growth is especially evident on this etched surface, as the growth occurs only within the inverted pyramids, with there being essentially no growth on the (111)B surface. Thus, the deposited precursors migrate along the (111)B surface to the etched inverted pyramids, where decomposition reactions release the constituent materials of the quantum dots. Although the low growth rate of the (111)B has been known for some time, its origin has only recently been conclusively demonstrated (Pelucchi *et al.* 2007) to be the vastly different decomposition rates of the metalorganic precursors (TMGa and TMAI). Figure 6.12(b) shows SEM images of the substrate after the growth of an AlGaAs/GaAs layer sequence. Each of the original (111)A planes develop into two vicinal (i.e. stepped) (111)A facets during growth. The emergence of vicinal facets is not altogether surprising in light of the evidence (Chua *et al.* 2008) to suggest that the decomposition reactions of the precursors are enhanced near steps.

The mechanism for the formation of these quantum dots is summarized in Fig. 6.13, which shows a cross-sectional AFM image, taken under ambient conditions, of an AlGaAs/GaAs growth sequence within an inverted pyramid. Since AlGaAs oxidizes faster than GaAs, the chemical differences of the two types of layers are observed as height differences in the AFM images. These images reveal a complex system of nanostructures that are formed during growth. There is a thickening of the GaAs layers near the three primary corners of the pyramids, between the (111)A sidewalls (Fig. 6.13(a)). This



**Fig. 6.13** (a) Cross-sectional AFM images showing the regular pattern of AlGaAs (light gray) and GaAs (dark gray) in the pyramidal holes in Fig. 6.12. The as-etched (111)A planes (marked “a” in the inset) are replaced during growth by vicinal (111)A planes “b,” “c,” and “b.” (b) The magnified view of the principal corner between facets “b” and “c” in the right inset demonstrates how the self-limited growth produces a vertical stacking of quantum wires and quantum dots. Reproduced with permission from A. Hartmann, Y. Ducommun, K. Leifer and E. Kapton, *J. Phys.: Condens. Matt.* **11**, 5901 (1999). Copyright 1999, Institute of Physics.

leads to the formation of quantum wires along these corners. A quantum dot is formed near the apex of an inverted pyramid where these quantum wires meet, near which there is also a thickening of the GaAs layers. The same type of mechanism is responsible for the formation of the corner quantum wires from quantum wells on the (111)A sidewalls. The radius of curvature reaches a self-limiting value during growth (see Fig. 6.9(c)), which allows quantum wires and quantum dots with effectively the same properties to be vertically stacked (Fig. 6.13(b)).

The self-limiting growth within the inverted pyramids also provides the mechanism for the confinement of carriers through the tapering of the GaAs regions near the apices of the pyramids, in analogy with that for V-groove quantum wires (Fig. 6.10). Moreover, because the formation of the quantum dots relies on chemical-potential gradients and decomposition-rate differences between facets, which results in facet-dependent growth rates, the dots have a more uniform size distribution than those produced solely by self-organization. Low-temperature (10 K) microphotoluminescence spectra of such arrays (Watanabe *et al.* 2004) show distinct luminescence from the quantum dots with a linewidth of 1–3 meV. Spectrally resolved images obtained from low-temperature cathodoluminescence reveal a uniform emission energy within an ensemble of 900 quantum dots.

## 6.5 Nanophotonics with quantum dots

Nanophotonics describes the generation (i.e. emission), propagation and detection of photons with nanoscale structures. The prospects of creating optical materials with novel properties and the possibility of fabricating densely integrated optical circuits has led to the investigation of a variety of concepts for nanophotonic materials. Photons are already the main carrier of information

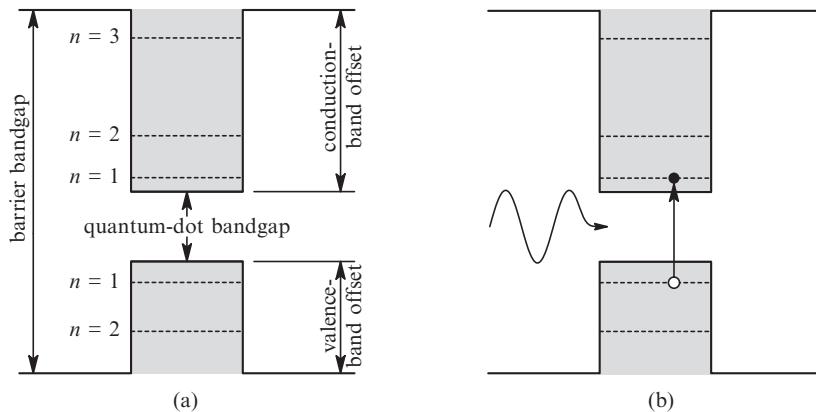
in modern telecommunications, albeit through conventional technologies such as fiber-optics and low index-contrast optical waveguides. More ambitious prospects for photonics-based information technologies stem from the fact that optically active nanostructures allow the storage of information in individual coherent states or in coherent superpositions. This opens up applications to quantum information processing if the coherent preparation and manipulation of the photons can be achieved. In this section, we describe the generation of quantum light by quantum dots.

Interest in the properties of quantum dots (and quantum wires) began in earnest in the 1980s with the realization that the dimensional confinement of the carriers beyond that offered by quantum wells held the promise for substantially improved laser characteristics. The principal advantages of using nanostructures as active media in lasers stem from the densities of states (Fig. 6.2), which enables the injected carriers to occupy narrow energy ranges near the valence- and conduction-band edges. This affords superior performance compared to bulk lasers in terms of higher material gain, lower threshold currents, reduced spectral broadening, and a substantially diminished temperature dependence, all of which are expected to improve with increasing dimensional confinement (Arakawa and Sakaki 1982; Ledentsov *et al.* 2000). Accordingly, much of the emphasis in the research and development of quantum dots has been in the arena of optics—and nanophotonics, in particular—for which there are abundant technological applications. However, the electronic and transport properties of quantum dots have also provided fertile ground for fundamental studies of correlated systems, as the references cited at the end of Section 6.1.3 indicate.

### **6.5.1 Optical properties of individual quantum dots**

Quantum dots are often referred to as “artificial atoms” because the carriers are confined in all three spatial directions, resulting in discrete energy levels (Fig. 6.2(c)), just as for atoms. But this analogy should not be taken too literally. Quantum dots are composed of (typically  $\sim 10^3\text{--}10^5$ ) atoms that are enclosed by heterogeneous interfaces, so the energy levels, carrier interactions, and the nature of the excitations are determined by factors that are fundamentally different from those in atoms. One consequence of these differences is that energy-level spacings in atoms are measured in electron-volts, while those in quantum dots are measured in millielectron-volts.

The basic premise of modelling the electronic states in quantum dots is that they are semiconductors in their own right in which the basic excitations are electrons and holes across an energy gap. This picture has provided the basis for much of our understanding of the properties of quantum wells and quantum wires (Johnson 2001) and has been used for detailed studies of quantum dots (Pryor 1999; Bester *et al.* 2007). The band-edge offsets of the surrounding semiconductor material confine the excited electrons and holes to the bound states of the quantum dots, as shown in Fig. 6.14(a). There is a finite number of bound states because of the finite size of the quantum dot and finite bounding potentials. The spectrum of an individual quantum dot, i.e. the number and



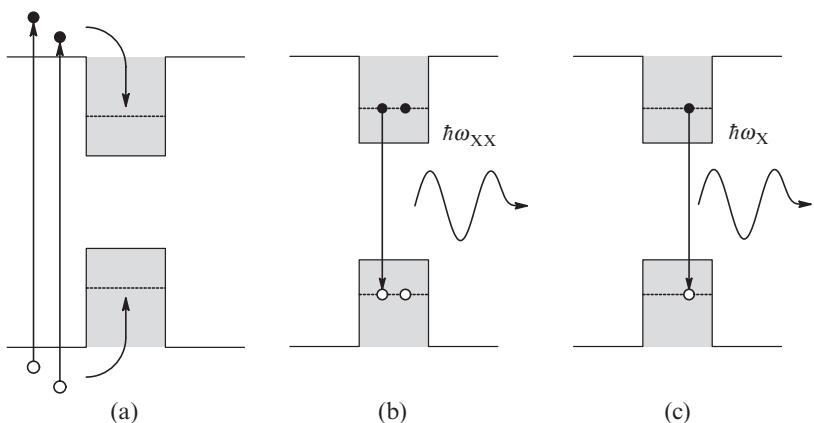
**Fig. 6.14** (a) The discrete energy levels of a semiconductor quantum dot (shaded region) obtained by surrounding a 3D island with a semiconductor that has a wider bandgap. The number of bound valence- and conduction-band states depends on factors such as the band offsets and the size of the confining region, as determined by elementary quantum mechanics. (b) The absorption of a photon causes an electron to be excited to the conduction level (filled circle), leaving a hole in the valence level (open circle). The attraction between the electron and hole leads to the formation of an exciton.

spacing of the energy levels, depends on several factors, including the size and shape of the quantum dot, and its composition. The dot-to-dot variation of any of these characteristics results in changes to the spectra of individual quantum dots, which causes an inhomogeneous broadening in the optical spectra of an array of dots, as discussed below.

In view of the foregoing, the illumination of a quantum dot by a pulse of light excites an electron to the conduction levels, leaving a hole in the valence levels (Fig. 6.14(b)). The Coulomb attraction between the electron and hole produces an exciton, i.e. a correlated electron–hole pair, that is confined to the dot. The recombination of the electron and hole produces a photon whose energy provides a measure of the energy-level spacings in the dot. This is the basis for optical characterization by photoluminescence (PL) (Kelsall *et al.* 2005). Electrons and holes are first excited by illumination with a laser. Photon energies are chosen so that absorption occurs in the surrounding material to produce large numbers of electrons and holes, which results in a strong PL signal. The migrating electrons and holes are captured by the attractive potential of the quantum dots, then quickly relax to the lowest-energy states (Fig. 6.14(b)). Recombination produces a photon whose energy is determined by a spectrometer. Depending on the size of the excited region, the PL signal embodies the contribution of many quantum dots, with fluctuations in size, shape, and composition causing an inhomogeneous broadening of a measured spectrum, or from individual dots, if a focused laser beam is used.

Suppose we consider a dot with only single spin-degenerate states in the valence and conduction bands. Then, upon illumination, this dot can trap two electrons and two holes from the surrounding material to form a biexcitonic state (Fig. 6.15(a)), which is an excitonic “molecule.” The biexciton is composed of two electrons and two holes, each of opposite spin, and is therefore in a spin-singlet state. Thus, direct recombination of the biexciton into the ground state of the quantum dot is optically forbidden, so the decay path is a cascade of two excitonic transitions. The biexciton decays when one of the electrons recombines with one of the holes, generating a first photon (Fig. 6.15(b)). This leaves a single electron–hole pair bound in an exciton, which also recombine to generate a second photon (Fig. 6.15(c)). The two photons have different

**Fig. 6.15** Schematic illustration of the biexciton cascade of a quantum dot. (a) The capture of two electrons and two holes by the dot from the surrounding material, which has been excited by a laser. The Coulomb attraction between the electrons and holes produces a biexciton. (b) The recombination of an electron–hole pair is accompanied by the emission of a biexcitonic photon with energy  $\hbar\omega_{XX}$ . (c) The recombination of the remaining electron–hole pair generates a second photon, called the excitonic photon, with energy  $\hbar\omega_X$ . The energies of the two photons are different because of the difference in the Coulomb interactions in their initial and final states.



energies because of the different Coulomb energies in their initial and final states. There can also be weaker lines resulting from the recombination of charged excitons that form intermittently when the dot captures an extra electron or hole (Landin *et al.* 1998). As the size of the quantum dot increases, there can be more bound states in the valence and conduction bands, resulting in a larger number of excitonic complexes that can be formed and, thus, a correspondingly richer optical spectrum.

The lifetime of the excitonic state, which is of the order of a nanosecond, is much longer than the duration of an exciting laser pulse and the lifetimes of the free electron–hole pairs in the surrounding semiconductor. Since the decay of this exciton is pre-dominantly radiative, i.e. accompanied by the emission of a photon, this provides the basis for single-photon emission from a quantum dot. This is discussed in the next section.

### 6.5.2 Classical and quantum light

Maxwell's equations describe the electromagnetic radiation from conventional sources of light. This is a *classical* picture in which light is an electromagnetic wave whose energy can take any (non-negative) value and whose energy and phase simultaneously have well-defined values. However, several applications in the emerging field of quantum information require strong quantum correlations between single photons, for which this classical description is inadequate. In the *quantum* theory of light the basic constituents are photons that have energies in units of  $\hbar\omega$ , where  $\omega$  is their frequency, whose energy and phase can no longer be measured simultaneously because of the number-phase Heisenberg uncertainty principle. Consider, for example, secure key distribution in quantum cryptography, which is dependent upon an optical source that emits a train of pulses that contain one and only one photon. Since any measurement unavoidably modifies the state of a quantum system, a third party cannot obtain information about the key without being detected, provided that the light pulses used for the transmission do not contain more than one photon. Another application of quantum light is quantum computing. The availability

of single-photon sources has been shown (Knill *et al.* 2001; Kok *et al.* 2007) to enable the implementation of quantum computation using only linear elements and photodetectors.

The statistics of photons are measured by the second-order correlation function  $g^{(2)}(\tau)$ , which is defined by

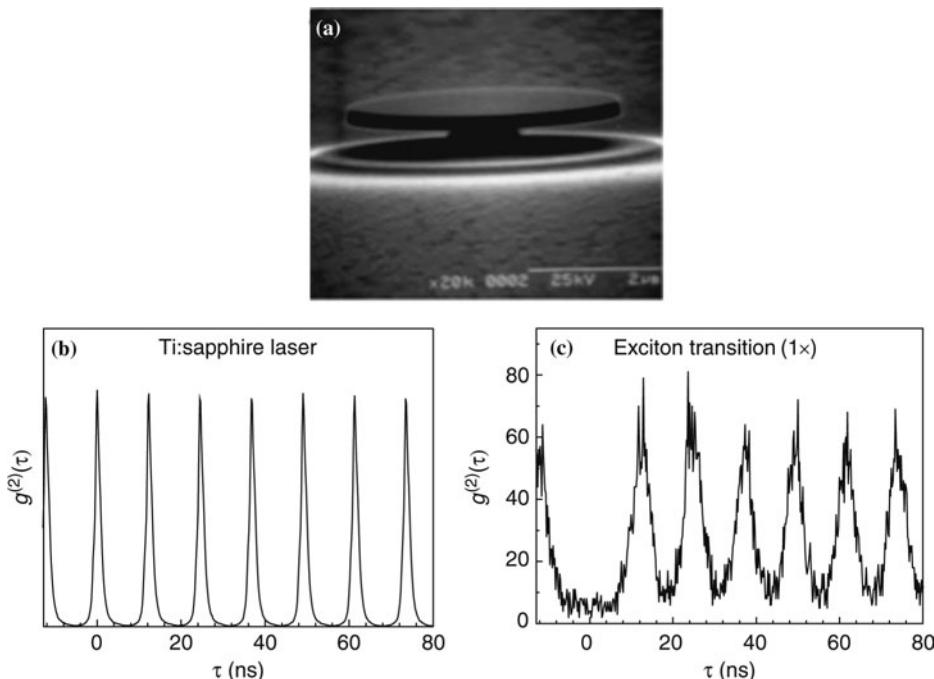
$$g^{(2)}(\tau) = \frac{\langle I(t)I(t+\tau) \rangle}{\langle I(t) \rangle^2}. \quad (6.8)$$

This function describes the correlations between light intensities separated by a time  $\tau$ , i.e. the likelihood of detecting a photon a time  $\tau$  after another photon was detected. Three regimes are identified by  $g^{(2)}(\tau)$ . For a source with  $g^{(2)}(\tau) = 1$ , there is no correlation between the photons. Continuous-light sources, such as lasers and light-emitting diodes (LEDs), fall into this category. If a source produces  $g^{(2)}(\tau) > 1$ , then there is an enhanced likelihood of finding two photons emitted within a time  $\tau$ . Such a source is said to be “bunched.” Alternatively, for a source with  $g^{(2)}(\tau) < 1$ , photons tend to be separated in time, and are said to be “antibunched.” This is the case for quantum light sources.

A driven quantum anharmonic system, such as a two-level atom, exhibits photon antibunching, where the emission of a photon reduces the emission probability of another photon immediately afterward. The basic principle of antibunching can be understood as follows. If a two-level atom emits a photon, this atom cannot emit another photon immediately afterward because it is in its ground state. The next photon can be emitted only after a waiting time that, in weak excitation conditions, is determined by its spontaneous emission time. The result is a time delay between successive photon emission events. Photon antibunching has been observed in a low-density vapor of sodium atoms (Kimble *et al.* 1977), a single ion stored in a radio-frequency trap (Diedrich and Walther 1987), a single pentacene molecule in a *p*-terphenyl host (Basché *et al.* 1992), a semiconductor quantum dot (Michler *et al.* 2000), and a single nitrogen-vacancy center in diamond (Brouri *et al.* 2000; Kurtsiefer *et al.* 2000). There is even a recent report (Högele *et al.* 2008) of photon antibunching in a single carbon nanotube. This finding is due to the fact that double occupancy of excited states is prohibited due to a number of effects. Electron–hole pairs are highly localized, and Auger processes, in which electron–hole pairs recombine without photon emission, are strong. Fewer than 1 in 20 events are observed as multiphoton emissions, making carbon nanotubes possible single-photon sources.

### 6.5.3 Single-photon sources

Figure 6.16(a) shows a microdisk on which there are InAs quantum dots produced by MBE (Michler *et al.* 2000). One of the quantum dots on the microdisk was excited using a Ti:sapphire laser, which is capable of producing femtosecond pulses of light at a rate of approximately 80 MHz (Fig. 6.16(b)). The photons emitted from the quantum dot and the time difference between emission events was recorded. The observed result (Fig. 6.16(c)) showed that there was essentially no peak at zero time difference between recorded emission events,



**Fig. 6.16** (a) A microdisk consisting of a 5- $\mu\text{m}$  diameter disk on a 0.5- $\mu\text{m}$   $\text{Al}_{0.65}\text{Ga}_{0.35}\text{As}$  post. The disk is composed of two GaAs layers between which is an InAs layer that contains the quantum dots. (b) The (unnormalized) correlation function  $g^{(2)}(\tau)$  in (8) for a Ti:sapphire laser that drives the microdisk. (c) Single quantum-dot excitonic emission under pulsed excitation conditions. This quantum-dot transition was out of resonance with the microdisk modes. From P. Michler, A. Kiraz, C. Becher, W.V. Schoenfeld, P.M. Petroff, L. Zhang, E. Hu, and A. Imamoğlu, *Science* **290**, 2282 (2000). Reprinted with permission from AAAS.

but a pronounced peak for every pulse interval thereafter. This indicates that the microdisk system emits single photons at regular intervals, since the emission of groups of photons would produce a peak at  $\tau = 0$ . Another noteworthy feature of the pulses in Fig. 6.16(c) is the reliability of the measured emission times of the photons after the excitation. The variations in emission times are reduced by the Purcell effect, which has a stabilizing effect on the spontaneous emission rate when the transition energy coincides with the energy of a cavity mode. Indeed, a major advantage of epitaxial quantum dots for single-photon generation is that they can be easily incorporated into cavities using standard growth and processing techniques, as the microdisk in Fig. 6.16(a) illustrates, and a variety of cavity structures has been fabricated (Shields 2007).

#### 6.5.4 Coupled quantum dots

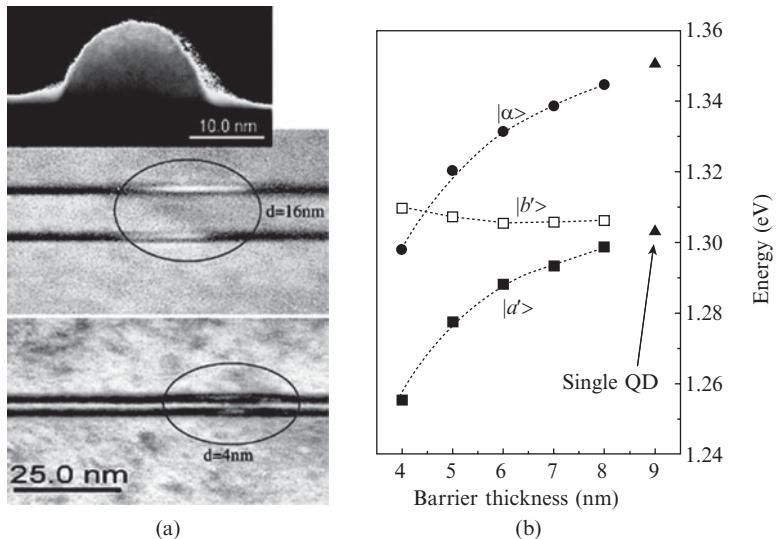
The next logical step after individual quantum dots is the study of systems composed of several tunnel-coupled quantum dots. To the extent that single quantum dots can be regarded as “artificial atoms,” two quantum dots can be coupled to form an “artificial molecule.” Depending on the strength of the tunnelling between the two dots, the coupled system forms either ionic-like

“bonds” if the tunnelling is weak, or covalent-like “bonds,” if the tunnelling is strong. In the ionic case, the electrons are localized on the individual dots and the binding occurs because a redistribution of the electrons leads to an attractive Coulomb interaction. In the covalent case, two electron states are quantum-mechanically coupled. The main requirement for covalent binding is that an electron can tunnel many times between the two dots in a phase-coherent way. Thus, the electron is not associated with either quantum dot, but must be regarded as a coherent wavefunction that is delocalized over the dot molecule.

Coupled quantum dots show fundamentally different behavior and provide new opportunities compared to single dots. Electrostatically coupled dots have been proposed for applications such as (classical) logic gates (Nomoto *et al.* 1996). But a more ambitious application of coupled quantum dots is in quantum computing. The theoretical possibility of performing certain tasks much more efficiently with a “quantum computer” instead of a “classical computer” has stimulated the search for physical realizations of the fundamental building block of a quantum computer: the quantum bit, or “qubit.” In principle, any two-level quantum system can be used as a qubit. One suggestion (Loss and DiVincenzo 1998) is an architecture based on coupled quantum dots in which an electron spin on each dot acts as a qubit, with entanglement of the qubits realized by coupling the two spins.

The basic component of a quantum processor is a quantum gate, which is used to entangle the states of two qubits. The critical requirements for a quantum gate are (i) that the coupling between the quantum dots is coherent, so that information is not lost during a computation, and (ii) that this coupling can be turned on and off externally to carry out the information exchange in the gate. Experiments that have demonstrated the feasibility of using pairs of stacked InAs quantum dots (see Fig. 6.7) as gates have been recently carried out by Bayer *et al.* (2001; 2003). Figure 6.17(a) shows two stacked quantum dots separated by 4 nm and 16 nm. The measurements were supported by quantum-mechanical calculations by diagonalizing the electron-hole Hamiltonian, which includes tunnelling and Coulomb effects, referred to a basis of four states in which an electron and hole can reside on either dot in the lowest energy states (which have zero orbital angular momentum). The luminescence emission lines of the quantum-dot pairs as a function of their separation are shown in Fig. 6.17(b). The optically active states, labelled  $|a'\rangle$  and  $|b'\rangle$ , are the excitonic analogs of symmetric and antisymmetric single-particle molecular states, with a splitting that is determined by tunnelling and Coulomb interactions. The  $|\alpha\rangle$  state originates from higher-lying states corresponding to non-vanishing orbital angular momentum, which were not included in the calculation. The observed behavior of the  $|a'\rangle$  and  $|b'\rangle$  states is in qualitative agreement with the model quantum-mechanical calculations. This suggests that the coupling between the dots in these experiments is inherently quantum mechanical and is, therefore, coherent. Additional experimental and theoretical results for the diamagnetic shifts and anticrossings in magneto-optical studies confirm that the coupling between the dots is indeed coherent (Ortner *et al.* 2003).

**Fig. 6.17** (a) Cross-sectional TEM images of pairs of InAs quantum dots (shown encircled) at separations of 4 nm and 16 nm. The upper panel shows an uncapped quantum dot produced under the same growth conditions. Reprinted with permission from S. Fafard, M. Spanner, J.P. McCaffrey, and Z.R. Wasilewski, *Appl. Phys. Lett.* **76**, 2268 (2000). Copyright 2000, American Institute of Physics. (b) The energies of emission lines in the coupled quantum dots as a function of their separation. The  $|\alpha\rangle$  state originates from higher-lying states. From M. Bayer, P. Hawrylak, K. Hinzer, S. Fafard, M. Korkusinski, Z.R. Wasilewski, O. Stern, and A. Forchel, *Science* **291**, 451 (2001). Reprinted with permission from AAAS.



## 6.6 Arrays of quantum dots

In the preceding section, we focused on the properties of single and double dots. Many of the studies reported on these systems have relied on the fabrication of arrays of self-assembled quantum dots as a starting point for further processing. An important advantage of such quantum dots for single-photon generation is that they can be incorporated into microcavities using standard semiconductor growth and processing techniques. Devices based on optical microcavities already find applications in a wide range of practical devices and play an important role in fundamental studies. For example, microcavities made of active III–V semiconductors control laser emission spectra to enable long-distance data transmission over optical fibers. They also ensure narrow read and write beams in compact disk (CD) and digital video disk (DVD) players. In this section, we discuss some of the applications where the properties of quantum-dot arrays are themselves of principal importance.

### 6.6.1 Quantum-dot computer memories

One application of quantum-dot arrays being researched on several fronts is their use in computer memories. To appreciate the potential benefits, we first consider the state-of-the-art. There are two basic issues: the retention of the memory, measured by the frequency with which the memory needs to be refreshed, and the accessibility of the information stored in the memory, which is measured by the time it takes to pass information. Dynamic random access memory (DRAM), the memory used by computers, is fast, with access times under 20 ns, but needs rapid refreshment, which increases power consumption. Hard disks, used for long-term information storage, also need to be rapidly refreshed, but have comparatively slow access rates. Flash

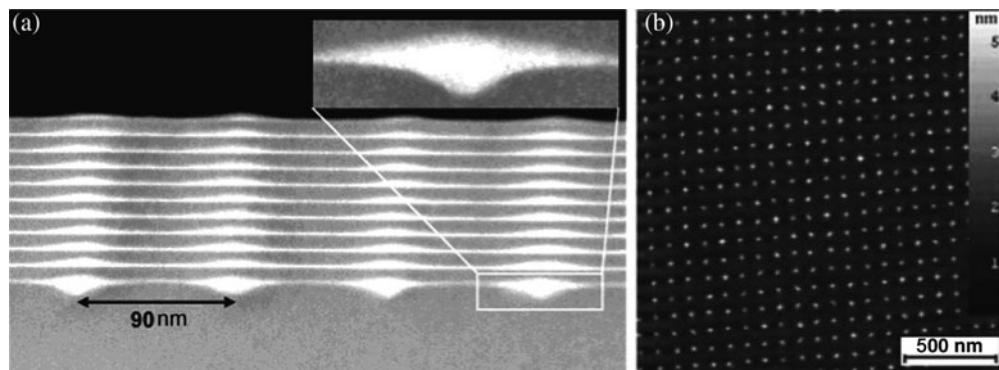
memories, which are used, for example, in memory sticks, digital cameras, and mobile telephones, have very long refresh times, but slow transmission times. These devices rely on a huge potential barrier created by a layer of SiO<sub>2</sub>. The probability of an electron tunnelling across this barrier, and thereby corrupting the stored information, is so low that the data remain intact for a time span that is measured in years. Correspondingly, however, the retrieval of these data relies on the transmission of electrons through the SiO<sub>2</sub> layer, which requires a substantial energy input that degrades the layer. Hence, there is a limited number of reliable cycles in flash memory.

Recent work (Marent *et al.* 2007; 2008) has suggested that using quantum dots as the fundamental storage units in a memory can remove the current limitations of retention and accessibility simultaneously. The basic operation of a storage-write-erase cycle of such a memory is as follows. Self-organized quantum dots that have been produced by the Stranski–Krastanov growth mode are placed in a p-n diode with the doping concentration adjusted so that, at zero external bias, the quantum dots are located within the depletion zone. This is the storage configuration. The width of the depletion zone can be varied by the application of an external bias. A forward bias reduces this width, in which case the quantum dots are now outside the depletion zone and able to capture carriers, and so become charged. Hence, in contrast to conventional flash memory operation, which has a fixed SiO<sub>2</sub> barrier height, rapid carrier capture into the quantum-dot states is induced by the external bias. To erase the information, a reverse bias is applied to the diode. The quantum dots are now inside the depletion region, with the strong electric field enabling tunnelling of the charge carriers from the dots through the confining barriers.

The performance of the memory is sensitive to the materials and compositions of the quantum dots and the matrix in which the dots are embedded. Data is stored by holes, which have a much higher effective mass than electrons, so their unintended tunnelling out of the quantum dots is unlikely. A storage time of 10<sup>6</sup> years is estimated for GaSb/AlAs quantum dots based on extrapolation of measured data. This reduces the refresh rate of the quantum-dot memory, as well as the magnitude of the tunnelling barrier, which can be overcome by applying an electric field, rendering the memory more robust against degradation over repeated cycles. Write times below 10 ns were obtained, but a reduction to below 1 ns was projected for improved device structures (Marent *et al.* 2007).

## 6.6.2 Three-dimensional quantum-dot crystals

The methods discussed in Sections 6.3 and 6.4 can be combined to produce a laterally ordered array of quantum dots within the first layer, which is then repeated within successive layers because of strain-induced preferential nucleation to produce a 3D *quantum-dot crystal*. Grützmacher *et al.* (2007) have fabricated such a material based on Si and Ge. Si(001) substrates were patterned with a periodic 2D array of pits by using extreme ultraviolet interference lithography and reactive-ion etching. Subsequent deposition by MBE of Ge produced preferential nucleation and growth of quantum dots in the pits,



**Fig. 6.18** (a) Cross-sectional TEM images of 10-period stacks of Ge islands and 10-nm Si spacer layers showing the vertical alignment of the quantum dots. The insert is a magnified image of a dot in the first layer. (b) AFM images of the topmost layer of the stack, showing the regularity of the quantum dot array. Reused with permission from D. Grützmacher, T. Fromherz, C. Dais, J. Stangl, E. Müller, Y. Ekinci, H.H. Solak, H. Sigg, R.T. Lechner, E. Wintersberger, S. Birner, V. Hoý, and G. Bauer, *Nano Lett.* **7**, 3150 (2007). Copyright 2007, American Chemical Society.

as in Fig. 6.11. Subsequent growth, interspersed with Si spacer layers, resulted in vertically aligned quantum dots, as shown in Fig. 6.18. The cross-sectional TEM image in Fig. 6.18(a) shows the integrity of the vertical alignment of the quantum dots. Figure 6.18(b) shows an AFM image of the top layer of quantum dots, which retain the periodicity and regularity of the original 2D array. The size distribution of the dots in the first layer is found to have less than 10% variations in the height and diameter, which persists to the quantum dots in the top layer. AFM scans over an area of  $8\ \mu\text{m} \times 8\ \mu\text{m}$  indicate that there are neither missing nor extra dots in the array (see Fig. 6.11). The structural regularity and low defect density of this 3D quantum-dot crystal suggests potential applications in Si-based optoelectronics, spintronics, and quantum computation.

## 6.7 Summary and outlook

Semiconductor quantum dots represent the ultimate degree of control that is attainable by modern fabrication technologies to produce materials that confine carriers in three spatial dimensions. Quantum dots combine basic characteristics of atoms, such as discrete energy spectra, with highly tunable properties, discrete excitonic transitions, large oscillator strengths, and strong luminescence, while allowing integration within solid-state systems. For these reasons, the optical properties of quantum dots have provided a rich arena for research for many applications ranging from telecommunications to quantum information.

There are two main strategies for producing epitaxial quantum dots: self-organization, possibly with some strategy to augment the lateral ordering by imposing variations of the surface chemical potential, and directed self-assembly on patterned substrates. Self-organization has the virtue of requiring no additional processing apart from standard surface preparation, but suffers from inherent statistical fluctuations in the positions, growth rates, and compositions within the quantum-dot ensemble. These can be alleviated to some

degree by growing multilayer stacks of quantum dots. Patterned substrates overcome this problem by providing pre-determined sites for the nucleation of quantum dots and MOVPE takes further advantage of this strategy by utilizing the facet dependence of the decomposition process to enable complex hybrid nanostructures to be fabricated in a controlled and reproducible manner.

Epitaxial techniques enable the engineering of charge and spin states and the wavefunctions of excitations within quantum dots and the tailoring of their optical absorption, emission, and refraction spectra. In particular, we have seen that the properties of carriers within quantum dots and other nanostructures are uncovering new physics in condensed-matter systems that is being harnessed for new applications. For example, cavity quantum electrodynamics effects can be investigated by using a single quantum dot embedded inside a photonic nanostructure, where both the carriers and photons are confined to submicrometer length scales in all three dimensions. Since the location of the quantum dot inside the cavity is fixed by the growth, this system is free from the stringent trapping requirements that limit its atomic counterpart. Using quantum dots in logic and memory elements that are required for quantum information processing represents an exciting application that may eventually see the integration of such functionality into a photonic integrated circuit.

But there are other applications that have shorter development horizons than quantum information. The recent discovery (Bowers *et al.* 2005) of CdSe quantum dots that emit white light has enormous potential for consumer electronics, where light-emitting diodes are already making inroads, while providing substantial savings in terms of power consumption compared to conventional lighting. Solar cells and photovoltaics are another area where quantum dots can provide a significant improvement over existing technology. Traditional solar cells are made from semiconductors, are expensive to produce and have an efficiency with a theoretical upper limit of 33% for the conversion of sunlight to electricity. But utilizing quantum dots increases this upper limit to near 60%, while providing appreciable savings in production costs (Asokan *et al.* 2007).

An altogether different class of semiconductors, the Group-III nitrides, have had an immediate technological impact in recent years. Interest in these materials began in the 1990s when a process developed by Nakamura *et al.* (1992) allowed the manufacture of LEDs and semiconductor diode lasers that operate at light wavelengths from the ultraviolet to the green. Today, GaN and InGaN LEDs are a multibillion dollar industry with applications in cell-phone back-lighting, traffic signals, and flat-panel displays (Ponce and Bour 1997), to name but a few. A natural extension of this work is the fabrication of quantum wires and quantum dots, again motivated by changes to the electronic states in these structures resulting from the additional confinement. Quantum dots based on Group-III nitride compounds, notably GaN (Pelekanos *et al.* 2005), have opened up new opportunities in the field of blue lasers for optical data storage and high-resolution printing. But challenges remain because of the lack of a well-developed methodology for the synthesis of these materials.

The application of colloidal quantum dots in the biological sciences has seen tremendous progress in recent years (Michalet *et al.* 2005) on a variety of fronts. Such quantum dots have a semiconductor core that is surrounded by a shell and encased in a biocompatible coating. Different coatings will bind to different receptors, which allows for site-specific binding and detection of cancer cells. Quantum dots are currently used for *in-vitro* tumor imaging in animals, with a recent use being the identification of lymph nodes in live animals by near-infra-red emission during surgery. However, there are toxicity issues that must be resolved before quantum dots can be used in humans. Quantum dots degrade under very high oxidative or photostresses, which can expose their toxic semiconductor cores. Nevertheless, quantum dots have far-reaching potential for the study of intracellular processes at the single-molecule level, high-resolution cellular imaging, tumor targeting, and diagnostics.

## References

- Alchalabi, K., Zimin, D., Kostorz, G., Zogg, H. *Phys. Rev. Lett.* **90**, 026104 (2003).
- Amar, J.G., Family, F. *Phys. Rev. Lett.* **74**, 2066 (1995).
- Aqua, J.-N., Frisch, T. *Phys. Rev. B* **78**, 121305 (2008).
- Arakawa, Y., Sakaki, H. *Appl. Phys. Lett.* **40**, 939 (1982).
- Arciprete, F., Placidi, E., Sessi, V., Fanfoni, M., Patella, F., Balzarotti, A. *Appl. Phys. Lett.* **89**, 041904 (2006).
- Asokan, S., Krueger, K.M., Colvin, V.L., Wong, M.S. *Small* **3**, 1164 (2007).
- Atkinson, P., Bremner, S.P., Anderson, D., Jones, G.A.C., Ritchie, D.A. *J. Vac. Sci. Technol. B* **24**, 1523 (2006).
- Avery, A.R., Dobbs, H.T., Holmes, D.M., Joyce, B.A., Vvedensky, D.D. *Phys. Rev. Lett.* **79**, 3938 (1997).
- Bartelt, M.C., Evans, J.W. *Phys. Rev. B* **46**, 12675 (1992).
- Bartelt, M.C., Evans, J.W. *Phys. Rev. B* **54**, R17359 (1996).
- Bartelt, M.C., Perkins, L.S., Evans, J.W. *Surf. Sci.* **344**, L1193 (1995).
- Basché, Th., Moerner, W.E., Orrit, M., Talon, H. *Phys. Rev. Lett.* **69**, 1516 (1992).
- Bauer, E. *Z. Krist.* **110**, 372 (1958).
- Bayer, M., Hawrylak, P., Hinzer, K., Fafard, S., Korkusinski, M., Wasilewski, Z.R., Stern, O., Forchel, A. *Science* **291**, 451 (2001).
- Bayer, M., Ortner, G., Forchel, A., Lyanda-Geller, Y.B., Reinecke, T.L., Hawrylak, P., Fafard, S., Wasilewski, Z.R. *Phys. Rev. Lett.* **90**, 086404 (2003).
- Beenakker, C.W. *Rev. Mod. Phys.* **69**, 731 (1997).
- Bell, G.R., Krzyzewski, T.J., Joyce, P.B., Jones, T.S. *Phys. Rev. B* **61**, R10551 (2000).
- Bester, G., Reuter, D., He, L., Zunger, A., Kailuweit, P., Wieck A.D., Zeitler, U., Maan, J.C., Wibbelhoff, O., Lorke, A. *Phys. Rev. B* **76**, 075338 (2007).
- Biasioli, G., Gustafsson, A., Leifer, K., Kapon, E. *Phys. Rev. B* **65**, 205306 (2002).
- Bimberg, D., Grundmann, M., Ledentsov, N.N. *Quantum Dot Nanostructures* (Wiley, Chichester, UK, 1999).

- Bowers, II, M.J., McBride, J.R., Rosenthal, S.J. *J. Am. Chem. Soc.* **127**, 15378 (2005).
- Bressler-Hill, V., Lorke, A., Varma, S., Petroff, P.M., Pond, K., Weinberg, W.H. *Phys. Rev. B* **50**, 8479 (1994).
- Brouri, R., Beveratos, A., Poizat, J.-P., Grangier, P. *Opt. Lett.* **25**, 1294 (2000).
- Bryant, G.W., Solomon, G.S. eds. *Optics of Quantum Dots and Wires* (Artech House, Norwood, MA, 2005).
- Cao, Y.G., Xie, M.H., Liu, Y., Hu, S.X., Ng, Y.F., Wu, H.S., Tong, S. Y. *Phys. Rev. B* **68**, 161304 (2003).
- Chakraborty, T. *Quantum Dots* (Elsevier, Amsterdam, The Netherlands, 1999).
- Cho, A. (ed.) *Molecular Beam Epitaxy* (American Institute of Physics, New York, 1994).
- Chua, A.L.-S., Pelucchi, E., Rudra, A., Dwir, B., Kapon, E., Zangwill, A., Vvedensky, D.D. *Appl. Phys. Lett.* **92**, 013117 (2008).
- Costantini, G., Manzano, C., Songmuang, R., Schmidt, O.G., Kern, K. *Appl. Phys. Lett.* **82**, 3194 (2003).
- Costantini, G., Rastelli, A., Manzano, C., Acosta-Diaz, P., Songmuang, R., Katsaros, G., Schmidt, O.G., Kern, K. *Phys. Rev. Lett.* **96**, 226106 (2006).
- Cronenwett, S.M., Oosterkamp, T.H., Kouwenhoven, L.P. *Science* **281**, 540 (1998).
- Cullis, A.G., Norris, D.J., Walther, T., Migliorato, M.A., Hopkinson, M. *Phys. Rev. B* **66**, 081305(R) (2002).
- Cullis, A.G., Norris, D.J., Walther, T., Migliorato, M.A., Hopkinson, M. in *Quantum Dots: Fundamentals, Applications, and Frontiers* (eds), Joyce, B.A., Kelires, P.C., Naumovets, A.G. and Vvedensky, D.D. (Springer, Dordrecht, The Netherlands, 2005), pp. 71–88.
- Diedrich, F., Walther, H. *Phys. Rev. Lett.* **58**, 203 (1987).
- Dingle, R., Wiegmann, W., Henry, C.H. *Phys. Rev. Lett.* **33**, 827 (1974).
- Eaglesham, D.J., Cerullo, M. *Phys. Rev. Lett.* **64**, 1943 (1990).
- Ebiko, Y., Muto, S., Suzuki, D., Itoh, S., Shiramine, K., Haga, T., Nakata, Y., Yokoyama, N. *Phys. Rev. Lett.* **80**, 2650 (1998).
- Ebiko, S., Muto, S., Suzuki, D., Itoh, S., Yamakoshi, H., Shiramine K., Haga, T., Unno, K., Ikeda, M. *Phys. Rev. B* **60**, 8234 (1999).
- Esaki, L., Tsu, R. *IBM J. Res. Develop.* **14**, 61 (1970).
- Evans, J.W., Thiel, P.A., Bartelt, M.C. *Surf. Sci. Rep.* **61**, 1 (2006).
- Fanfoni, M., Placidi, E., Arciprete, F., Orsini, E., Patella, F., Balzarotti, A. *Phys. Rev. B* **75**, 245312 (2007).
- Georgsson, K., Carlsson, N., Samuelson, L., Seifert, W., Wallenberg, L.R. *Appl. Phys. Lett.* **67**, 2981 (1995).
- Goldman, R.S. *J. Phys. D: Appl. Phys.* **37**, R163 (2004).
- Grützmacher, D., Fromherz, T., Dais, C., Stangl, J., Müller, E., Ekinci, Y., Solak, H.H., Sigg, H., Lechner, R.T., Wintersberger, E., Birner, S., Holý, V., Bauer, G. *Nano Lett.* **7**, 3150 (2007).
- Hanson, R., Kouwenhoven, L.P., Petta, J.R., Tarucha, S., Vandersypen, L.M. *Rev. Mod. Phys.* **79**, 1217 (2007).

- Hartmann, A., Ducommun, Y., Leifer, K., Kapon, E. *J. Phys.: Condens. Matter* **11**, 5901 (1999).
- Hartmann, A., Loubies, L., Reinhardt, F., Kapon, E. *Appl. Phys. Lett.* **71**, 1314 (1997).
- Herring, C. in *The Physics of Powder Metallurgy* (ed.), Kingston, W.E. (McGraw-Hill, New York, 1951), 143–179.
- Högele, A., Galland, C., Winger, M., Imamoğlu, A. *Phys. Rev. Lett.* **100**, 217401 (2008).
- Ichimura, A., Cohen, P.I. *Reflection High-energy Electron Diffraction* (Cambridge University Press, Cambridge, UK, 2004).
- Jensen, K.F. in *Chemical Vapor Deposition Principles and Applications* (eds), Hitchman, M.L. and Jensen, K.F. (Academic, London, 1993), 31–90.
- Johnson, E.A. in *Low-dimensional Semiconductor Structures: Fundamentals and Device Applications* (eds), Barnham, K. and Vvedensky, D.D. (Cambridge University Press, Cambridge, UK, 2001), 56–123.
- Joyce, B.A. *Rep. Prog. Phys.* **48**, 1637 (1985).
- Joyce, B.A., Kelires, P.C., Naumovets, A.G., Vvedensky, D.D. (ed.) *Quantum Dots: Fundamentals, Applications, and Frontiers* (Springer, Dordrecht, The Netherlands, 2005).
- Joyce, B.A., Vvedensky, D.D. *Mater. Sci. Eng. R* **46**, 127 (2004).
- Joyce, P.B., Krzyzewski, T.J., Bell, G.R., Joyce, B.A., Jones, T.S. *Phys. Rev. B* **58**, R15981 (1998).
- Kastner, M.A. *Ann. Phys. (Leipzig)* **9**, 885 (2000).
- Kelsall, R.W., Hamley, I.W., Geoghegan, M. *Nanoscale Science and Technology* (Wiley, Chichester, UK, 2005).
- Kimble, H.J., Dagenais, M., Mandel, L. *Phys. Rev. Lett.* **39**, 691 (1977).
- Kitamura, M., Nishioka, M., Ohinowo, J., Arakawa, Y. *Appl. Phys. Lett.* **66**, 3663 (1995).
- Knill, E., Laflamme, R., Milburn, G. *Nature* **409**, 46 (2001).
- Koduvely, H.M., Zangwill, A. *Phys. Rev. B* **60**, R2204 (1999).
- Kok, P., Munro, W.J., Nemoto, K., Ralph, T.C., Dowling, J.P., Milburn, G.J. *Rev. Mod. Phys.* **79**, 135 (2007).
- Krzyzewski, T.J., Joyce, P.B., Bell, G.R., Jones, T.S. *Phys. Rev. B* **66**, 201302 (2002).
- Kurtsiefer, C., Mayer, S., Zarda, P., Weinfurter, H. *Phys. Rev. Lett.* **85**, 290 (2000).
- Landin, L., Miller, M.S., Pistol, M.-E., Pryor, C.E., Samuelson, L. *Science* **280**, 262 (1998).
- Ledentsov, N.N., Grundmann, M., Heinrichsdorff, F., Bimberg, D., Ustinov, D.M., Zhukov, A.E., Maximov, M.V., Alferov Zh.I., Lott, J. A. *IEEE J. Sel. Top. Quantum Electron.* **6**, 439 (2000).
- Lee, H., Johnson, J.A., Speck, J.S., Petroff, P.M. *J. Vac. Sci. Technol. B* **18**, 2193 (2000).
- Leonard, D., Krishnamurthy, M., Reaves, C.M., Denbaars, S.P., Petroff, P.M. *Appl. Phys. Lett.* **63**, 3203 (1993).
- Leroy, F., Renaud, G., Letoublon, A., Lazzari, R., Mottet, C., Goniakowski, J. *Phys. Rev. Lett.* **95**, 185501 (2005).

- Liao, X.Z., Zou, J., Cockayne, D.J., Leon, R., Lobo, C. *Phys. Rev. Lett.* **82**, 5148 (1999).
- Loss, D., DiVincenzo, D.P. *Phys. Rev. A* **57**, 120 (1998).
- Lytvyn, P.M., Strel'chuk, V.V., Kolomys, O.F., Prokopenko, I.V., Valakh, M. Ya., Mazur Yu. I., Wang Zh. M., Salamo, G.J. *Appl. Phys. Lett.* **91**, 173118 (2007).
- Madhukar, A. *Surf. Sci.* **132**, 344 (1983).
- Madhukar, A., Rajkumar, K.C. *Appl. Phys. Lett.* **57**, 2110 (1990).
- Maksym, P.A., Imamura, H., Mallon, G.P., Aoki, H. *J. Phys.: Condens. Matter* **12**, R299 (2000).
- Marent, A., Geller, M., Schliwa, A., Feise, D., Pötschke, K., Bimberg, D., Akçay, N., Öncan, N. *Appl. Phys. Lett.* **91**, 242109 (2007).
- Marent, A., Geller, M., Bimberg, D. *Microelectron. J.* doi:10.1016/j.mejo.2008.06.056 (2008).
- Marzin, J.Y., Garard, J.M., Izrael, A., Barrier, D., Bastard, G. *Phys. Rev. Lett.* **73**, 716 (1994).
- Masumoto, Y., Takagahara, T. (eds) *Semiconductor Quantum Dots* (Springer, Berlin, 2002).
- Michalet, X., Pinaud, F.F., Bentolila, L.A., Tsay, J.M., Doose, S., Li, J.J., Sundaresan, G., Wu, A.M., Gambhir, S.S., Weiss, S. *Science* **307**, 538 (2005).
- Michler, P. (ed.) *Single Quantum Dots: Fundamentals, Applications and new Concepts* (Springer, New York, 2004).
- Michler, P., Imamoğlu, A., Mason, M.D., Carson, P.J., Strouse, G.F., Buratto, S.K. *Nature* **406**, 968 (2000).
- Migliorato, M.A., Cullis, A.G., Fearn, M., Jefferson, J.H. *Phys. Rev. B* **65**, 11536 (2002).
- Miller, M.S., Weman, H., Pryor, C.E., Krishnamurthy, M., Petroff, P.M., Kroemer, H., Merz, J.L. *Phys. Rev. Lett.* **68**, 3464 (1992).
- Moison, J.M., Houzay, F., Barthe, F., Leprince, L., André, E., Vatel, O. *Appl. Phys. Lett.* **64**, 18 (1994).
- Mui, D.S., Leonard, D., Coldren, L.A., Petroff, P.M. *Appl. Phys. Lett.* **66**, 1620 (1995).
- Mulheran, P.A., Blackman, J.A. *Phys. Rev. B* **53**, 10261 (1996).
- Murray, C.B., Norris, D.J., Bawendi, M.G. *J. Am. Chem. Soc.* **115**, 8706 (1993).
- Nakamura, S., Mukai, T., Senoh, M., Iwasa, N. *Jap. J. Appl. Phys. Part 2*, **31**, L139 (1992).
- Nakamura, Y., Schmidt, O.G., Jin-Phillip, N.Y., Kiravittaya, S., Mueller, C., Eberl, K., Graebeldinger, H., Schweizer, H. *J. Cryst. Growth* **242**, 339 (2002).
- Nikolić, K., MacKinnon, A. *Phys. Rev. B* **50**, 11008 (1994).
- Niu, X.B., Lee, Y.-J., Caflisch, R.E., Ratsch, C. *Phys. Rev. Lett.* **101**, 086103 (2008).
- Nomoto, K., Ugajin, R., Suzuki, T., Hase, I. *J. Appl. Phys.* **79**, 291 (1996).
- Ogawa, T., Takagahara, T. *Phys. Rev. B* **44**, 8138 (1991).
- Ortner, G., Bayer, M., Larionov, A., Timofeev, V.B., Forchel A., Lyanda-Geller, Y.B., Reinecke, T.L., Hawrylak, P., Fafard, S., Wasilewski, Z. *Phys. Rev. Lett.* **90** 086404 (2003).
- Ozdemir, M., Zangwill, A. *J. Vac. Sci. Technol. A* **10**, 684 (1992).
- Pashley, D.W., Neave, J.H., Joyce, B.A. *Surf. Sci.* **476**, 35 (2001).

- Pelekanos, N.T., Dialynas, G.E., Simon, J., Mariette, H., Daudin, B. *J. Phys.: Conf. Ser.* **10**, 61 (2005).
- Pelucchi, E., Watanabe, S., Leifer, K., Dwir, B., Zhu, Q., De Los Rios, P., Kapon, E. *Nano Lett.* **7**, 1282 (2007).
- Pimpinelli, A., Villain, J. *Physics of Crystal Growth* (Cambridge University Press, Cambridge, UK, 1998).
- Placidi, E., Arciprete, F., Fanfoni, M., Patella, F., Orsini, E., Balzarotti, A. *J. Phys.: Condens. Matter* **19**, 225006 (2007).
- Ponce, F.A., Bour, D.P. *Nature* **386**, 351 (1997).
- Pryor, C. *Phys. Rev. B* **60**, 2869 (1999).
- Ratsch, C., Šmilauer, P., Zangwill, A. *Surf. Sci.* **314**, L937 (1994a).
- Ratsch, C., Zangwill, A., Šmilauer, P., Vvedensky, D.D. *Phys. Rev. Lett.* **72**, 3194 (1994b).
- Ratsch, C., Šmilauer, P., Zangwill, A., Vvedensky, D.D. *Surf. Sci.* **329**, L599 (1995).
- Ravisharan, A., Liu, C.-P., Kim, J., Cahill, D.G., Gibson, J.M. *Phys. Rev. B* **63**, 125314 (2001).
- Raymond, S., Fafard, S., Charbonneau, S., Leon, R., Petroff, P.M., Merz, J.L. *Phys. Rev. B* **52**, 17238 (1995).
- Richter, W. *Appl. Phys. A* **75**, 129 (2002).
- Rossi, F., Molinari, E. *Phys. Rev. B* **53**, 16462 (1996).
- Scherer, A., Craighead, H.G. *Appl. Phys. Lett.* **49**, 1284 (1986).
- Schmidt, O.G. (ed.) *Lateral Alignment of Epitaxial Quantum Dots* (Springer, Berlin, 2007).
- Shchukin, V.A., Bimberg, D. *Rev. Mod. Phys.* **71**, 1125 (1999).
- Shields, A.J. *Nature Photon.* **1**, 215 (2007).
- Singh, J. *Physics of Semiconductors and their Heterostructures* (MacGraw-Hill, New York, 1993).
- Skolnick, M.S., Mowbray, D.J. *Annu. Rev. Mater. Res.* **34**, 181 (2004).
- Stangl, J., Holý, V., Bauer, G. *Rev. Mod. Phys.* **76**, 725 (2004).
- Strel'chuk, V.V., Lytvyn, P.M., Kolomys, A.F., Valakh, M. Ya., Mazur Yu. I., Wang Zh., M., Salamo, G.J. *Semiconductors* **41**, 73 (2007).
- Stringfellow, G.B. *Organometallic Vapor-phase Epitaxy* (Academic, Boston 1989).
- Taylor, J.P., Hugill, K.J., Vvedensky, D.D., MacKinnon, A. *Phys. Rev. Lett.* **67**, 2359 (1991).
- Teichert, C., Lagally, M.G., Petcolas, L.J., Bean, J.C., Tersoff, J. *Phys. Rev. B* **53**, 16334 (1996).
- Tersoff, J., Teichert, C., Lagally, M.G. *Phys. Rev. Lett.* **76**, 1675 (1996).
- Tu, Y., Tersoff, J. *Phys. Rev. Lett.* **93**, 216101 (2004).
- Ustinov, V.M., Zhukov, A.E., Egorov, A.Y., Maleev, N.A. *Quantum Dot Lasers* (Oxford University Press, New York, 2003).
- van der Wiel, W.G., De Franceschi, S., Elzerman, J.M., Fujisawa, T., Tarucha, S., Kouwenhoven, L.P. *Rev. Mod. Phys.* **75**, 1 (2003).

- Venables, J.A. *Introduction to Surface and Thin Film Processes* (Cambridge University Press, Cambridge, UK, 2000).
- Voigtländer, B. *Surf. Sci. Rep.* **43**, 127 (2001).
- Vouilloz, F., Oberli, D.Y., Dupertuis, M.-A., Gustafsson, A., Reinhardt, F., Kapon, E. *Phys. Rev. B* **57**, 12378 (1998).
- Vvedensky, D.D. *Phys. Rev. B* **62**, 15435 (2000).
- Walther, T., Cullis, A.G., Norris, D.J., Hopkinson, M. *Phys. Rev. Lett.* **86**, 2381 (2001).
- Watanabe, S., Pelucchi, E., Dwir, B., Baier, M.H., Leifer, K., Kapon E. *Appl. Phys. Lett.* **84**, 2907 (2004).
- Xie, Q., Madhukar, A., Chen, P., Kobayashi, N.P. *Phys. Rev. Lett.* **75**, 2542 (1995).

# 7

# Intersublevel quantum-dot infrared photodetectors

7.1 Introduction	244
7.2 Infrared photon absorption	247
7.3 Some metrics for photon detectors	254
7.4 Experimental single-pixel quantum-dot infrared photodetectors	260
7.5 Device characteristics	270
7.6 Toward quantum-dot focal plane array imagers	282
7.7 Challenges and prospects for high-performance detectors and arrays	288
Acknowledgments	290
References	290

*E. Towe and D. Pal*

## 7.1 Introduction

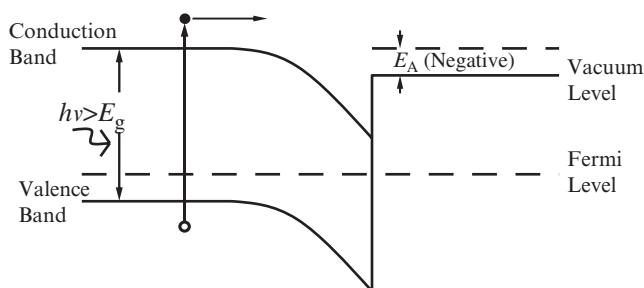
The goal of this chapter is to provide the basic principles of semiconductor quantum-dot infrared photodetectors based on conduction-band intersublevel transitions. Sufficient background material is discussed to enable an appreciation of the subtle differences between quantum-well and quantum-dot devices. No attempt is made to discuss details or subtleties of device design. However, some discussion of device results from our own work and that of others is presented. The field of semiconductor quantum-dot nanostructures in general, and quantum-dot devices in particular, is currently a very dynamic field of research. Any attempt on our part to be exhaustive would be futile as advances are being made almost daily.

### 7.1.1 Basic detection approaches

Infrared radiation is broadly defined to be that portion of the electromagnetic spectrum that stretches from about 0.75 to 1000  $\mu\text{m}$ . There are two distinct methods for detecting radiation in this portion of the spectrum: the first is by monitoring a change in temperature of an object that has absorbed radiant infrared energy, and the second is by converting the energy directly to electrical form by generating electronic charge. The former method is often referred to as thermal detection. It relies on detectors called bolometers. The latter approach is photon detection—otherwise known as photodetection. Within photodetection, there are photoconductive and photovoltaic detectors. In this chapter, we focus primarily on photon detection. In particular, on how quantum-confined size effects in semiconductors are used in photon detection. We begin by providing a brief overview of infrared detection to set the context within which semiconductor quantum structures are discussed.

### 7.1.2 Semiconductor photon detectors

The objective, in the detection of infrared radiation by semiconductors, is to arrange for incident electromagnetic energy to interact with a semiconductor medium in such a way that electronic charge carriers in the semiconductor absorb the radiant energy. If the absorbing charge carriers acquire sufficient energy to dislodge them from their bound equilibrium states, they may be excited into states in the conduction (or valence) bands of the semiconductor where they become mobile free carriers (electrons in the conduction band or holes in the valence band). Depending on the nature of the arrangement in the semiconductor film in the structure, an internal or external electric field may act to collect the mobile charge carriers as electrical current, thus making evident the detection of the radiation. If the charge carriers are not collected as current but are instead photoemitted, then the process must necessarily involve (i) excitation of the charge carriers by absorption of photons (to become photoelectrons), (ii) diffusion of the photoelectrons to a surface where they can be emitted, and (iii) escape of the photoelectrons from the surface to the surrounding environment (typically a vacuum) where they may then be collected. This method of photon detection is identical to a phenomenon first explained by Einstein in 1905 in terms of electron emission induced by absorption of incident radiant energy. In such a scheme, an appropriate semiconductor is deposited on a substrate to form a cathode. The principle of operation of this mode of detection comes closest to the mechanism that is the focus of this chapter: intersublevel detection. Intersublevel electronic transitions in quantum-dot nanostructures can be used to detect incident infrared radiation. We illustrate in Fig. 7.1, the conceptual aspects of photoemission. The illustration shows the energy band structure of a p-type semiconductor with a surface layer of an alkali metal such as cesium. The role of the surface layer is to reduce the electron affinity energy,  $E_A$ —defined here as the energy difference between the bottom of the conduction band of the semiconductor and the vacuum level (where the electron is essentially free). For cesium on top of a p-type semiconductor, the electron affinity is negative; ensuring that excited electrons can indeed be collected as evidence of any absorbed infrared radiation. In a detection process that uses the photoemission mechanism, the peak wavelength of the detected radiation is determined by the bandgap of the semiconductor,  $E_G$ . By choosing appropriately narrow-bandgap materials, one can detect infrared radiation. However, the choice of materials is rather limited



**Fig. 7.1** Schematic band structure of narrow-gap p-type semiconductor, illustrating the photoemission process used in some infrared detectors.

for the range of infrared wavelengths of interest, particularly in the spectral region between 3 and 15  $\mu\text{m}$ . This is the range where quantum-confined size effects in semiconductor nanostructures can play an important role in detection of infrared radiation.

#### 7.1.2.1 *Detection through band-to-band transitions*

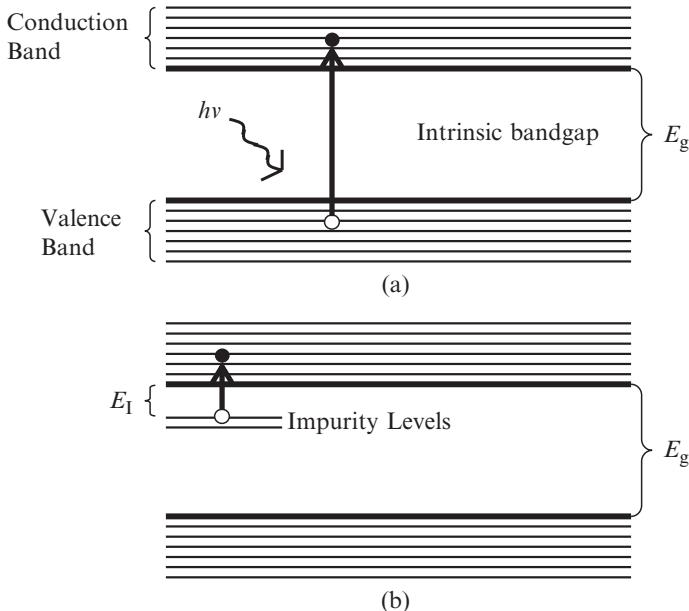
To use semiconductors for detecting radiant energy, one relies on the concept of electron transitions between energy levels or bands in the semiconductor. The connection between the transition energy and the wavelength of the radiant energy to be detected can be appreciated from expression

$$\Delta E = h\nu = \frac{hc}{\lambda} = \frac{1.24}{\lambda} \text{ eV}, \quad (7.1)$$

where  $\Delta E$  is the transition energy,  $h$  is Planck's constant,  $c$  is the velocity of light, and  $\lambda$  is the wavelength of the radiation in microns. If the radiant energy is absorbed by electrons in semiconductors, the energy of the transition that the electrons undergo after absorption is inversely proportional to the wavelength of the absorbed radiation. Wavelengths in the infrared region of the electromagnetic spectrum correspond to small energy transitions within a semiconductor. To be specific, electrons absorbing radiant energy in the infrared spectral region between 0.75 and 1000  $\mu\text{m}$  would only be able to make transitions with energies in the range between 1.24 meV to 1.65 eV. The most interesting range of the infrared spectrum for civilian and military applications is from 2 to 15  $\mu\text{m}$ , corresponding roughly from 83 to 620 meV. Since thermal energy at room temperature is about 25 meV, most infrared detectors in this spectral range are usually operated below room temperature; this is achieved by thermoelectric or cryogenic cooling. The longer the peak wavelength of operation (corresponding to smaller transition energies), the lower the operating temperature of the device must be. An operating temperature as low as 40 K is sometimes necessary for some long-wave infrared devices. In the spectral range between 2 to 15  $\mu\text{m}$ , special semiconductors with energy bandgaps corresponding to these wavelengths are used. The most common direct bandgap semiconductor with these bandgaps is mercury cadmium telluride (MCT— $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ ) (Destefanis *et al.* 1998; Tribolet *et al.* 1998; Gordon *et al.* 2006; Piotrowski *et al.* 2007). The bandgap of this material can be continuously adjusted to correspond to wavelengths from about 0.7 to 25  $\mu\text{m}$ . Detectors made from such semiconductors are called *intrinsic detectors*, a reference to the fact that carriers in bound energy states at the valence band of the (narrow gap) semiconductor absorb incident radiant energy and make transitions to states in the conduction band, where they are mobile and can be collected as current. The energy structure involved in the transitions is schematically shown in Fig. 7.2(a).

#### 7.1.2.2 *Detection via “impurity level to conduction band” transition*

It is also possible to make infrared detectors from semiconductors with normal bandgaps—meaning bandgaps not in the typical infrared range ( $>0.75$  eV). However, because the bandgaps of such semiconductors are usually too wide to allow absorption of infrared energy, one can “dope” the semiconductor



**Fig. 7.2** (a) Energy structure of a semiconductor showing the principle of band-to-band carrier transition in an intrinsic detector, and (b) impurity level to a band detection scheme in an extrinsic device.

to introduce impurity states with energy levels that are a few millielectron volts below the bottom of the conduction band. The energy structure of such a scheme is shown in Fig. 7.2(b). Any incident infrared photons can then be absorbed by the outer shell electrons of the donor impurity atoms. The absorbed energy is sufficient to cause the electrons to make transitions to the conduction band itself. Once in the conduction band, the now mobile electrons can be collected as current—manifesting the detection of the infrared radiation. Photon detectors made according to this principle are called *extrinsic detectors*, where the extrinsic refers to the addition of impurities to an otherwise pure semiconductor.

## 7.2 Infrared photon absorption

Infrared photon absorption is best understood by considering general light-absorption processes in semiconductors. Absorption entails the interaction of light (photons) with matter. At the microscopic level, the process can be described through an interaction Hamiltonian. In the single-particle picture, the Schrödinger equation that describes an electron in a solid (such as a semiconductor) can be written as

$$H_0(r)\Psi_n(r) = E_n \Psi_n(r), \quad (7.2)$$

where  $\Psi_n(r)$  is the electron wavefunction and  $E_n$  the energy eigenvalue. The unperturbed Hamiltonian is  $H_0(r) = \frac{\vec{p}^2}{2m^*} + V(r)$ , where  $m^*$  is the effective mass of a charge carrier and  $V(r)$  the potential energy. When light is incident onto a semiconductor and is absorbed by an electron, the electron momentum  $\vec{p}$  must now be written as  $(\vec{p} + e\vec{A})$ , where  $e$  is the magnitude of the electron

charge and  $\vec{A}(r, t)$  is a magnetic vector potential related to the incident light field. The Hamiltonian describing the electron in the presence of the light field can now be written as

$$H(r) = \frac{(\vec{p} + e\vec{A})^2}{2m^*} + V(r) = H_0(r) + H'(r, t), \quad (7.3)$$

where  $H'(r, t)$  is the perturbation Hamiltonian due to the interaction of the electron with the light. The perturbation (or interaction) Hamiltonian can be approximated as

$$H'(r, t) \approx \frac{e\vec{p} \cdot \vec{A}(r, t)}{m^*}, \quad (7.4)$$

where we have assumed that  $\vec{p}$  commutes with  $\vec{A}$  so that that  $\vec{p} \cdot \vec{A} = \vec{A} \cdot \vec{p}$ ; furthermore, we have neglected the quadratic term ( $e^2 A^2 / 2m^*$ ) in the expansion of the kinetic-energy term because we are only interested in linear absorption. The electric field of the light interacting with the electron is related to the magnetic vector potential through Maxwell's equations. In the Coulomb gauge, where one chooses the scalar electric potential to be zero, the relationship between the electric field and the magnetic vector potential is

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t}. \quad (7.5)$$

For a time-harmonic electric field of the form

$$\vec{E} = \hat{e} E_0 \cos(\omega t - \vec{q} \cdot \vec{r}), \quad (7.6)$$

where  $\hat{e}$  is the unit polarization vector of the light and  $\vec{q}$  the wave vector of the light, we can write the magnetic vector potential thus:

$$\vec{A}(r, t) = -\hat{e} E_0 \int \cos(\omega t - \vec{q} \cdot \vec{r}) dt = -\frac{\hat{e} E_0}{\omega} \sin(\omega t - \vec{q} \cdot \vec{r}), \quad (7.7)$$

which is constituted of two counterpropagating plane waves of the form

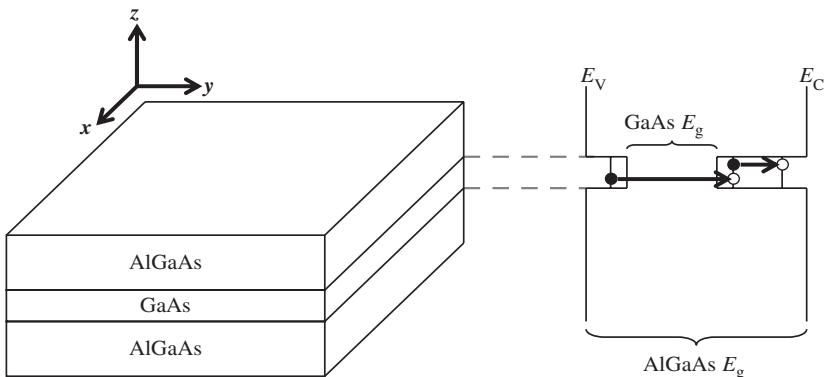
$$\frac{j\hat{e} E_0}{2\omega} \{ \exp(j[\omega t - \vec{q} \cdot \vec{r}]) - \exp(-j[\omega t - \vec{q} \cdot \vec{r}]) \}. \quad (7.8)$$

For the assumed time-harmonic electric field given in eqn (7.6), the corresponding magnetic field can be derived from one of Maxwell's equations; the Poynting flux (or optical intensity) of the incident light is then calculated as

$$\langle S(r, t) \rangle = \langle \frac{1}{2} \text{Re}[\vec{E}(r, t) \times \vec{H}^*(r, t)] \rangle = \frac{\tilde{n}\varepsilon_0 c E_0^2}{2}, \quad (7.9)$$

where  $\tilde{n}$  is the index of refraction,  $c$  the velocity of light, and  $\varepsilon_0$  the permittivity of free space.

In a heterostructure such as a quantum well, where a small bandgap semiconductor of type I (Sai-Halasz *et al.* 1977) is sandwiched between two large bandgap semiconductors, electrons (holes) can be confined in the small-gap material. The confining potential for electrons (holes) is formed by the bandgap



**Fig. 7.3** Schematic layer structure of a GaAs quantum well sandwiched between two (Al,Ga)As barrier layers. The relevant energy band structure of the ‘‘particle-in-a-box’’ model for the confined carriers is shown on the right of the illustration.

discontinuity in the conduction (valence) band at the heterointerface of the semiconductors. In the GaAs/(Al,Ga)As materials system, for example, the potential discontinuity is determined by the composition of the (Al,Ga)As barrier material. Figure 7.3 schematically shows the epitaxial layer structure of a GaAs quantum well sandwiched between two (Al,Ga)As barrier layers (on the left) and the energy scheme (on the right). Possible photon-absorption mechanisms are indicated with arrows in the schematic on the right; a simple two-band model is assumed.

Quantum-confined size effects force electrons (holes) to acquire quantized energy levels in the  $z$ -direction, which is usually the epitaxial film growth direction. In the  $x-y$ -plane of the quantum-well thin film, the electrons (holes) retain their plane-wave character. The solutions of the Schrödinger equation in the structure depicted in Fig. 7.3 lead to energy eigenvalues with corresponding eigenfunctions. By applying appropriate boundary conditions at the heterointerfaces, the energy eigenvalues are found to be of the form

$$E_{nk_{||}} = E_n + \frac{\hbar^2 k_{||}^2}{2m^*}. \quad (7.10)$$

The subband energies,  $E_n$ , depend on the potential (shape and magnitude),  $V(z)$ , in the confinement direction. The electron wave vector,  $k_{||}^2 = k_x^2 + k_y^2$ , represents the plane-wave nature of the electron (hole) motion in the film  $x-y$ -plane. The details of how to compute the energy eigenvalues and the associated eigenfunctions are well documented in the literature and will not be repeated here (Weisbuch *et al.* 1991; Harrison *et al.* 1999). In the following, we are only interested in the nature of the optically induced transitions between eigenenergies.

The absorption coefficient,  $\alpha$  (1/cm), is the fraction of photons absorbed per unit distance. It is usually expressed as a ratio of the number of photons absorbed per unit volume per second divided by the number of photons incident onto the material per unit area per second. Quantitatively,

$$\alpha(\hbar\omega) = \frac{R_{\text{if}}}{(S/\hbar\omega)}, \quad (7.11)$$

where  $R_{\text{if}}$  is the electron transition rate (which, assuming that a single photon generates a single electron, is effectively the number of photons absorbed per unit volume per second). The optical intensity,  $S(\text{watts}/\text{cm}^2)$ , divided by the photon energy,  $\hbar\omega$ , gives the number of photons per unit area. Using the expression for the optical intensity,  $S$ , derived earlier—in eqn (7.9)—in terms of the magnitude of the incident electric field, the absorption coefficient can be written as

$$\alpha(\hbar\omega) = \frac{2\hbar\omega R_{\text{if}}}{\tilde{n}\varepsilon_0 c E_0^2}. \quad (7.12)$$

Photon absorption is clearly proportional to the optically induced electron transition rate from a valence band to a conduction band, or to the transition rate from an energy level within a valence (conduction) to another higher level within the valence (conduction) band. Such transitions for a typical quantum structure are shown in Fig. 7.3. The electron transition rate,  $R_{\text{if}}$ , is usually related to the Fermi golden rule that can be derived from the time-dependent Schrödinger equation through a perturbation theory analysis (Parker *et al.* 2005). The resulting Fermi golden rule is

$$R_{\text{if}} = \frac{2\pi}{\hbar} | \langle \Psi_i | H'(r, t) | \Psi_f \rangle |^2 \delta(E_f - E_i - \hbar\omega). \quad (7.13)$$

Observe that the interaction Hamiltonian derived earlier features prominently in the Fermi relation. From eqn (7.4) and eqn (7.8) we can rewrite the interaction Hamiltonian as

$$H'(r, t) = H'(r) e^{j\omega t} + H'^*(r) e^{-j\omega t}, \quad (7.14)$$

where  $H'(r) = \frac{jeE_0}{2m^*\omega} e^{j\vec{q} \cdot \vec{r}} \hat{e} \cdot \vec{p}$ . Note that  $H'^*(r)$  is simply the complex conjugate of  $H'(r)$ . With this new definition of the interaction term, we can rewrite the Fermi result as

$$R_{\text{if}} = \frac{2\pi}{\hbar} \frac{e^2 E_0^2}{4(m^*\omega)^2} | \langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle |^2 \delta(E_f - E_i - \hbar\omega). \quad (7.15)$$

The most important term in this expression is the transition matrix element,  $\langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle$ , which links an optically induced electron transition from an initial state,  $\Psi_i$ , to a final state  $\Psi_f$ .

In eqn (7.15), we have used the fact that in the infrared ( $\lambda > 1 \mu\text{m}$ ), the photon wavevector  $\vec{q}$  can be neglected so that  $\exp(j\vec{q} \cdot \vec{r}) \rightarrow 1$  in eqn (7.14) since the photon wavelength  $\lambda \sim 2\pi/q$  is much larger than any characteristic lattice period over which the electron–photon interaction occurs. This is basically an alternative restatement of the well-known dipole approximation (Parker *et al.* 2005).

In practical semiconductor materials, there are usually a number of initial states occupied by electrons that absorb incident light. Let the probability that these states are occupied be  $f_i$ . For induced transitions to occur, there must also be a corresponding set of empty states to which electrons can make a transition; these would be the final states, and the probability that they are occupied can be represented by  $f_f$ . The probability that these states are empty is given by  $1 - f_f$ . The probability distribution function,  $f_i$ , for the filled states

and  $1 - f_f$  for the empty states are the usual Fermi–Dirac distribution functions (Parker *et al.* 2005; Yariv *et al.* 2007). The overall transition rate must take into account these probabilities. The modified transition rate expression is therefore

$$R_{if} = \frac{2}{V} \sum_{if} \frac{2\pi}{\hbar} \frac{e^2 E_0^2}{4(m^* \omega)^2} |\langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle|^2 \times f_i(1 - f_f) \delta(E_f - E_i - \hbar\omega), \quad (7.16)$$

where we must sum over all the initial and final electron states. A factor of 2 for electron spin degeneracy has been included and the transition rate is now expressed per unit volume, where the symbol  $V$  represents the volume. The absorption coefficient for the heterostructure we have been discussing can now be obtained by combining eqns (7.12) and (7.16) to yield

$$\alpha(\hbar\omega) = \frac{\pi e^2}{\tilde{n}\epsilon_0 c (m^*)^2 \omega} \frac{2}{V} \sum_{if} |\langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle|^2 \times f_i(1 - f_f) \delta(E_f - E_i - \hbar\omega). \quad (7.17)$$

### 7.2.1 Optical matrix element for interband and intersubband transitions

The general matrix element for electronic transitions induced by light absorption, as extracted from the Fermi result (which also appears in the expression for the absorption coefficient) is

$$M = \langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle. \quad (7.18)$$

The initial and final state functions  $\Psi_i(r)$  and  $\Psi_f(r)$ , can be separated into components such that

$$\Psi_i(r) = u_i(r)\varphi_i(r), \quad (7.19)$$

where  $u_i(r)$  is the cell periodic Bloch functions for a particular band, and  $\varphi_i(r)$  is a slowly varying envelope function. For a quantum-well structure, the envelop function can be split further into

$$\varphi_i(r) = \frac{\exp(j\vec{k}_{||} \cdot \vec{r}_{||})}{\sqrt{A}} Z(z), \quad (7.20)$$

where  $A$  is the area of the sample used to normalize the wavefunction. The plane-wave component of the function  $\varphi_i(r)$  describes electron motion in the  $xy$ -plane with  $\vec{k}_{||} = \hat{x}k_x + \hat{y}k_y$  and  $\vec{r}_{||} = \hat{x}x + \hat{y}y$ ;  $Z(z)$  is the wavefunction describing confined electron motion in the  $z$ -direction.

The properties of Bloch functions and the slowly varying envelope functions allow us to separate the transition matrix element into two components as follows:

$$\begin{aligned} \langle \Psi_i | \hat{e} \cdot \vec{p} | \Psi_f \rangle &= \hat{e} \cdot \langle u_i^n | \vec{p} | u_f^{n'} \rangle \langle \varphi_i^m | \varphi_f^{m'} \rangle \\ &\quad + \hat{e} \cdot \langle u_i^n | u_f^{n'} \rangle \langle \varphi_i^m | \vec{p} | \varphi_f^{m'} \rangle. \end{aligned} \quad (7.21)$$

The labels  $n$  and  $n'$ , and  $m$  and  $m'$  refer to band and subband indices. For a two-band model of the energy structure of a semiconductor (where  $n$  = valence band and  $n' =$  conduction band), the first term of the matrix element on the right-hand side of eqn (7.21) describes *interband* transitions between the valence and the conduction band. The matrix element of the Bloch functions,  $\hat{e} \cdot \langle u_i^n | \vec{p} | u_f^{n'} \rangle$ , dictates the polarization selection rules that govern the transitions (Coon *et al.* 1984). The overlap integral of the envelope functions,  $\langle \varphi_i^m | \varphi_f^{m'} \rangle$ , gives the selection rules for the hole or electron quantum numbers in the subbands. When the transitions of interest are within the same band (intersubband), then  $n = n'$  and the first term on the right of eqn (7.21) vanishes. In this case we are then primarily concerned with the second term of eqn (7.21). This term can be rewritten as

$$\begin{aligned} \langle \varphi_{k_{||}}^m | \hat{e} \cdot \vec{p} | \varphi_{k_{||}}^{m'} \rangle &= \frac{1}{A} \iiint e^{-jk_{||} \cdot r_{||}} Z_m^*(z) \\ &\times [e_x p_x + e_y p_y + e_z p_z] e^{jk_{||}' r_{||}'} Z_{m'}(z) dx dy dz, \end{aligned} \quad (7.22)$$

where  $\hat{e} \cdot \vec{p} = (\hat{x}e_x + \hat{y}e_y + \hat{z}e_z) \cdot (\hat{x}p_x + \hat{y}p_y + \hat{z}p_z) = e_x p_x + e_y p_y + e_z p_z$ . The only term on the right-hand side of eqn (7.22) that contributes a non-zero value is the one proportional to  $e_z$ . All other terms evaluate to zero except when  $m = m'$  and  $k_{||} = k_{||}'$ , which is equivalent to saying that the light frequency is zero. The equality of the inplane wave vectors simply implies that only vertical transitions are allowed in  $k_{||}$ -space; in other words, there is a  $k_{||}$ -selection rule imposed. But in this case there would be no light-induced electron transition since  $m = m'$  implies no change in subband. This leads us to conclude that intersubband absorption is only possible in quantum wells when the light polarization vector is aligned along the  $z$ -direction of the film, which is usually the epitaxial growth direction. This property of quantum wells has important practical consequences for intersubband detectors of infrared light. We discuss these consequences in some detail in a later section.

### 7.2.2 Matrix element for intersublevel transitions

For a conceptually simple—albeit unrealistic—picture of a quantum dot, we imagine a three-dimensional cube of a low-bandgap semiconductor inserted in a matrix of a relatively larger bandgap. One could, in principle, make such a structure by embedding a cube of a small-gap semiconductor, such as InAs, inside a relatively wider-gap GaAs or (Al,Ga)As semiconductor material to form a quantum cube structure. Real, epitaxially synthesized quantum dots are not usually cubic in shape. Depending on the prevailing synthesis conditions, they can be pyramidal or lens-shaped. For the sake of the discussion here, we will continue to regard the dots as cuboidal. A solution of the Schrödinger equation in the Cartesian coordinate system for the quantum cube would lead to eigenfunctions of the form

$$\Phi_i(r) = \frac{1}{\sqrt{V}} u_i^n(r) X_i^n(x) Y_i^n(y) X_i^n(z), \quad (7.23)$$

where, as before, the  $u_i^n(r)$  are the Bloch wavefunctions for band  $n$  and  $V$  is the volume of the sample. The functions  $X_i^n(x)$ ,  $Y_i^n(y)$  and  $Z_i^n(z)$  are slowly varying envelope functions that describe electron confinement in three spatial directions. It will be instructive to assume that the box-type dots we are discussing have asymmetric spatial dimensions ( $d_x = d_y \neq d_z$ ). As before, one can form the optical matrix element for optically induced transitions from initial states in the valence band to final states in the conduction band (interband), and from low-energy levels in a band to other higher levels within the same band (intraband). Here, we are only interested in the character of the intraband transitions and will ignore the interband transitions as they vanish, for reasons already discussed for the quantum-well case.

Since electrons in a quantum cuboid do not enjoy freedom of motion, and are restricted to having only quantized energy levels in all three spatial directions, any transitions they make within a band are clearly between levels in subbands: these are *intersublevel* transitions. The optical matrix element for such transitions can be written (ignoring the factor  $1/V$ ) as

$$\begin{aligned} <\Phi_i^m|\hat{e}\cdot\vec{p}|\Phi_f^{m'}> &= \iiint X_i^{*m}(x)Y_i^{*m}(y)Z_i^{*m}(z) \\ &\times [e_x p_x + e_y p_y + e_z p_z] X_f^{m'}(x)Y_f^{m'}(y)Z_f^{m'}(z) dx dy dz. \end{aligned} \quad (7.24)$$

The terms of the integral on the right-hand side of eqn (7.24) that evaluate to non-zero essentially determine the transitions that are allowed (thus dictating the polarization selection rules). Clearly, the allowed transitions are those for which the light-polarization vector is non-zero, i.e.  $\hat{e} \neq 0$  in any particular spatial direction, and where the difference in state quantum numbers for the allowed transitions in that direction is  $\Delta m = 1$ . The latter condition reflects the fact that the magnitude of the oscillator strength for a transition between neighboring states is large compared to that for transitions between states not close to each other (West *et al.* 1985). For confinement in the  $z$ -direction, this means a transition is possible between two states when the difference between the quantum numbers of the states in the  $z$ -direction is  $\Delta m_{\perp} = 1$ . In the  $xy$ -plane of the cuboid, a transition is possible along two distinct directions:  $x$  or  $y$ ; thus  $\Delta m_{||} = 1$ . This discussion can be summarized by stating that integrals for the matrix element,  $\tilde{m}_z$ , of the form

$$\tilde{m}_z = j\hbar \int Z_i^{*m}(z) \left[ e_z \frac{\partial}{\partial z} \right] Z_f^{m'}(z) dz, \quad (7.25)$$

are non-zero for allowed transitions in each spatial direction. Equation (7.25) is written for the  $z$ -direction where  $m - m' = \Delta m_{\perp} = 1$ . If we assume the envelope functions of the cuboidal quantum dot are sinusoids of the form  $\sin(\frac{m_z \pi z}{d_z})$  (where  $d_x = d_y \neq d_z$ ), then for an infinite potential barrier bounding the cuboid, the quantum numbers for the first few allowed states are shown in Table 7.1.

From Table 7.1, one observes that a transition from the ground level, (1,1,1), to level (1,2,1) or (2,1,1) is allowed for light polarized in the plane of the cuboid (in the  $e_x$  (or  $e_y$ ) direction since  $\Delta m_{||} = 1$ ). Similarly, a transition is

**Table 7.1** Quantum numbers and levels of degeneracy.

Quantum numbers ( $m_x, m_y, m_z$ )	Level degeneracy $g$
(1,1,1)	2
(1,2,1) (2,1,1)	4
(1,1,2)	2
(2,2,1)	2

allowed from the ground level, (1,1,1), to the level (1,1,2) for light polarized perpendicular to the cuboid plane (in the  $e_z$ -direction since  $\Delta m_{\perp} = 1$ ). Several other combinations of quantum numbers for different polarizations can be determined from the table (and its extended version to states of higher quantum numbers).

The possible combinations of the two basic conditions for allowed transitions lead to an interesting conclusion: that light of any polarization incident onto a properly prepared quantum cuboid can induce an intersublevel transition. This follows from the fact that light of any polarization can be resolved into distinct components in the spatial directions of confinement for an electron in a quantum cuboid. This result has important practical implications for intersublevel photodetectors. We explain this and what we mean by a “properly prepared” quantum cuboid in a later section.

## 7.3 Some metrics for photon detectors

### 7.3.1 Overview on metrics

The discussion in the previous section provided some background on the mechanism by which infrared light is absorbed (and hence detected) by semiconductor nanostructures. To use these nanostructures in practical detection of light, one must engineer device structures that incorporate basic elements of the absorption process to allow quantitative determination of the incident light. The number of photons incident onto a detector cannot be directly determined. The light (photons) must first be transduced into a measurable signal form. Whenever possible, the transduction must minimize injection of unwanted signals. These requirements define engineering constraints for photon detection. Stated explicitly, photon detection entails creation of an arrangement of semiconductor structures that allow optimal measurement of the amount of light absorbed. One can elaborate on this statement by defining quantitative metrics that evaluate the success of the measurement process. There are a number of parameters and concepts used in characterization of photon detectors. We discuss some of the most important in the following.

#### 7.3.1.1 Responsivity

Photon detectors are used to convert input radiant energy to an output signal of some convenient form. To characterize this function, a fundamental parameter called the responsivity,  $\mathfrak{R}$ , has been defined. This is the ratio of the output signal to the radiant input. The most convenient output signal is usually of electrical form—either current or voltage. We will use current in our subsequent discussion of infrared responsivity. When a photon is absorbed by a semiconductor in a detector structure, an electron at an initially low-energy state (usually the ground state) gets excited to a higher-energy state, where it may become a free and mobile charge carrier. With appropriate Ohmic contacts deployed, this charge,  $q$ , may be collected as current in an external circuit. Let

the efficiency of the process for an incident photon to generate an electron be  $\eta$ . This efficiency is also called the quantum efficiency. For an incident input light beam of intensity  $S$  (watts/cm<sup>2</sup>), one can determine the number of photons in the beam by dividing the intensity by the energy of a photon. For light of wavelength  $\lambda$ , the number of photons in the beam is  $S/h\nu = S\lambda/hc$ , where  $h$  is Planck's constant and  $c$  the velocity of light. The output current generated by this number of photons can be written as

$$i_p = \eta\gamma(q/hc)\lambda SA_d, \quad (7.26)$$

where  $A_d$  is the sensitive area of the detector; the parameter  $\gamma$  is a ratio of the number of electrons collected as current to the number of electrons generated by photoabsorption. It is often called the photoconductive gain (which can also be defined as the ratio of the excited electron lifetime to the transit time:  $\gamma = \tau_e/\tau_T$ ). The photoconductive gain can also be related to other device parameters such as the mean-free distance an electron travels without collision,  $l$ , and the total length of the active region,  $L$ . The gain is then linked to these parameters by the ratio  $\gamma = l/L$ . The ratio of the output current to the input radiant power is the responsivity; by rearranging eqn (7.26), we obtain the responsivity as

$$\mathfrak{R} = \frac{i_p}{SA_d} = \eta\gamma(q/hc)\lambda. \quad (7.27)$$

Since current is measured in amperes and optical power in watts, the equation above indicates that the units of responsivity are amperes/watt.

Responsivity is clearly a function of the quantum efficiency and the incident light wavelength. The quantum efficiency accounts for how much of the incident radiant flux actually enters the semiconductor and contributes to generating photoelectrons. Because light is usually incident from a medium of a different index of refraction from that of the absorbing semiconductor material, a fraction of it,  $R$ , will be reflected so that only  $(1 - R)$  enters the semiconductor to generate photoelectrons. If the absorbing region of the detector is of thickness  $d$ , then the fraction of light that generates electrons in the active volume can be given as

$$\beta = \frac{\int_0^d \exp(-\alpha z) dz}{\int_0^\infty \exp(-\alpha z) dz} = 1 - \exp(-\alpha d). \quad (7.28)$$

This factor and the fraction of light that is not reflected contribute to the quantum efficiency. The *effective* quantum efficiency of the “photon-to-collected current” process must also include a factor,  $p$ , which is the probability that a generated photoelectron contributes to a collected photocurrent. The *effective* quantum efficiency can be written as

$$\eta = (1 - R)[1 - \exp(-\alpha d)]p. \quad (7.29)$$

To improve detector responsivity, there are at least two design parameters in eqn (7.29) within our control. The first is the reflectivity; the surface reflectivity of the detector can be minimized by applying appropriate antireflection coatings tailored to the specific spectral band of interest. The second parameter is the absorbance ( $\alpha d$ ). One can either increase the absorption coefficient or increase the thickness of the active region up to a certain point. As we will discuss later, increasing the thickness of the active region may not necessarily lead to an increase in responsivity. For intersublevel infrared detectors, eqn (7.17) suggests an approach to increasing the absorption coefficient. The most direct approach is to examine the Fermi–Dirac distribution factors in eqn (7.17). These factors are controlled via the quasi-Fermi level that depends on the density of extrinsic dopant impurities in the active medium. It is these dopants that determine the number of electrons available to absorb incident infrared photons. A judicious engineering tradeoff is usually necessary. Too large a concentration of dopants can lead to “dark” electrons that contribute noise rather than signal photocurrent; too low a concentration reduces absorption. These factors will be discussed in a subsequent section on device results. It should be pointed out that increasing the dopant density only makes sense when there is a corresponding increase in the density of dots in the active region.

One of the unusual properties of quantum-dot nanostructures that distinguishes them from quantum-well films is that they are akin to atoms; they are, in fact, often called artificial atoms. This means strict rules exist on how many electrons can occupy an energy level. For the quantum dot, the ground state can only accommodate two electrons; the first excited state can have four electrons; and the second excited state can accommodate up to six electrons. This manner of occupation constrains the density of dopant impurities that can be incorporated in the active region of a quantum-dot detector structure (Pan *et al.* 1998).

### 7.3.1.2 Noise and the concept of signal-to-noise ratio

In the context of infrared detectors, noise is unwanted electrical output. It is unavoidable and one generally strives to minimize it so that it does not obscure the desirable signal. Some noise sources are fundamental and cannot be minimized or eliminated. These sources include (i) vibrations of the atoms in the lattice of the semiconductor out of which the detector is fabricated; (ii) fluctuations of the arrival rate of photons, and (iii) random motions of the electrons in the detector structure.

Since noise is random, deviating below or above an average signal, the usual definition of it is to consider the root-mean-square (rms) deviation as the measure of its magnitude. Components of noise appear at specific frequencies. Because of this, limiting the measurement bandwidth to a frequency range within the vicinity of the signal can help to reduce the total noise. When noise is spread uniformly over a wide frequency range, it is said to be *white noise*: this is in analogy to the fact that white light is made up of all the colors of the spectrum. In this case, the total noise is then generally proportional to the square root of the noise bandwidth. Incidentally, it can also

be shown that noise is proportional to the square root of the active area of a detector.

The concept of a signal-to-noise ratio is used to describe the “cleanliness” of a given signal level. It is the signal current divided by the rms noise current.

#### 7.3.1.3 Noise equivalent power

Noise equivalent power (NEP) is a measure of the ultimate sensitivity of a given detector. It is defined as the optical power that needs to be incident on a detector to cause a signal-to-noise ratio of 1 ( $\text{SNR} = i_p/i_n = 1$ ). Quantitatively, we can write

$$\text{NEP} = \frac{SA_d}{i_p/i_n} = \frac{i_n}{i_p/SA_d} = \frac{i_n}{\mathfrak{R}}, \quad (7.30)$$

where all the variables have the same meaning as before, except the variable  $i_n$ , which is the rms noise current. From eqn (7.30), we observe that the noise equivalent power is the ratio of the rms noise current to the responsivity, and it has the units of watts.

#### 7.3.1.4 Detectivity

The most sensitive detectors tend to have the smallest values of NEP (as defined above). Noise depends on the size (area) of a detector as well as the bandwidth of measurement. We cannot therefore use the NEP as a metric to compare different detectors (with varying sizes or characterized at different bandwidths). However, if one normalizes the noise current by the area ( $A_d$ ) of the detector and the bandwidth ( $\Delta f$ ) at which it is measured, the new  $\text{NEP}^*$  can be used as a metric for comparison. Normalizing the right-hand side of eqn (7.30) renders it to the form

$$\text{NEP}^* = \frac{i_n}{\mathfrak{R}} \frac{1}{\sqrt{A_d \Delta f}}. \quad (7.31)$$

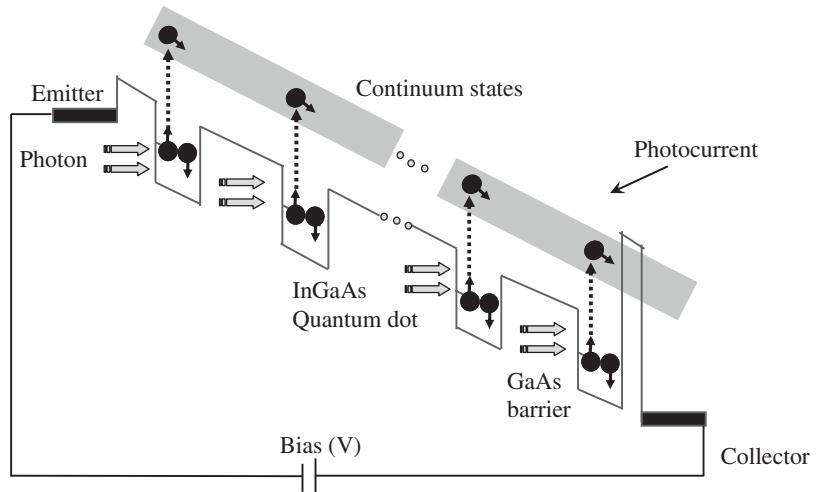
For very sensitive detectors, the numerical values of  $\text{NEP}^*$  can be absurdly small; because of this, Jones (1957) suggested the use of the reciprocal of  $\text{NEP}^*$  as a better measure of the sensitivity of a detector. He called this the  $D^*$  (D-star), which is now widely known as the specific detectivity. By inverting eqn (7.31), we obtain the specific detectivity as:

$$D^* = \frac{1}{\text{NEP}^*} = \frac{\mathfrak{R} \sqrt{A_d \Delta f}}{i_n}. \quad (7.32)$$

The units of D-star are Jones or  $\text{cm Hz}^{1/2}/\text{watt}$ .

### 7.3.2 Quantum-dot intersublevel photon detectors

The idea of the quantum-dot intersublevel photon detector is a natural evolution of a concept previously used in the quantum-well intersubband detector. In its simplest embodiment, the concept adopts a scheme where photoexcited charge carriers in high-energy states—for example, in the continuum of a (valence or conduction) band—are collected as photocurrent that results from

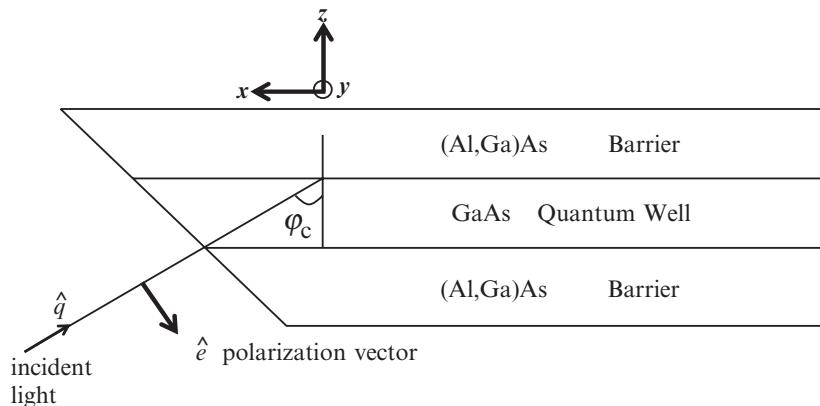


**Fig. 7.4** Conduction-band energy structure (under bias) for intersubband (intersublevel) infrared absorption scheme used in quantum-structure-based detection.

absorption of infrared light. The carriers, which are initially confined in the ground state of a potential well in a dot or a thin film, absorb light with energy slightly larger than the difference between the ground state and the next excited state above the well. If the excited state is above or near the top of the potential well, the carriers may be free and can be collected as current. This situation is illustrated in Fig. 7.4 for carriers initially confined in a conduction-band potential of a dot (or film). Such is the scheme used in quantum-structure-based infrared detection.

The major difference between a quantum-dot and a quantum-well detector is the degree of spatial confinement for the charge carriers. In quantum dots, carriers are confined in all three spatial dimensions. In quantum wells, on the other hand, they are only confined in one spatial dimension and are free to move in the other two dimensions. This difference is at the heart of the selection mechanism that determines the type of light absorbed by a quantum-well or quantum-dot structure.

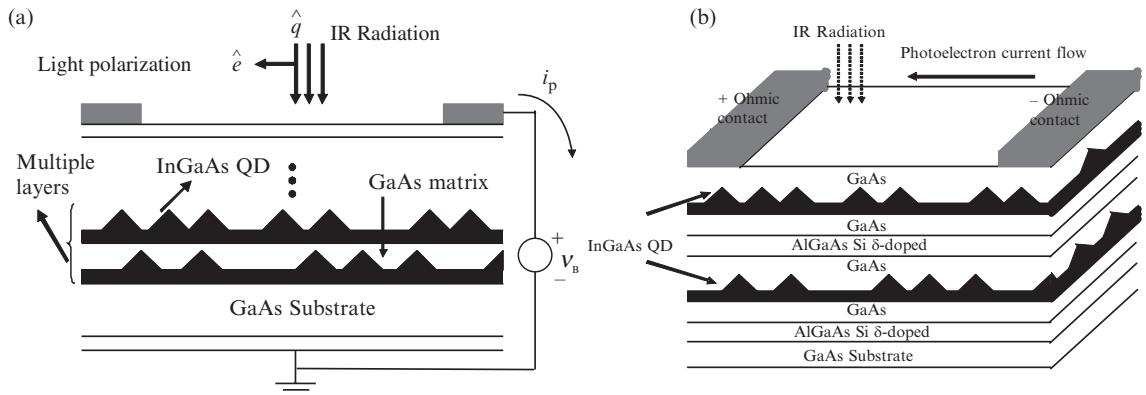
As previously discussed in an earlier section, electron transitions caused by light absorption between two energy subbands in a quantum well are only possible when the polarization of the light is parallel to the confinement direction. This arrangement is impossible to achieve in a practical device without resorting to light-manipulation schemes. This is because the light to be absorbed (sensed) is generally incident perpendicularly onto the detector surface. The transverse nature of an electromagnetic light field means that its polarization is parallel to the surface of the detector. Detectors fabricated from quantum-well structures are not sensitive to normal-incidence light. The way to make quantum wells absorb light is to bevel the sample as shown in Fig. 7.5. In this arrangement, the light-polarization vector  $\hat{e}$  is parallel to the bevelled surface of the structure containing the quantum well. This configuration allows the light polarization to have a component normal to the quantum-well plane, thus satisfying the condition for a fraction of the light to be absorbed in the quantum-well structure.



**Fig. 7.5** Illustration of bevelled-geometry configuration of a sample to enable detection of light by quantum-well structures in the plane of the sample. Quantum wells in the plane of a sample do not absorb light that is perpendicularly incident to the sample surface.

The generalization of this configuration for a quantum-well detector structure involves “corrugating” the surface of the sample so that normal-incidence light can be scattered in the “right direction” to satisfy the condition for absorption. Invariably, the corrugation is a bevelled grating structure on the surface of the detector (Chen *et al.* 1997a; Choi *et al.* 2003, 2005). This is the so-called grating coupling scheme. There are two major issues with this method of coupling light to a detector. Inclusion of the grating structure requires additional fabrication steps and, gratings are spectrally sensitive structures. The latter point means that for multispectral detectors, it may be difficult to fabricate structures that couple the full bandwidth for which a detector array may be designed for.

For a quantum-dot detector structure, the three-dimensional confinement of carriers creates the right set of conditions for absorption of light of any polarization. These conditions, which have been discussed earlier, guarantee that normal-incidence light is absorbed. There are generally two basic device geometries used in quantum-dot detectors. The first geometry is illustrated in Fig. 7.6(a). Here, multiple periods of quantum-dot layers are stacked to form the device structure. Contacts for collecting the photocurrent from the device are formed at the bottom and top of the structure. Photocurrent generated in such a structure is collected vertically. The second device geometry, shown in Fig. 7.6(b), differs from the first in that the generated photocurrent is collected laterally (horizontally). Each structure has its advantages and disadvantages. Most of the work reported in the literature has been focused on the vertical transport structure. One of the main drawbacks of this device is the fact that carriers must tunnel through multiple barriers before they are collected as current. If there are defects in the barriers, they could cause scattering and hence a reduction in carrier mobility. On the other hand, stacking the dots allows one to increase the number of dots and hence carriers that can potentially absorb light. The lateral transport detector whose structure is reminiscent of a modulation-doped field effect transistor (Mimura *et al.* 2005) has the virtue that dopant impurities and their associated free electrons occupy different spatial regions. There is a channel to which electrons that have absorbed light can be excited. This channel forms the photocurrent transport layer. This device geometry



**Fig. 7.6** Device geometries used for quantum-dot infrared detection in (a) vertical current-collection device structure, and (b) lateral current-collection structure similar to the modulation-doped, high electron mobility field effect transistor.

could potentially lead to large photoconductive gains for the detector. The gain in this case is given by

$$\gamma = \frac{\mu E \tau_e}{L}, \quad (7.33)$$

where  $\mu$  and  $\tau_e$  are the mobility and lifetime of electrons excited from a quantum-dot ground state to a continuum in the channel transport layer;  $E$  is the static electric field applied across the electrodes, and  $L$  is the interelectrode separation. The disadvantage of this device structure is that the number of periods of dot layers has to be limited to a reasonable number because the field effect necessary for carrier transport is not as effective at dot layers deep in the structure. Note that the gain, as defined here, is identical to the one discussed earlier. In fact, it is trivial to observe that the ratio  $L/\mu E = \tau_T$  is the transit time across the device length. By rewriting eqn (7.33) in terms of the transit time, we arrive at our earlier expression for the photoconductive gain.

To date, quantum-dot infrared detectors have not lived up to their promise. When compared to quantum-well detectors, their performance metrics have not yet surpassed those of the quantum-well detectors. At best, they may be comparable. However, quantum-dot detectors do have the desirable quality of being able to absorb normal-incidence light. Moreover, recent results also show that these devices are capable of operating at temperatures well above 77 K.

## 7.4 Experimental single-pixel quantum-dot infrared photodetectors

### 7.4.1 Quantum dots and their epitaxial synthesis

We define a practical quantum dot as a nanostructure whose spatial extent is of the order of the de Broglie wavelength in all three spatial dimensions. For simplicity, a de Broglie wavelength is defined with reference to an electron in the material out of which the dot is made. The lateral extent of most practical

dots is about several nanometers; the exact size depends on the effective mass of the electron that determines the de Broglie wavelength in the material. Epitaxial semiconductor quantum dots are typically comprised of a material of low energy bandgap surrounded by a matrix of another material with a relatively large bandgap.

Beginning in the 1980s, several researchers, notably Arakawa and Sakaki (Arakawa *et al.* 1982; Sakaki *et al.* 1985) studied the impact of three-dimensional spatial confinement of electrons on the properties of semiconductors. Several approaches to making the confining structures or dots were proposed. The approach that appeared to be most feasible was to take thin-film quantum-well structures (such as AlGaAs/GaAs/AlGaAs films), pattern them, and then etch “tiny” boxes of pillars on them to create three-dimensional confinement for the carriers. Confinement in the vertical direction was a result of the thin-film quantum-well structure, while confinement in the lateral inplane directions was a result of the lithographically defined box sides or pillar walls. Imperfections of the etched facets on these “quantum boxes” led to defects of several kinds; the most deleterious are dangling bonds that contribute to non-radiative surface recombination.

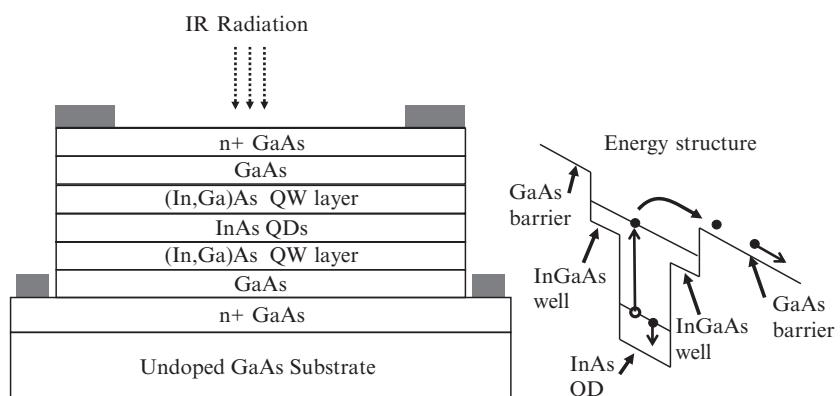
Today, the simplest approach to creating epitaxial quantum dots is via assisted self-assembly, also known as Stranski–Krastanow epitaxial growth mode (Stranski *et al.* 1937). In this approach, two semiconductors with different inplane lattice constants are grown one-on-top-of-the-other. In a typical implementation, a semiconductor with a large inplane lattice constant is grown on top of another with a smaller lattice constant. The simplest example is that of attempting to grow InAs on top of GaAs. The inplane lattice constant mismatch for these materials is about 7%. As long as the thickness of the InAs layer is below a certain critical thickness, the mismatch is accommodated by elastic strain. Beyond that critical limit, the compressive strain in the film is relieved by the formation of coherent islands. The islands are quantum dots if their dimensions satisfy the de Broglie criterion. These islands can be defect free and suitable for device work. The most common epitaxial techniques for growing quantum dots are molecular beam epitaxy (Leonard *et al.* 1993) and metalorganic chemical vapor deposition (Petroff *et al.* 1994).

## 7.4.2 Overview on mid-wave and long-wave quantum-dot infrared photodetectors

Depending on the particular field of application, the infrared region of the electromagnetic spectrum is subdivided into three separate bands. These bands are (i) the short-wave infrared, (ii) the mid-wave infrared, and (iii) the long-wave infrared. There is no general agreement on the exact demarcation of the bands. For the discussions here, we will take the short-wave infrared region to span from 0.75 to 2.5  $\mu\text{m}$ ; the mid-wave infrared region is generally accepted to be from 3–5  $\mu\text{m}$ , and the long-wave infrared is from 8–12  $\mu\text{m}$ . The mid-wave and long-wave infrared are defined to be coincident with atmospheric windows of low absorption through 300 meters of humid air at sea level (Towe *et al.* 2000).

A number of materials exist for readily making detectors in the near-infrared. It is in the mid-wave and long-wave region of the spectrum that one is always looking for potentially better materials than what is currently available. In these regions of the spectrum, engineered quantum structures (such as quantum wells and quantum dots) have proven to be worth investigating. In engineered quantum structures, the attraction is the ease with which the peak wavelength of detection can often be precisely designed. This is particularly true for thin-film quantum-well structures. Here, exquisite control of the thickness of the well size gives one a handle on the desirable wavelength (or energy of transition). Coupled with the ability to control the energy-band discontinuity for the conduction (or valence) band at the heterointerfaces of the semiconductors, one has complete control of the depth of the potential in which carriers can be confined. The immediate consequence of this is control of the number and location of the energy levels for the desirable intersubband transitions. The major undesirable aspect of the one-dimensional quantum well is the fact that infrared light incident on the structure perpendicularly cannot be absorbed. The reasons for this have already been discussed.

Quantum-dot nanostructures offer the desirable feature of being sensitive to normal-incidence light. However, their one major disadvantage is the lack of control of the peak wavelength of detection. At this stage of their development, it is quite challenging to make dots of uniform sizes at will. Size dispersion tends to make their absorption spectra broad. The basic device configurations for quantum-dot infrared detectors have already been discussed. In the general discussions that follow, we will focus on the vertical device geometry of Fig. 7.6. In this configuration, the structure usually has two alternative formats. The first format is that shown in Fig. 7.6(a), and the second is shown in Fig. 7.7. The first device characteristics of quantum-dot intersublevel detectors in the format of Fig. 7.6(a) were reported by Pan *et al.* (1998). This report stimulated a spate of activity in quantum-dot-based infrared detectors. In the format of Fig. 7.7, the quantum-dot nanostructures embedded inside a quantum-well potential to create what is often called a “dots-in-a-well” structure. The reason for embedding the dots inside the well is to achieve better control of the



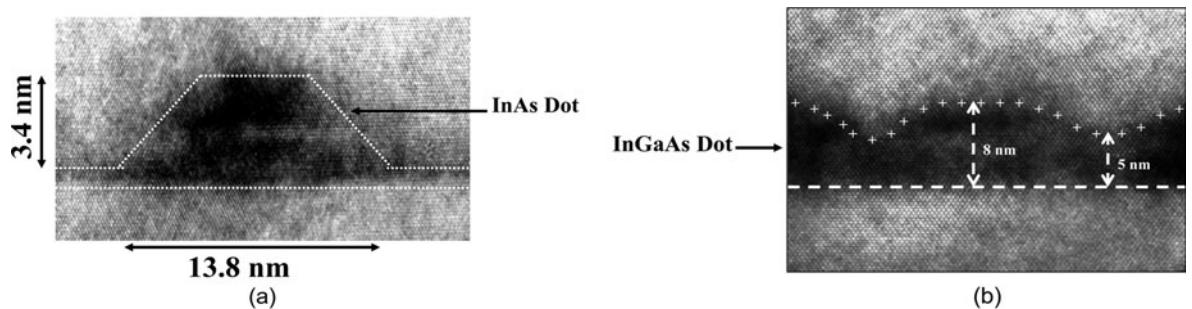
**Fig. 7.7** Layer configuration of the “dots-in-a-well” device structure. The conduction-band scheme of a single period of the structure is shown on the right.

location of the excited state(s). Energy eigenvalues in rectangular wells can be precisely controlled through the thickness of the well. As shown in Fig. 7.7, the well in which the dots are embedded can be either symmetric or asymmetric. This structure, however, no longer constitutes a quantum-dot device structure; it is a quantum-well/quantum-dot hybrid device structure better described as a quantum-structure detector. For both types of device structures, the objective is to control the transition energy for electrons initially bound in a quantum-dot potential. Variations of these device structures are generally used in the design of infrared quantum-dot photodetectors. In the following, we provide an overview discussion of various device characteristics to illustrate the breadth of what has been achieved.

Usually, one of the first key parameters to be measured in quantum-dot infrared detectors is the spectral response; this is reported as a relative response. The reason for this is partly due to the fact that most spectrometers only allow convenient measurement of relative response curves, but not absolute spectral responses. In order to determine the absolute spectral response, a relative response curve must be combined with blackbody responsivity data. Otherwise, a calibrated detector should be used to transform the relative responsivity curve to one with absolute spectral responsivity values.

At this point in their development, many researchers have reported photoresponse spectra of state-of-the-art quantum-dot infrared photodetectors. These results depend on several of the physical parameters of the dot as well as the device structure. Chen *et al.*, for example, showed the effect of (GaAs and AlGaAs) barrier layers on the peak detection wavelength of undoped InAs quantum-dot infrared photodetectors (Chen *et al.* 2001). The peak wavelength, when a GaAs barrier layer was used, was found to be located at about  $7.2\text{ }\mu\text{m}$ ; it shifted to about  $6.0\text{ }\mu\text{m}$  when an (Al,Ga)As barrier was used. This blue shift is attributed to the higher confining potential of the (Al,Ga)As barrier layer. In another experiment, it was demonstrated that the peak wavelength could be tailored by using  $\text{In}_x\text{Ga}_{1-x}\text{As}$  strain-relieving quantum wells (Kim *et al.* 2001). In the context of Fig. 7.7, the  $\text{In}_x\text{Ga}_{1-x}\text{As}$  strain-relieving layer basically serves to modify the potential of the InAs quantum dot. In the specific experiment reported, a 20-monolayer thick  $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$  cap layer on top of the InAs quantum dots led to a peak photoresponse wavelength of about  $9\text{ }\mu\text{m}$ . However, when a GaAs cap was used, the wavelength shifted to a range between 5 and  $7\text{ }\mu\text{m}$ .

The physical structure of epitaxial quantum dots, which depends on many factors, is either pyramidal or lens-shaped. We show in Figs. 7.8(a) and (b), the typical shapes of InAs and InGaAs quantum dots on GaAs substrates. InAs dots tend to be pyramidal with truncated tops, while InGaAs dots are lens-shaped. This shape asymmetry causes some technical problems, which will become clear shortly. The lens-shaped dots are typically broader in the lateral  $xy$ -plane than in the vertical  $z$ -direction. In the plane, they measure anywhere from about 15 to 20 nm; in the vertical growth direction, they are usually about 2–8 nm. These dimensions are typical for InAs and InGaAs dots on GaAs substrates. Other materials combinations may have slightly different dimensions. The point, however, is that the asymmetry leads to different size quantum wells confining electrons in the plane and in the growth directions.



**Fig. 7.8** Transmission electron micrographs of the physical structure of (a) a pyramidal InAs, and (b) a lens-shaped (In,Ga)As quantum dot. Figure (a) was taken from Pal *et al.* (2001) *Appl. Phys. Lett.* **78**, 4133 with permission of AIP.

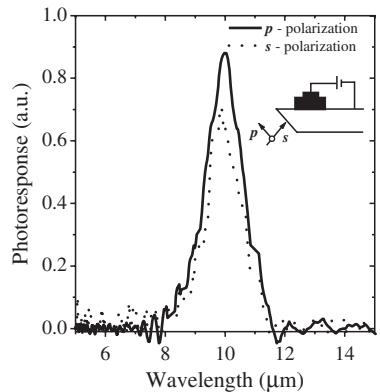
The narrow size of the well in the vertical direction may allow only a single confined energy level in that direction. In the  $xy$ -plane however, there may be several confined energy levels allowed. The tight electron confinement in the growth direction leads to strong coupling of  $z$ -polarized (or  $p$ -polarized) light and hence absorption. The allowed transition for the growth-direction quantum well is from the lone ground energy level to the continuum. For the  $xy$ -plane quantum well, a transition may occur from the ground level to one of the bound excited states or to the continuum. The strongest transitions are those for which the quantum number difference,  $\Delta m = 1$ , as discussed previously. The light that induces transitions in the  $x$ - $y$ -plane is  $s$ -polarized, originating from normal-incidence light. These observations imply that most present-day quantum-dot detectors absorb both  $s$ - and  $p$ -polarized light. Systematic experiments to study this behavior have been reported. Photoresponse spectra from the mid-wave to the long-wave infrared region, showing absorption of both  $p$ - and  $s$ -polarized light have been reported by Aslan *et al.* (2003). These measurements, which were conducted at 6 K, showed that InAs/GaAs quantum-dot infrared photodetectors responded to  $s$ -polarized light in the spectral region between  $14 \mu\text{m}$  and  $18 \mu\text{m}$ , and to  $p$ -polarized light at about  $6 \mu\text{m}$ . The observed response at the two distinct spectral regions is evidence of the asymmetry of the wells in the plane of the dot and in the growth direction. This observation is characteristic of a behavior that can be explained by the physical structure of most epitaxial dots. A way around this problem is to strive to create dots that are uniform in size and that confine only one energy level in the plane of the sample as well as in the vertical direction. Even if this condition is met, one may still not eliminate the sensitivity of dots to  $p$ -polarized light.

Pal and Towe have reported an attempt to create uniform dots that do not exhibit the multispectral behavior discussed above (Pal *et al.* 2003). One of the reasons behind their experiment was to synthesize dots that absorb light polarized along the plane of the quantum dots—the so-called  $s$ -polarized light. A few other papers in the literature have reported on the polarization characteristics of (In,Ga)As quantum dots (Chua *et al.* 1998; Xu *et al.* 1998). The majority of results indicate that for incident light coupled to a multiperiodic quantum-dot sample fashioned into a waveguide structure (such as the zigzag

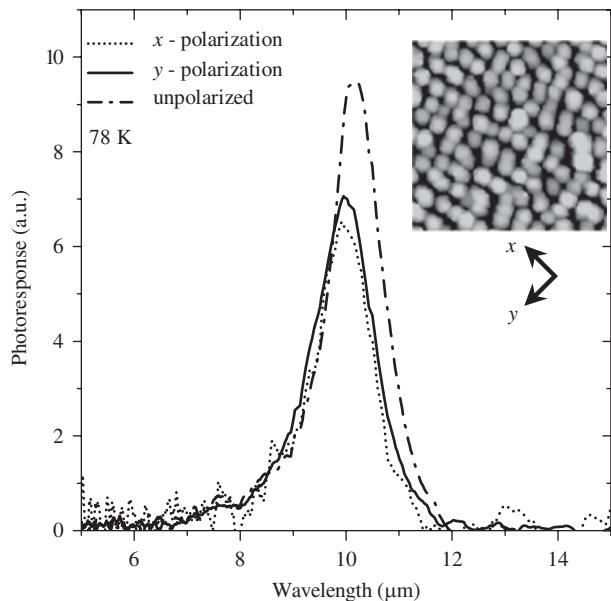
configuration depicted in the inset of Fig. 7.9), the majority of the absorption ( $> 80\%$ ) is due to light polarized in the direction perpendicular to the dot layers. This is inconvenient and undesirable; it would be better if the absorption were pre-dominantly due to *s*-polarized light, which corresponds to normal-incidence light. For quantum-dot structures grown on [001]-oriented GaAs (or InP) substrates, one would also desire that the absorption of *s*-polarized light be independent of direction in the plane of the dots. This independence or lack of it is related to the inplane dot-shape symmetry.

In the bevelled wedge light-coupling configuration, there are two possible polarizations of incoming light that can cause absorption. For *s*-polarization (which is often called transverse-electric or TE-polarization), the optical electric field is parallel to the dot-layer plane. For *p*-polarization (also known as transverse-magnetic or TM-polarization), the  $45^\circ$ -angle wedge, causes 50% of the magnitude of the optical electric field to be along the growth direction, i.e. perpendicular to the dot-layer plane; the other 50% is directed inplane. Under this circumstance, one can use a polarizer to select the desired polarization of the incident light. Figure 7.9 shows the photoresponse spectra for intersublevel (intraband) transitions recorded at 78 K for an (In,Ga)As/GaAs quantum-dot structure. The spectra are normalized to the corresponding polarized spectral intensity of the incident infrared beam. Note the strong signatures for both the *s*- and *p*-polarizations. This indicates that the photoresponse arising from the intersublevel absorption is independent of the polarization of excitation, as expected and previously discussed for three-dimensionally confined electrons. The strong photoresponse for the *s*-polarization is simply an affirmation of earlier results demonstrating the suitability of (In,Ga)As/GaAs quantum-dot structures for normal-incidence infrared photodetectors.

In a separate but related experiment, the top sides of some detectors were illuminated with normal-incidence *s*-polarized light. In one case the light is specifically aligned to be polarized along the [110] crystallographic direction, while in another it is set to be along the [−110] direction; these are orthogonal inplane directions. Figure 7.10 shows the intersublevel photoresponse for light polarized along the [110] (*x*-axis) and the [−110] (*y*-axis) directions. The response for unpolarized light is also shown. This measurement was performed at 78 K with an applied bias voltage of about 1 V. The main spectral photoresponse for all three cases peaks at about  $10 \mu\text{m}$ . The spectral spread, defined as  $\Delta\lambda/\lambda$ , is about 11.3%. This narrow linewidth is an indication of fairly uniform dot sizes. The width also points to a transition from a confined ground-state level to a bound excited state. The amplitudes of the photoresponse for light polarized along the [110] and the [−110] directions are almost identical, suggesting that the shape of the dots in the *xy*-plane is symmetric. This is consistent with the atomic force microscope (AFM) image shown in the inset of Fig. 7.10. Weber *et al.* (1999) however, indicate that this may not necessarily be true for InAs/InAlAs quantum dots grown on InP substrates. Their work shows that these type of dots exhibit an inplane shape anisotropy. The anisotropy, in their case, led to a carrier confinement that was pre-dominantly in the [110] direction, resulting in photoresponse spectra that were strong for light polarized pre-dominantly in the [110] direction.



**Fig. 7.9** Normalized spectral photoresponse of long-wave (In,Ga)As/GaAs quantum-dot structure, for both *s*- and *p*-polarized light. Figure taken from Pal *et al.* (2003), *Appl. Phys. Lett.* **83**, 4634 with permission from AIP.



**Fig. 7.10** Spectral photoresponse for an (In,Ga)As quantum-dot structure to normal-incident light whose polarization has been resolved into two inplane and orthogonal directions labelled  $x$  and  $y$ . The figure was taken from Pal *et al.* (2003), *Appl. Phys. Lett.* **83**, 4634 with permission from AIP.

As indicated above, the spectral width of the response depends on the composition and size of dots. The magnitude of the electron-confining potential determines the number of energy levels in the dot; this in turn determines the type of intersublevel transitions that may occur. Bound-to-bound transitions, for example, tend to lead to narrow spectral widths; bound-to-continuum states, on the other, usually result in broader spectral responses. This behavior is similar to what has been observed in quantum-well infrared detector studies (Levine 1993). An example that illustrates this point was reported by Jiang *et al.* (2004b). They showed that the spectral width of a mid-wave infrared (In,Ga)As quantum-dot device responding at about  $4.7\text{ }\mu\text{m}$ , was about 14%. No appreciable changes of the peak wavelength, spectral shape, and cutoff wavelength were observed as the temperature was increased from 30 K to 160 K (Jiang *et al.* 2004b). We will return to this point later in the discussion. The spectral width of 14% indicates transitions between two bound states. Previous studies on quantum-well-type infrared detectors showed that—in general—for  $\Delta\lambda/\lambda_{\text{peak}}$  of less than 15%, the transitions are between bound states. Spectral widths larger than 15% tend to indicate bound-to-continuum transitions (Levine *et al.* 1992; Liu *et al.* 1993).

Although it has been challenging to control the size of quantum dots in detector structures, one of the desirable capabilities would be to vary the dot size in a layer-by-layer fashion to allow multicolor detection. Since this has been difficult to achieve controllably, the alternative has been to use the “dots-in-a-well” scheme where one can controllably confine one or more excited states in a thin-film well. Multicolor detectors using this scheme are very common. A typical device might contain an InAs/(In,Ga)As dots-in-a-well structure where the quantum well is deliberately made wide to allow more than one excited state. Krishna *et al.* have used this scheme to make a

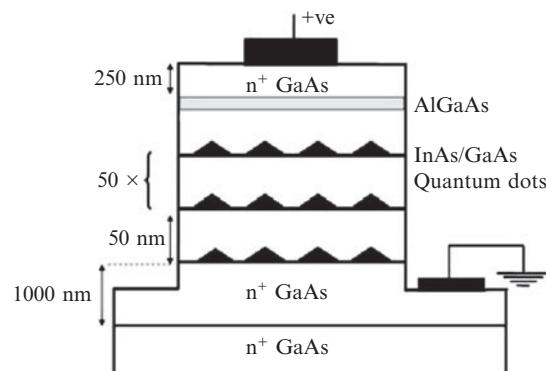
two-color detector where the (In,Ga)As quantum well is about 9 nm (Krishna *et al.* 2003). Their device responded at around  $4.2\text{ }\mu\text{m}$  and  $7.6\text{ }\mu\text{m}$ ; the low-wavelength response was attributed to a bound-to-continuum transition, while the longer-wavelength response was thought to be due to a transition from a bound-to-bound state. Another route to multicolor detection is to use the quantum-confined Stark-effect (Miller *et al.* 1984). Kim and Harris reported a voltage-tuneable two-color quantum-dot detector. Their device was able to demonstrate tuneable two-color operation in the mid-wave and long-wave infrared. At zero bias, the device operated in the mid-wave infrared. At negative applied voltages, it operated at 5.5 and  $9.2\text{ }\mu\text{m}$  (Kim *et al.* 2004). Multicolor operation in quantum-dot detectors can also be obtained simply by taking advantage of the dot-size distribution if there is a distinct clustering of dots by size. Dot-size dispersions have typically been reported to be in the range of  $\pm 10\%$ . There have also been reports where the distribution in a given ensemble may be bimodal (Pal *et al.* 2002). If this is the case, one may use such a distribution to make a two-color detector. Chen *et al.* appear to have taken advantage of this feature in a three-color infrared detector (Chen *et al.* 2005). Their device, using InAs quantum dots covered with an (Al,Ga)As layer in a GaAs matrix, responds at 6 and  $10\text{ }\mu\text{m}$ . The transitions responsible for these signatures are probably due a bound-to-bound and a bound-to-continuum state transitions in one set of dots with small sizes. A peak at around  $16\text{ }\mu\text{m}$  was also reported; this one could be attributed to transitions from ground states to continuum states in a set of larger dots because of the spectral width of the response. When the samples were annealed, the spectra changed. This was probably due to a change in the sizes of the dots because of the annealing.

In addition to being sensitive to normal-incidence light, quantum dots have been expected to lead to infrared detectors that are capable of high-temperature operation. This expectation was speculated to be due to a lack of strong coupling between confined electrons and a phonon bath. This so-called phonon bottleneck remains controversial (Benisty *et al.* 1991; Inoshita *et al.* 1997). However, several examples of quantum-dot detectors operating at relatively high temperatures have been shown. Bhattacharya *et al.* (2005), for example, reported a broad spectral photoresponse stretching from about 5 to beyond  $20\text{ }\mu\text{m}$  from a tunnel-barrier (In,Ga)As quantum-dot photodetector that operated at room temperature (Bhattacharya *et al.* 2005). This device avoids the usual problem that both photoelectrons and dark electrons must travel through the same path by using sequential tunnelling for the photoelectrons. The tunnel barriers block dark electrons with broad energy distributions. Another approach to achieving the same aim is to use multiple periods of quantum-dot layers to increase the absorption cross-section and hence photocurrent, and to simultaneously block the dark electrons by using GaAs or (AlGa)As barriers. Such an approach is reported in a contribution by Chakrabarti *et al.* where an InAs/GaAs quantum-dot detector with 70 periods is discussed (Chakrabarti *et al.* 2004a). This particular device operated up to 150 K. The use of many periodic layers of the basic quantum-dot strucure, coupled with dark current blocking barriers, appears to work in increasing the high-temperature limit of quantum-dot detectors. The number of periods

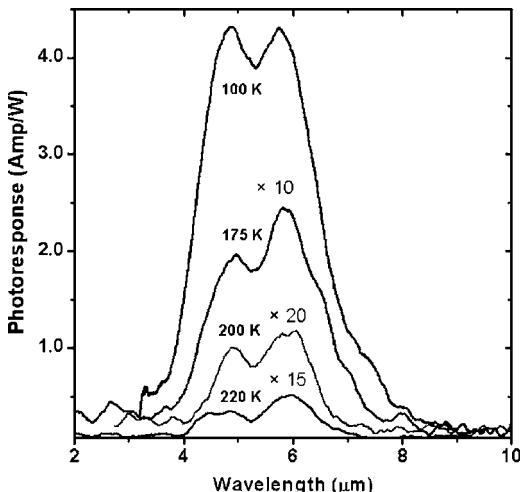
and the type and size of the dark current-blocking barrier must be optimally determined to minimize defect generation in the structure.

The magnitude of dark current to be contended with generally depends on the density of excess carriers not involved in the photoelectron-generation process. The excess density depends on the dopant impurity concentration. There are several ways by which carriers are introduced into dots: (i) the dots themselves may be doped, (ii) the (GaAs or AlGaAs) barrier layers adjacent to the dots are doped, or (iii) the entire active volume is uniformly doped. It appears that uniform doping of multiple periods of InAs/GaAs quantum dots in the active volume of a detector can lead to devices capable of operating up to a temperature of  $\sim 220$  K. This approach has been used by Pal and Towe (2006a). The uniform doping must take into account the volume density of dots to minimize the density of excess carriers that could result and potentially lead to additional dark current. The uniformly doped detector structure is a good prototype for improving overall dark current uniformity; this is an essential first step toward high-quality focal-plane imaging arrays. Because of its importance, we provide some salient details. A typical InAs/GaAs quantum-dot structure can be grown by solid-source molecular beam epitaxy. The details are as follows: on top of a highly doped GaAs substrate, a  $0.75\text{-}\mu\text{m}$  GaAs ( $n = 2 \times 10^{18} \text{ cm}^{-3}$ ) layer is grown at a substrate temperature of  $580^\circ\text{C}$ . This layer is followed by another GaAs layer with a thickness of 50 nm; then 50 periods of InAs/GaAs quantum-dot layers are grown at a substrate temperature of  $490^\circ\text{C}$ . After the final GaAs barrier layer in the InAs/GaAs superlattice is grown, the substrate temperature is ramped up to  $580^\circ\text{C}$ , whereupon an undoped, 40-nm  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$  dark current-blocking layer is grown. Finally, a  $0.25\text{-}\mu\text{m}$  Si-doped ( $n = 2 \times 10^{18} \text{ cm}^{-3}$ ) GaAs layer is grown for the top Ohmic contact. The entire InAs/GaAs quantum-dot superlattice structure is uniformly doped with Si to achieve an electron density of about  $2 \times 10^{16} \text{ cm}^{-3}$ ; this translates roughly into having 2–3 electrons per dot. The desired goal is to have two electrons occupy the ground state. A schematic for the device structure described above is shown in Fig. 7.11.

The normalized spectral response of the device just discussed was measured for operating temperatures from 100 to 220 K. These characteristics are shown in Fig. 7.12. Two peaks are evident in the spectrum; one is located at the

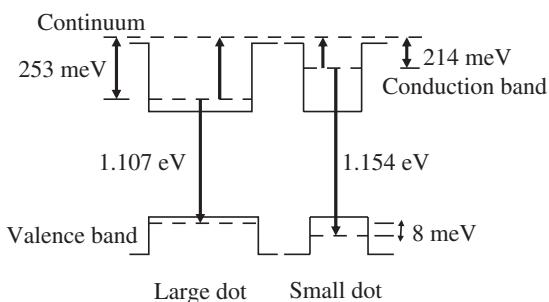


**Fig. 7.11** Schematic device structure of an InAs/GaAs quantum-dot infrared photodetector with an  $(\text{Al},\text{Ga})\text{As}$  dark current-blocking layer. The current-blocking layer makes it possible for the device to operate at relatively high temperatures.



**Fig. 7.12** Photoresponse spectra of mid-wave InAs/GaAs quantum-dot detector capable of operating up to 220 K. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

wavelength of about  $5.8 \mu\text{m}$ , and the other at  $4.9 \mu\text{m}$ . These peaks probably originate from transitions in dots that belong to an ensemble with a bimodal size distribution. The peak at  $5.8 \mu\text{m}$  is probably a result of transitions from ground-state levels in small-size dots to the continuum; the other peak, the one at  $4.9 \mu\text{m}$ , could be the signature of transitions from the ground state to the continuum in larger size dots. For both peaks, a clean photoresponse spectrum continues to be observed up to the highest temperature of about 220 K. No appreciable change in the peak wavelengths is observed as the operating temperature is raised from 175 to 220 K. The spectral insensitivity of the peaks to temperature indicates the quantum nature of the transition; this is in contrast to what one generally expects from band-to-band transitions that are strongly temperature dependent. At an operating temperature of 100 K, we obtain a spectral spread of  $\Delta\lambda/\lambda_{\text{average}} \sim 47\%$ . This is consistent with transitions from bound levels in the dots to the continuum (Levine 1993). The band-to-band emission characteristics of the dots were probed by the technique of photoluminescence; the possible location of the energy levels extracted from the spectra and correlated with some calculations are shown in Fig. 7.13. The conduction-band information is consistent with the intersublevel transition energies of the device discussed above. Other researchers have also reported similar transitions thought to be due to dot-size dispersion (Chen *et al.* 2002, 2005).



**Fig. 7.13** Schematic of the possible energy structure of a dot ensemble with a bimodal size distribution. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

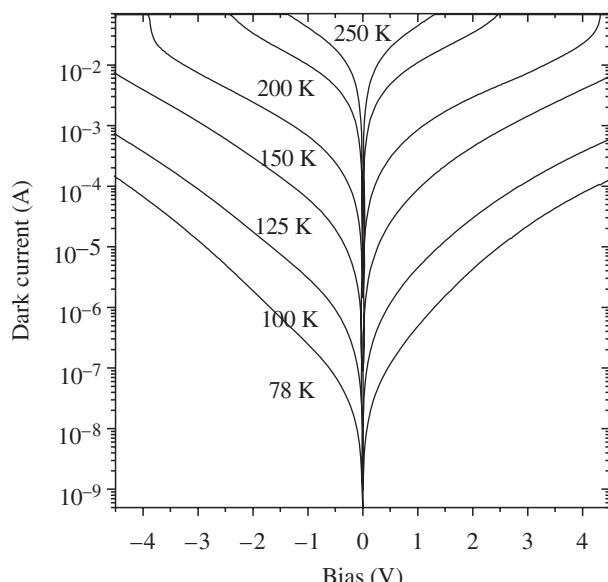
## 7.5 Device characteristics

### 7.5.1 Current–voltage characteristics of quantum-dot detectors

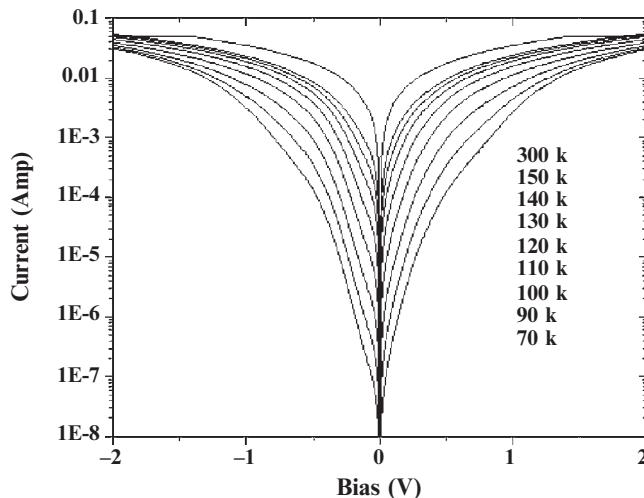
The dark current–voltage ( $I$ – $V$ ) characteristics of a detector generally give information on the variation of the mobility and carrier lifetimes in the detector; these are somewhat dependent on the applied bias voltage. The characteristics are also a subtle statement of Ohm’s law. One can usually determine from dark current–voltage measurements if a device is behaving “properly” as a detector or not by simply observing if the characteristics obey the usual Ohm’s law or not. If they do, then the device is not a quantum-dot detector. When measured without illumination, the characteristics can also give an indication of the level of “noise” inherent in the devices.

In a quantum-dot infrared photodetector, dark current noise originates from several sources that include (a) thermionic emission, (b) thermally assisted tunnelling of dark electrons, and (c) temperature-independent tunnelling. In thermionic emission processes, carriers from within dots are thermally excited to continuum states from where they contribute to dark current. The current due to thermionic emission processes depends exponentially on temperature. Thermionic processes generally dominate above 50 K. In thermally assisted tunnelling, carriers are thermally excited into higher energy levels within a dot, just below the top of the conduction band; from there, they can much more readily tunnel through the triangular tip of the potential barrier into states that support conduction. Finally, in temperature-independent tunnelling, carriers in adjacent dots may tunnel from one dot to another, and perhaps from or to states in dot-wetting layers; these carriers eventually reach conduction states where they contribute to dark current.

In Fig. 7.14, we illustrate typical dark current–voltage characteristics of a standard mid-wave InAs/GaAs quantum-dot detector with an (Al,Ga)As dark



**Fig. 7.14** Current–voltage characteristics of an InAs/GaAs mid-wave quantum-dot detector (in the dark) with an (Al,Ga)As dark current-blocking layer. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.



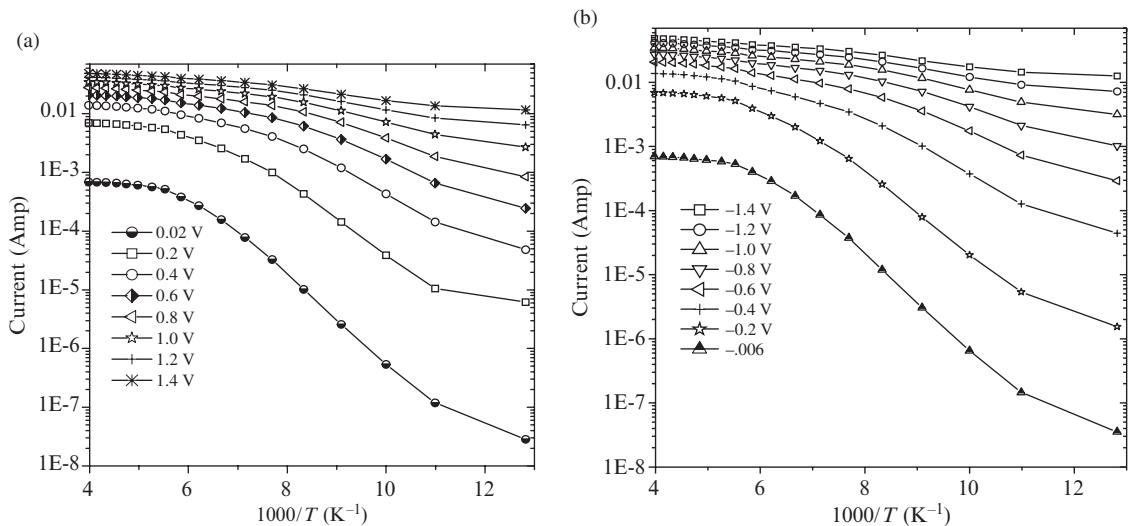
**Fig. 7.15** Current–voltage characteristics of a long-wave (In,Ga)As/GaAs quantum-dot detector (in the dark) at various operating temperatures.

current-blocking barrier layer for temperatures ranging from 78 to 250 K. At a bias voltage of about 0.5 V, the dark current increases by over five orders of magnitude from  $5.2 \times 10^{-8}$  A at 78 K to about 0.028 A at 250 K. The relatively low dark currents in this particular device over the measurement temperature range, and in the bias voltage range of  $\pm 4.5$  V, are primarily due to the wide-bandgap current-blocking layer in the structures. The increase of dark current with increasing operating temperature is due to thermionic emission of carriers from the dots.

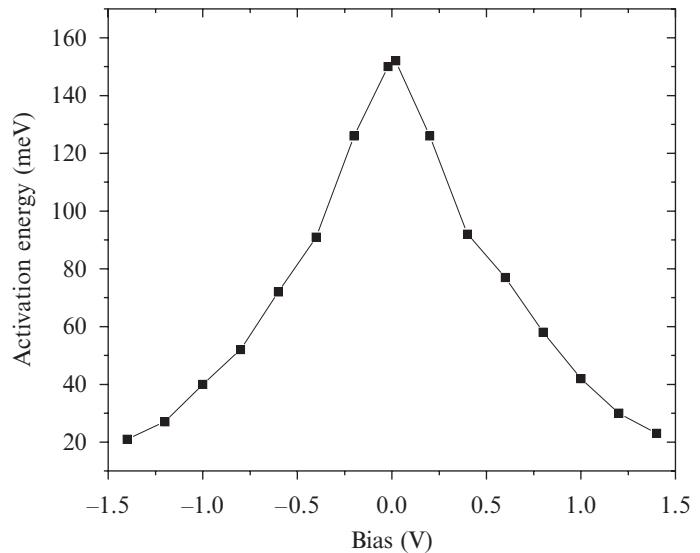
Dark current–voltage characteristics for long-wave (In,Ga)As/GaAs detectors tend to be a little noisier than those for the short-wave detector shown in Fig. 7.14. Measured characteristics for an (In,Ga)As/GaAs detector are shown in Fig. 7.15 for a similar temperature range, but for a bias voltage range of  $\pm 2.0$  V. At the bias point of 0.25 V at 300 K and then at 78 K, the dark current decreases by three orders of magnitude. This is probably due to carrier freeze-out in the dots. One can examine the thermally activated processes in this device structure by constructing an Arrhenius plot of the dark current for a fixed bias voltage point, but at varying temperatures. Plots of this kind are shown in Figs. 7.16(a) and (b) for a positive and a negative bias point, respectively. Activation energies for the thermal process involved can be extracted from the linear portions of the plots. The variation of such activation energy as a function of bias voltage is shown in Fig. 7.17. For this particular device, one finds that the activation energy decreases from a high of 155 to a low of 20 meV in the voltage range from 0.02 to 1.4 V. The plot displays an obvious symmetry; this is a direct consequence of the symmetry inherent to the device structure used (which is shown in Fig. 7.18).

### 7.5.2 Quantum-dot detector peak responsivity

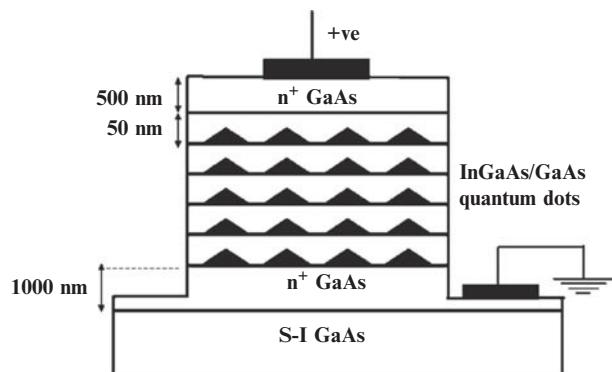
Detector responsivity—as defined earlier—is a measure of the transfer function of the device. It is generally wavelength dependent for all photon detectors as shown in eqn (7.27). For quantum-dot detectors, it is measured



**Fig. 7.16** Temperature dependence of the dark current of a long-wave (In,Ga)As/GaAs quantum-dot detector for (a) positive and (b) negative bias points.



**Fig. 7.17** Variation of thermal activation energy as a function of bias voltage for a long-wave (In,Ga)As/GaAs quantum-dot detector.



**Fig. 7.18** Schematic layer structure of a long-wave (In,Ga)As/GaAs quantum-dot detector.

at the peak wavelength of response for the device under test. One then usually quotes it as the “peak responsivity”, which still depends on other important device structure parameters, as well as on materials properties. The role of a device designer is therefore to optimize the peak responsivity, among other important device parameters. We discuss the various factors that can influence peak responsivities of quantum-dot detectors below.

Using undoped quantum dots in the active volume of a detector may often lead to a relatively high responsivity. In such device structures, the dots are populated by electrons that transfer from ionized dopant impurities at the top and bottom contact layers that are usually heavily doped n-type. The electron transfer occurs because the dots represent the lowest potential-energy regions in the device structure. Several researchers, including us, have reported the successful use of this approach in devices with good results (Raghavan *et al.* 2002; Ye *et al.* 2002; Kim *et al.* 2004). In the typical device, such as that reported Ye *et al.*, the structure consists of InAs dot-layers capped with (In,Ga)As stress-relieving material. The entire active region is sandwiched between heavily doped GaAs contact layers. Because of dot-size dispersion, the device exhibited a multicolor response. The peak responsivities at the two prominent wavelengths were fairly high. For increasing bias, there was a corresponding increase in responsivity that was approximately exponential. The response increased by four orders of magnitude from a low value of  $10^{-4}$  A/W at a bias of slightly less than 0.5 V to about 1 A/W at  $-1.3$  V.

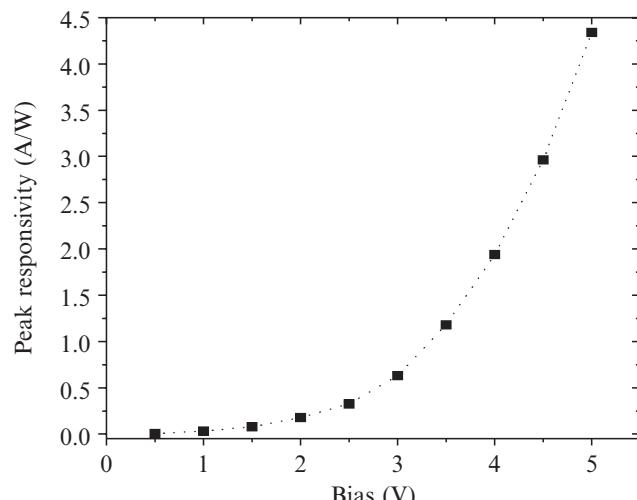
The majority of quantum-dot photodetectors reported are based on confinement of electrons in conduction-band potential wells. The electron-type detector works rather well for a number of reasons. These include a large conduction-band heterointerface energy discontinuity for electron confinement, and a high electron mobility. However, a valence-band potential-well photodetector, based on holes would probably work, but at much lower cryogenic temperatures because of the lower confining potentials for holes. A rather interesting quantum-dot detector that includes p-type doping was reported by Chou *et al.* This device is not strictly a hole-type detector. It consists of multiple periods of InAs dots in the active region that are doped n-type, but surrounded by GaAs barriers that are doped p-type (Chou *et al.* 2005). The authors claim that the p-type layers were included to provide a mechanism for controlling the Fermi level in the device. Presumably, by lowering the effective Fermi level, one should see a decrease in dark current. The reported results, however, do not offer much hope. The devices were characterized at 10 K, a temperature that is much lower than the usual one of 77 K. A peak responsivity of about 0.5 A/W at  $6\ \mu\text{m}$  was measured at 10 K. Even though this number may appear reasonable, the recorded photoresponse spectrum was extremely noisy (with what appeared to be a SNR  $\sim 2$  to 3), making it impossible to extract a detectivity value.

The highest peak responsivity ever reported for quantum-dot detectors, to the best of the authors’ knowledge, was observed in (In,Ga)As/GaAs quantum-dot detector structures with lateral carrier transport. Lateral transport devices are expected to have large photoconductive gains—given by

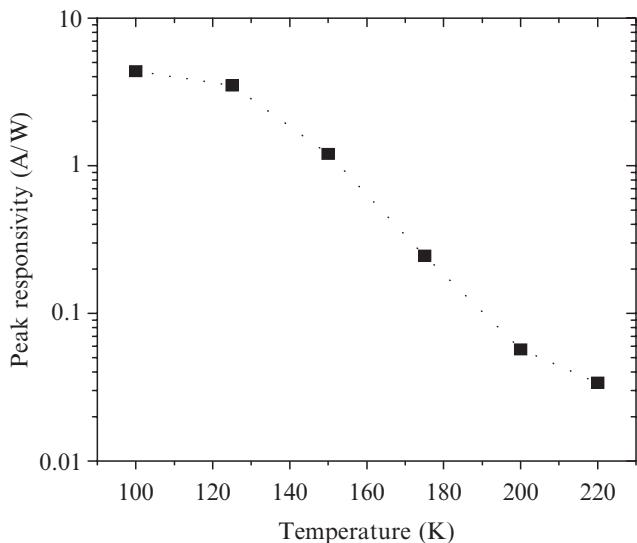
eqn (7.22)—that potentially have beneficial effects on responsivity. In a device of this type, discussed by Chu *et al.*, a responsivity of about 11.2 A/W at a bias of 1.7 V was measured. The measurement temperature and the chopping frequency of the incident radiation were 30 K and 11 Hz, respectively (Chu *et al.* 2001). The (quantum well) channel transport layer in this device offers high electron mobility as well as a region with low scattering cross-section.

We now discuss some details on some aspects of detector responsivity. These measurements are typically carried out using a calibrated blackbody source. The temperature of the blackbody source can be maintained at about 500 °C for the mid-wave and long-wave infrared radiation. One typically chops the radiation from the blackbody source at a frequency of about 507 Hz; this frequency is deliberately chosen not to be a multiple of 60 Hz (the ac line frequency) to distinguish the photocurrent signal from the power-line noise. The photocurrent response is obtained by averaging the detector signal over a 50-Hz bandwidth centered around the chopper frequency. The noise signal should also be measured in the same bandwidth range and chopper frequency (507 Hz). We have carried out these measurements on a number of (In,Ga)As quantum-dot detector samples with and without current-blocking layers. The typical device structure is as illustrated in Fig. 7.18, with minor deviations in some cases to suit the measurement objective. The results of these measurements are discussed in the following.

Peak responsivities at various levels of bias for the mid-wave device operating at 5.8 μm are shown in Fig. 7.19. This device was measured at 100 K. The responsivity increases from about 32 mA/W to 4.3 A/W when the bias is raised from 1 to 5 V. Although this device was expected to have higher responsivities, the (Al,Ga)As current-blocking layer in it blocks both the dark and photoelectrons, thus reducing the overall photocurrent. The low dark current in the device, however, allows one to apply relatively high bias voltages that apparently lead to high peak responsivities. Other mechanisms that contribute



**Fig. 7.19** Variation of peak responsivity with bias for an InAs/GaAs quantum-dot detector with a dark current-blocking (Al,Ga)As layer. The measurement temperature is 100 K. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

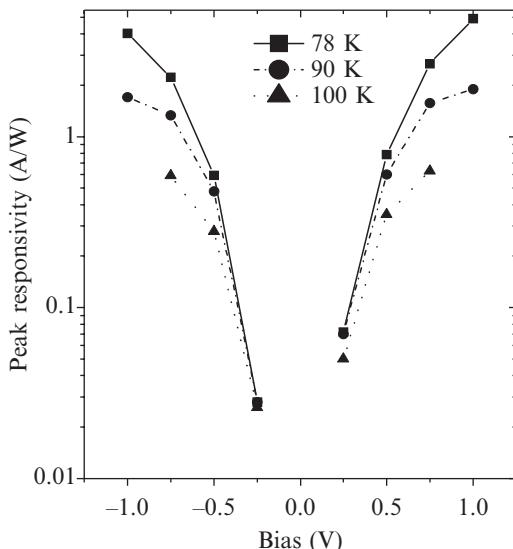


**Fig. 7.20** Variation of peak responsivity with operating temperature for an InAs/GaAs quantum-dot detector with a dark-current blocking (Al,Ga)As layer. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

to the high responsivity include a faster carrier escape time, high field-assisted carrier transport, and perhaps avalanche multiplication due to the high bias voltages.

To compare peak responsivities at different temperatures, it is desirable to maintain the same level of dark current in the devices being compared at the different temperatures. This can be done by appropriately adjusting the bias voltages of the devices for each temperature of operation. We illustrate in Fig. 7.20, the measurement of peak responsivities (at constant dark current) for different temperatures. Observe that the peak responsivity decreases from a high value of about 4.3 A/W at 100 K to a low value of 34 mA/W at 220 K as the temperature is raised. The decrease of peak responsivity with increasing temperature is due to the reduction of the number of light-absorbing confined electrons in the dots because of thermal emission. Thermal ejection of carriers from the dots decreases the number available for photoexcitation, leading to a reduction in photocurrent. Furthermore, the decrease of electron mobility with increasing temperature may also contribute to the decrease in photoresponse.

Detector responsivity is a function of both the operating temperature and bias voltage. Both variables affect the number of electrons participating in photocurrent generation. The illustration in Fig. 7.21 shows how peak responsivity varies as a function of bias at different temperatures. The measurements are for a long-wave infrared (In,Ga)As/GaAs quantum-dot photodetector with a structure similar to that shown in Fig. 7.18. For the operating temperature of 78 K, the peak responsivity increases from about 72 mA/W to 4.9 A/W when the bias voltage is raised from 0.25 to 1.0 V. For negative bias, the responsivity increases from 28 mA/W to 4.02 A/W as the bias voltage is varied from -0.25 to -1.0 V. At the single bias point of 0.75 V, the peak responsivity decreases from 2.67 A/W to 0.63 A/W when the temperature is increased from 78 K to 100 K. Clearly, carriers that would have contributed to the photoelectron



**Fig. 7.21** Peak responsivity of a long-wave (In,Ga)As/GaAs quantum-dot detector measured at different operating temperatures. Figure taken from Pal *et al.* (2003), *Appl. Phys. Lett.* **83**, 4634 with permission from AIP.

population are being thermally ejected from the quantum dots as the temperature is increased. The thermal loss of carriers is reflected in lowered responsivity (Kim *et al.* 1998). The data in Fig. 7.21 indicate an asymmetry of response with bias polarity. This is likely due to the asymmetric shape of the dots along the growth direction; the asymmetry could also be due to the wetting layers under the dots. Because of this, electrons in dots encounter different barrier heights—depending on whether transport is toward the top or bottom contact.

### 7.5.3 Measured noise characteristics

We discussed earlier that noise is comprised of unwanted random signals arising from various mechanisms in the detector. The various sources of noise depend on the detector type. For most applications, only a few sources dominate the performance of infrared detectors. For example, Johnson noise is present in all resistor-type detectors (such as the photoconductive type), whereas in an ideal photovoltaic detector, this type of noise is less of a problem. The dominant noise in photovoltaic detectors is shot noise, which is due to the random arrival rate of photons, and hence carriers to the junction of the photovoltaic structure. Since a quantum-dot detector is a photoconductive device, the main noise types are Johnson noise and the so-called generation–recombination noise.

Detailed noise measurements are essential for complete characterization of detectors. One of the key performance metrics of quantum-dot detectors depends on knowledge of the noise magnitude. The detectivity, as discussed earlier, is a measure of the normalized signal-to-noise-ratio. Its value can be increased by reducing the noise. It is thus important to understand the noise characteristics of quantum-dot detectors in order to find ways to reduce it.

To date, only a few researchers have carried out extensive studies of noise characteristics in quantum-dot photodetectors. The most detailed report is by Ye *et al.* (2003). These researchers studied the noise spectral density of unintentionally doped InAs quantum-dot detectors as a function of frequency for various operating temperatures. It is found that for frequencies below 2 Hz, the dominant noise is of  $1/f$  type. Beyond 2 Hz, the noise is independent of frequency, indicating a generation–recombination-type noise similar to that in bulk photoconductive devices, as discussed above. This observation means that the generation and trapping processes may be the dominant noise sources in quantum-dot photodetectors. Furthermore, the noise floor increases with increasing temperature; this is to be expected as dark current increases with temperature.

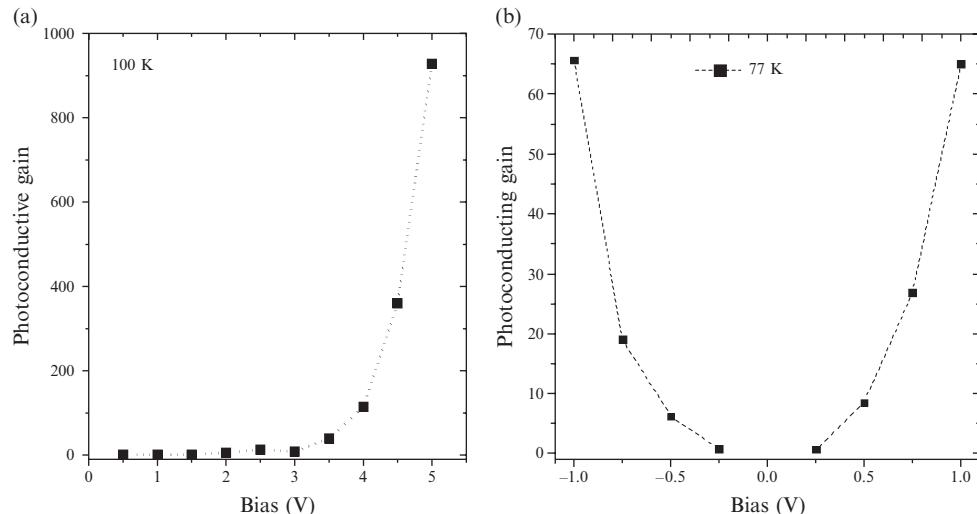
Noise measurements are also important for determining detector gain. The dark current,  $I_d$ , in a detector can be written as

$$I_d = qG_{\text{th}}\gamma V, \quad (7.34)$$

where  $q$  is the electron charge as before,  $G_{\text{th}}$  is the thermal generation rate,  $\gamma$  the gain (as defined before), and  $V$  the volume of the detector active region. Since generation and recombination can be assumed to be uncorrelated Poisson processes, the noise current in the device can be written as

$$i_n^2 = 4q^2\gamma^2 G_{\text{th}}V\Delta f = 4qI_d\gamma\Delta f, \quad (7.35)$$

where we have inserted the expression for the dark current in eqn (7.34) into eqn (7.35), and the factor  $\Delta f$  is the noise current measurement bandwidth. Equation (7.35) can be used to determine gain in quantum-dot detectors. In fact, several researchers have reported results based on this expression (Ye *et al.* 2003). The validity of its usage is based on an approximation that assumes that in conventional photoconductors, noise gain is equal to photoconductive gain. Quantum-dot detectors tend to have very large gains that are attributed to longer carrier lifetimes in the excited state due to suppressed electron–phonon scattering. However, avalanche multiplication can also lead to high gains in quantum-dot detector structures. For unintentionally doped InAs quantum-dot structures, gains as high as 750 at relatively low bias have been reported. The corresponding electric field in the particular structure used was about 2.3 kV/cm (Ye *et al.* 2003). We illustrate in Fig. 7.22(a), the typical variation of photoconductive gain as a function of bias voltage in an InAs/GaAs detector with an (Al,Ga)As current-blocking layer. The schematic of the device structure, shown in Fig. 7.11, has been discussed before. As shown in the illustration, the gain increases from 1 to 928 when the bias voltage is raised from 1 V to 5 V in Fig. 7.22(a). The onset of the rapid rise of gain at about 4 V is thought to be due to avalanche multiplication. Similar large gains have been observed in (In,Ga)As/(In,Ga)P quantum-dot photodetectors (Jiang *et al.* 2004b). Long-wave infrared (In,Ga)As/GaAs quantum-dot detectors do not tend to have gains as large as those for the mid-wave infrared devices. Gains for the long-wave infrared detectors tend to be modest as shown in Fig. 7.22(b), which shows photoconductive gain measured at 77 K. As in the mid-wave



**Fig. 7.22** Photoconductive gain as a function of bias for (a) an InAs/GaAs quantum-dot detector with an (Al,Ga)As dark-current-blocking layer measured at 100 K, and for (b) long-wave (In,Ga)As/GaAs detector measured at 77 K. Figure (a) taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

infrared case, the gain increases with increasing bias voltage. The data in this figure shows that gain is asymmetric with bias polarity—an indication of the inherent asymmetry of the physical shape of the dot discussed earlier.

#### 7.5.4 Measured peak detectivity

Detectivity is perhaps the most widely used evaluation metric for quantum-dot detectors. As previously defined in eqn (7.32), it gives an indication of the sensitivity of the device. Experimental values of detectivity are usually reported for a peak wavelength. This is consistent with our discussion on responsivity whose value is generally measured at the peak photoresponse wavelength. Accurate measurement of the noise current in a detector is an important determinant of detectivity. One could argue that the two most important measurements for quantum-dot detectors are the peak responsivity and the noise-current measurements. These determine the detectivity.

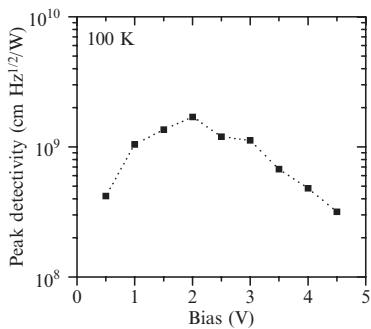
Detectivities ranging from  $10^7$  to  $10^{11} \text{ cm Hz}^{1/2}/\text{W}$  for quantum-dot infrared photodetectors have been reported (Raghavan *et al.* 2002; Jiang *et al.* 2004b; Kim *et al.* 2004). To the best of our knowledge, the highest peak detectivity (of about  $3 \times 10^{11} \text{ cm Hz}^{1/2}/\text{W}$ ) was reported for an unintentionally doped InAs/InGaAs/GaAs quantum-dot photodetector operated at 78 K (Kim *et al.* 2004). The most likely contributor to such a high detectivity value is low dark current. The unintentionally doped active region probably generated very few “dark” mobile electrons. The majority of mobile electrons in the structure were therefore photoelectrons that directly contributed to high responsivity. In the reported device, the detectivity increased with increasing bias voltage; this was probably due to increased field-dependent electron mobility, leading to an increased peak responsivity. We should point out that responsivity does

not increase indefinitely with bias; up to a point, it begins to decrease with continued increase in bias. This is because at the higher bias voltages, the dark current increases more rapidly than the photocurrent.

A number of techniques have been explored for increasing the responsivity of quantum-dot detectors. The most straightforward approach is to simply increase the number of periods of the dot layers. This is an optimization game. Simply increasing the number of periods of dot-layers may not necessarily lead to higher detectivities. A large number of periods may actually lead to dislocations that are deleterious to electron transport or lead to the creation of recombination centers for photoelectrons. The number of periods of quantum-dot layers that seems optimal is somewhere between 50–70. No rigorous optimization calculations have been carried out. Experiments with structures containing 70 periods have been reported. One example, by Chakrabarti *et al.*, uses 70 periods of InAs/GaAs quantum dots in a structure with an (Al,Ga)As current-blocking layer; the reported peak detectivity was about  $10^{11} \text{ cm Hz}^{1/2}/\text{W}$  at 100 K (Chakrabarti *et al.* 2004b). The large number of quantum-dot layers and the current-blocking layer played an important role in achieving the high peak detectivity. As already discussed, the key is to increase the density of photoelectrons generated while minimizing the transport of “dark” electrons. When the operating temperature in the Chakrabarti device was raised to 200 K, the peak detectivity decreased to  $\sim 6 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ .

Since peak responsivity depends on bias (as discussed earlier), one expects that detectivity will show a similar behavior. We show in Fig. 7.23, typical peak detectivities measured at 100 K for a mid-wave infrared InAs/GaAs quantum-dot detector at different bias voltages. The maximum peak detectivity is about  $2 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ ; this occurs for a bias voltage of about 2 V. The rise of detectivity with increasing bias voltage from 0.5 V to 2 V is due to a corresponding increase in responsivity. Its decrease as the bias voltage is increased beyond 2 V is probably because of the rapid increase of the noise current with increasing bias voltage. We believe that increasing the areal dot density, which is directly related to the number of “good” electrons available to absorb infrared light would increase the photoelectron current; this in turn would improve the detectivity. These considerations must not be taken in isolation; one should, for example, simultaneously strive to get a better handle on the dopant impurity concentration. By providing an optimum number of dopant impurities, one guarantees that the right number of electrons are available to populate the dots while minimizing the number of excess electrons that contribute to noise during device operation.

High-temperature ( $>77 \text{ K}$ ) operation of long-wave infrared detectors is an important requirement that has often dictated the choice of materials for the devices. Several researchers have reported operation beyond 77 K (Stiff *et al.* 2001; Tang *et al.* 2001). It is important to note that while operation at high temperatures is in itself an important attribute, it is closely intertwined with the wavelength of operation. It is generally easier to design and fabricate short-wave infrared ( $<7 \mu\text{m}$ ) quantum-dot photodetectors that operate at relatively high temperatures ( $>77 \text{ K}$ ) than those that operate at long-wave infrared wavelengths ( $>7 \mu\text{m}$ ). This fact, which is often glossed over, is simply a constraint imposed by the thermal energy  $kT$ , at the operating temperature

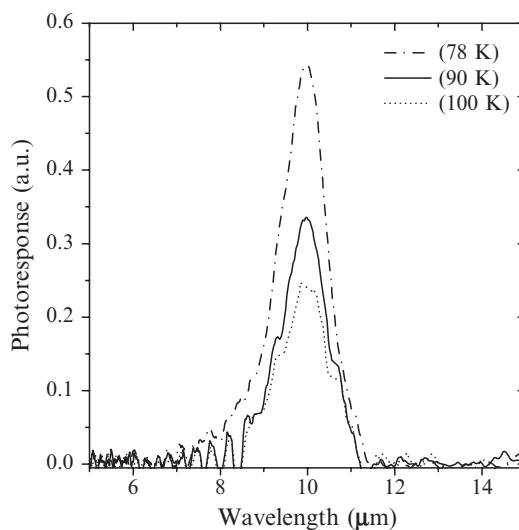


**Fig. 7.23** Bias-dependent peak detectivity of an InAs/GaAs quantum-dot detector with a current-blocking layer measured at 100 K. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.

and its relationship to the equivalent transition energy,  $\Delta E$ , for the chosen wavelength,  $\lambda$ , and the effective confinement potential,  $\Delta V_{\text{eff}}^{\text{conf}}$ .

For short-wave infrared, the energy difference between levels is generally large and the probability of thermally depopulating filled ground states of electrons that could fill empty upper states is low. At long-wave infrared wavelengths (corresponding to narrower energy differences between levels), the probability of filling excited states with thermal electrons gets larger, thus limiting operation at higher temperatures for the long-wave infrared devices. Another way to think about this is to consider the value of thermal energy,  $kT$ , where  $k$  is the Boltzmann constant and  $T$  the temperature in degrees absolute. When this number is small compared to the separation between levels, carriers are unlikely to be thermally ejected from their low energy levels. On the other hand, if  $kT$  is comparable to the energy-level separation, then carriers are more likely to be thermally ejected out of a level.

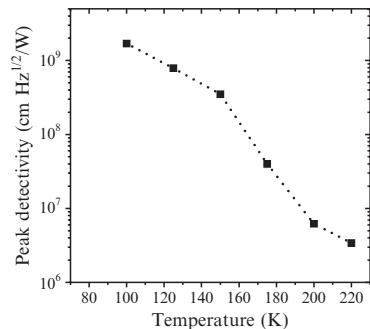
The temperature-dependent intersublevel photoresponse of long-wave infrared  $\text{In}_{0.23}\text{Ga}_{0.77}\text{As}/\text{GaAs}$  quantum-dot infrared detectors has been investigated by Pal *et al.* (2003). Spectra measured at different temperatures in the normal-incidence configuration are shown in Fig. 7.24 at three different temperatures: 78, 90 and 100 K; these spectra were measured under an applied bias voltage of about 0.75 V. It is evident that the intensity of the photoresponse decreases as temperature is raised. The explanation is that as temperature is raised, more thermally excited electrons become available to occupy unfilled energy levels in the dots; as a consequence, the population of empty levels that would have been available to photoelectrons is decreased, leading to a decrease in the intensity of the absorption (Ye *et al.* 2002a). The clear photoresponse spectrum at 100 K indicates that long-wave detectors, without (AlGa)As blocking barrier layers, can be operated at around 100–105 K. While there is a notable decrease in photoresponse with temperature, there is no change of peak wavelength. As already discussed, the peak wavelength depends on the



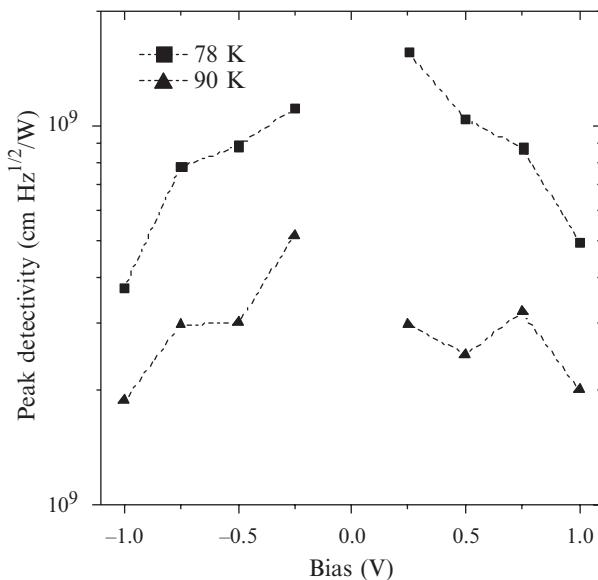
**Fig. 7.24** Temperature dependency of the photoresponse spectra for an (In,Ga)As/GaAs quantum-dot detector at a bias of 0.75 V. Figure taken from Pal *et al.* (2003), *Appl. Phys. Lett.* **83**, 4634 with permission from AIP.

energy-level separation,  $\Delta E$ , between the ground and first excited energy level. The quantized energy levels within the dot may move toward lower energy levels with an increase in temperature; however, the separation between the levels remains the same as observed in the photoresponse spectrum.

The foregoing discussion provided the context within which the temperature dependence of photorepsonses should be viewed. The essential point to keep in mind is that the spectral region of operation is a major determinant of the temperature behavior of a quantum-dot detector. We now show in Fig. 7.25 how the peak detectivity of a mid-wave infrared InAs/GaAs quantum-dot device varies as a function of temperature. The maximum peak detectivity for this particular device is about  $2 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$  at 100 K; this detectivity decreases to about  $3.5 \times 10^6 \text{ cm Hz}^{1/2}/\text{W}$  when the temperature is increased to 220 K. The decrease in detectivity is due to a number of reasons, the most dominant being the increased noise current and the reduced peak responsivity at elevated temperatures. Detectivities for a long-wave infrared InGaAs/GaAs device at different bias voltages and temperatures are shown in Fig. 7.26. The detectivity values are about  $1.6 \times 10^9$  and  $1.1 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$  at 0.25 and  $-0.25 \text{ V}$ , respectively, at 78 K. These values are reduced with an increase in temperature as well as bias voltage. They are reduced, for example, to  $4.9 \times 10^8$  and  $3.7 \times 10^8 \text{ cm Hz}^{1/2}/\text{W}$  from  $1.6 \times 10^9$  and  $1.1 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ , when the bias voltage is increased from 0.25 to 1 and  $-0.25$  to  $-1.0 \text{ V}$ , respectively. When the temperature is increased from 78 to 90 K, the detectivity at a bias of 0.25 V is reduced from  $1.6 \times 10^9$  to  $3 \times 10^8 \text{ cm Hz}^{1/2}/\text{W}$ . And for a negative bias of  $-0.25 \text{ V}$ , the detectivity decreases from  $1.6 \times 10^9$  to  $5.2 \times 10^8 \text{ cm Hz}^{1/2}/\text{W}$  with a corresponding increase in temperature from 78 K to 90 K. As we have already discussed, the decrease in detectivity with increasing temperature and bias is a result of increasing dark current.



**Fig. 7.25** Temperature dependence of peak detectivity of an InAs/GaAs quantum-dot detector with a dark-current-blocking layer. Figure taken from Pal *et al.* (2006a) *Appl. Phys. Lett.* **88**, 153109 with permission from AIP.



**Fig. 7.26** Peak detectivity values of an (In,Ga)As/GaAs long-wave quantum-dot detector at 78 and 90 K as bias voltage is varied from  $-1.0$  to  $+1.0 \text{ V}$ .

### 7.5.5 Measured quantum efficiency

The external quantum efficiency or conversion efficiency is defined as the ratio of the total number of carriers collected at the device contacts to the total number of photons incident on the device. This quantity can be extracted by rearranging eqn (7.27). There appears to be no consistent definition of it in the literature. From eqn (7.27), one can write it as

$$\eta = \Re \frac{hc}{\gamma \lambda q}. \quad (7.36)$$

Levine (Levine *et al.* 1992) defines the external quantum efficiency as comprised of two components: one component  $\eta_0$ , from optical absorption, and a factor,  $p$ , from the probability that a photoexcited electron will escape from a potential well and contribute to the photocurrent. This is the same probability factor included in eqn (7.29). According to this definition,  $\eta = \eta_0 p$ . The optical absorption factor can be extracted from eqn (7.29); it includes a factor for surface reflectivity and another for the absorption coefficient. By grouping the terms in eqn (7.29) according to this definition, we arrive at Levine's two components for the efficiency

$$\eta = \overbrace{\eta_0}{(1 - R)(1 - \exp(-\alpha d)} p = \eta_0 p. \quad (7.37)$$

There are very few reported measurements of the external quantum efficiency for quantum-dot infrared detectors. The values that have been reported have tended to be lower than those for quantum-well infrared detectors. Reported values for quantum-well infrared detectors, for example, have ranged from 8 to 25%. In general, the quantum efficiency for quantum-well-and quantum-dot-based infrared detectors has been consistently lower than what has been measured for band-to-band infrared detectors made from mercury cadmium telluride semiconductor. Since quantum efficiency evidently depends on responsivity—as can be seen from eqn (7.36)—it is also a function of bias voltage. This dependency has been shown for InAs quantum-dot detectors with  $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$  strain-relief layers. It was found that the quantum efficiency increased with increasing bias voltage up to a maximum about 9% (Ye *et al.* 2002b). A related parameter is the internal quantum efficiency; this is typically a much smaller fraction of the external quantum efficiency. Reported values of it have been below 4% (Raghavan *et al.* 2002; Lim *et al.* 2005).

## 7.6 Toward quantum-dot focal plane array imagers

infrared photodetectors, and the focal plane arrays that are often fabricated from them, have many important commercial and military applications. These include, among others, medical thermography, environmental monitoring, and night vision imaging. As we have already mentioned, mercury cadmium telluride is the dominant materials system from which advanced and extremely

sensitive focal plane arrays are made today. There are several very good reasons why—the most important being the fact that the bandgap of MCT can be continuously tuned to allow very efficient detection of infrared light from about  $0.7\text{ }\mu\text{m}$  to  $25\text{ }\mu\text{m}$ . This is in spite of difficulties in growing defect-free, large-area MCT layers on compatible substrates. An emerging approach to making focal plane arrays is to use quantum structures made from Group III–V semiconductor alloys; for example, GaAs/AlGaAs. Excellent focal plane arrays have been demonstrated using GaAs/AlGaAs quantum-well infrared photodetectors operating in the mid-wave infrared and long-wave infrared regions of the electromagnetic spectrum (Gunapala *et al.* 1997, 2003). An important advantage of the quantum-well-based focal plane arrays is that one can leverage well-established and developed technology infrastructure for manufacturing Group III–V compound semiconductors. There is, however, one disadvantage: quantum-well devices are not sensitive to normally incident light. The reasons for this have already been discussed. To make quantum-well structures absorb normal incidence light, special gratings or corrugations on the array surface must be fabricated; these reduce the light-coupling efficiency, and furthermore, add to the fabrication complexity of the focal plane arrays.

The successes in making normal-incidence, single-pixel, quantum-dot infrared detectors has motivated several groups to attempt making focal plane arrays. These successes are based on characteristics that arise due to three-dimensional confinement of carriers (Pan *et al.* 1998; Harrison 1999; Phillips *et al.* 2002). The properties of quantum dots that make them attractive for focal plane arrays include (i) sensitivity to normal-incidence illumination (Pal *et al.* 2003), (ii) potential ease of tunability of the peak operating wavelength, (iii) a predicted high responsivity due to a long electron lifetime in the excited states (Bockelmann *et al.* 1990; Boucaud *et al.* 2003) and (iv) a potential for high-temperature operation (Bhattacharya *et al.* 2005).

The first generation of quantum-dot imaging arrays were based on rather simple configurations. Stiff *et al.* for example, fabricated a  $13 \times 13$  InAs/GaAs quantum-dot detector array with which they were able to demonstrate raster scanning of images of heated objects (Stiff-Roberts *et al.* 2002). In another report, Jiang *et al.* used a modest size array of detectors fabricated from a InGaAs/InGaP quantum-dot sample to image a hot soldering iron (Jiang *et al.* 2004a). The quality of images obtained by these preliminary arrays was inferior to what is possible with commercial-quality quantum-well infrared photodetector focal plane array cameras. In more recent work, larger size focal plane arrays have been made and some improvement in image quality has been demonstrated (Krishna *et al.* 2005). The best imaging camera to date was fabricated from a sample that was comprised of InAs/InGaAs quantum dots embedded inside GaAs quantum-well structures (Gunapala *et al.* 2007). Fairly good images, comparable to those obtainable with focal plane cameras made from quantum-well structures, have been demonstrated. The camera was operated at  $60\text{ K}$ , and the noise equivalent temperature difference was about  $40\text{ mK}$ . However, a grating structure was used to enhance the absorption. In view of the fact that dots were embedded inside quantum-well structures, the intraband transitions involved were hybrid in kind. That is: an electron in a

bound state in a quantum dot made a transition to a bound or continuum state in a quantum-well film. The use of grating structures on the array surfaces negates the perceived advantage that the quantum-dot-based device is supposed to have over the old quantum-well-based devices. Under these circumstances, it is not clear what the benefits of using the dots are when the imaging performance of the hybrid structures is not any better than that of the old quantum-well-based infrared focal plane array cameras. An exception is perhaps when one can operate the quantum-dot-based focal plane array at high temperatures. High-temperature operation (up to 120 K) was recently reported in a focal plane array that was fabricated from a sample with dots-in-a-well structure. The sample was comprised of InAs dots embedded inside InGaAs wells that are in turn surrounded by InAlAs barriers. The structure was grown on an InP substrate (Tsao *et al.* 2007). The high-temperature operation was possible because the peak detection wavelength of the array was in the mid-wave infrared region, at about 4  $\mu\text{m}$ . All reported imaging arrays have so far been based on “dots-in-a-well” hybrid structure.

### 7.6.1 Quantum-dot uniformity requirements

The discussion in the previous section makes it clear that more development work is still required before high-quality imaging can be obtained from pure quantum-dot-based focal plane arrays. At a minimum, it is essential that the quantum-dot devices that form the focal plane arrays have uniform characteristics; this is in addition to having high responsivities and detectivities. Before fabrication of any focal plane array, it is generally useful to study the statistics of the characteristics of individual devices in order to assess their uniformity. This is particularly important in the case of quantum dots because of the rather random manner in which they are formed. The stochastic manner in which they are formed can often lead to unacceptably high variations in dot size, areal density, or mole fraction composition. This variability in basic structural properties affects device characteristics. Furthermore, the nature of the electron energy-level occupation rules can cause a non-uniform spatial carrier distribution in a quantum-dot detector, and hence in a focal plane array. Pal and Towe have studied characteristics of individual devices in a linear array of InAs/GaAs quantum-dot photodetectors and statistically analyzed the data in order to understand the distribution of the results (Pal and Towe 2006b).

The basic device structure used in the array was grown by solid-source molecular beam epitaxy. On top of a highly doped n-type GaAs substrate, a relatively thin GaAs buffer layer (0.75  $\mu\text{m}$ ) was grown at about 580 °C; this layer is doped with Si to achieve a free carrier density of about  $n = 2 \times 10^{18} \text{ cm}^{-3}$ . The buffer is then followed by a 50-nm layer of GaAs on top of which are grown fifty periods of InAs/GaAs quantum-dot layers. The fifty periods form the active region of the device and are grown at a substrate temperature of about 500 °C. The entire active region (which is composed of GaAs barriers and the InAs quantum-dot arrays) is uniformly doped with Si such that the resulting ionized electron density is about  $2 \times 10^{16} \text{ cm}^{-3}$ , which is expected

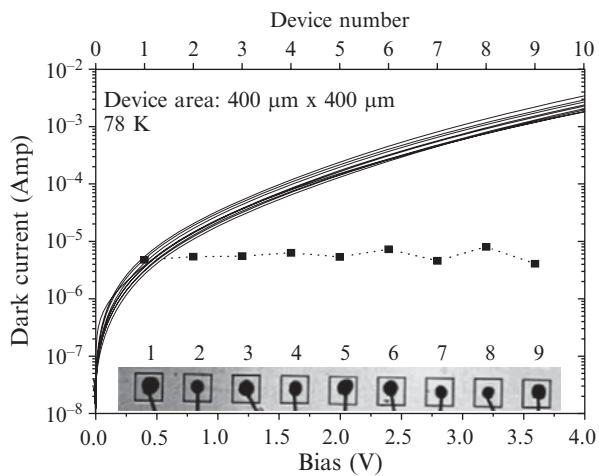
to yield 2 to 3 electrons per dot. After growth of the active region, the substrate temperature was raised to 580 °C. Following growth of the last GaAs barrier layer of the active region, an undoped 40-nm Al<sub>0.3</sub>Ga<sub>0.7</sub>As layer is grown; this serves as a dark-current-blocking layer. Finally, a 0.25-μm GaAs-layer doped with Si ( $n = 2 \times 10^{18} \text{ cm}^{-3}$ ) was grown as a top Ohmic contact. The entire device structure was grown while the substrate holder was rotated at 25 revolutions per minute to ensure uniformity.

From atomic force microscopy images of uncapped samples, the estimated average surface density of the quantum dots was about  $3 \times 10^{10} \text{ cm}^{-2}$ . Basic photoluminescence characterization showed that the dots were of high quality, emitting efficiently at room temperature (300 K) with a dominant luminescence peak located at 1.08 eV.

After the basic sample characterization, a linear array of nine devices was fabricated using conventional photolithography and wet-chemical etching. Each device was about 400 μm × 400 μm and the pitch of the array was about 700 μm. The length of the array was about 6 mm. This length could accommodate about 150 devices if the pitch of the array was reduced to 40 μm, and if each device in the array were to measure about 30 μm × 30 μm, which is a typical size for a pixel in some commercially available focal plane arrays. Characterization data from a single large-size device measuring 400 μm × 400 μm could be considered as the average of the data from 100 elements of a 10 × 10 array. After the lithographic process, top and bottom Ohmic contacts to the array were formed by annealing a composite layer of gold/germanium/gold: Au (20 nm)/Ge (60 nm)/Au (200 nm). The sample was then mounted in a leadless chip carrier. Each individual device in the array was bonded (by a gold wire) to pads on the chip carrier. For each device, the optically active, as well as the total device area were measured using a high-resolution optical microscope. Because of the imprecision of wet-chemical etching, the sizes of the devices in the array were not identical. It was therefore necessary to use the measured device sizes for the calculation of the parameters of the array.

The first set of electrical measurements consisted of dark current–voltage characteristics; these measurements were obtained by using a semiconductor parameter analyzer. Infrared spectral characteristics and responsivities of the devices were measured under normal-incidence illumination. These measurements were performed by using a monochromator and a calibrated pyroelectric detector. The noise characteristics of the devices were measured using a network signal analyzer.

We illustrate in Fig. 7.27, the low-temperature (78 K) dark  $I$ – $V$  characteristics of the nine devices in the array. An image of the array is also shown in the inset of the figure. Notice the tight clustering of the current–voltage characteristics for all nine devices. For the range of bias voltages used in the measurements, there is a fairly narrow spread in the dark current for all devices in the array. To examine this behavior further, we measured the dark-current level for each of the nine devices when the bias voltage across it was 0.5 V. The plot of dark-current level against device number—shown in the figure—remains relatively flat (see the square symbols in the figure). In fact, for the 0.5 V bias, the dark current varies between  $4.1 \times 10^{-6} \text{ A}$  and  $8.1 \times 10^{-6} \text{ A}$ ,



**Fig. 7.27** Current–voltage characteristics (in the dark) of a linear array of quantum-dot detectors. A photograph of the array investigated is shown in the inset. A plot of the dark current (shown with square symbols) at a fixed bias of 0.5 V is given as a function of the device location in the array (top axis). Figure taken from Pal *et al.* (2006b) *J. Appl. Phys.* **100**, 084322 with permission from AIP.

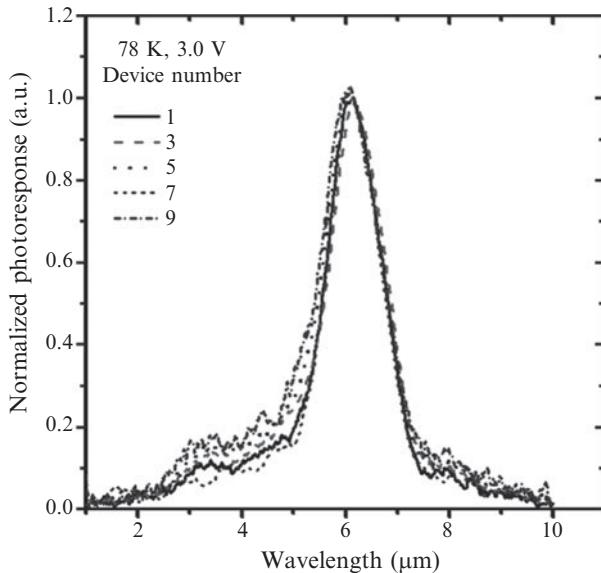
with a calculated average of about  $5.7 \times 10^{-6}$  A. For a scaled-down pixel size of about  $30\text{ }\mu\text{m} \times 30\text{ }\mu\text{m}$ , this current would only be about 32 nA at 78 K.

Perhaps a more enlightening measure of the dispersion of characteristics for the nine devices is the standard deviation—defined as the square root of the variance (Martin *et al.* 1971), which for the measured currents can be written as

$$\bar{I}_{\sigma}^2 = (1/N) \sum_{n=1}^N (I_n - \bar{I})^2, \quad (7.38)$$

where  $N$  is the number of devices,  $I_n$  the measured current for the  $n$ th device, and  $\bar{I}$  is the arithmetic mean of the measured currents. On this basis, the calculated standard deviation for the dark current is  $1.2\text{ }\mu\text{A}$ . Another group has also reported the individual dark current–voltage characteristics of photodetectors in a  $4 \times 4$  array at 300 K (Stiff-Roberts *et al.* 2002). The standard deviation for their particular devices at the nominal bias of  $\sim 0.1$  V was about  $3.1 \times 10^{-5}$  A, this is about 25 times higher than the deviation observed in our work (at 78 K). This statistical concept of variation, and the associated measure of the average: the mean, can be extended to analyze other key parameters of all detectors in an array. This has been done for a few more parameters in the linear array under discussion here.

The photoresponse spectra for all nine devices in the array were measured at 78 K. For clarity, the normalized spectra of only five of the devices are shown in Fig. 7.28. Note that the peak wavelength of each device is located at about  $5.6\text{ }\mu\text{m}$ . Furthermore, the spectra of the devices almost lie on top of one another, indicating the good uniformity of the quantum dots in the structure. Since the energy levels in the dots are dictated by the size of the dots, for intersublevel transitions to yield an identical absorption peak, the size distribution of the dots in each device of the array must be nearly the same. Except for the shoulder around  $3\text{ }\mu\text{m}$ , all the spectra show one photoresponse peak; the shoulder could be a result of bound-to-continuum transitions in a

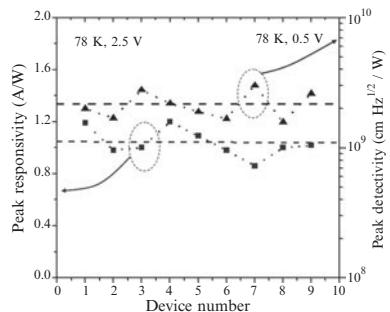


**Fig. 7.28** Low-temperature normalized photoresponse of five of the nine devices in the linear array whose current–voltage characteristics are shown in Fig. 7.27. These spectra were measured at 78 K. Figure taken from Pal *et al.* (2006b) *J. Appl. Phys.* **100**, 084322 with permission from AIP.

set of relatively large-size dots. The full width at half-maximum intensity ( $\Delta\lambda$ ) of the spectra is about  $1.3 \mu\text{m}$ , giving a ratio  $\Delta\lambda/\lambda \sim 23\%$ ; this value indicates that the absorption transitions are from energy levels in the dots to the continuum, as we have already discussed (Levine 1993).

The other key parameter of the array is the peak responsivity. For the individual devices in the array, we show in Fig. 7.29 (left axis), the low-temperature (78 K) peak responsivity. This quantity varies between  $0.86 \text{ A/W}$  and  $1.2 \text{ A/W}$ . Its average value is about  $1 \text{ A/W}$ , and is indicated by the dotted line in the figure. The calculated standard deviation of the responsivity is about  $0.10 \text{ A/W}$ . The deviation could be a result of a variation in the total number of the dots in a device or the number of carriers in each dot. It is not possible to achieve uniform doping across a large-size wafer. As a consequence, variations in the number of electrons supplied to a dot in a device will occur. Furthermore, the voltage dropped across a device or the electric field in a device in the array will vary because of variations in the dot density and number of electrons in a dot. This electric-field variation will lead to a variation in peak responsivity. The deviation in responsivity could probably be reduced so that when the array is used in imaging, the performance across the array is uniform.

The detectivity, which was defined earlier as a measure of the signal-to-noise ratio, primarily depends on responsivity, the noise current, the noise-measurement bandwidth, and the optically active area of the device. The peak detectivity for each device (1 to 9) in the array was measured at low temperature (78 K). This measurement was carried out when the applied bias voltage across each device was about 0.5 V. The results of the measurement are shown in Fig. 7.29 (right axis) for each device (1 to 9). The detectivity varies between  $1.6 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$  and  $3 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ , with the average being about  $2.2 \times 10^9 \text{ cm Hz}^{1/2}/\text{W}$ . The calculated standard deviation is about  $5.0 \times 10^8 \text{ cm Hz}^{1/2}/\text{W}$ . As discussed, the spread in the important



**Fig. 7.29** Peak responsivity (left axis) and detectivity (right axis) for the linear array of detectors as a function of device location in the array at 78 K. Figure taken from Pal *et al.* (2006b) *J. Appl. Phys.* **100**, 084322 with permission from AIP.

device parameters of an array should be minimized in order to achieve good uniformity across the array. For detectivity in particular, this means a tight control on the responsivity and the dark current.

The variation in device characteristics primarily depends on the variation of quantum-dot density and the number of dopant atoms in the structure. In order to understand the variation of dot density, we examined several atomic force microscope images of scan areas from different regions of the samples. For two areal scan sizes of  $2\text{ }\mu\text{m} \times 2\text{ }\mu\text{m}$  and  $3\text{ }\mu\text{m} \times 3\text{ }\mu\text{m}$ , the standard deviations of dot densities were found to be 2.3% and 3%, respectively. Even though the deviations appear to be consistent with the fairly narrow distribution range of the characteristics we have discussed, one should not extrapolate the results to typical pixel sizes of  $30\text{ }\mu\text{m} \times 30\text{ }\mu\text{m}$  in a real array. It would be better to study characteristics of individual pixels in an array to understand the effect of the variations. These could then be correlated with measurements of dot density over fairly large scan areas (on the order of a pixel), and the dopant areal density. The insight gained in such studies would be a better guide for high-performance focal plane array development.

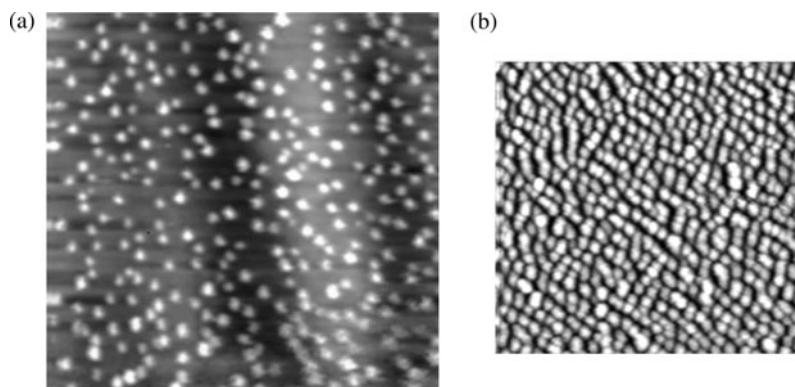
## 7.7 Challenges and prospects for high-performance detectors and arrays

Quantum-dot nanostructures have many potential applications that include—besides focal plane array imaging—new light sources (such as single-photon lasers, polariton lasers, light-emitting diodes for high-efficiency lighting, plasmonic sources), quantum-dot solar-electric conversion devices, data-storage devices, and spectral tagging devices, as well as devices for quantum computing (Ledentsov *et al.* 1998; Maximov *et al.* 2000; Sellin *et al.* 2001; Toth *et al.* 2001). The major challenge in creating high-quality quantum-dot based devices is rooted in the synthesis process by which the dots are made. Epitaxial synthesis of dots is based on a strain mechanism during what is essentially a self-assembly process that researchers have very little control over. This means it is difficult to control the shape, size, composition, and charge carrier density per dot. Since being able to supply a specific number of electrons per dot is critical, this becomes one of the key impediments to creating structures with reproducible properties. This problem is exacerbated by the need to maintain a uniform and precise distribution of dopant impurities (hence free charge carriers) over an entire 3- or 4-inch-diameter wafer during the synthesis process. The difficulties in synthesis arise because of the lack of precise control over the many kinetic, thermodynamic, as well as chemical parameters on which strain depends. Ultimately, these difficulties translate into electronic and optical properties that are difficult to predict or model theoretically (Nishi *et al.* 1994; Jain *et al.* 1996; Spencer *et al.* 1997; Tersoff *et al.* 1998). To illustrate, consider that the detection and emission wavelengths of a quantum-dot structure depend on the location of the quantized energy levels within a confining potential; this depends on the shape, size, and chemical composition of the dot. In a study designed to understand this dependency, Pal *et al.* (2002)

investigated the impact of dot-size variation on absorption and emission characteristics of (In,Ga)As quantum dots. The study only amplified the magnitude of the task at hand. Although many researchers have developed theoretical models that are supposed to shed light on the properties of dots, none of these models can satisfactorily predict or explain the dependency of measured properties on the shape, size, and composition of the dots (Priester *et al.* 1995; Daruka *et al.* 1997a,b; Chen *et al.* 1997b; Venezuela *et al.* 1998; Lang *et al.* 2003).

To further illustrate the difficulties, we discuss some results from our own work. Our experience so far has been that there exist some non-trivial differences in results from different groups that seemingly use similar or in some cases identical growth conditions. The shape, size, and composition of dots are very sensitive to growth temperature, number of monolayers deposited, composition of the dot material, composition of the cap layer, as well as any layer beneath the dot. Furthermore, it is not easy to reproduce synthesis conditions used by other groups on identical pieces of equipment; sometimes even the same group is unable to reproduce its own results!

As an example, we consider the growth of binary (InAs) and ternary (InGaAs) quantum dots on a GaAs substrate by the method of molecular beam epitaxy. The two experiments were carried out at exactly the same substrate temperature. After growth of the samples, we performed atomic force microscopy on them. The topographic images of the samples are shown in Figs. 7.30(a) and (b); the first image is a micrograph of InAs dots grown on top of GaAs. The scan area is about  $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m}$  and the dot density is around  $3 \times 10^{10}\text{ cm}^{-2}$ . The second image in Fig. 7.30(b), shows  $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$  dots on a GaAs substrate grown at exactly the same substrate temperature as that used during the growth of the InAs dots. The density of the  $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$  is about  $6\text{--}8 \times 10^{10}\text{ cm}^{-2}$ . These images show the contrast in the binary and ternary dots. The population of the binary dots is sparse, while that of the ternary dots is denser. The ternary dots are also larger. Transmission electron microscopy studies indicate that the InAs dots tend to have pyramidal shapes (as shown in Fig. 7.8(a)). On the other hand, ternary or quaternary dots are usually lens-shaped (see Fig. 7.8(b)).



**Fig. 7.30** An atomic force microscope image of (a) InAs quantum dots in a scan area of  $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m}$ , and (b) a similar image for  $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$  quantum dots in a scan area of the same size.

### 7.7.1 The promise of quantum-dot nanostructures for infrared detection

While the discussion in the last section may have appeared to overemphasize the challenges that have yet to be solved in the synthesis, theoretical modelling, and simulation of dots, we should point out that the advantages and benefits of using quantum dots in devices have been abundantly demonstrated. It is therefore only a matter of time before quantum dots are widely used in practical devices—particularly optoelectronic devices.

For infrared detection, the three-dimensional quantum-confinement of carriers in dots embedded in active regions of devices endows the devices with advantages absent in those based on conventional bulk, band-to-band structures, two-dimensional quantum-well structures, or the emergent type-II superlattice structures. Some of these advantages include (i) absorption of normal incidence light, (ii) reduction of thermally generated carriers because of quantum-mechanical restrictions originating from the nature of electron–phonon coupling in such structures, and (iii) long excited-state carrier lifetimes. In spite of the challenges already discussed, use of quantum dots has resulted in the demonstration of robust mid-wave and long-wave infrared detectors. The natural extension of the single-pixel devices into arrays has already shown that respectable focal plane array cameras can be made. In fact, some of the arrays have shown potential for high-temperature operation ( $> 150\text{ K}$ ). The ultimate promise of quantum-dot-based detection is infrared cameras operating at or close to room temperature.

## Acknowledgments

Portions of this work were performed under the sponsorship of the U.S. Army Research Office, the U.S. Army Research Laboratory, and the Defense Advanced Research Projects Agency. We gratefully acknowledge the sustained support of the following past and present program officers during the course of the work: Dr. Mitra Dutta (ARO), Dr. William Clark (ARO), Dr. Dan Radack (DARPA), and Mr. Ray Balcerak (DARPA). We are also grateful to Steve Kennerly (at the Army Research Laboratory) for sharing his vast experience in infrared measurements and for his helpful collaboration. Finally, we acknowledge contributions to this work by the following past graduate students and postdoctoral research associates: Dong Pan, Liang Chen, V. Gabriella Storelu, and Dmitry Firsov.

## References

- Arakawa, Y., Sakaki, H., *Appl. Phys. Lett.* **40**, 939–941 (1982).
- Aslan, B., Liu, H.C., Korkusinski, M., Cheng, S.J., Hawrylak, P. *Appl. Phys. Lett.* **82**, 630–632 (2003).
- Benisty, H., Sotomayortorres, C.M., Weisbuch, C. *Phys. Rev., B* **44**, 10945–10948 (1991).

- Bhattacharya, P., Su, X.H., Chakrabarti, S., Ariyawansa, G., Perera, A.G.U. *Appl. Phys. Lett.* **86**, 191106 (2005).
- Bockelmann, U., Bastard, G. *Phys. Rev. B* **42**, 8947–8951 (1990).
- Boucaud, P., Sauvage, S. *Comptes Rendus Physique* **4**, 1133–1154 (2003).
- Chakrabarti, S., Stiff-Roberts, A.D., Bhattacharya, P., Gunapala, S., Bandara, S., Rafol, S.B., Kennerly, S.W. *IEEE Photon. Technol. Lett.* **16**, 1361–1363 (2004a).
- Chakrabarti, S., Stiff-Roberts, A.D., Bhattacharya, P.B., Kennerly, S.W. *Electronics Lett.* **40**, 197–198 (2004b).
- Chen, C.J., Choi, K.K., Chang, W.H., Tsui, D.C. *Appl. Phys. Lett.* **71**, 3045–3047 (1997a).
- Chen, K.M., Jesson, D.E., Pennycook, S.J., Thundat, T., Warmack, R.J. *Phys. Rev. B* **56**, R1700–R1703 (1997b).
- Chen, S.D., Chen, Y.Y., Lee, S.C. *Appl. Phys. Lett.* **86**, (2005).
- Chen, Z.H., Baklenov, O., Kim, E.T., Mukhametzhanov, I., Tie, J., Madhukar, A., Ye, Z., Campbell, J.C. *Infr. Phys. Technol.* **42**, 479–484 (2001).
- Chen, Z.H., Kim, E.T., Madhukar, A. *Appl. Phys. Lett.* **80**, 2490–2492 (2002).
- Choi, K.K., Lin, C.H., Leung, K.M., Tamir, T., Mao, J., Tsui, D. C., Jhabvala, M. *Infr. Phys. Technol.* **44**, 309–324 (2003).
- Choi, K.K., Monroy, C., Goldberg, A., Dang, G., Jhabvala, M., La, A., Tamir, T., Leung, K.M., Majumdar, A., Li, J.J., Tsui, D.C. *Infr. Phys. Technol.* **47**, 76–90 (2005).
- Chou, S.T., Tsai, C.H., Wu, M.C., Lin, S.Y., Chi, J.Y. *IEEE Photon. Technol. Lett.* **17**, 2409–2411 (2005).
- Chu, L., Zrenner, A., Bichler, M., Abstreiter, G. *Appl. Phys. Lett.* **79**, 2249–2251 (2001).
- Chua, S.J., Xu, S.J., Zhang, X.H., Wang, X.C., Mei, T., Fan, W. J., Wang, C.H., Jiang, J., Xie, X.G. *Appl. Phys. Lett.* **73**, 1997–1999 (1998).
- Coon, D.D., Karunasiri, R.P.G. *Appl. Phys. Lett.* **45**, 649–651 (1984).
- Daruka, I., Barabasi, A.L. *Phys. Rev. Lett.* **79**, 3708–3711 (1997a).
- Daruka, I., Barabasi, A.L. *Phys. Rev. Lett.* **78**, 3027–3027 (1997b).
- Destefanis, G., Audebert, P., Mottin, E., Rambaud, P. *J. Crys. Growth* **185**, 1288–1292 (1998).
- Gordon, N.T., Lees, D.J., Bowen, G., Phillips, T.S., Haigh, M., Jones, C.L., Maxey, C.D., Hipwood, L., Catchpole, R.A. *J. Electro. Mater.* **35**, 1140–1144 (2006).
- Gunapala, S.D., Bandara, S.V., Hill, C.J., Ting, D.Z., Liu, J.K., Rafol, S.B., Blazejewski, E.R., Mumolo, J.M., Keo, S.A., Krishna, S., Chang, Y.C., Shott, C.A. *J. IEEE Quantum Electron.* **43**, 230–237 (2007).
- Gunapala, S.D., Bandara, S.V., Liu, J.K., Rafol, S.B., Mumolo, J.M. *IEEE Transactions on Electron Dev.* **50**, 2353–2360 (2003).
- Gunapala, S.D., Liu, J.K., Park, J.S., Sundaram, M., Shott, C.A., Hoelter, T., Lin, T.L., Massie, S.T., Maker, P.D., Muller, R. E., Sarusi, G. *IEEE Transactions on Electron Dev.* **44**, 51–57 (1997).
- Harrison, P. *Quantum Wells, Wires, and Dots* (Wiley, Chichester, 1999).
- Inoshita, T., Sakaki, H. *Phys. Rev. B* **56**, R4355–R4358 (1997).
- Jain, S.C., Willander, M., Maes, H. *Semicond. Sci. Technol.* **11**, 641–671 (1996).
- Jiang, J., Mi, K., Tsao, S., Zhang, W., Lim, H., O’Sullivan, T., Sills, T., Razeghi, M., Brown, G.J., Tidrow, M.Z. *Appl. Phys. Lett.* **84**, 2232–2234 (2004a).

- Jiang, J., Tsao, S., O'Sullivan, T., Zhang, W., Lim, H., Sills, T., Mi, K., Razeghi, M., Brown, G.J., Tidrow, M.Z. *Appl. Phys. Lett.* **84**, 2166–2168 (2004b).
- Jones, R.C. *Proc. IRIS* **2**, 9–122 (1957).
- Kim, E.T., Chen, Z.H., Madhukar, A. *Appl. Phys. Lett.* **79**, 3341–3343 (2001).
- Kim, E.T., Madhukar, A., Ye, Z.M., Campbell, J.C. *Appl. Phys. Lett.* **84**, 3277–3279 (2004).
- Kim, S., Mohseni, H., Erdtmann, M., Michel, E., Jelen, C., Razeghi, M. *Appl. Phys. Lett.* **73**, 963–965 (1998).
- Kim, S.M., Harris, J.S. *IEEE Photon. Technol. Lett.* **16**, 2538–2540 (2004).
- Krishna, S., Forman, D., Annamalai, S., Dowd, P., Varangis, P., Tumolillo, T., Gray, A., Zilko, J., Sun, K., Liu, M.G., Campbell, J., Carothers, D. *Appl. Phys. Lett.* **86**, (2005).
- Krishna, S., Raghavan, S., von Winckel, G., Rotella, P., Stintz, A., Morath, C.P., Le, D., Kennerly, S.W. *Appl. Phys. Lett.* **82**, 2574–2576 (2003).
- Lang, C. *Mater. Sci. Technol.* **19**, 411–421 (2003).
- Ledentsov, N.N., Ustinov, V.M., Shchukin, V.A., Kop'ev, P.S., Alferov, Z.I., Bimberg, D. *Semicond.* **32**, 343–365 (1998).
- Leonard, D., Krishnamurthy, M., Reaves, C.M., Denbaars, S.P., Petroff, P.M. *Appl. Phys. Lett.* **63**, 3203–3205 (1993).
- Levine, B.F. *J. Appl. Phys.* **74**, R1–R81 (1993).
- Levine, B.F., Zussman, A., Kuo, J.M., Dejong, J. *J. Appl. Phys.* **71**, 5130–5135 (1992).
- Lim, H., Zhang, W., Tsao, S., Sills, T., Szafraniec, J., Mi, K., Movaghar, B., Razeghi, M. *Phys. Rev. B* **72**, (2005).
- Liu, H.C. *J. Appl. Phys.* **73**, 3062–3067 (1993).
- Martin, B.R. *Statistics for Physicists* (Academic Press, London, 1971).
- Maximov, M.V., Ledentsov, N.N., Ustinov, V.M., Alferov, Z.I., Bimberg, D. *J. Electro. Mater.* **29**, 476–486 (2000).
- Miller, D.A.B., Chemla, D.S., Damen, T.C., Gossard, A.C., Wiegmann, W., Wood, T.H., Burrus, C.A. *Phys. Rev. Lett.* **53**, 2173–2176 (1984).
- Mimura, T. *Jap. J. Appl. Phys. Part I-Regular Papers Brief Communications & Review Papers* **44**, 8263–8268 (2005).
- Nishi, K., Yamaguchi, A.A., Ahopelto, J., Usui, A., Sakaki, H. *J. Appl. Phys.* **76**, 7437–7445 (1994).
- Pal, D., Towe, E., Chen, S. *Appl. Phys. Lett.* **78**, 4133 (2001).
- Pal, D., Chen, L., Towe, E. *Appl. Phys. Lett.* **83**, 4634–4636 (2003).
- Pal, D., Stoleru, V.G., Towe, E., Firsov, D. *Jap. J. Appl. Phys. Part I-Regular Papers Short Notes & Review Papers* **41**, 482–489 (2002).
- Pal, D., Towe, E. *Appl. Phys. Lett.* **88**, 153109 (2006a).
- Pal, D., Towe, E. *J. Appl. Phys.* **100**, 084322 (2006b).
- Pan, D., Towe, E. *Electron. Lett.* **34**, 1883–1884 (1998).
- Pan, D., Towe, E., Kennerly, S. *Appl. Phys. Lett.* **73**, 1937–1939 (1998).
- Parker, M.A. *Phys. Optoelectron* (CRC Press, Boca Raton, 2005).
- Petroff, P.M., Denbaars, S.P. *Superlattices and Microstructures* **15**, 15–21 (1994).
- Phillips, J. *J. Appl. Phys.* **91**, 4590–4594 (2002).

- Piotrowski, A., Madejczyk, P., Gawron, W., Klos, K., Pawluczyk, J., Rutkowski, J., Piotrowski, J., Rogalski, A. *Infr. Phys. Technol.* **49**, 173–182 (2007).
- Priester, C., Lannoo, M. *Phys. Rev. Lett.* **75**, 93–96 (1995).
- Raghavan, S., Rotella, P., Stintz, A., Fuchs, B., Krishna, S., Morath, C., Cardimona, D.A., Kennerly, S.W. *Appl. Phys. Lett.* **81**, 1369–1371 (2002).
- Sai-Halasz, H.G., Tsu, T., Esaki, L. *Appl. Phys. Lett.* **30**, 651–653 (1977).
- Sakaki, H., Arakawa, Y., Nishioka, M., Yoshino, J., Okamoto, H., Miura, N. *Appl. Phys. Lett.* **46**, 83–85 (1985).
- Sellin, R.L., Ribbat, C., Grundmann, M., Ledentsov, N.N., Bimberg, D. *Appl. Phys. Lett.* **78**, 1207–1209 (2001).
- Spencer, B.J., Tersoff, J. *Phys. Rev. Lett.* **79**, 4858–4861 (1997).
- Stiff, A.D., Krishna, S., Bhattacharya, P., Kennerly, S.W. *J. IEEE Quantum Electron.* **37**, 1412–1419 (2001).
- Stiff-Roberts, A.D., Chakrabarti, S., Pradhan, S., Kochman, B., Bhattacharya, P. *Appl. Phys. Lett.* **80**, 3265–3267 (2002).
- Stranski, I.N., Krastanow, L. *Sitzungsberichte, d. Akad. D. Wissenschaften in Wein, Abt. IIb*, **146**, 797–810 (1937).
- Tang, S.F., Lin, S.Y., Lee, S.C. *Appl. Phys. Lett.* **78**, 2428–2430 (2001).
- Tersoff, J. *Phys. Rev. Lett.* **81**, 3183–3186 (1998).
- Toth, G., Lent, C.S. *Phys. Rev. A* **63**, (2001).
- Towe, E., Pan, D. *IEEE Journal of Selected Topics in Quantum Electronics* **6**, 408–421 (2000).
- Tribolet, P., Chatard, J.P., Costa, P., Manissadjian, A. *J. Crys. Growth* **185**, 1262–1271 (1998).
- Tsao, S., Lim, H., Zhang, W., Razeghi, M. *Appl. Phys. Lett.* **90**, (2007).
- Venezuela, P., Tersoff, J. *Phys. Rev. B* **58**, 10871 LP - 10874 (1998).
- Weber, A., Gauthier-Lafaye, O., Julien, F.H., Brault, J., Gendry, M., Desieres, Y., Benyattou, T. *Appl. Phys. Lett.* **74**, 413–415 (1999).
- Weisbuch, C., Vinter, B. *Quantum Semiconductor Structures* (Harcourt, San Diego, 1991).
- West, L.C., Eglash, S.J. *Appl. Phys. Lett.* **46**, 1156–1158 (1985).
- Xu, S.J., Chua, S.J., Mei, T., Wang, X.C., Zhang, X.H., Karunasiri, G., Fan, W.J., Wang, C.H., Jiang, J., Wang, S., Xie, X.G. *Appl. Phys. Lett.* **73**, 3153–3155 (1998).
- Yariv, A., Yeh, P. *Optical Electronics in Modern Communications* (Oxford University Press, Oxford, 2007).
- Ye, Z.M., Campbell, J.C., Chen, Z.H., Kim, E.T., Madhukar, A. *J. IEEE Quantum Electron.* **38**, 1234–1237 (2002a).
- Ye, Z.M., Campbell, J.C., Chen, Z.H., Kim, E.T., Madhukar, A. *J. Appl. Phys.* **92**, 4141–4143 (2002b).
- Ye, Z.M., Campbell, J.C., Chen, Z.H., Kim, E.T., Madhukar, A. *Appl. Phys. Lett.* **83**, 1234–1236 (2003).

# 8

# Nanoionics and its device applications

*T. Hasegawa, K. Terabe, T. Sakamoto, and M. Aono*

8.1 Introduction	294
8.2 Materials	295
8.3 Solid electrochemical reaction	296
8.4 Fundamentals of an atomic switch	299
8.5 New types of atomic switches	301
8.6 Applications of atomic switches	306
8.7 Summary and conclusion	309
References	310

## 8.1 Introduction

The study of ionic conductivity has a long history. More than 400 years ago, there was a report that silver whiskers grow from  $\text{Ag}_2\text{S}$ , which is one of the key materials mentioned in this chapter (Ercker 1574). A systematic study of ionic conductivity by Faraday in the early nineteenth century established the thermodynamical theory of ionic conduction as a key characteristic of solid electrolytes (Faraday 1833). Many extensive studies carried out subsequently paved the way for the application of ionic conductive materials to devices such as physicochemical sensors and batteries. Electronic devices using a solid electrolyte, such as a two-terminal switch, were developed more than 30 years ago (Hirose *et al.* 1976). However, ionic conductivity has not been used so much in the field of electronic devices, especially for commercial ones. This must be because the semiconductor transistor provides almost all the functions required for various applications in today's sophisticated information-oriented society.

The performance of the semiconductor transistor can be improved simply by downsizing its device structure, and the development of miniaturization techniques has led to the dramatic progress of the semiconductor transistor. However, due to the minimum feature size that comes into the picture at the nanoscale, we will face the physical limit of the semiconductor transistor within a few decades. This opens up the possibilities for new nanodevices to be used as actual components in commercial applications, which has markedly accelerated the research in nanoelectronics.

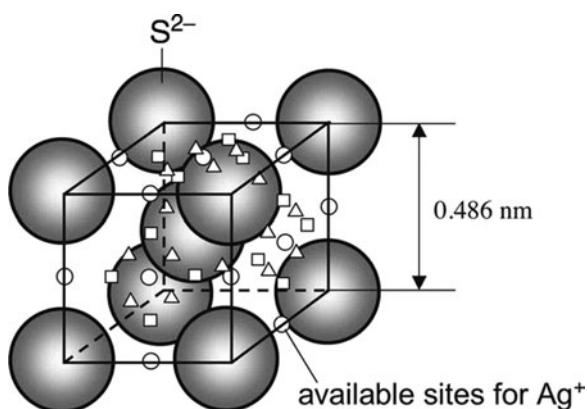
The concept of the atomic switch was introduced by Eigler *et al.* in 1991. They succeeded in transferring a Xe atom between the apex of a tip and a sample surface by using a scanning tunnelling microscope (STM). Depending on the adsorption site of the Xe atom, the tunnelling resistance of the system changed by an order of magnitude. Consequently, the system worked as a resistive switch using atomic motion. This chapter introduces a new type of nanoionics-based atomic switch with a brief description of key materials, fundamental principles, and applications.

## 8.2 Materials

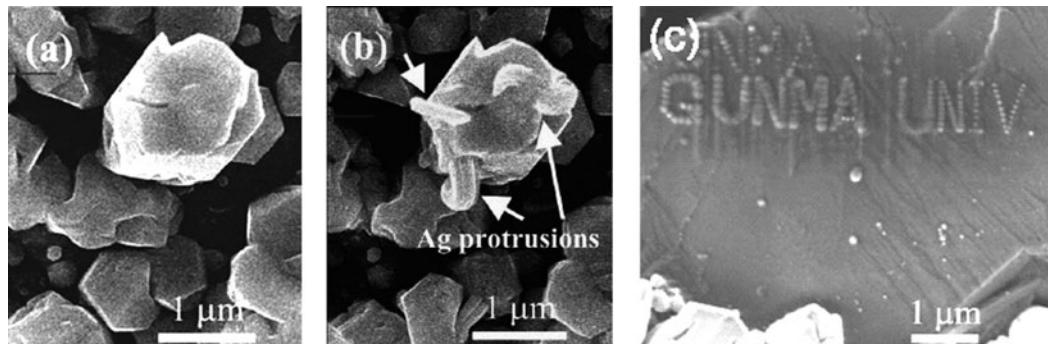
Electrical conduction is a phenomenon caused by a flow of charged particles, such as electrons, holes, and ions. In metals, the conductivity is due to the flow of electrons. On the other hand, the conductivity of ionic materials is due to the flow of ions. The ionic conductive materials are classified into two categories in terms of the charged particles: solid electrolytes in which only ions contribute to the current flow and mixed electronic and ionic conductors in which both electrons and ions contribute to the current flow (Kudo *et al.* 1990). Examples of solid electrolytes are  $ZrO_2$ ,  $HfO_2$ ,  $CeO_2$ ,  $AgI$ , and  $CuI$  and examples of mixed conductors are  $Ag_2S$ ,  $Cu_2S$ , and  $Ag_2Se$ . Both types of ionic materials can be used to configure an atomic switch. Ionic conductive materials with very low conductivity, such as  $Ta_2O_5$ , which is not usually regarded as an ionic conductive material, can also be used for making an atomic switch.

The unit cell of  $Ag_2S$  crystal, which was used as the key material for making the first atomic switch (Terabe *et al.* 2001) is shown in Fig. 8.1. In a rigid structure formed by sulfur anions ( $S^{2-}$ ), silver cations ( $Ag^+$ ) can migrate by moving among the adsorption sites. Ionic conductivity due to  $Ag^+$  diffusion has been reported to be of the order of  $\Omega^{-1} cm^{-1}$  at elevated temperatures such as  $200^\circ C$  (Rickert 1982), where  $Ag_2S$  crystal exists in a high-temperature form ( $\beta$ - $Ag_2S$ ). Below the phase-transition temperature ( $179^\circ C$ ),  $Ag_2S$  crystal exists in a low-temperature phase with lower ionic conductivity.

Electrically,  $Ag_2S$  is an n-type semiconductor, and it is correctly denoted  $Ag_{2+\delta}S$ . Increasing the temperature causes an increase in excess  $Ag^+$  cations, namely  $\delta$  of  $Ag_{2+\delta}S$ , in the equilibrium condition. The increase in  $\delta$  makes both the electronic and ionic conductivities higher. On the other hand,  $Cu_{2-\delta}S$ , which was also used to make an atomic switch (Sakamoto *et al.* 2003), is a p-type semiconductor. Therefore, a decrease in the number of  $Cu^+$  cations increases both the ionic and electronic conductivities. From the viewpoint of applications, being able to control the conductivity is useful. This is because the operating parameters, such as the threshold bias voltage (switching bias voltage) and switching speed, are determined by the ionic conductivity,



**Fig. 8.1** Structure of  $Ag_{2+\delta}S$  crystal.  $Ag^+$  cations migrate in a rigid lattice formed by  $S^{2-}$  anions by moving among the available sites.



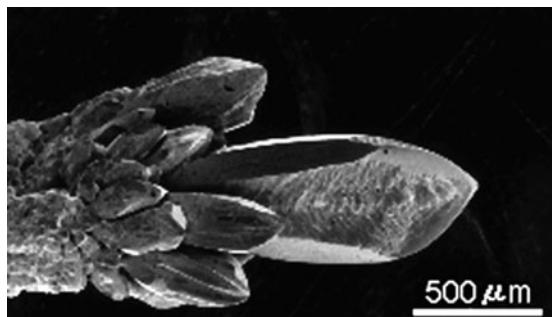
**Fig. 8.2** SEM images of  $\text{Ag}_{2+\delta}\text{S}$  crystal. (a) in initial electron-beam (EB) irradiation, (b) Ag nanowires formation after long EB irradiation, and (c) patterned array of Ag nanodots on the surface of  $\text{Ag}_{2+\delta}\text{S}$  crystal (Sone *et al.* 2006. Copyright Elsevier.)

i.e.  $\delta$ , which can be controlled by changing the film growth conditions. The device parameters can also be controlled by using different kinds of ionic conductive materials, as described later.

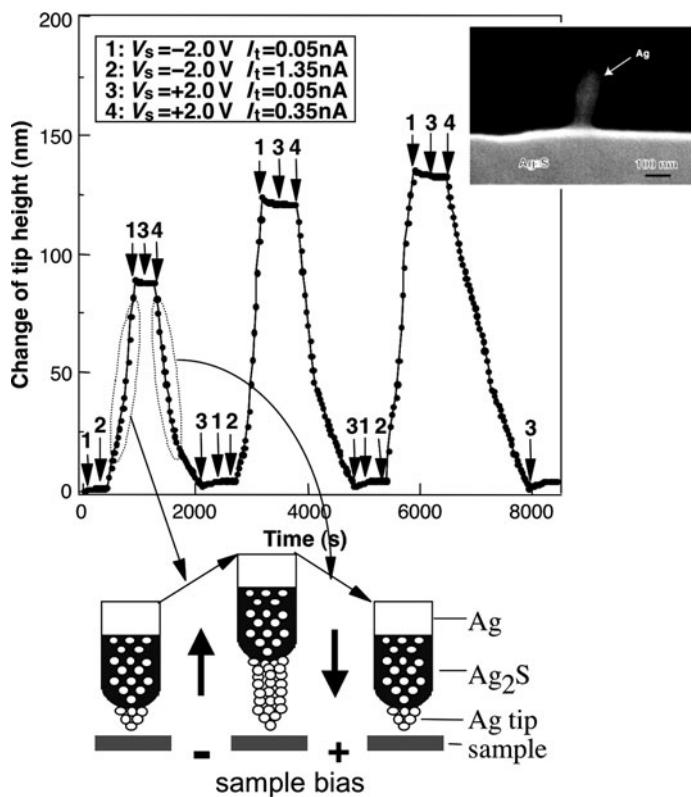
### 8.3 Solid electrochemical reaction

The key phenomenon for atomic switch operation is metal filament formation and annihilation using a solid electrochemical reaction. This section reviews the controlled growth and annihilation of metal filaments by scanning tunnelling microscopy (STM). Injecting electrons into a  $\text{Ag}_2\text{S}$  crystal causes the reduction of  $\text{Ag}^+$  cations ( $\text{Ag}_{(\text{Ag}_2\text{S})}^+ + \text{e}^- \rightarrow \text{Ag}_{(\text{cluster})}$ ), resulting in the precipitation of Ag atoms at the surface of the  $\text{Ag}_2\text{S}$  crystal. For instance, electron-beam irradiation in a scanning electron microscope (SEM) makes Ag nanowires grow on a  $\text{Ag}_2\text{S}$  surface, as shown in Fig. 8.2 (Sone *et al.* 2006). By controlling the position and irradiation time of the electron beam, one can form Ag nanodots in a patterned structure, as shown in Fig. 8.2c. The smallest dot size (diameter) of 25 nm was achieved in an experiment corresponds to an ultrahigh density of about 600 Gb/in<sup>2</sup> in a storage device with a dot pitch of 30 nm.

Although SEM can only inject electrons into a  $\text{Ag}_2\text{S}$  crystal, STM can both inject electrons into and extract electrons from the crystal if we change the polarity of the bias voltage applied between the tip and sample. A needle-like single crystal of  $\text{Ag}_2\text{S}$  was made by sulfurising the Ag wire with sulfur vapor in an evacuated glass ampoule. When the direction of sulfur vapor flow was controlled, a tip of  $\text{Ag}_2\text{S}$  was formed at the end of the Ag wire (Fig. 8.3). When a  $\text{Ag}_2\text{S}$  tip was used, the growth and shrinkage of a Ag protrusion was controlled by STM (Terabe *et al.* 2002); namely, both the reduction and oxidation processes of Ag cations/atoms ( $\text{Ag}_{(\text{Ag}_2\text{S})}^+ + \text{e}^- \leftrightarrow \text{Ag}$ ) can be induced by changing the direction of the tunnelling current. An example of controlled growth and shrinkage of a Ag protrusion is shown in Fig. 8.4. In this experiment, the bias voltage ( $V_s$ ) and tunnelling current ( $I_t$ ) were changed among four conditions. The Ag protrusion grew at  $V_s = -2.0\text{ V}$  and  $I_t = 1.35\text{ nA}$  (condition 2) and shrank at  $V_s = +2.0\text{ V}$  and  $I_t = 0.35\text{ nA}$



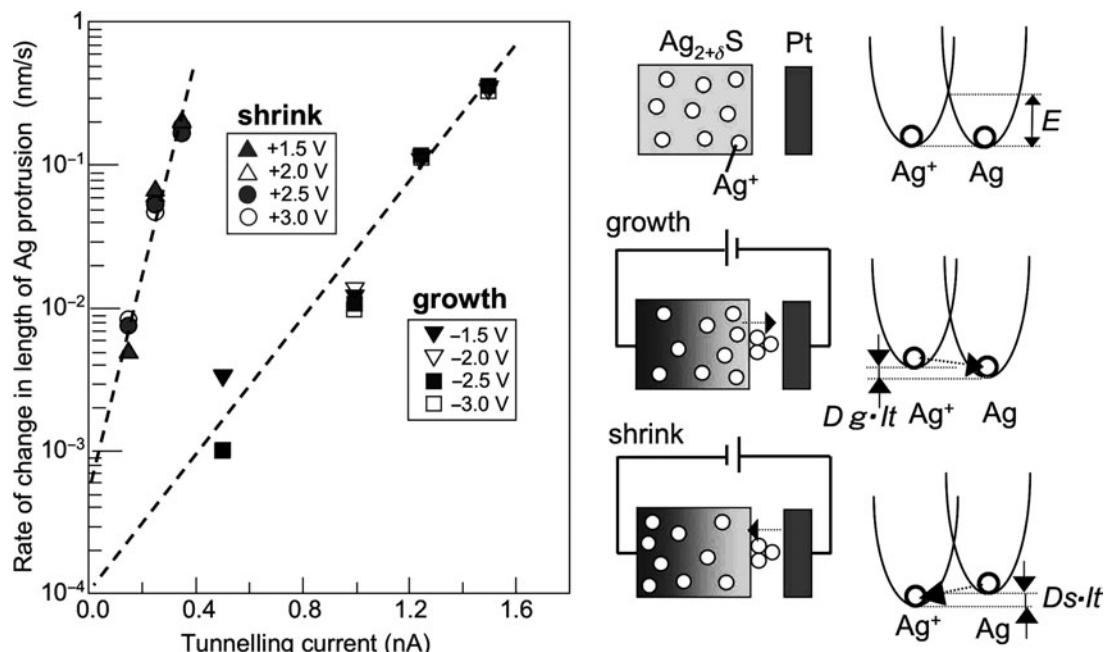
**Fig. 8.3** SEM image of needle-like  $\text{Ag}_{2+\delta}\text{S}$  crystal grown on Ag wire.



**Fig. 8.4** Controlled growth and shrinkage of Ag protrusion on the surface of  $\text{Ag}_{2+\delta}\text{S}$  crystal using STM. Inset is an SEM image of the Ag protrusion grown in the experiment. (Terabe *et al.* 2002. Copyright AIP.)

(condition 3). Nothing was observed to happen in conditions 1 and 4. This means that the growth and shrinkage had a threshold bias voltage and current. Hence, the state of the switch, i.e. ON or OFF, can be read without changing the status of the device by using a bias voltage that is lower than the threshold bias voltage. This characteristic is also the origin of the non-volatility of the atomic switch.

The rates of growth and shrinkage of a Ag protrusion with respect to the bias voltage and tunnelling current of the STM are shown in Fig. 8.5. The rates did not depend on the bias voltage, but did depend on the tunnelling current. They became larger exponentially as the tunnelling current ( $I_t$ ) increased, as

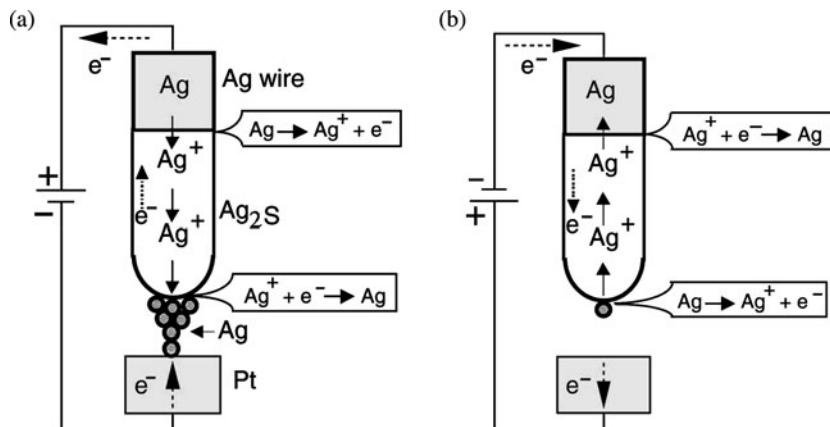


**Fig. 8.5** Growth and shrinkage speed of Ag protrusion controlled by STM. (Terabe *et al.* 2002. Copyright AIP.)

expressed by

$$\frac{dl}{dt} = A \exp\left(\frac{-E + DIt}{k_B T}\right),$$

where  $E$  is the activation energy of the reduction and oxidation of a Ag atom,  $k_B$  is the Boltzmann constant,  $T$  is temperature, and  $A$  and  $D$  are coefficients. This relation can be explained by Tafel's theory (Tafel 1905). In the experiment, one Ag atom precipitated with  $10^5$  electrons. Differences in the slope of the growth and shrinkage speeds gave different coefficients, i.e.  $D_s$  (for shrinkage) and  $D_g$  (for growth), which means that the electrochemical potential changes are different between shrinkage and growth. This difference was caused mainly by changes in the concentration of Ag<sup>+</sup> cations at the surface region, as shown in Fig. 8.5. The Ag substrate, on which the Ag<sub>2+δ</sub>S electrode was formed, not only made the electrochemical potentials of Ag<sup>+</sup> in the Ag<sub>2+δ</sub>S and Ag on the surface equal, but also made them equal at the interface between the Ag substrate and the Ag<sub>2+δ</sub>S electrode (top right in Fig. 8.5). When a bias voltage was applied, the electrochemical potential changed according to the polarity of the applied bias voltage. Namely, the concentration increased in the growth mode and decreased in the shrinkage mode. Since Ag<sub>2+δ</sub>S is an n-type semiconductor, increasing the excess Ag<sup>+</sup> cations is energetically easier than decreasing them.

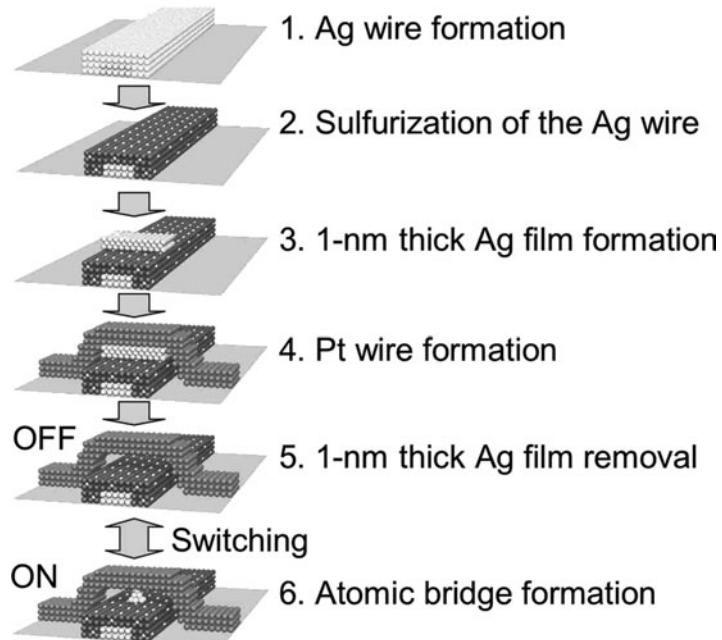


**Fig. 8.6** Schematic diagram of operating mechanism of atomic switch. (a) Turning on process and (b) turning off process.

## 8.4 Fundamentals of an atomic switch

The atomic switch consists of an electrode made of ionic and electronic mixed conductors and a counterelectrode made of metal arranged with a spacing of about 1 nm between them (Terabe *et al.* 2001). The nanogap is a key structure and is necessary for a tunnelling current to flow between the two electrodes in order to cause the reduction of  $\text{Ag}^+$  cations during filament formation. A schematic diagram showing the operating mechanism of the atomic switch is shown in Fig. 8.6. Here,  $\text{Ag}_{2+\delta}\text{S}$  is used as the mixed conductor and Pt as the counterelectrode. When a positive bias is applied to the  $\text{Ag}_{2+\delta}\text{S}$  electrode,  $\text{Ag}^+$  cations in the  $\text{Ag}_{2+\delta}\text{S}$  crystal migrate towards the surface of the crystal. At the surface of the  $\text{Ag}_{2+\delta}\text{S}$  electrode,  $\text{Ag}^+$  cations are reduced by electrons coming from the Pt electrode, resulting in the precipitation of Ag atoms. The precipitated Ag atoms grow as a protrusion on the surface, which finally forms a bridge between the two electrodes. Thus, the switch is turned on. Oxidation of Ag atoms at the interface between the  $\text{Ag}_{2+\delta}\text{S}$  crystal and Ag wire supplies  $\text{Ag}^+$  cations to the  $\text{Ag}_{2+\delta}\text{S}$  crystal, which maintains the concentration of  $\text{Ag}^+$  cations in the  $\text{Ag}_{2+\delta}\text{S}$  crystal. This guarantees the stability of the precipitated Ag atoms, and the atomic switch keeps its state even after the supply of power has been cut off, i.e. it is non-volatile. Without the Ag nanowire at the bottom of the  $\text{Ag}_{2+\delta}\text{S}$  crystal, the precipitation of Ag atoms decreases the concentration of  $\text{Ag}^+$  cations in the  $\text{Ag}_{2+\delta}\text{S}$  crystal, which makes the electrochemical potential in the  $\text{Ag}_{2+\delta}\text{S}$  crystal much lower than that of Ag on a surface. Therefore, the Ag protrusion disappears as soon as the power supply for turn-on is cut off and the concentration recovers to the equilibrium condition (Terabe *et al.* 2002).

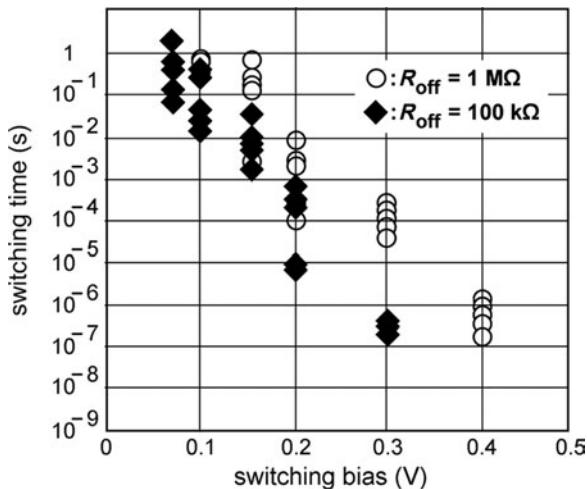
If a bias voltage with the opposite polarity is applied, the precipitated Ag atoms are oxidized and incorporated into the  $\text{Ag}_{2+\delta}\text{S}$  crystal as  $\text{Ag}^+$  cations. This annihilates the grown Ag protrusion, resulting in the atomic switch being turned off. In the  $\text{Ag}_{2+\delta}\text{S}$  crystal,  $\text{Ag}^+$  cations migrate towards the interface between the  $\text{Ag}_{2+\delta}\text{S}$  crystal and the Ag wire, and they are reduced to Ag atoms at the interface. Thus, the switching operation is achieved simply by changing the polarity of the applied bias voltage.



**Fig. 8.7** Fabrication process of atomic switch in cross-bar structure using conventional miniaturization techniques.

The operation of the atomic switch was confirmed by using STM during the initial stages of development. This is because a 1-nm gap, which is a key structure of the atomic switch, can easily be formed using an STM. That is, a  $\text{Ag}_{2+\delta}\text{S}$  tip is moved close to a Pt surface with a certain bias voltage and a certain tunnelling current using the STM feedback loop. The tip position was then fixed with a certain tunnelling gap, and the operation was achieved. Later, a fabrication method using conventional miniaturization techniques was developed (Terabe *et al.* 2005). The 1-nm gap between the  $\text{Ag}_{2+\delta}\text{S}$  and Pt electrodes is formed by using the solid electrochemical reaction after fabrication of the device structure in the on-state. The fabrication process is shown schematically in Fig. 8.7. A Ag nanowire is formed on an insulator, the surface of the Ag nanowire is sulfurized, and a 1-nm thick Ag film is then deposited on the  $\text{Ag}_{2+\delta}\text{S}$ -covered Ag nanowire at the place where the wire is to be crossed by the Pt wire. Pt deposition to form the crossing nanowire completes the fabrication process. The application of a positive bias voltage to the Pt nanowire oxidizes Ag atoms to  $\text{Ag}^+$  cations, which are incorporated into the  $\text{Ag}_{2+\delta}\text{S}$  layer. Then, the 1-nm gap appears, and the switch is in the off-state.

The dependence of the switching time on the switching bias voltage of a  $\text{Ag}_{2+\delta}\text{S}$  atomic switch is shown in Fig. 8.8. The time taken for the resistance of the switch to decrease to  $12.9\text{ k}\Omega$  from certain values of off-resistance ( $1\text{ M}\Omega$  and  $100\text{ k}\Omega$ ) was measured, where  $12.9\text{ k}\Omega$  has been reported to be the resistance of a single atomic contact (Ohnishi *et al.* 1998). As expected from the operating mechanism involving a solid electrochemical reaction with



**Fig. 8.8** Switching-time dependence on switching bias voltage and initial off-resistance. (Tamura *et al.* 2006. Copyright JSAP.)

a certain activation energy, the switching time becomes shorter exponentially as the switching bias voltage increases.

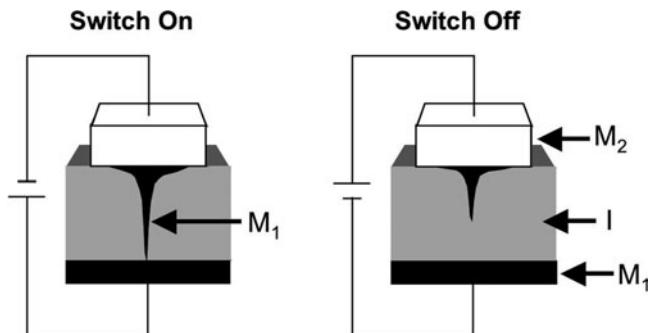
The characteristics of the atomic switch are listed in Table 8.1. Because the operating mechanism is different from that of conventional semiconductor devices, the atomic switch has unique characteristics. For instance, it can be regarded as a mechanical switch utilizing atomic movement. Even though an atom is much heavier than an electron, downsizing of the device to the nanoscale enables fast switching comparable with that of the semiconductor transistor. Non-volatility and low on-resistance are advantages for device application.

## 8.5 New types of atomic switches

Initially, the atomic switch was developed as a two-terminal switch with a vacuum gap between the mixed-conductor electrode and metal counter-electrode, as described above. However, different types of atomic switches have recently been developed. One is a gapless two-terminal atomic switch (Fig. 8.9). This has a metal, ionic conductor, metal (MIM) structure, where one of the metals ( $M_1$ ) should be electrochemically active and the other ( $M_2$ ) should be electrochemically inert. When a positive bias voltage is applied to

**Table 8.1** Characteristics of the atomic switch.

Atomic movement is controlled.	High-speed switching.
Electrons are controlled in conventional semiconductor devices.	$\sim 1$ GHz is expected.
Small size. less than $10 \times 10 \times 10 \text{ nm}^3$ .	Simple structure. low fabrication cost.
Low power consumption. $\sim 1 \text{ nW}$	Low on-resistance. suitable for high-speed switching.
Non-volatility. Data is kept forever without power.	

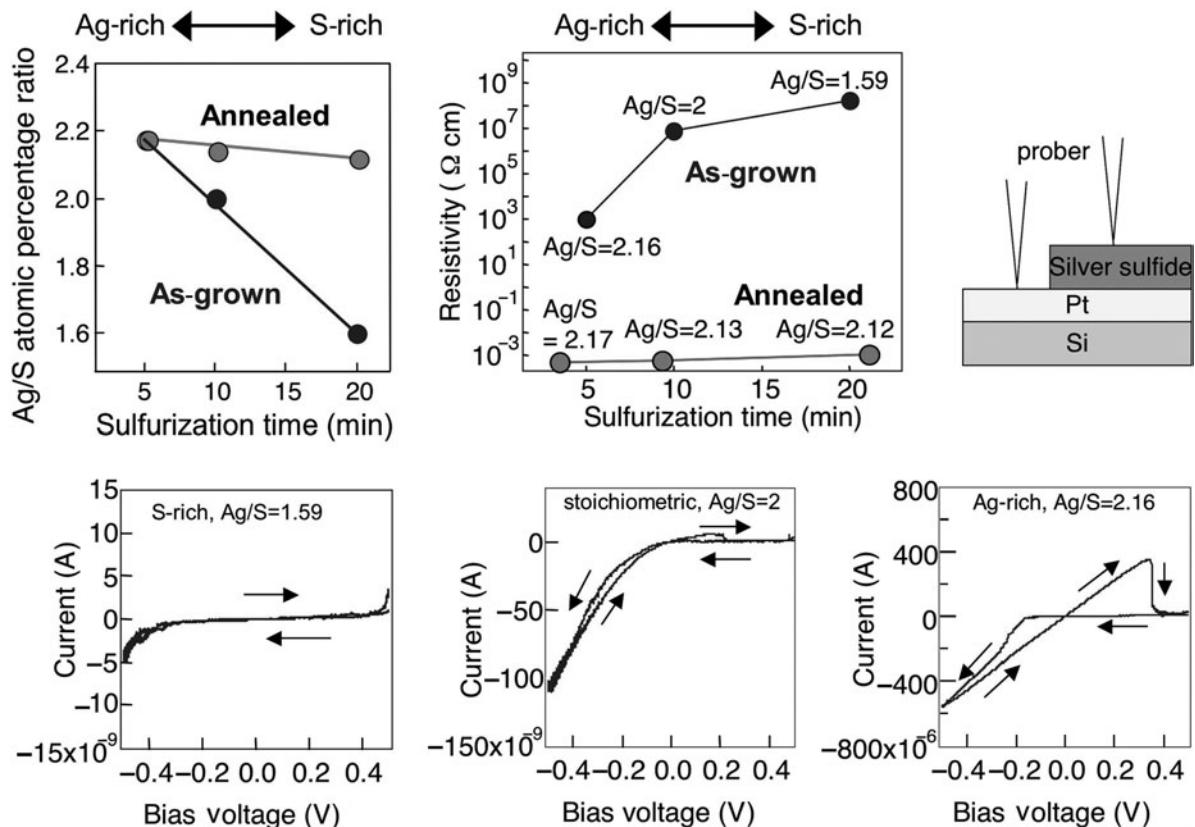


**Fig. 8.9** Schematic diagram of another type of atomic switch. A metal filament of M<sub>1</sub> is formed in the ionic conductive material (I).

the M<sub>1</sub> electrode, its atoms dissolve into the ionic conductor. The dissolved metal cations (M<sub>1</sub><sup>+</sup>) migrate towards the inert electrode (M<sub>2</sub>), where they are reduced to form a metal filament growing toward the M<sub>1</sub> electrode. A bias voltage of the opposite polarity oxidizes metal atoms in the filament and causes them to dissolve again. Consequently, the filament disappears, which turns the switch off. The operation has been confirmed for Cu<sub>2-δ</sub>S (Sakamoto *et al.* 2003), Ag<sub>2+δ</sub>S (Liang *et al.* 2005), Ag-Ge-Se (Kozicki *et al.* 1999), Ge<sub>0.2</sub>S<sub>0.8</sub> (Kozicki *et al.* 2005), Ta<sub>2</sub>O<sub>5</sub> (Sakamoto *et al.* 2007), and AgI (Liang *et al.* 2007). These gapless MIM switches also use a solid electrochemical reaction, so they can also be called gapless atomic switches.

There is also a device having a sandwich structure of metal-insulator-metal (MIM) called a resistive random access memory (ReRAM). Though its structure is similar to that of the gapless atomic switch, its operating mechanism is completely different because the atomic switch makes a metal filament in the ionic conductor using ionic diffusion, while ReRAM makes a conductive path in an insulator (Waser *et al.* 2007).

Since the gapless atomic switch forms a metal filament in an ionic conductive material, the characteristics of the ionic conductive material must affect the switching properties. This effect was reported in detail for an atomic switch made of Ag<sub>2+δ</sub>S by Kundu *et al.* in 2006. They made a variety of Ag<sub>2+δ</sub>S films on a Pt electrode by changing the sulfurization condition and temperature in the film growth. The S-rich (Ag/S = 1.59), stoichiometric (Ag/S = 2), and Ag-rich (Ag/S = 2.16) films were grown as shown in Fig. 8.10. The resistivity increased from Ag-rich to S-rich, as expected for the n-type Ag<sub>2+δ</sub>S semiconductor. Postdeposition annealing of the samples improved the conductivity of all the films. This is because, although the as-grown films had the α-phase structure with higher resistivity, the annealed films contained a mixture of acanthite α-phase and highly conductive argentite β-phase. Electrical measurement, performed as schematically shown in Fig. 8.10, revealed that only the Ag-rich film showed good switching properties among the as-grown films. This can be explained as follows. A supersaturation of Ag<sup>+</sup> cations is required to cause the precipitation of Ag atoms at the interface between the Pt electrode and the Ag<sub>2+δ</sub>S film. In the experiment, a W probe was used, and there was no supply of Ag<sup>+</sup> cations into the Ag<sub>2+δ</sub>S film. Therefore, there were not enough Ag<sup>+</sup> cations to cause supersaturation in the S-rich and stoichiometric films.

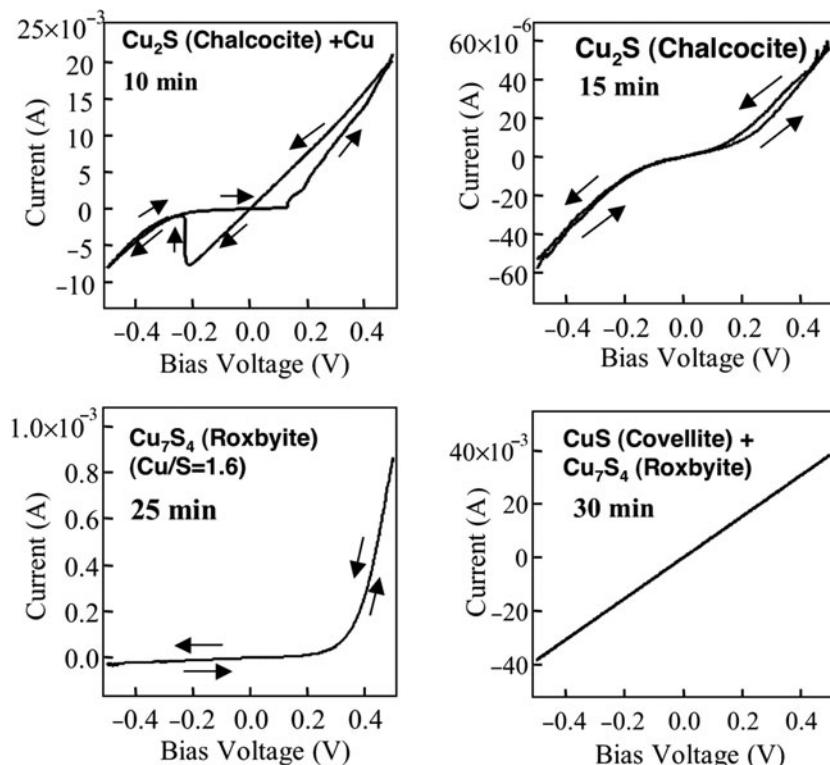


**Fig. 8.10** Properties of gapless  $\text{Ag}_{2+\delta}\text{S}$  atomic switch. Only the Ag-rich  $\text{Ag}_{2+\delta}\text{S}$  film showed good switching properties. (Kundu *et al.* 2006. Copyright AIP.)

All of the annealed films became Ag-rich and showed good switching properties. However, due to the phase transition to the  $\beta$ -phase with higher conductivity, the ON/OFF ratio became smaller than that of the as-grown Ag-rich film.

The effect of sulfurization conditions on the structural and electrical properties has been examined for  $\text{Cu}_{2-\delta}\text{S}$  films with various phases such as mixed-metallic Cu-chalcocite, chalcocite, roxybite, and roxybite-covellite (Kundu *et al.* 2008). The Cu/S atomic percentage ratio of the films decreased with increasing sulfurization time, and films with various compositions, such as Cu-rich, stoichiometric  $\text{Cu}_2\text{S}$  with underlying unreacted Cu, pure stoichiometric, and S-rich films, were formed. Similar to the results for  $\text{Ag}_{2+\delta}\text{S}$  films, only the Cu-rich film (Cu-chalcocite) showed good switching properties, as shown in Fig. 8.11. The results for  $\text{Cu}_{2-\delta}\text{S}$  films also suggest that a supersaturation of metal cations initiates metal filament growth to turn on the switch. Therefore, by using a chalcogenide film formed on a metal electrode that can supply metal cations into the chalcogenide film, one can make an atomic switch that works well.

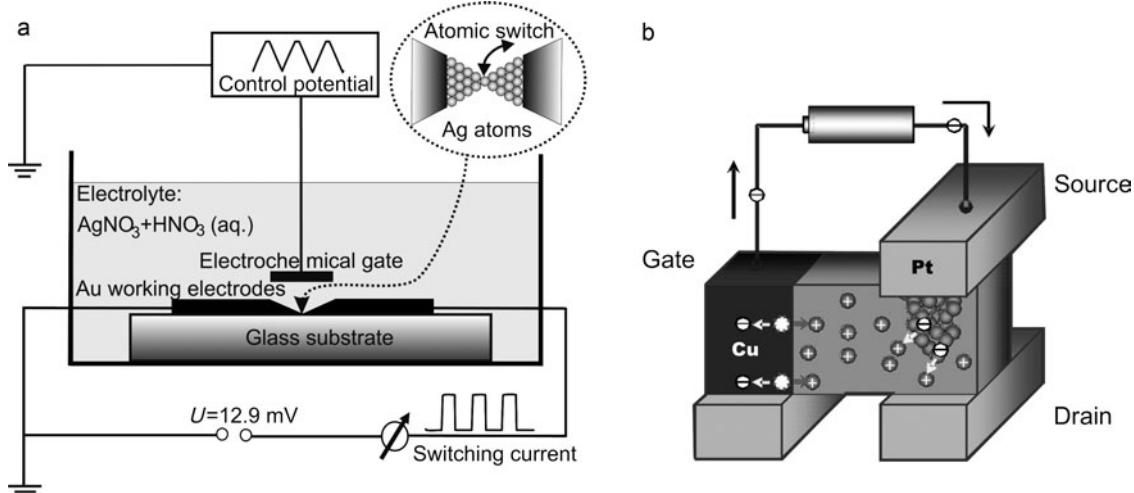
A three-terminal atomic switch, in which the control gate is separated from the signal line, has been developed. In electrochemistry, controlling the



**Fig. 8.11** Properties of gapless  $\text{Cu}_{2-\delta}\text{S}$  atomic switch. Only the Cu-rich  $\text{Cu}_{2-\delta}\text{S}$  film showed good switching properties. (Kundu *et al.* 2008. Copyright AIP.)

reduction and oxidation in an electrolyte (solution) by using a gate electrode is a conventional technique. Actually, this technique makes it possible to control the formation and annihilation of a metal filament between two Au electrodes in an electrolyte such as  $\text{AgNO}_3 + \text{HNO}_3$ , as schematically shown in Fig. 8.12(a) (Xie *et al.* 2004). An atomic bridge formed by precipitated Ag atoms shows a quantized conductance. Precise control of the gate voltage has led to switching among the quantized conductances. In other words, one can control the number of atoms forming a point contact.

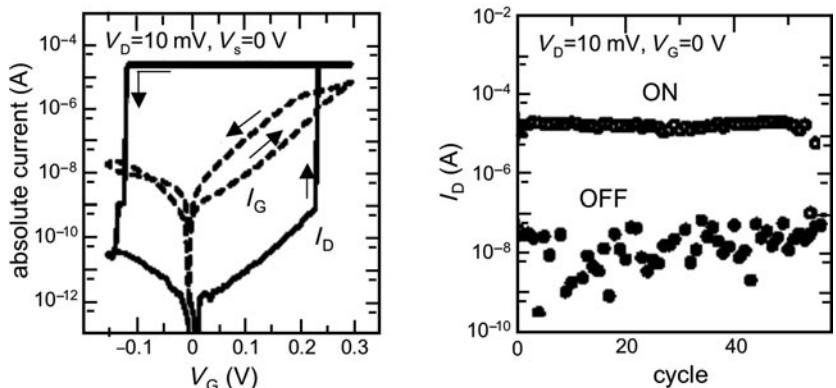
In 2006, the operation of a three-terminal atomic switch using  $\text{Cu}_{2-\delta}\text{S}$  as a solid electrolyte was demonstrated. Cu bridge formation and annihilation between the source electrode (Pt) and drain electrode (Cu), was controlled by a gate electrode (Cu) as schematically shown in Fig. 8.12(b). When a positive bias voltage was applied to the gate electrode,  $\text{Cu}^+$  cations were supplied into the  $\text{Cu}_{2-\delta}\text{S}$  from the gate electrode by the oxidation of Cu atoms at the surface of the gate electrode.  $\text{Cu}^+$  cations migrated towards the source and the drain electrodes to be reduced at the two electrodes. The precipitated Cu atoms formed a metal atomic bridge between the source and drain electrodes to make a current path. When a negative bias voltage was applied to the gate electrode, Cu atoms were extracted from the atomic bridge as a result of their oxidation to  $\text{Cu}^+$  cations. The atomic bridge then became thinner and eventually disappeared to turn off the switch. Here, the arrangement of the three electrodes should be designed carefully to avoid a metal filament



**Fig. 8.12** Three-terminal-type atomic switch. (a) Atomic-bridge formation in electrolyte (in liquid) (Xie *et al.* 2004. Copyright APS.) and (b) that in solid electrolyte.

forming between the gate electrode and the other two electrodes. This can usually be achieved by designing the device structure so that the distance between the source and drain electrodes is much smaller than those between the gate and the other two electrodes. The initialization process, in which a Cu metal atomic bridge is formed between the source and drain electrodes by using the electrodes as a two-terminal atomic switch, is helpful to make the effective distance between the source and the drain electrodes smaller. For this initialization process in the experiment performed by Sakamoto *et al.*, the drain electrode was made of Cu.

A three-terminal atomic switch was fabricated by depositing a 40-nm thick  $\text{Cu}_{2-\delta}\text{S}$  film on 120-nm thick Cu drain and gate electrodes (Banno *et al.* 2006). The source electrode was formed by electron-beam lithography to restrict the contact area between the top source electrode and the  $\text{Cu}_{2-\delta}\text{S}$  film. The results of operating the three-terminal atomic switch using  $\text{Cu}_{2-\delta}\text{S}$  are shown in Fig. 8.13. The drain current ( $I_D$ ) increased at a gate voltage ( $V_G$ ) of 0.22 V and the on-resistance was estimated to be of the order of  $100\ \Omega$ . The resistance rose to the order of  $100\ \text{M}\Omega$  at  $V_G$  of  $-0.13\ \text{V}$ . Consequently, the on/off ratio was about  $10^6$ . The linear relationship between the gate current ( $I_G$ ) and gate voltage indicates that no filaments were formed between the gate electrode and the other two electrodes. For the first three-terminal atomic switch, the number of switching cycles was of the order of ten, as shown in Fig. 8.13. The resistance of the on-state was of the order of a few kilohms, and that of the off-state was  $10\text{--}100\ \text{M}\Omega$ . The linear relation between the drain current ( $I_D$ ) and the source voltage supports the conjecture that a metal filament was formed as a conductive path. An inplane three-terminal atomic switch was also developed by Sakamoto *et al.* in 2007 in order to confirm metal-filament formation between the source and drain electrodes. The device clearly showed that a Cu filament grew between the source and drain electrodes. The experiments have clearly demonstrated the operation of the three-terminal atomic switch



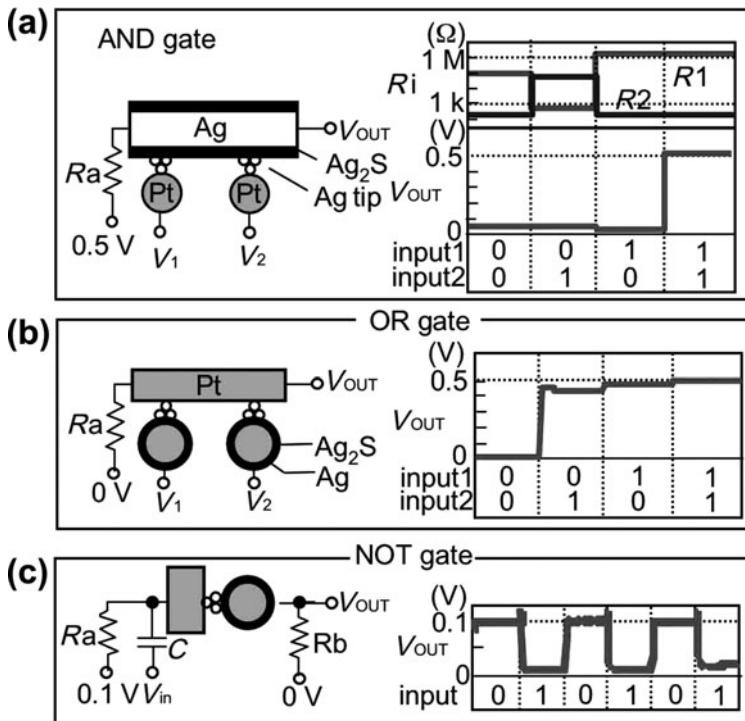
**Fig. 8.13** Operating results for three-terminal atomic switch using solid electrolyte. (Banno *et al.* 2006. Copyright IEICE.)

using a solid electrolyte and its operating mechanism. The development of the three-terminal atomic switch has widened the possibility of the atomic switch.

## 8.6 Applications of atomic switches

The novel characteristics of the atomic switch can improve the performance of present-day electronic devices and could lead to the development of new types of electronic devices. The atomic switch can provide both memories and logic cells, which are the two main elements of electronic devices. For instance, all of the three basic logic gates were configured using atomic switches as shown in Fig. 8.14 (Terabe *et al.* 2005). An AND gate was configured by two atomic switches in a cross-bar structure, in which the formation and annihilation of each atomic switch was controlled by two input signals applied to the two Pt electrodes. Interestingly, the two atomic switches were operated in combination with each other. For instance, the resistance of both atomic switches changed when only input 1 was changed from (0, 0) to (0, 1), where the numbers in parentheses are (input 1, input 2). Consequently, the output signal measured at the end of the Ag<sub>2</sub>S wire shows the AND logic operation. By using two atomic switches in a cross-bar structure, but one Pt wire and two Ag<sub>2</sub>S wires, an OR gate was also configured, in which the output signal became low when both inputs were at a low level. The non-volatility of the atomic switch enables it to be configured as a NOT gate; this is very different from a volatile switch such as a diode because diodes can configure only AND and OR gates; they cannot configure a NOT gate. The fact that the atomic switch can configure all three of the basic logic gates, i.e. AND, OR, and NOT gates, means that any kind of logic circuit can be configured using atomic switches. For instance, the quantized conductance shown by the atomic switch can be used to configure an adder circuit (Terabe *et al.* 2005).

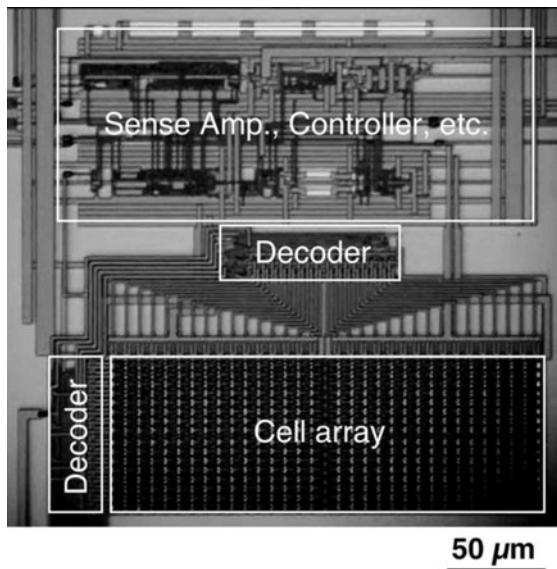
The small size and non-volatility of the atomic switch are useful for making a non-volatile memory. Unlike the semiconductor transistor, which is volatile in nature, the atomic switch can store information by itself. A switching



**Fig. 8.14** Logic gates configured by atomic switches. (Terabe *et al.* 2005. Copyright NPG.)

speed of around the gigahertz level can be expected from the electrochemical reaction rates. A micrograph of a developed 1-kb non-volatile memory chip is shown in Fig. 8.15 (Kaeriyama *et al.* 2005). In the cell array, one thousand gapless atomic switches were made using Cu<sub>2</sub>S as a solid electrolyte. The peripheral circuits, such as a driver circuit and address decoders for writing and reading the atomic switches, were also fabricated using 0.25-μm CMOS (complementary metal-oxide-semiconductor) technology. To select a switch (bit) to be written or read, an access transistor was used in this architecture. The selected bit was alternately programmed to be in the on- and off-states with a programming voltage of 1.1 V. Since this programming voltage is almost the same as those of CMOS devices, the atomic switch can be easily introduced into conventional CMOS devices to improve their performance.

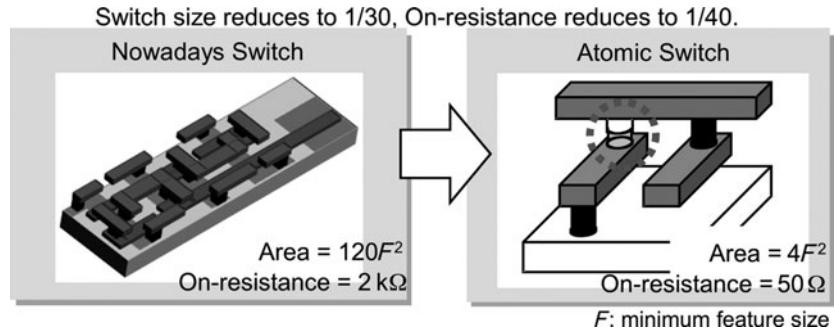
One of the most promising applications of the atomic switch is a program switch for reconfigurable large-scale integrated circuits (LSIs) such as the field programmable gate array (FPGA), which can provide many different functions by changing the circuits under the control of program switch circuits. However, the switching circuit used in the conventional FPGA is composed of a pass transistor combined with either a static random access memory (SRAM) cell or a flip-flop circuit, which makes the switching circuit as large as  $120 F^2$  ( $F$ : minimum feature size). Consequently, today's FPGAs have a large area for the switching circuits, which restricts the performance of the device. The fact that the on-resistance of the switching circuit is as high as 1–2 kΩ



**Fig. 8.15** 1-kb non-volatile memory chip using atomic switches. (Kaeriyama *et al.* 2005. Copyright IEEE.)

restricts the number of switches in a single chip. As a result of these characteristics of present-day switching circuits, the FPGA must use coarse-grained logic cells.

The atomic switch, which is a small non-volatile switch, works as a switching circuit by itself. Since it can be formed at the cross-point of two wires, the size of the switching circuit is  $4F^2$ , which is 1/30th the size of the conventional switching circuit. The on-resistance of the atomic switch is of the order of  $10\ \Omega$ , which reduces the on-resistance of the switching circuit to 1/40th, as shown in Fig. 8.16. Consequently, a much larger number of switching circuits can be used in a single chip of the FPGA, so a large logic cell can be divided into fine logic cells to improve the logical efficiency. Because the number of functions is determined by the number of logic cells and program switches, the atomic switch enables the development of a new type of a programmable logic device, called the cell-based integrated circuit (CBIC). The characteristics of



**Fig. 8.16** Switching circuits using semiconductor transistors and ones using atomic switches.

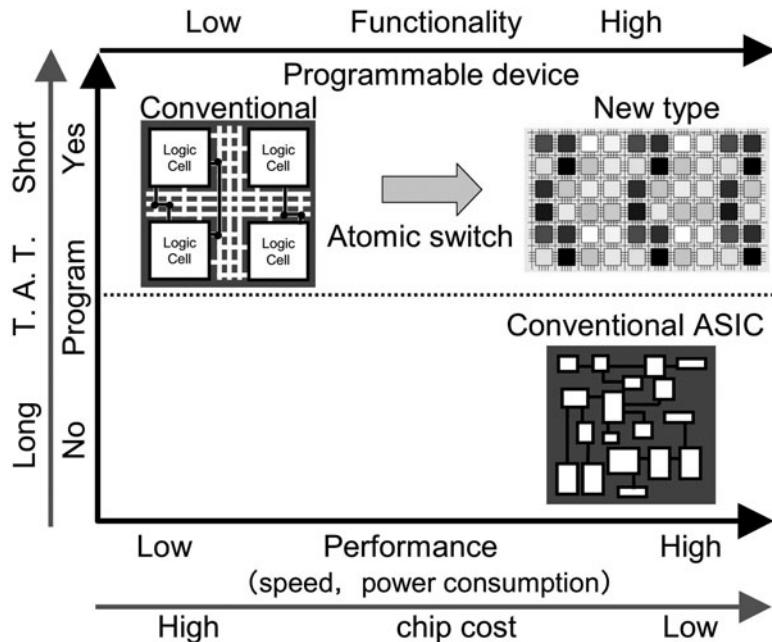


Fig. 8.17 Comparison of programmable devices with ASICs.

two types of integrated circuits (ICs), i.e. application-specific integrated circuit (ASIC) and programmable logic device are shown in Fig. 8.17. When the atomic switch is used as the program switching circuit of a programmable logic device, a new programmable device can be made that has comparable performance to an ASIC. A  $4 \times 4$  cross-bar switching circuit made of Cu<sub>2</sub>S atomic switches was formed on a conventional CMOS device, and its operation was confirmed by Kaeriyama *et al.* (2005).

## 8.7 Summary and conclusion

Nanoionics phenomena and their applications for making new types of electronic devices are introduced. Using ionic conductive materials, metal-filament growth and shrinkage are controlled by a solid electrochemical reaction, which can be used for making a novel switching device in nanoscale. Because we are approaching the physical limit of downsizing in semiconductor transistors, the nanoionics-based switching device has attracted much attention as an emerging device for replacing the semiconductor transistor.

The nanoionics-based switching device, i.e. the atomic switch, has novel characteristics such as small size, low power consumption, non-volatility, and low on-resistance. The characteristics enable us to improve the performance of present-day electronic devices. A 1-kb non-volatile memory chip using atomic switches has been developed to demonstrate the possibility to improve the performance of conventional CMOS devices. The atomic switch

also enables development of conceptually new types of electronic devices, such as programmable cell-based integrated circuits (CBIC), which will have comparable performance with the application-specific integrated circuits (ASICs).

Quantized conductance and learning ability of the atomic switch can be used for developing new types of computer architectures, such as neural networks. Thus, the atomic switch can be anticipated to be one of the key technologies for achieving the advanced information society.

## References

- Banno, N., Sakamoto, T., Iguchi, N., Kawaura, H., Kaeriyama, S., Mizuno, M., Terabe, K., Hasegawa, T., Aono, M. *IEICE Trans. Electron.* E89-C **1492** (2006).
- Banno, N., Sakamoto, T., Hasegawa, T., Terabe, K., Aono, M. *Jpn. J. Appl. Phys.* **45**, 3666 (2006).
- Eigler, D.M., Lutz, C.P., Rudge, W.E. *Nature* **352**, 600 (1991).
- Ercker, L. *Treaties on Ores and assaying* (translated by Sisco, A.G., Smith, C.S.), University of Chicago, **177**, 1951 (1574).
- Faraday, M. *Philos. Trans. Roy. Soc.* **123**, 507 (1833).
- Hirose, Y., Hirose, H. *J. Appl. Phys.* **47**, 2767 (1976).
- Kaeriyama, S., Sakamoto, T., Sunamura, H., Mizuno, M., Kawaura, H., Hasegawa, T., Terabe, K., Nakayama, T., Aono, M. *IEEE, J. Solid-State Circuits* **40**, 168 (2005).
- Kozicki, M.N., Yun, M., Hilt, L., Singh, A. *Electrochem. Soc.* **298** (1999).
- Kozicki, M.N., Balakrishnan, M., Gopalan, C., Ratnakumar, C., Mitkova, M. in *Proc. Non-Volatile Memory Technology Symposium (NVMTS) 2005*, **7** (2005).
- Kudo, T., Fueki, K. *Solid State Ionics* (Kodansha, Tokyo, 1990).
- Kundu, M., Terabe, K., Hasegawa, T., Aono, M. *J. Appl. Phys.* **99**, 103501 (2006).
- Kundu, M., Hasegawa, T., Terabe, K., Aono, M. *J. Appl. Phys.* **103**, 073523 (2008).
- Liang, Ch., Terabe, K., Hasegawa, T., Negishi, R., Tamura, T., Aono, A. *Small* **10**, 971 (2005).
- Liang, Ch., Terabe, K., Tsuruoka, T., Osada, M., Hasegawa, T., Aono, M. *Adv. Funct. Mater.* **17**, 1466 (2007).
- Ohnishi, H., Kondo, Y., Takayanagi, K. *Nature* **395**, 780 (1998).
- Rickert, H. *Electrochemistry of Solids—an Introduction* (Springer-Verlag, 1982).
- Sakamoto, T., Sunamura, H., Kawaura, H., Hasegawa, T., Nakayama, T., Aono, M. *Appl. Phys. Lett.* **82**, 3033 (2003).
- Sakamoto, T., Lister, K., Banno, N., Hasegawa, T., Terabe, K., Aono, M. *Appl. Phys. Lett.* **91**, 092110 (2007).
- Sone, H., Tamura, T., Miyazaki, K., Hosaka, S. *Microelectron. Engng.* **831**, 487 (2006).
- Tafel, J. *Z. Phys. Chem.* **54**, 641 (1905).
- Tamura, T., Hasegawa, T., Terabe, K., Nakayama, T., Sakamoto, T., Sunamura, H., Kawaura, H., Hosaka, S., Aono, M. *Jpn. J. Appl. Phys.* **45**, L364 (2006).

Terabe, K., Nakayama, T., Hasegawa, T., Aono, M. *Appl. Phys. Lett.* **80**, 4009 (2002).

Terabe, K., Hasegawa, T., Nakayama, T., Aono, M. *Riken Rev.* **37**, 7 (2001).

Terabe, K., Hasegawa, T., Nakayama, T., Aono, M. *Nature* **433**, 47 (2005).

Waser, R., Aono, M. *Nature Mater.* **6**, 833 (2007).

Xie, F.-Q., Nittler, L., Obermair, Ch., Schimmel, Th. *Phys. Rev. Lett.* **93**, 128303 (2004).

# 9

# Molecular electronics based on self-assembled monolayers

9.1 Introduction	312
9.2 Nanofabrication for molecular devices	313
9.3 Molecular tunnelling barrier	319
9.4 Molecular semiconducting wire	321
9.5 Molecular rectifying diode	323
9.6 Molecular switches and memories	326
9.7 Molecular transistor	330
9.8 Conclusion	331
Acknowledgments	332
References	332

*D. Vuillaume*

## 9.1 Introduction

Since the first measurement of electron tunnelling through an organic monolayer in 1971 (Mann and Kuhn 1971), and the *gedanken* experiment of a molecular current-rectifying diode in 1974 (Aviram and Ratner 1974), molecular-scale electronics have attracted a growing interest, both for basic science at the nanoscale and for possible applications in nanoelectronics. In the first case, molecules are quantum objects by nature and their properties can be tailored by chemistry, opening avenues for new experiments. In the second case, molecule-based devices are envisioned to complement silicon devices by providing new functions or already existing functions at a simpler process level and at a lower cost by virtue of their self-organization capabilities, moreover, they are not bound to von Neuman architecture and this may open the way to other architectural paradigms.

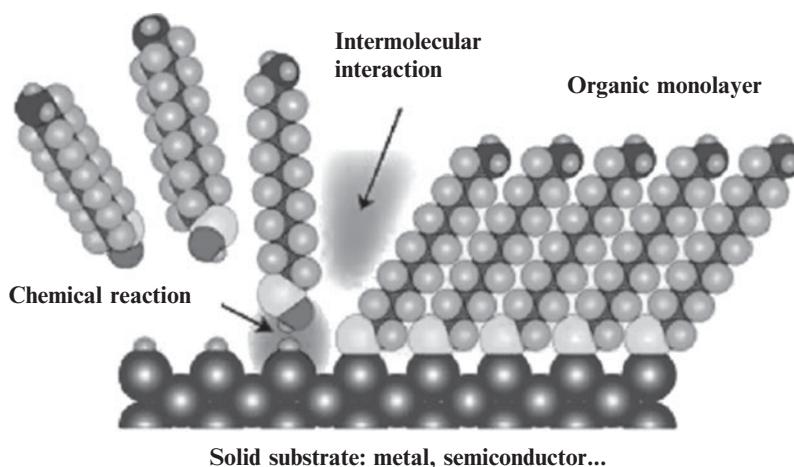
Molecular electronics, i.e. the information processing at the molecular scale, becomes more and more investigated and envisioned as a promising candidate for the nanoelectronics of the future. One definition is “information processing using photo-, electro-, iono-, magneto-, thermo-, mechano- or chemicoactive effects at the scale of structurally and functionally organized molecular architectures” (adapted from Lehn 1988). In the following, we will consider devices based on organic molecules with size ranging from a single molecule to a monolayer. This definition excludes devices based on thicker organic materials referred to as organic electronics. Two studies laid the foundation of this molecular-scale electronics field. In 1971, Mann and Kuhn were the first to demonstrate tunnelling transport through a monolayer of aliphatic chains (Mann and Kuhn 1971). In 1974, Aviram and Ratner theoretically proposed the concept of a molecular rectifying diode where an acceptor–bridge–donor (A–b–D) molecule can play the same role as a semiconductor

p-n junction (Aviram and Ratner 1974). Since then, many groups have reported on the electrical properties of molecular-scale devices from single molecules to monolayers.

After a brief overview of the nanofabrication of molecular devices, we review in this chapter, the electronic properties of several basic devices, from simple molecules such as molecular tunnel junctions and molecular wires, to more complex ones such as molecular rectifying diodes, molecular switches and memories.

## 9.2 Nanofabrication for molecular devices

To measure the electronic transport through an organic monolayer, we need a test device that is as simple as possible. The generic device is a metal/monolayer/metal or metal/molecules/metal (MmM) junction (for simplicity, we will always use this term and acronym throughout the chapter even if the metal electrode is replaced by a semiconductor). Organic monolayers and submonolayers (down to single molecules) are usually deposited on the electrodes by chemical reactions in solution or in gas phase using molecules of interest bearing a functional moiety at the ends that is chemically reactive to the considered solid surface (for instance, a thiol group on metal surfaces such as Au, silane group on oxidized surfaces, etc.), see Fig. 9.1. However, Langmuir–Blodgett (LB) monolayers have also been used for device applications since the 1970s (for a review see Ulman 1991). Some important results are, for instance, the observation of a current-rectification behavior through LB monolayers of hexadecylquinolinium tricyanoquinodimethanide (Ashwell *et al.* 1990; Martin *et al.* 1993; Geddes *et al.* 1995; Metzger *et al.* 1997; Vuillaume *et al.* 1999; Metzger *et al.* 2001; Xu *et al.* 2001) and the fabrication of molecular switches based on LB monolayers of catenanes (Collier *et al.* 1999, 2000; Pease *et al.* 2001; Chen *et al.* 2003a,b). The second method deals with monolayers of organic molecules chemically grafted on solid substrates, also called self-assembled monolayers (SAM) (Ulman 1991). Many reports



**Fig. 9.1** A schematic description of the formation of an organic monolayer on a solid substrate, showing the chemical reaction between a functionalized end of the molecule and the substrate, and the interactions between adjacent molecules (from [www.mtl.kyoto-u.ac.jp/groups/sugimura-g/index-E.html](http://www.mtl.kyoto-u.ac.jp/groups/sugimura-g/index-E.html), slightly modified).

in the literature concern SAMs of thiol-terminated molecules chemisorbed on gold surfaces, and to a less extent, molecular-scale devices based on SAMs chemisorbed on semiconductors, especially silicon. Silicon is the most widely used semiconductor in microelectronics. The capability to modify its surface properties by the chemical grafting of a broad family of organic molecules (e.g. modifying the surface potential Bruening *et al.* 1994; Cohen *et al.* 1997, 1998) is the starting point for making almost any tailored surfaces useful for new and improved silicon-based devices. Between the end of the silicon roadmap and the envisioned advent of fully molecular-scale electronics, there may be a role played by such hybrid-electronic devices (Compano *et al.* 2000; Joachim *et al.* 2000). The use of thiol-based SAMs on gold in molecular-scale electronics is supported by a wide range of experimental results on their growth, structural and electrical properties (see a review by Schreiber 2000). However, SAMs on silicon and silicon dioxide surfaces were less studied and were more difficult to control. This has resulted in an irreproducible quality of these SAMs with large time-to-time and lab-to-lab variations. This feature may explain the smaller number of attempts to use these SAMs in molecular-scale electronics than for the thiol/gold system. Since the first chemisorption of alkyltrichlorosilane molecules from solution on a solid substrate (mainly oxidized silicon) introduced by Bigelow *et al.* (1946) and later developed by Maoz and Sagiv (1984), further detailed studies (Brzoska *et al.* 1992, 1994; Parikh *et al.* 1994; Allara *et al.* 1995) have led to a better understanding of the basic chemical and thermodynamical mechanisms of this self-assembly process. For a review on these processes, see Ulman (1991) and Schreiber (2000).

In their pioneering work, Mann and Kuhn used a mercury drop to contact the monolayer (Mann and Kuhn 1971), and this technique is still used nowadays (Rampi *et al.* 1998; Holmlin *et al.* 2001; Selzer *et al.* 2002a,b) at the laboratory level as an easy technique for a quick assessment of the electrical properties. Several types of MmM junctions have been built. The simplest structure consists of depositing the monolayer onto the bottom electrode and then evaporating a metal electrode on top of the monolayer through a masking technique. These shadow-masks are fabricated from metal or silicon nitride membranes and the dimensions of the holes in the mask may range from a few hundred  $\mu\text{m}$  to a few tens of nanometers. Chen and coworkers (Chen *et al.* 1999, 2000) have used nanopores (about 30 nm in diameter in a silicon nitride membrane), in which a small numbers of molecules are chemisorbed to fabricate these MmM junctions. From  $\sim 10^{10}$  to  $\sim 10^2$  molecules can be measured in parallel with these devices. The critical point deals with the difficult problem of making a reliable metal contact on top of an organic monolayer. Several studies (Jung and Czanderna 1994; Herdt and Czanderna 1995; Konstadinidis *et al.* 1995; Jung *et al.* 1996; Fisher *et al.* 2000, 2002) have analyzed (by X-ray photoelectron spectroscopy, infra-red spectroscopy, ...) the interaction (bond insertion, complexation ...) between the evaporated atoms and the organic molecules in the SAM. When the metal atoms are strongly reactive with the end-groups of the molecules (e.g. Al with COOH or OH groups, Ti with COOCH<sub>3</sub>, OH or CN groups...) (Jung and Czanderna 1994; Herdt and Czanderna 1995; Konstadinidis *et al.* 1995; Jung *et al.* 1996; Fisher *et al.*

2000, 2002), a chemical reaction occurs forming a molecular overlayer on top of the monolayer. This overlayer made of organometallic complexes or metal oxides may perturb the electronic coupling between the metal and the molecule, leading, for instance, to partial or total Fermi-level pinning at the interface (Lenfant *et al.* 2006). In some cases, if the metal chemically reacts with the end-group of the molecule (e.g. Au on thiol-terminated molecules), this overlayer may further prevent the diffusion of metal atoms into the organic monolayer (Aswal *et al.* 2005). The metal/organic interface interactions (e.g. interface dipole, charge transfer,...) are very critical and they have strong impacts on the electrical properties of the molecular devices. Some reviews are given in Cahen *et al.* (2005) and Kahn *et al.* (2003). If the metal atoms are not too reactive (e.g. Al with CH<sub>3</sub> or OCH<sub>3</sub> ...) (Jung and Czanderna 1994; Herdt and Czanderna 1995; Konstadinidis *et al.* 1995; Jung *et al.* 1996; Fisher *et al.* 2000, 2002), they can penetrate into the organic monolayer, diffusing to the bottom interface where they can eventually form an adlayer between this electrode and the monolayer (in addition to metallic filamentary short circuits). In a practical way for device application using organic monolayers, the metal evaporation is generally performed onto a cooled substrate ( $\sim$ 100 K). It is also possible to intercalate blocking baffles on the direct path between the crucible and the sample, or/and to introduce a small residual pressure of inert gas in the vacuum chamber of the evaporator (Okazaki and Sambles 2000; Metzger *et al.* 2001; Xu *et al.* 2001). These techniques allow the energy of the metal atoms arriving on the monolayer surface to be reduced, thus reducing the damage.

To avoid these problems, alternative and soft metal-deposition techniques were developed. One called nanotransfer printing (nTP), has been described and demonstrated (Loo *et al.* 2003). Nanotransfer printing is based on soft lithographic techniques used to print patterns with nanometric resolution on solid substrates (Xia and Whitesides 1998). The principle is briefly described as follows. Gold electrodes are deposited by evaporation onto an elastomeric stamp and then transferred by mechanical contact onto a thiol-functionalized SAM. Transfer of gold is based on the affinity of this metal for thiol function –SH forming a chemical bond Au–S. Loo *et al.* (2003) have used the nTP technique to deposit gold electrodes on alkane dithiol molecules self-assembled on gold or GaAs substrates. Nanotransfer printing of gold electrodes was also deposited onto oxidized silicon surface covered by a monolayer of thiol-terminated alkylsilane molecules (Loo *et al.* 2002; Guerin *et al.* 2007). Soft depositions of pre-formed metal electrodes, e.g. lift-off float-on (LOFO) (Vilan and Cahen 2002), have also been developed. Recently, another solution has been proposed in which a thin conducting polymer layer has been intercalated as a buffer layer between the organic monolayer and the evaporated metal electrode (Akkerman *et al.* 2006). It was also reported to use metallic electrode made of a 2D network of carbone nanotubes (He *et al.* 2006). Finally, another solution to avoid problems with metal evaporation is to cover a metal wire (about 10  $\mu$ m in diameter) with a SAM and then to bring this wire in contact with another wire (crossing each other) using the Laplace force (Kushmerick *et al.* 2002a,b). About 10<sup>3</sup> molecules can be contacted by this way.

At the nanometer scale, the top electrode can also be a STM tip. The properties of a very small number of molecules (a few tens down to a single molecule) can be measured. If one assumes that an intimate contact is provided by the chemical grafting (in the case of a SAM) at one end of the molecules on the bottom electrode, the drawback of these STM experiments is the fact that the electrical “contact” at the other end occurs through the air gap between the SAM surface and the STM tip (or vacuum in the case of an UHV-STM). This leads to a difficult estimate of the true conductance of the molecules, while possible through a careful data analysis and choice of experimental conditions (Bumm *et al.* 1999; Labonté *et al.* 2002). Recently, some groups have used a conducting-atomic force microscope (C-AFM) as the upper electrode (Wold and Frisbie 2000, 2001; Wold *et al.* 2002). In that case, the metal-coated tip is gently brought into a mechanical contact with the monolayer surface (this is monitored by the feedback loop of the AFM apparatus) while an external circuit is used to measure the current–voltage curves. The advantage over the STM is twofold, (i) tip-surface position control and current probing are physically separated (while the same current in the STM is used to control the tip position and to probe the electronic transport properties), (ii) under certain conditions, the molecules may be also chemically bonded to the C-AFM tip at the mechanical contact (Cui *et al.* 2001). The critical point of C-AFM experiments is certainly the very sensitive control of the tip load to avoid excessive pressure on the molecules (Son *et al.* 2001) (which may modify the molecule conformation and thus its electronic transport properties, or can even pierce the monolayer). On the other hand, the capability to apply a controlled mechanical pressure on a molecule to change its conformation is a powerful tool to study the relationship between conformation and electronic transport (Moresco *et al.* 2001). A significant improvement has been demonstrated by Xu and Tao (2003) to measure the conductance of a single molecule by repeatedly forming a few thousand Au–molecule–Au junctions. This technique is a STM-based break junction, in which molecular junctions are repeatedly formed by moving the STM tip back and forth into and out of contact with a gold surface in a solution containing the molecules of interest. A few molecules, bearing two chemical groups at their ends, can bridge the nanogap formed when moving back the tip from the surface. Due to the large number of measurements, this technique provides statistical analysis of the conductance data. This technique has been recently used to obtain new insights into the electronic transport through molecular junctions, e.g. on the analysis of the variability of the conductance (Ulrich *et al.* 2006; Venkataraman *et al.* 2006b), on the role of the chemical link between the molecule and the metal electrode (Chen *et al.* 2006; Venkataraman *et al.* 2006b) (for instance, it has been shown that the amine group gives a better-defined conductance than thiol (Venkataraman *et al.* 2006b)), on the influence of the atomic configuration of the chemical link (Li *et al.* 2006). Changes in the electrical conductance of a single molecule as a function of a chemical substitution (Venkataraman *et al.* 2007) and a conformational change were also evidenced (Venkataraman *et al.* 2006a).

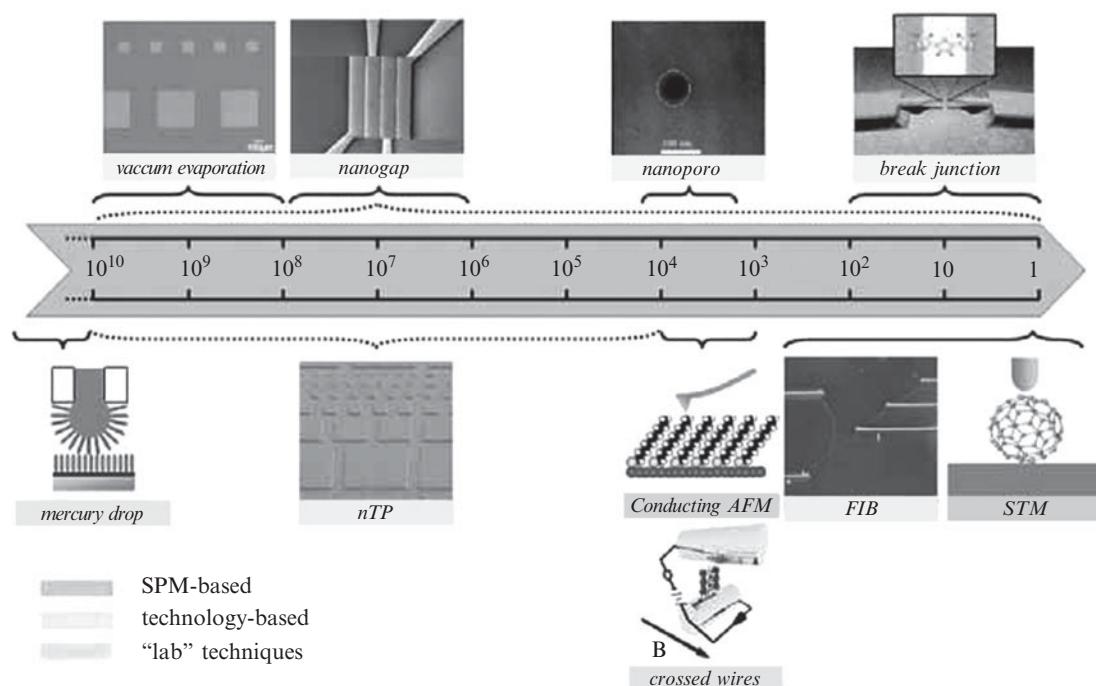
The second type of MmM junction uses a “planar” configuration (two electrodes on the same surface). The advantage over a vertical structure is

the possibility to easily add a third gate electrode (3-terminal device) using a bottom gate transistor configuration. The difficulties are (i) to make these electrodes with a nanometer-scale separation; (ii) to deposit molecules into these nanogaps. Alternatively, if the monolayer is deposited first onto a suitable substrate, it would be very hard to pattern, with a nanometer-scale resolution, the electrodes on top of it. The monolayers have to withstand, without damage, a complete electron-beam patterning process for instance. This has been proved possible for SAMs of alkyl chains (Collet and Vuillaume 1998; Collet *et al.* 2000) and alkyl chain functionalized by  $\pi$ -conjugated oligomers (Mottaghi *et al.* 2007) used in nanoscale (15–100 nm) devices. However, recently developed soft lithographies (microimprint contact...) can be used to pattern organic monolayers or to pattern electrodes on these monolayers (Xia and Whitesides 1998). Nowadays, 30-nm width nanogaps are routinely fabricated by e-beam lithography and 5-nm width nanogaps are attainable with a lower yield (a few tens of per cent) (Bezryadin and Dekker 1997; Cholet *et al.* 1999; Guillorn *et al.* 2000). However, these widths are still too large compared to the typical molecule length of 1–3 nm.

The smallest nanogaps ever fabricated have a width of about 1 nm. A metal nanowire is e-beam fabricated and a small gap is created by electromigration when a sufficiently high current density is passing through the nanowire (Park *et al.* 1999). These gold nanogaps were then filled with few molecules (bearing a thiol group at each end) and Coulomb-blockade and Kondo effects were observed in these molecular devices (Liang *et al.* 2002; Park *et al.* 2002). A second approach is to start by making two electrodes spaced by about 50–60 nm, then to gradually fill the gap by electrodeposition until a gap of a few nanometers has been reached (Li *et al.* 2000; Boussaad and Tao 2002; Kervennic *et al.* 2002). Recently, carbon nanotubes (CNT) have been used as electrodes separated by a nanogap (< 10 nm) (Guo *et al.* 2006). The nanogap is obtain by a precise oxidation cutting of the CNT, and the two facing CNT ends that are now terminated by carboxylic acids, are covalently bridged by molecules of adapted length derivatized with amine groups at the two ends. It is also possible to functionalize the molecular backbone for further chemical reactions, allowing the electrical detection of molecular and biological reactions at the molecule scale (Guo *et al.* 2006, 2007). Another approach is to use a breaking junction, bridged by a few dithiol-terminated molecules. Reed and coworkers (Reed *et al.* 1997) and Kergueris and coworkers (Kergueris *et al.* 1999) have used these breaking junctions to fabricate and to study some MmM junctions based on dithiolbenzene and bisthiolterthiophene, respectively, and this technique was further used with other short oligomers (Reichert *et al.* 2002; Weber *et al.* 2002). However, these MmM breaking junctions are not stable over a very long period of time (no more than 20–30 min) while the vertical MmM junctions and the “planar” ones based on nanofabricated nanogaps are stable over months. Weber *et al.* reported some improvements allowing stable MmM breaking-junction measurements at low temperature (Reichert *et al.* 2003; Elbing *et al.* 2005). Finally, we mention that Au nanoparticles (NP) can be used to connect a few molecules, these NP (tens of nm in diameter) being themselves deposited between electrodes or contacted with a STM (Cui *et al.* 2001; Dadosh *et al.* 2005; Long *et al.*

2005). Microspheres metallized by Ni/Au can also be magnetically trapped between microlithographically patterned electrodes covered by a monolayer of molecules forming two molecular junctions in series (Long *et al.* 2005). These approaches allow measuring a small number of molecules and avoid the difficult fabrication of gaps a few nm in size. A very recent review on how to electrically connect molecules and organic monolayers is given by Haick and Cahen (2008).

To conclude this section, many technological solutions are available to measure the electronic transport properties of molecular monolayers with lateral extension from a few molecules to  $\sim 10^{10}$  (Fig. 9.2). A comparison between electrical measurements at the molecular scale and those on macroscopic devices will be helpful to understand the effect of intermolecular interactions on the transport properties. As a result of these various approaches for making the organic monolayers and the MmM junctions, the nature of the interfaces, and thus the electronic coupling between the molecules and the electrodes are largely dependent on the experimental conditions and protocols. This feature requires a multi-test-bed approach to assess the intrinsic properties of the molecular devices and not of the contacts (Szuchmacher Blum *et al.* 2005). In the following sections, we illustrate and discuss the effects of this molecule/electrode coupling on the electronic transport properties of some molecular devices.

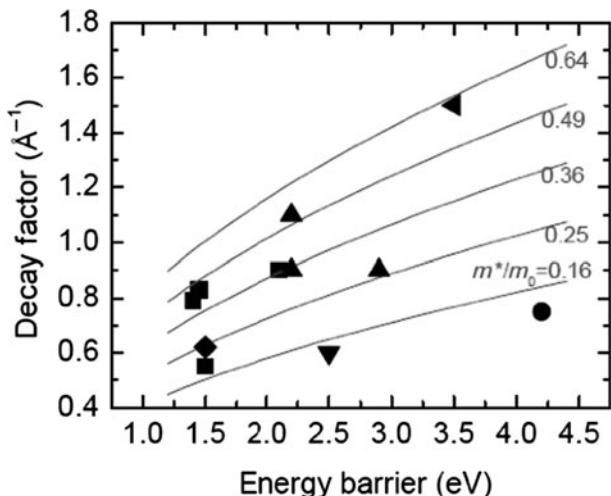


**Fig. 9.2** A schematic overview of the different test-beds used to electrically contact organic molecules. The scale gives the approximate number of molecules contacted from monolayer (left) to single molecule (right). The techniques are (from left to right, upper part of the figure): micrometer-scale metal evaporation, nanogap patterned by e-beam lithography, nanopores, break-junction, and (from left to right, lower part of the figure): mercury drop, nanotransfer printing, conducting AFM, crossed wires, metal deposition by FIB, STM (courtesy of S. Lenfant, IEMN-CNRS).

### 9.3 Molecular tunnelling barrier

It has long been recognized that a monolayer of alkyl chains sandwiched between two metal electrodes acts as a tunnelling barrier. Mann and Kuhn (1971); Polymeropoulos and Sagiv (Polymeropoulos 1977; Polymeropoulos and Sagiv 1978) have demonstrated that the current through LB monolayers of alkyl chains follows the usual distance-dependent exponential law,  $I = I_0 \exp(-\beta d)$ , where  $d$  is the monolayer thickness and  $\beta$  is the distance decay rate. They have found  $\beta \sim 1.5 \text{ \AA}^{-1}$ . More recently, we found (Lenfant 2001)  $\beta \sim 0.7\text{--}0.8 \text{ \AA}^{-1}$  for n<sup>+</sup>-Si/native SiO<sub>2</sub>/SAM of alkyl-1-enyl trichlorosilane/metal (Au or Al) junctions and Whitesides' group (Holmlin *et al.* 2001) found  $\beta \sim 0.9 \text{ \AA}^{-1}$  for Hg/SAM of alkylthiol/Ag junctions. All these experiments were done with macroscopic-size electrodes. Data taken for alkanethiols in a nanopore junction gave  $\sim 0.8 \text{ \AA}^{-1}$  (Wang *et al.* 2003). Recently, C-AFM experiments were also done addressing the properties of a small number of molecules. Again, a tunnelling law was observed with  $\beta \sim 0.9\text{--}1.4 \text{ \AA}^{-1}$  for Au/SAM of alkylthiols/Au-covered AFM tip junctions (Wold and Frisbie 2000, 2001; Sakaguchi *et al.* 2001; Engelkes *et al.* 2004). A smaller value ( $\beta \sim 0.5 \text{ \AA}^{-1}$ ) was reported for Au/SAM of alkyldithiol/Au-covered AFM tip junctions (Cui *et al.* 2002), but another study reported no significant variation of  $\beta$  between alkanethiols and alkanedithiols, but only a contact resistance 1 or 2 decades lower for the alkanedithiols. A more complete review of these data and others is given by (Salomon *et al.* 2003). The  $\beta$  value is related to the tunnelling barrier height ( $\Delta$ ) at the molecule/electrode interface and to the effective mass ( $m^*$ ) of carriers in the monolayer,  $\beta = \alpha(m^*/m_0)^{1/2} \Delta^{1/2}$ , with  $m_0$  the rest mass of the electron and  $\alpha = 4\pi(2m_0e)^{1/2}/h = 10.25 \text{ eV}^{-1/2} \text{ nm}^{-1}$  ( $e$  is the electron charge and  $h$  the Planck constant). The tunnelling barrier height may be measured independently by internal photoemission experiment (IPE) (Powell 1970) where carriers in one of the electrodes are photoexcited over the tunnelling barrier and collected at the other electrode (under a small applied dc bias). The threshold energy of the exciting photons allows the measurement of  $\Delta$ . We have found an electron tunnelling barrier of about 4.3–4.5 eV at the silicon/native SiO<sub>2</sub>/SAM and aluminum/SAM interfaces in the case of densely packed, well-ordered, SAMs of alkyl chains (Boulaas *et al.* 1996), a larger value than  $\sim 1.4$  to 3 eV found in other experiments on LB monolayers and alkylthiol SAM on Au (Mann and Kuhn 1971; Polymeropoulos and Sagiv 1978; Holmlin *et al.* 2001; Wang *et al.* 2003). This high value ( $\sim 4.5$  eV) is in agreement with theoretical calculations (Vuillaume *et al.* 1998). For the same alkyl chains directly chemisorbed on Si (no native oxide), lower values have been reported from a combination of electrical ( $\sim 1\text{--}1.5$  eV) and UPS/IPES (2.5–3.5 eV) experiments (Salomon *et al.* 2005, 2007). The discrepancy between electrical and spectroscopy data is due to the fact that charge-carrier transport is dominated by the presence of interface states localized between the molecular HOMO (highest-occupied molecular orbital) and LUMO (lowest-unoccupied molecular orbital) and the Si band edges (Salomon *et al.* 2007).

**Fig. 9.3** Left: Tunnel decay factor–energy barrier plot for several molecular tunnel junctions: (■) metal–alkylthiol or dithiol–metal (Au or Hg) junctions (Cui *et al.* 2001; Holmlin *et al.* 2001; Wang *et al.* 2003), (▲) Au-alkylthiol or dithiol-Au C-AFM junctions (Wold and Frisbie 2001; Beebe *et al.* 2002; Wold *et al.* 2002; Engelkes *et al.* 2004), (◀) LB monolayer (Mann and Kuhn 1971), (◆) Si-alkyl-Hg junction (Salomon *et al.* 2005, 2007), (●) Si-native SiO<sub>2</sub>-alkylsilane-Al junction (Boulaas *et al.* 1996; Vuillaume *et al.* 1998), (▼) Si-native SiO<sub>2</sub>-mercaptopropyltrimethoxysilane-Au junction (Aswal *et al.* 2005). Lines are calculated according to the classical equation (see text) for different values of the effective mass.



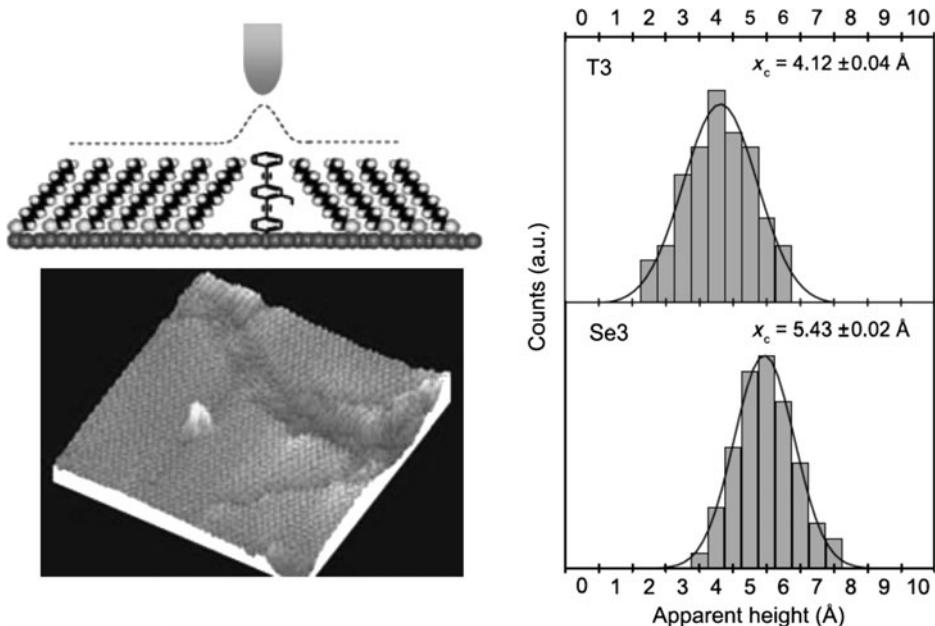
These puzzling data may be rationalized if we consider the nature of the molecule/electrode coupling. Figure 9.3 shows some of these data in a  $\beta - \Delta$  plot. The smallest  $\beta$  and  $\Delta$  values are obtained for a good or “intimate” coupling at both the two electrodes. This is the case for SAM of alkylthiols chemisorbed at the two electrodes (Holmlin *et al.* 2001; Cui *et al.* 2002) and for SAM chemisorbed at one end and contacted at the other one by an evaporated metal (Lenfant 2001). This is also the case for alkyl chains directly attached to Si without native oxide between the substrate and the molecules (Salomon *et al.* 2005, 2007). The largest values are obtained when at least one coupling is weak, as is the case for physisorbed LB monolayers (Mann and Kuhn 1971; Polymeropoulos 1977; Polymeropoulos and Sagiv 1978) and SAM mechanically contacted by C-AFM tip (Wold and Frisbie 2000, 2001) or chemisorbed on the native oxide of the Si substrate (Boulaas *et al.* 1996; Vuillaume *et al.* 1998). In this latter case, the top metal electrode (Al or Au) was also weakly coupled with the CH<sub>3</sub>-terminated molecules. The tunnel barrier height is lowered (2.2–2.5 eV) (Aswal *et al.* 2005) if Au is used as the top electrode on thiol-terminated SAM of alkyl chains still grafted on naturally oxidized Si, probably due to a better molecule/metal coupling through the S–Au chemical link. This feature reveals that the nature of the molecule–electrode coupling strongly changes the electronic properties of the molecules. The HOMO–LUMO gap of the molecule, and therefore the tunnel barrier height, may be reduced by several eV for a chemisorbed molecule on metal compared to the gas-phase molecule (Vondrak *et al.* 1999). Charge transfer and interface dipole also move the position of the molecular orbitals with respect to the Fermi energy of the electrodes. A review on these phenomena is given by Kahn *et al.* (2003) and Cahen *et al.* (2005). The molecule–electrode contact is a key parameter in the overall transport properties of the MmM junctions. It was demonstrated that the conductance of a MmM junction is increased when the molecule is chemisorbed at its two ends (via a thiol link on gold for instance) compared to the situation when only one end is chemically connected

to one electrode. An increase by a factor  $10^3$  was observed for a monolayer of octadecanedithiol molecules as compared to a monolayer of octadecanethiol (Cui *et al.* 2001; Beebe *et al.* 2002). Another experimental evidence is given by a comparison of two systems (Hg-S-alkyl and Hg/alkyl) where the sulfur-linked molecules showed a better electrical conductivity (Selzer *et al.* 2002a).

Finally, these tunnel junctions are also good prototypical devices to study more detailed phenomena such as: electron–molecular vibration coupling using inelastic electron tunnel spectroscopy (IETS) (Kushmerick *et al.* 2004; Wang *et al.* 2004; Petit *et al.* 2005; Aswal *et al.* 2006; Long *et al.* 2006; Beebe *et al.* 2007), current-induced local heating in a molecular junction (Huang *et al.* 2006), dynamical charge fluctuations using noise measurements (Clement *et al.* 2007) and spin-polarized transport (Petta *et al.* 2004; Wang and Richter 2006). Beyond the first results, more such experiments are now required to achieve a good agreement between a variety of different results, as well as with theoretical predictions. These approaches open a very interesting pathway toward a better understanding of electronic transport in molecular junctions.

## 9.4 Molecular semiconducting wire

Contrary to the case of fully saturated alkyl chains, short oligomers of  $\pi$ -conjugated molecules are considered as the prototype of molecular semiconducting wires. At low bias, when the LUMO and HOMO of the molecules are not in resonance within the Fermi energy window opened between the two electrodes by the applied bias, the conduction is still dominated by tunnelling. However, the decay factor  $\beta$  is lower than in the case of alkyl chains (see *supra*), typically  $\beta \sim 0.2$  to  $0.6 \text{ \AA}^{-1}$ . This is related to the lower HOMO–LUMO gap of the  $\pi$ -conjugated molecules ( $\sim 2\text{--}4 \text{ eV}$ , typically, against  $8\text{--}9 \text{ eV}$  for alkyl chains), and therefore to a lower energy barrier for charge injections. A detailed comparison of transport properties between saturated and  $\pi$ -conjugated molecules is given in Salomon *et al.* (2003) Bumm and coworkers (1996) have studied the conductivity of prototypes of molecular wires. A few molecules of di(phenylene-ethynylene)benzenethiolate were inserted in a SAM of dodecanethiols (which are insulating molecules), and the difference in conductivity was investigated using the tip of a STM. With a STM working at a constant current, the tip is retracted when passing over a more conducting molecule than the surrounding matrix of alkyl chains. Thus, the apparent amplitude height in the STM image is directly related to the conducting behavior of these molecules. Patrone and coworkers (2002, 2003b) have repeated these experiments for thiolterthiophene molecules, another prototype of molecular wires (Fig. 9.4). However, as explained *supra*, the drawback of these experiments is the fact that the electrical “contact” at the upper end of the molecules occurs through the air gap between the SAM surface and the STM tip (or vacuum in the case of an UHV-STM). This leads to a difficult estimation of the true conductance of the molecules. Reed *et al.* (1997), Kergueris *et al.* (1999) and Weber *et al.*



**Fig. 9.4** Top left, schematic view of a mixed monolayer where a few “conducting” molecules (dithiol-terthiophene) are intercalated into “insulating” ones (alkanethiol) used for STM measurements. Bottom left, STM image ( $28\text{ nm} \times 28\text{ nm}$ ). The bump in the image is due to a higher current when the tip is passing over the more conducting terthiophene molecules. The background corresponds to the tunnelling current through the alkanethiols (Patrone *et al.* 2002, 2003b). Right, comparison of the apparent height (which is related to the molecular conductance) measured on the STM images for the S- and Se-linked terthiophene molecules—T3 and Se3, respectively (histogram taken from many measurements). Copyright (2002) with permission from Elsevier.

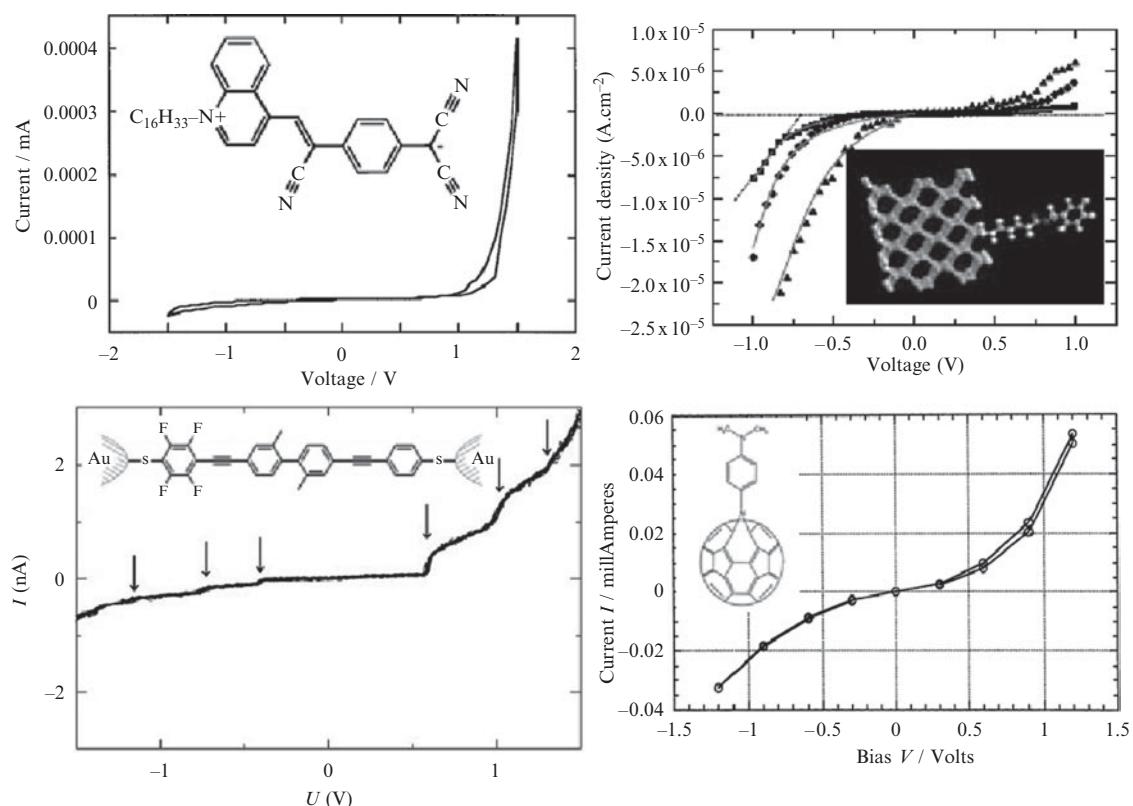
(Reichert *et al.* 2002; Weber *et al.* 2002; Elbing *et al.* 2005) have used breaking junctions to fabricate and to study some MmM junctions based on short conjugated oligomers. The current–voltage curves are strongly non-linear with steps (peaks in the first derivative) corresponding to resonant charge carrier transfer through the molecular orbitals (MO) of the molecules. The measured conductance corresponds to the conductance through the molecules and the conductance of the molecule/electrode contact. Thus, the influence of the chemical link between the molecules and the electrode is of prime importance. A change from an asymmetric to a symmetric current–tension (with respect to the bias polarity) curve was observed when comparing MmM junctions of SAMs of monothiolate and dithiolate oligo(phenylene ethynylene) molecules (Kushmerick *et al.* 2002b). The current increases by about a factor 10 when a sulfur atom attaches the molecule to the gold electrode compared to a mechanical contact. Today, the thiol group is the most used link to gold. However, theoretical calculations have recently predicted that selenium (Se) and tellurium (Te) are better links than sulfur (S) for the electronic transport through MmM junctions based on phenyl-based molecular wires (Yaliraki *et al.* 1999; Di Ventra and Lang 2001). This was recently demonstrated in a series of experiments using SAMs made of bisthiol- and biselenol-terthiophene molecules inserted in a dodecanethiol matrix (Patrone *et al.* 2002, 2003a). Using both STM in ambient air and UHV-STM, the apparent height of the molecular wires above

the dodecanethiol matrix (as in the Bumm *et al.* work quoted above, Bumm *et al.* 1996) is used to compare the electron transfer through the terthiophene molecule linked to the gold surface by S or Se atoms. Whatever the experimental conditions (air or UHV, tip–substrate bias, tunnel current set-point), the Se-linked molecules always appear higher in the STM images than the ones with a S linker. This feature directly demonstrates that a Se atom provides a better electron coupling between the gold electrode and the molecular wire than a S atom does (at least for the terthiophene molecule used in these experiments). From UPS experiments, this was attributed to a reduction of the energy offset between the highest occupied molecular orbital (HOMO) of the molecules (these molecules are mainly a better hole-transport material than an electron-transport material) and the Fermi energy of the gold electrode (Patrone *et al.* 2002, 2003b). This offset reduction is in agreement with theory (Yaliraki *et al.* 1999; Di Ventra and Lang 2001). Similarly, comparing the electron transport through SAMs of alkylthiols and alkyl-isonitriles (C-AFM measurements), it was established that the contact resistance for the Au/CN link is about 10% lower than for the Au/S interface (Beebe *et al.* 2002). Further experiments have shown that: (i) amine group ( $\text{NH}_2$ ) give better controlled conductance variability than thiol (SH) and isonitrile (CN) (Venkataraman *et al.* 2006b) and (ii) the interface contact resistance is lower for amine than for thiol (Chen *et al.* 2006). Further experiments are now required to deeply investigate all possible anchoring atom/electrode couples (S, Se, Te, CN, COOH, etc . . . , on one side and Au, Ag, Pt, Pd, for instance, on the other side) and to determine to what extent the conclusions drawn for a peculiar molecule are valid for any other ones. With all these data to hand, one would optimize the design of future devices for molecular electronics. Electron–molecular vibronic coupling in short semiconducting oligomers has also been recently studied by IETS (Kushmerick *et al.* 2004; Long *et al.* 2006) as for alkane molecules, as well as thermoelectricity in these molecular junctions (Reddy *et al.* 2007). In this latter case, the Seebeck coefficient of the single molecules has been determined, as well as a clear evidence of hole transport through the junctions. This result allows exploration of thermoelectric energy conversion at the molecular scale.

## 9.5 Molecular rectifying diode

A basic molecular device is the electrical current rectifier based on suitably engineered molecules. This molecular diode is the organic counterpart of the semiconductor p-n junction. At the origin of this idea, Aviram and Ratner (AR) proposed in 1974 to use D– $\sigma$ –A molecules where D and A are, respectively, electron donor and acceptor, and  $\sigma$  is a covalent “sigma” bridge (Aviram and Ratner 1974). Several molecular rectifying diodes were synthesized based on this AR paradigm, with donor and acceptor moieties linked by a short  $\sigma$  or even  $\pi$  bridge (Ashwell *et al.* 1990; Metzger and Panetta 1991; Martin *et al.* 1993; Metzger *et al.* 1997, 2001; Metzger 1999; Vuillaume *et al.* 1999; Xu *et al.* 2001). This D–b–A (b=bridge) group is also  $\omega$ -substituted by an alkyl chain to allow a monolayer formation by the

Langmuir–Blodgett (LB) method and this LB monolayer is then sandwiched in a metal/monolayer/metal junction. The first experimental results were obtained with the hexadecylquinolinium tricyanoquinodimethane molecule ( $C_{16}H_{33}-Q-3CNQ$  for short)—Fig. 9.5 (Ashwell *et al.* 1990; Martin *et al.* 1993; Metzger *et al.* 1997, 2001; Vuillaume *et al.* 1999; Xu *et al.* 2001). However, the chemical synthesis of this molecule was not obvious with several routes leading to erratic and unreliable results. A more reliable synthesis was reported with a yield of 59% (Metzger *et al.* 1997). More recently, other D–b–A molecules have been synthesized and tested (Baldwin *et al.* 2002; Metzger *et al.* 2003) showing rectification with a ratio up to  $\sim 2 \times 10^4$ . We can also mention some other approaches using D–A diblock co-oligomers (Ng *et al.* 2002) or CNT asymmetrically functionalized by D and A moieties at their ends (Wei *et al.* 2006) with a rectification ratio of  $\sim 10^3$  in this latter case. Even if these results represent an important progress to achieve molecular electronics, the physical mechanism responsible for the rectification is not



**Fig. 9.5** Typical current–voltage characteristics of some molecular rectifying diodes. From top to bottom: LB monolayer of D– $\pi$ –A molecules (tricyanoquinodimethane) between metal electrodes, from Metzger *et al.* (1997); Vuillaume *et al.* (1999),  $\sigma - \pi$  molecule grafted on Si ( $\sigma$  is alkyl chain and  $\pi$  groups are thiophene ( $\blacktriangle$  and  $\blacklozenge$ ) and phenyl ( $\blacksquare$ )), from Lenfant *et al.* (2003, 2006), D–A molecule inserted in a break junction (at 30 K in this latter case), from Elbing *et al.* (2005) (copyright 2005, National Academy of Science, USA), and D–b–A (dimethylanilinoazafullerene) LB monolayers, from Metzger *et al.* (2003).

clear. One critical issue is to determine if the AR model can be applied to C<sub>16</sub>H<sub>33</sub>-Q-3CNQ because it is a D- $\pi$ -A molecule (Metzger *et al.* 1997), and due to the  $\pi$  bridge, the HOMO and LUMO may be more delocalized than expected in the AR model. On the theoretical side, these molecular diodes are complex systems, characterized by large and inhomogeneous electric fields, which result from the molecular dipoles in the monolayer, the applied bias and the screening induced by the molecules themselves and the metallic electrodes. A theoretical treatment of these effects requires a self-consistent resolution of the quantum-mechanical problem, including the effect of the applied bias on the electronic structure. Combining *ab initio* and semi-empirical calculations, it was shown (Krzeminski *et al.* 2001) that the direction of easy current flow (rectification current) depends not only on the placement of the HOMO and LUMO relative to the Fermi levels of the metal electrodes before bias is applied, but also on the shift induced by the applied bias: this situation is more complex than the AR mechanism, and can provide a rectification current in an opposite direction. The electrical rectification results from the asymmetric profile of the electrostatic potential across the system (Krzeminski *et al.* 2001; Stokbro *et al.* 2003). In other words, this means that the molecule is more strongly coupled with one electrode than with the other one (more closer to one of the electrodes due to the presence of the alkyl chain). The alkyl tail in the C<sub>16</sub>H<sub>33</sub>-Q-3CNQ molecule plays an important role in this asymmetry, and it was predicted (Krzeminski *et al.* 2001) a symmetric current-voltage curve in the case of molecules without the alkyl chain. This asymmetry effect was further theoretically studied more extensively (Kornilovitch *et al.* 2002; Taylor *et al.* 2002). Generally speaking, any asymmetric coupling of the molecules with the electrodes or any asymmetry in the molecule will result in a rectification effect (Datta *et al.* 1997; Elbing *et al.* 2005)—Fig. 9.5. This emphasizes the importance of the electrostatic potential profile in a molecular system and suggests that this profile can be chemically engineered to build new devices. For instance, based on these considerations, we have recently reported an experimental demonstration of a simplified and more robust synthesis of a molecular rectifier with only one donor group and an alkyl spacer chain (Lenfant *et al.* 2003, 2006). We have used a sequential self-assembly process (chemisorption directly from solution) on silicon substrates. We have analyzed the properties of these molecular devices as a function of the alkyl chain length and for ten different donor groups. We have obtained rectification ratios up to 37 (Fig. 9.5). We have shown that rectification occurs from resonance through the HOMO of the  $\pi$ -group in good agreement with our calculations and internal photoemission spectroscopy. However, improvements are still required to suppress Fermi-level pinning at the molecule/metal interface (Lenfant *et al.* 2006) and to allow a clear design and tuning of the electrical behavior of the molecular diode through the right choice of the chemical nature of the molecule. This approach will allow us to fabricate molecular rectifying diodes compatible with silicon nanotechnologies for future hybrid circuitries. Finally, more efforts have also been put forward to design and synthesis new D-b-A molecules not affected by the presence of an asymmetric alkyl chain (see Fig. 9.5 for one example) (Baldwin *et al.* 2002; Metzger *et al.* 2003; Honciuc *et al.* 2007).

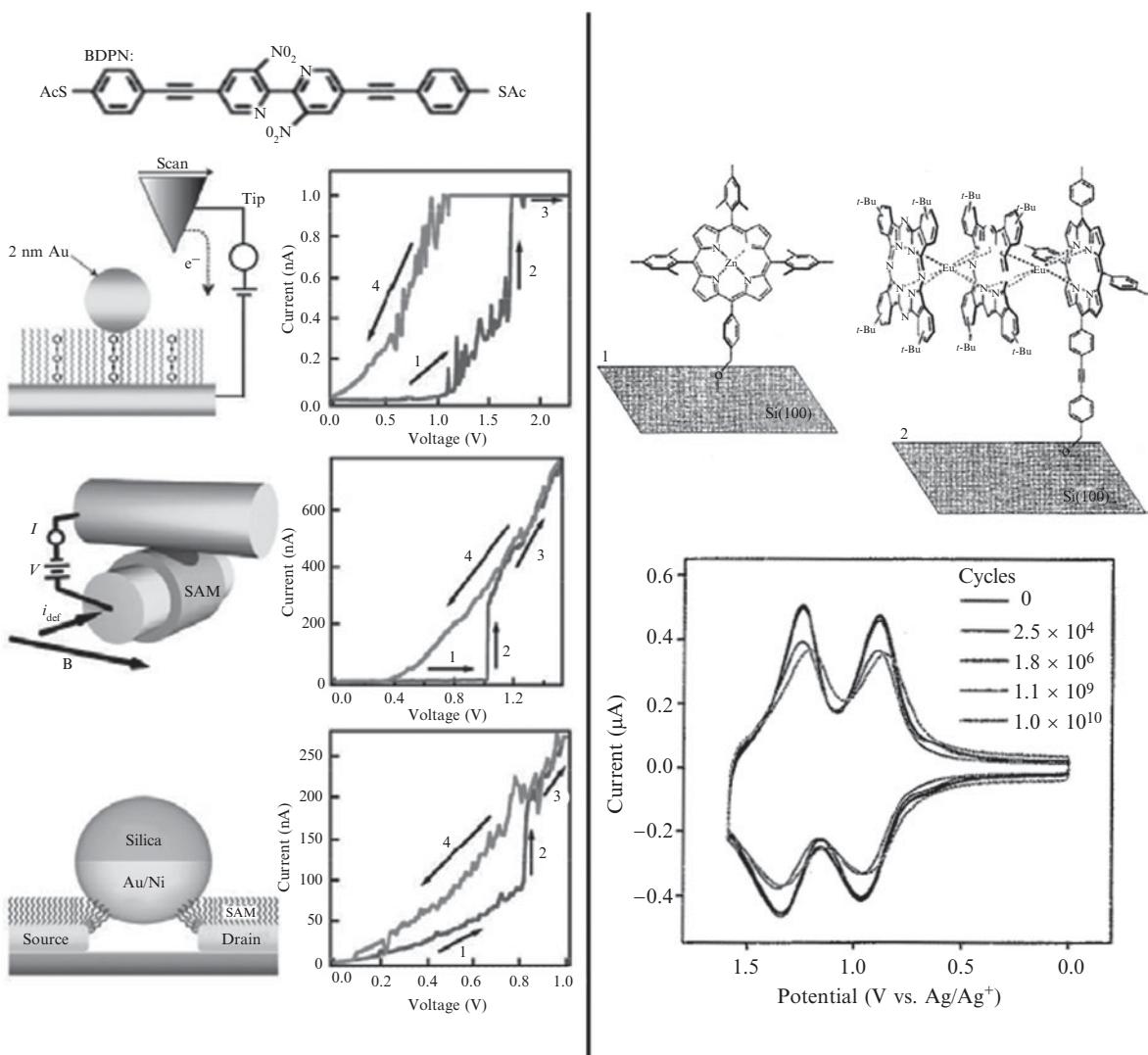
## 9.6 Molecular switches and memories

Molecular switches and memories were also suggested in the early stage of the molecular electronics history (Aviram 1988; Aviram *et al.* 1988, 1989). We generally distinguish three approaches called “conformational memory”, “charge-based memory” and “RTD-based memory” (RTD is resonant tunnelling diode). The first one relies on the idea to store a data bit on two bistable conformers of a molecule; the second on different redox states and the third on a negative differential resistance (NDR) due to resonant tunnelling through molecular orbitals.

### 9.6.1 Conformational memory

One of the most interesting possibilities for molecular electronics is to take advantage of the soft nature of organic molecules. Upon a given excitation, molecules can undergo conformational changes. If two different conformations are associated with two different conductivity levels of the molecule, this effect can be used to make molecular switches and memories. Such an effect is expected in  $\pi$ -conjugated oligomers used as molecular wires, if one of the monomers is twisted away from a planar conformation of the molecule (Venkataraman *et al.* 2006a). Twisting one monomer breaks the conjugation along the backbone, thus reducing the charge-transfer efficiency along the molecule. This has been experimentally observed for a small molecular wire where the central unit was substituted with redox moieties. With the nanopore configuration to fabricate the MmM junction, Chen and coworkers (Chen *et al.* 1999, 2000) have observed that molecules with a nitroamine redox center (2'-amino-4, 4'-di(ethynylphenyl)-5'-nitro-1-benzenethiol) exhibit a negative differential resistance behavior. In other words, they have observed that for a certain voltage range (typically between 1.5 and 2.2 V) applied on the MmM junction, the conductivity of the junction increased by a factor  $10^3$  (At 60 K, while the on/off ratio dropped to 1 at about 140 K. Other molecules with some changes of the redox moieties have exhibited on/off ratio of about 1.5 at RT (Chen *et al.* 2000)). They have also reported the feasibility of molecular random access memory cell using these molecules (Reed *et al.* 2001). The switching behavior of these compounds inserted in an alkanethiol SAM was also observed by STM (Donhauser *et al.* 2001). To separate the intrinsic behavior of the molecules from the molecule/metal interface, the same types of molecules have been measured on various test-beds (Fig. 9.6) (Szuchtmacher Blum *et al.* 2005). These experiments demonstrated a clear bias-induced switching, while with a large statistical variability. However, it is not firmly established that this switching behavior is solely due to the molecules. Recently, Lindsey's group showed that another possible mechanism is a random and temporary break in the chemical link between the molecule and the gold surface (Ramachandran *et al.* 2003) and this point is still a subject of debate.

Catenane and rotaxane are a class of molecules synthesized to exhibit a bistable behavior. In brief, these molecules are made of two parts, one allowed



**Fig. 9.6** Left: Current–voltage characteristics of bipyridyl-dinitro oligophenylene-ethynylene dithiol connected by Au electrodes using different test-beds (top to bottom): Au nanoparticle with STM, crossed wires put in contact by the Lorentz force and Ni/Au metallized microsphere used as a magnetic bead junction. These experiments demonstrate a clear bias-induced switching behavior, while with a large variability. From Szuchmacher Blum *et al.* (2005); (reprinted by permission from Macmillan Publishers Ltd: *Nature Materials*, copyright 2005). Right: Typical redox molecules (porphyrin derivatives) attached to a silicon substrate used in a charge-based molecular memory device and its electrical response as a function of the number of write/erase cycles. This electrochemical response show 2 redox states that can be used to implement a multilevel memory, from Liu *et al.* (2003) reprinted with permission from AAAS.

to move around or along the other one (e.g. a ring around a rod, two interlocked rings). These molecules adopt two different conformations depending on their redox states, changing the redox state triggers the displacement of the mobile part of the structure to minimize the total energy. This kind of molecule was used to build molecular memories. A MmM junction using a LB monolayer of these molecules mixed with phospholipid acid showed a clear electrical bistable behavior at room temperature (Collier *et al.* 2000, 2001;

Pease *et al.* 2001). A voltage pulse of about 1.5–2 V was used to switch the device from the “off” state to its “on” state. The state was read at a low bias (typically 0.1–0.2 V). The on/off ratio was about a few tens. A pulse in reverse bias (−1.5 to −2 V) returned the device to the “off” state. Using these molecular devices, Chen and coworkers (Chen *et al.* 2003a,b) have demonstrated a 64-bit non-volatile molecular memory cross-bar with an integration density of 6.4 Gbit/cm<sup>2</sup> (a factor ∼10 larger than the current state-of-the-art silicon memory chip). The fabrication yield of the 64-bit memory is about 85%, the data retention is about 24 h and about 50–100 write/erase cycles are possible before the collapse of the on/off ratio to 1. Recently, a 160-kbit based on the same class of molecules has been reported, patterned at a 33-nm pitch ( $10^{11}$  bits/cm<sup>2</sup>) (Green *et al.* 2007). About 25% of the tested memory points passed an on/off ratio larger than 1.5 with an average retention time of ∼1 h. However, it has also been observed that similar electrical switching behaviors can be obtained without such a class of bistable molecules (i.e. using simple alkyl chains instead of the rotaxanes) (Stewart *et al.* 2004). The switching behavior is likely due to the formation and breaking of metallic microfilaments introduced through the monolayer during the top-metal evaporation. The presence of such filaments is not systematic (see discussion *supra*), however, caution has to be taken before to definitively ascribe the memory effect as entirely due to the presence of the molecules. While having rather poor performances at the moment, these demonstrations allow us to envision the coming era of hybrid electronics, where molecular cross-bar memories like these, will be addressed by a multiplexer/demultiplexer and so one fabricated with standard semiconductor CMOS technologies (Chen *et al.* 2003a). The advantages of such molecular cross-bar memories are (i) a low cost, (ii) a very high integration density, (iii) a defect-tolerant architecture, (iv) an easy postprocessing onto a CMOS circuitry and (v) a low power consumption. For instance, it has been measured that an energy of ∼50 zJ (or ∼0.3 eV) is sufficient to rotate the dibutyl-phenyl side group of a single porphyrin molecule (Loppacher *et al.* 2003). This is ∼10<sup>4</sup> lower than the energy required to switch a state-of-the art MOSFET, and near the  $kT\ln 2$  (2.8 zJ at 300 K, or 0.017 eV) thermodynamic limit.

### 9.6.2 Charge-based memory

The redox-active molecules, such as metallocene, porphyrin and triple-decker sandwich coordination compounds attached on a silicon substrate have been found to act as charge-storage molecular devices (Li *et al.* 2002; Roth *et al.* 2002, 2003; Liu *et al.* 2003). The molecular memory works on the principle of charging and discharging of the molecules into different chemically reduced or oxidized (redox) states. It has been demonstrated that porphyrins (i) offer the possibility of multibit storage at a relatively low potentials (below ∼1.6 V), (ii) can undergo trillions of write/read/erase cycles, (iii) exhibit charge-retention times that are long enough (minutes) compared with those of semiconductor DRAM (tens of ms) and (iv) are extremely stable under harsh conditions (400 °C—30 min) and therefore

meet the processing and operating conditions required for use in hybrid molecule/silicon devices (Liu *et al.* 2003). Moreover, the same principle works with semiconducting nanowires dressed with redox molecules in a transistor configuration (Duan *et al.* 2002; Li *et al.* 2004a,b). Optoelectronic memories have also been demonstrated with polymer-functionalized CNT transistors (Star *et al.* 2004; Borghetti *et al.* 2006). However, in all cases, further investigations on the search of other molecules and, understanding the factors that control parameters such as, charge transfer rate, which limit write/read times, and charge retention times, which determines refresh rates, are needed.

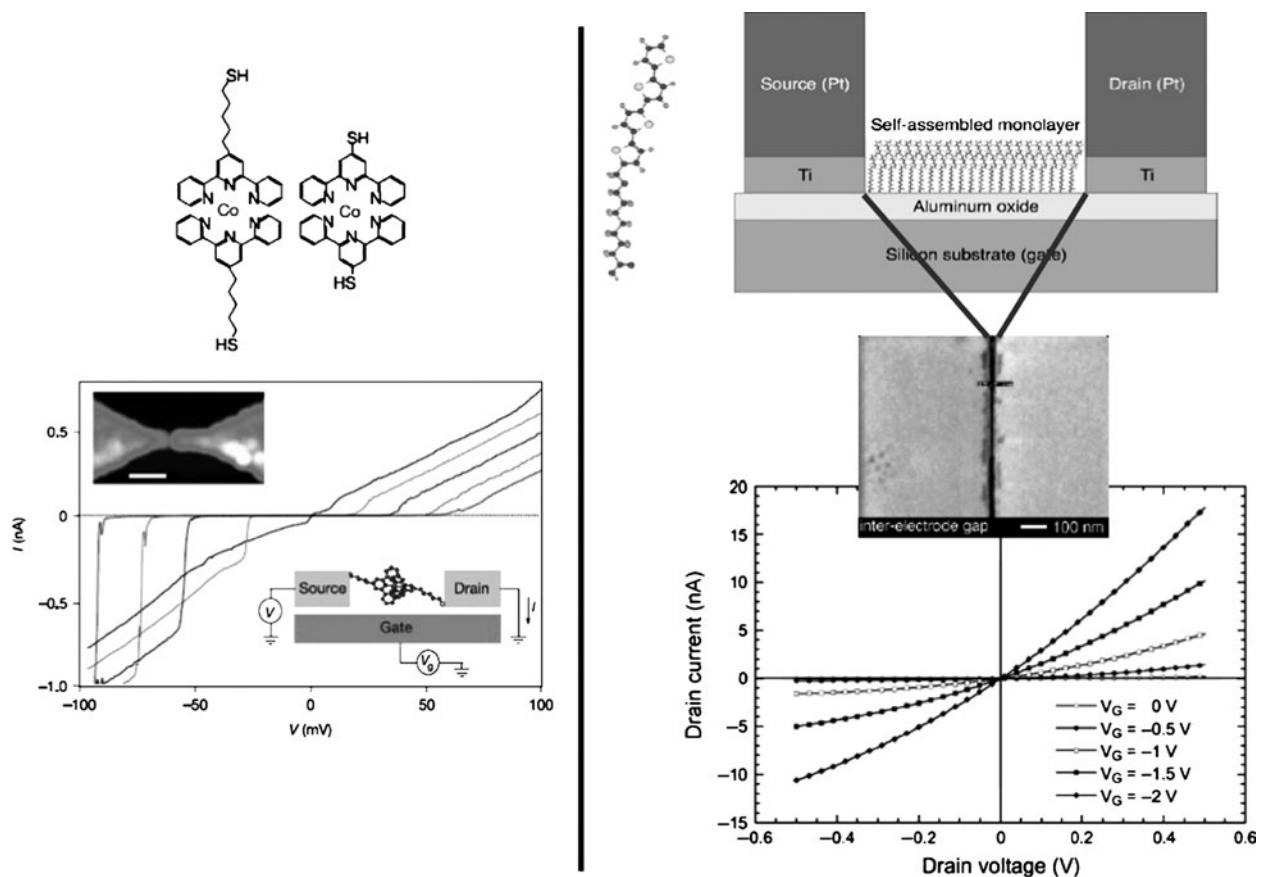
### 9.6.3 RTD-based memory

Memory can also be implemented from RTD devices following cell architecture already used for semiconductor devices. Memory cell-based on RTD can be set up with 2 RTD and 2 transistors in a cross-bar architecture (Van Der Wagt *et al.* 1998). The advantages compared to “resistive” and “capacitive” molecular memories are fast switching times and possible long retention times. RTD devices are characterized by a NDR behavior in their current-voltage curves. Many papers reported NDR behavior through molecular junctions (Chen *et al.* 1999, 2000; Gorman *et al.* 2001; Amlani *et al.* 2002; Kratochvilova *et al.* 2002; Rawlett *et al.* 2002; Le *et al.* 2003; Li *et al.* 2003) with peak-to-valley ratio from about 1.5 to 5 (at room temperature). However, a NDR may be also induced by other physical phenomena such as conformational changes already discussed *supra* or thiol-gold bond fluctuations (Ramachandran *et al.* 2003). The principle of a RTD molecular device is similar to that of his solid-state counterpart (a potential well separated from the electrodes by two tunnel barriers). In the molecular analog, the barriers should consist of aliphatic chains (of variable length) and the well should be made up of a short conjugated oligomer. Even if NDR behavior has been observed from STM results on a single molecule attached to Si (Guisinger *et al.* 2004) and has been ascribed to resonance through the molecular orbitals in agreement with a theoretical result (Rakshit *et al.* 2004), this interpretation has been ruled out both experimentally (Pitters and Wolkow 2006) and theoretically (Quek *et al.* 2007). In detailed STM experiments, Pitters and Wolkow showed that the NDR behavior might be explained by random changes of the conformation of molecules on the surface (molecular rearrangement, desorption and/or decomposition) and not by a resonant tunnelling through the molecule orbitals. These random phenomena are triggered by inelastic interactions between the molecular vibrations and electrons passing through the molecules. Resonant tunnelling was also theoretically questioned by Quek *et al.* (2007) using density-functional theory and a many-electron GW self-energy approach. They showed, for the specific molecule (e.g. cyclopentene on silicon) that the frontier energy levels do not move with the applied electric field, thus the molecular orbitals could not align with the silicon energy bands at certain applied bias. In conclusion, the exact origin of the molecular NDR behavior is still an

open question, and therefore the RTD molecular device was not yet clearly demonstrated.

## 9.7 Molecular transistor

A true transistor effect (i.e. the current through 2 terminals of the device controlled by the signal applied on a third terminal) embedded in a single three-terminal molecule (e.g. a star-shaped molecule) has not yet been demonstrated. To date, only hybrid-transistor devices have been studied. The typical configuration consists of a single molecule or an ensemble of molecules (monolayer) connected between two source and drain electrodes separated by a nanometer-scale gap, separated from an underneath gate electrode by a thin dielectric film—Fig. 9.7. At a single-molecule level (single-molecule transistor), these devices have been used to study Coulomb-blockade effects and Kondo effects at very low temperature. For instance, Coulomb blockade (electrons flowing one-by-one between source and drain through the molecule due to electron–electron Coulomb repulsion, the molecule acting as a quantum dot) was observed for molecules such as fullerene ( $C_{60}$ ) and oligo-phenyl-vinylene (OPV) weakly coupled to the source–drain electrodes. (Park *et al.* 2000; Kubatkin *et al.* 2003). In this latter case, up to 8 successive charge states of the molecule have been observed. With organo-metallic molecules bearing a transition metal, such as cobalt terpiridynil complex and divanadium complex, Kondo resonance (formation of a bound state between a local spin on the molecule, or an island, or a quantum dot, and the electrons in the electrodes leading to an increase of the conductance at low bias, around zero volts) has also been observed in addition to Coulomb blockade (Liang *et al.* 2002; Park *et al.* 2002). Kondo resonance is observed when increasing the coupling between the molecule and the electrodes (for instance by changing the length of the insulating tethers between the metal ion and the electrodes). At a monolayer level, self-assembled monolayer field effect transistors (SAMFET) have been demonstrated at room temperature (Tulevski *et al.* 2004; Mottaghi *et al.* 2007). The transistor effect is observed only if the source and drain length is lower than about 50 nm, that is, more or less matching the size of domains with well-organized molecules in the monolayer. This is mandatory to enhance  $\pi$  stacking within the monolayer and to obtain a measurable drain current. SAM of tetracene (Tulevski *et al.* 2004), terthiophene and quaterthiophene (Mottaghi *et al.* 2007) derivatives have been formed in this nanogap. Under this condition, a field effect mobility of about  $3.5 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  was measured for a SAMFET made with a quaterthiophene (4T) moiety linked to a short alkyl chain (octanoic acid) grafted on a thin aluminum oxide dielectric (Fig. 9.7). This value is on a par with those reported for an organic transistor made of thicker films of evaporated 4T ( $10^{-3}$  to  $10^{-2} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ ) (Mottaghi *et al.* 2007). The on/off ratio was about  $2 \times 10^4$ . For some devices, a clear saturation of the drain current vs. drain voltage curve has been observed, but usually, these output characteristics display a superlinear behavior. This feature has been explained by a gate-induced lowering of the charge injection energy barrier at the source/organic channel interface (Collet *et al.* 2000).



**Fig. 9.7** Left: Structure of the Co-terpyridyl complex molecules, AFM image of the source–drain nanogaps ( $\sim 1\text{--}2\text{ nm}$ ) made by electromigration, and typical  $I$ – $V$  with Coulomb-blockade gaps measured at  $100\text{ mK}$  for various gate voltage, and a schematic diagram of the device, from Park *et al.* (2002, reprinted by permission from Macmillan Publishers Ltd: *Nature*, copyright 2002). Right: Schematic diagram of the SAMFET and the 4T-octanoic acid molecule, SEM image of the 16-nm source–drain gap, and typical drain current–drain voltage curve for various gate voltages measured at  $300\text{ K}$ , from Mottaghi *et al.* (2007); Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

## 9.8 Conclusion

We have described several functions and devices that have been studied at the molecular scale: tunnel barrier, molecular wire, rectifying and NDR diodes, bistable devices and memories. However, a better understanding and further improvements of their electronic properties are mandatory and need to be confirmed. These results suffer from a large dispersion and more efforts are now required to improve reproducibility and repeatability. For viable applications, more efforts are also mandatory to test the integration of molecular devices with silicon-CMOS electronics (hybrid molecular–CMOS nanoelectronics). Moreover, most of these devices are 2-terminal, what about a true/fully molecular 3-terminal device? We have also pointed out that the molecule–electrode coupling and conformation strongly modify the molecular-scale device properties. Molecular engineering (changing ligand atoms for example) may be used

to improve or adjust the electrode–molecule coupling. Nevertheless, a better control of the interface (energetics and atomic conformation) is still compulsory. Beyond the study of single or isolated devices, more studies towards molecular architectures and circuits are required. Up to now, mainly the “cross-bar” architecture has been studied. Is it sufficient? More, new architectures must be explored (e.g. non-von Neuman, neuronal, quantum computing...). Open questions concern the right approaches for intermolecular device connections and nano-to-microconnections, the interface with the outer-world, hybridation with CMOS and 3D integration (Goldstein and Budiu 2001; Tour *et al.* 2002; Dehon *et al.* 2003; Likharev and Strukov 2005). To go beyond the CMOS limits probably relies on devices not working with electrical charges. Molecular devices using other state variables (e.g. spin, molecule conformation,...) to code a logic state are still challenging and exciting objectives. Finally, other reviews, current status and challenges on charge transfer on the nanoscale can be found (Nitzan 2001; Adams *et al.* 2003; Nitzan and Ratner 2003; Tao 2006).

## Acknowledgments

The studies done at IEMN were financially supported by CNRS, ministry of research, ANR-PNANO, IRCICA, EU-FEDER Region Nord-Pas de Calais and IFCPAR. I thank all the colleagues in the “molecular nanostructures and devices” group at IEMN and many others outside our group for fruitful collaborations.

## References

- Adams, D.M., Brus, L., Chidsey, C.E.D., Creager, S., Creutz, C., Kagan, C.R., Kamat, P.V., Lieberman, M., Lindsay, S.M., Marcus, R.A., Metzger, R.M., Michel-Beyerle, M.E., Miller, J.R., Newton, M.D., Rolison, D.R., Sankey, O., Schanze, K.S., Yardley, J., Zhu, X. Charge transfer on the nanoscale: Currents status. *J. Phys. Chem. B* **107**, 6668 (2003).
- Akkerman, H.B., Blom, P.W.M., De Leeuw, D.M., De Boer, B. Towards molecular electronics with large-area molecular junctions. *Nature* **441**, 69 (2006).
- Allara, D.L., Parikh, A.N., Rondelez, F. Evidence for a unique chain organization in long chain silane monolayers deposited on two widely different solid substrates. *Langmuir* **11**, 2357 (1995).
- Amlani, I., Rawlett, A.M., Nagahara, L.A., Tsui, R.K. An approach to transport measurements of electronic molecules. *Appl. Phys. Lett.* **80**, 2761 (2002).
- Ashwell, G.J., Sambles, J.R., Martin, A.S., Parker, W.G., Szablewski, M. Rectifying characteristics of Mg/(C<sub>16</sub>H<sub>33</sub>-Q3CNQ LB film)/Pt structures. *J. Chem. Soc. Chem. Commun.* **19**, 1374 (1990).
- Aswal, D.K., Lenfant, S., Guerin, D., Yakhmi, J.V., Vuillaume, D. A tunnel current in self-assembled monolayers of 3-mercaptopropyltrimethoxysilane. *Small* **1**, 725 (2005).
- Aswal, D.K., Petit, C., Salace, G., Guérin, D., Lenfant, S., Yakhmi, J.V., Vuillaume, D. Role of interfaces on the direct tunnelling and the inelastic tunnelling behaviors through metal/alkylsilane/silicon junctions. *Phys. Status Solidi. A* **203**, 1464 (2006).

- Aviram, A. Molecules for memory, logic, and amplification. *J. Am. Chem. Soc.* **110**, 5687 (1988).
- Aviram, A., Joachim, C., Pomerantz, M. Evidence of switching and rectification by a single molecule effected with a scanning tunnelling microscope. *Chem. Phys. Lett.* **146**, 490 (1988).
- Aviram, A., Joachim, C., Pomerantz, M. Errata on “Evidence of switching and rectification by a single molecule effected by a scanning tunnelling microscope”. *Chem. Phys. Lett.* **162**, 416 (1989).
- Aviram, A., Ratner, M.A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277 (1974).
- Baldwin, J.W., Amaresh, R.R., Peterson, I.R., Shumate, W.J., Cava, M.P., Amiri, M.A., Hamilton, R., Ashwell, G.J., Metzger, R. M. Rectification and nonlinear optical properties of a Langmuir-Blodgett monolayer of a pyridinium dye. *J. Phys. Chem. B* **106**, 12158 (2002).
- Beebe, J.M., Engelkes, V.B., Miller, L.L., Frisbie, C.D. Contact resistance in metal-molecule-metal junctions based on aliphatic SAMs: Effects of surface linker and metal work function. *J. Am. Chem. Soc.* **124**, 11268 (2002).
- Beebe, J.M., Moore, H.J., Lee, T.R., Kushmerick, J.G. Vibronic coupling in semi-fluorinated alkanethiol junctions: Implications for selection rules in inelastic electron tunnelling spectroscopy. *Nano Lett.* **7**, 1364 (2007).
- Bezryadin, A., Dekker, C. Nanofabrication of electrodes with sub-5 nm spacing for transport experiments on single molecules and metal clusters. *J. Vac. Sci. Technol. B* **15**, 793 (1997).
- Bigelow, W.C., Pickett, D.L., Zisman, W.A. *J. Colloid Sci.* **1**, 513 (1946).
- Borghetti, J., Derycke, V., Lenfant, S., Chenevier, P., Filoromo, A., Goffman, M., Vuillaume, D., Bourgoin, J.-P. Optoelectronic switch and memory devices based on polymer-functionalized carbon nanotube transistors. *Adv. Mater.* **18**, 2535 (2006).
- Boulas, C., Davidovits, J.V., Rondelez, F., Vuillaume, D. Suppression of charge carrier tunnelling through organic self-assembled monolayers. *Phys. Rev. Lett.* **76**, 4797 (1996).
- Boussaad, S., Tao, N.J. Atom-size gaps and contacts between electrodes fabricated with a self-terminated electrochemical method. *Appl. Phys. Lett.* **80**, 2398 (2002).
- Bruening, M., Moons, E., Yaron-Marcovitch, D., Cahen, D., Libman, J., Shanzer, A. Polar ligand adsorption controls semiconductor surface potentials. *J. Am. Chem. Soc.* **116**, 2972 (1994).
- Brzoska, J.B., Ben Azouz, I., Rondelez, F. Silanization of solid substrates: a step toward reproducibility. *Langmuir* **10**, 4367 (1994).
- Brzoska, J.B., Shahidzadeh, N., Rondelez, F. Evidence of a transition temperature for optimum deposition of grafted monolayer coatings. *Nature* **360**, 719 (1992).
- Bumm, L.A., Arnold, J.J., Cygan, M.T., Dunbar, T.D., Burgin, T.P., Jones II, L., Allara, D.L., Tour, J.M., Weiss, P.S. Are single molecular wires conducting? *Science* **271**, 1705 (1996).
- Bumm, L.A., Arnold, J.J., Dunbar, T.D., Allara, D.L., Weiss, P.S. Electron transfer through organic molecules. *J. Phys. Chem. B* **103**, 8122 (1999).
- Cahen, D., Kahn, A., Umbach, E. Energetics of molecular interfaces. *Mater. Today*, July/August, 32 (2005).

- Chen, F., Li, X., Hihath, J., Huang, Z., Tao, N.J. Effect of anchoring groups on single-molecule conductance: comparative study of thiom-, amine-, and carboxylic-acid-terminated molecules. *J. Am. Chem. Soc.* **128**, 15874 (2006).
- Chen, J., Reed, M.A., Rawlett, A.M., Tour, J.M. Large on-off ratios and negative differential resistance in a molecular electronic device. *Science* **286**, 1550 (1999).
- Chen, J., Wang, W., Reed, M.A., Rawlett, A.M., Price, D.W., Tour, J.M. Room-temperature negative differential resistance in nanoscale molecular junctions. *Appl. Phys. Lett.* **77**, 1224 (2000).
- Chen, Y., Jung, G.-Y., Ohlberg, D.A.A., Li, X., Stewart, D.R., Jeppesen, J.O., Nielsen, K.A., Stoddart, J.F., Williams, R.S. Nanoscale molecular-switch crossbar circuits. *Nanotechnol.* **14**, 462 (2003a).
- Chen, Y., Ohlberg, D.A.A., Li, X., Stewart, D.R., Williams, R.S., Jeppesen, J.O., Nielsen, K.A., Stoddart, J.F., Olynick, D.L., Anderson, E. Nanoscale molecular-switch devices fabricated by imprint lithography. *Appl. Phys. Lett.* **82**, 1610 (2003b).
- Cholet, S., Joachim, C., Martinez, J.P., Rousset, B. Fabrication of co-planar metal-insulator-metal solid state nanojunction down to 5 nm. *Eur. Phys. J. Appl. Phys.* **8**, 139 (1999).
- Clement, N., Pleutin, S., Seitz, O., Lenfant, S., Vuillaume, D.  $1/f^{\gamma}$  tunnel current noise through Si-bound alkyl monolayers. *Phys. Rev. B* **76**, 205407 (2007).
- Cohen, R., Bastide, S., Cahen, D., Libman, J., Shanzer, A., Rosenwaks, Y. Controlling surfaces and interfaces of semiconductors using organic molecules. *Opt. Mater.* **9**, 394 (1998).
- Cohen, R., Zenou, N., Cahen, D., Yitzchaik, S. Molecular electronic tuning of Si surfaces. *Chem. Phys. Lett.* **279**, 270 (1997).
- Collet, J., Tharaud, O., Chapoton, A., Vuillaume, D. Low-voltage, 30 nm channel length, organic transistors with a self-assembled monolayer as gate insulating films. *Appl. Phys. Lett.* **76**, 1941 (2000).
- Collet, J., Vuillaume, D. A nano-field effect transistor with an organic self-assembled monolayer as gate insulator. *Appl. Phys. Lett.* **73**, 2681 (1998).
- Collier, C.P., Jeppesen, J.O., Luo, Y., Perkins, J., Wong, E.W., Heath, J.R., Stoddart, J.F. Molecular-based electronically switchable tunnel junction devices. *J. Am. Chem. Soc.* **123**, 12632 (2001).
- Collier, C.P., Mattersteig, G., Wong, E.W., Luo, Y., Beverly, K., Sampaio, J., Raymo, F., Stoddart, J.F., Heath, J.R. A [2]catenane-based solid state electronically reconfigurable switch. *Science* **289**, 1172 (2000).
- Collier, C.P., Wong, E.W., Belohradsky, M., Raymo, F.M., Stoddart, J.F., Kuekes, P.J., Williams, R.S., Heath, J.R. Electronically configurable molecular-based logic gates. *Science* **285**, 391 (1999).
- Compano, R., Molenkamp, L., Paul, D.J. Technology roadmap for nanoelectronics. Brussels, European Commission, IST programme, Future and Emerging Technologies (2000).
- Cui, X.D., Primak, A., Zarate, X., Tomfohr, J., Sankey, O.F., Moore, A.L., Moore, T.A., Gust, D., Harris, G., Lindsay, S.M. Reproducible measurement of single-molecule conductivity. *Science* **294**, 571 (2001).
- Cui, X.D., Primak, A., Zarate, X., Tomfohr, J., Sankey, O.F., Moore, A.L., Moore, T.A., Gust, D., Nagahara, L.A., Lindsay, S.M. Changes in the electronic properties of a molecule when it is wired into a circuit. *J. Phys. Chem. B* **106**, 8609 (2002).

- Dadosh, T., Gordin, Y., Krahne, R., Khrivrich, I., Mahalu, D., Frydman, V., Sperling, J., Yacoby, A., Bar-Joseph, I. Measurement of the conductance of single conjugated molecules. *Nature* **436**, 677 (2005).
- Datta, S., Tian, W., Hong, S., Reifenberger, R., Henderson, J.I., Kubiak, C.P. Current-voltage characteristics of self-assembled monolayers by scanning tunnelling microscopy. *Phys. Rev. Lett.* **79**, 2530 (1997).
- Dehon, A., Lincoln, P., Savage, J.E. Stochastic assembly of sublithographic nanoscale interfaces. *IEEE Trans. Nanotechnol.* **2**, 165 (2003).
- Di Ventra, M., Lang, N.D. Transport in nanoscale conductors from first principles. *Phys. Rev. B* **65**, 045402 (2001).
- Donhauser, Z.J., Mantooth, B.A., Kelly, K.F., Bumm, L.A., Monnell, J.D., Stapleton, J.J., Price, D.W., Rawlett, A.M., Allara, D.L., Tour, J.M., Weiss, P.S. Conductance switching in single molecules through conformational changes. *Science* **292**, 2303 (2001).
- Duan, X., Huang, Y., Lieber, C.M. Nonvolatile memory and programmable logic from molecule-gated nanowires. *Nano Lett.* **2**, 487 (2002).
- Elbing, M., Ochs, R., Koentopp, M., Fischer, M., Von Hänisch, C., Weigend, F., Evers, F., Weber, H.B., Mayor, M. A single-molecule diode. *Proc. Natl. Acad. Sci. USA* **102**, 8815 (2005).
- Engelkes, V.B., Beebe, J.M., Frisbie, C.D. Length-dependent transport in molecular junctions based on SAMs of alkanethiols and alkanedithiols: Effects of metal work function and applied bias on tunnelling efficiency and contact resistance. *J. Am. Chem. Soc.* **126**, 14287 (2004).
- Fisher, G.L., Hooper, A.E., Opila, R.L., Allara, D.L., Winograd, N. The interaction of vapor-deposited Al atoms with COOH groups at the surface of a self-assembled alkanethiolate monolayer on gold. *J. Phys. Chem. B* **104**, 3267 (2000).
- Fisher, G.L., Walker, A.V., Hooper, A.E., Tighe, T.B., Bahnck, K.B., Skriba, H.T., Reinard, M.D., Haynie, B.C., Opila, R.L., Winograd, N., Allara, D.L. Bond insertion, complexation and penetration pathways of vapor-deposited aluminium atoms with HO- and CH<sub>3</sub>O-terminated organic monolayers. *J. Am. Chem. Soc.* **124**, 5528 (2002).
- Geddes, N.J., Sambles, J.R., Martin, A.S. Organic molecular rectifiers. *Adv. Mater. Opt. Electron.* **5**, 305 (1995).
- Goldstein, S.C., Budiu, M. Nanofabrics: spatial computing using molecular electronics. Int. Symp. on Computer Architecture, Göteborg, Sweden (2001).
- Gorman, C.B., Carroll, R.L., Fuierer, R.R. Negative differential resistance in patterned electroactive self-assembled monolayers. *Langmuir* **17**, 6923 (2001).
- Green, J.E., Choi, J.W., Boukai, A., Bunimovich, Y., Johnston-Halperin, E., Delonno, E., Luo, Y., Sherrif, B.A., Xu, K., Shin, Y.S., Tseng, H.-R., Stoddart, J.F., Heath, J.R. A 160-kilobit molecular electronic memory patterned at 10<sup>11</sup> bits per square centimeter. *Nature* **445**, 414 (2007).
- Guerin, D., Merckling, C., Lenfant, S., Wallart, X., Vuillaume, D. Silicon-molecules-metal junctions by transfer printing: chemical synthesis and electrical properties. *J. Phys. Chem. C* **111**, 7947 (2007).
- Guillorn, M.A., Carr, D.W., Tiberio, R.C., Greenbaum, E., Simpson, M.L. Fabrication of dissimilar metal electrodes with nanometer interelectrode distance for molecular electronic device characterization. *J. Vac. Sci. Technol. B* **18**, 1177 (2000).

- Guisinger, N.P., Greene, M.E., Basu, R., Baluch, A.S., Hersam, M.C. Room temperature negative differential resistance through individual organic molecules on silicon surfaces. *Nano Lett.* **4**, 55 (2004).
- Guo, X., Small, J.P., Klare, J.E., Wang, Y., Purewal, M.S., Tam, I.W., Hong, B.H., Caldwell, R., Huang, L., O'Brien, S., Yan, J., Breslow, R., Wind, S.J., Hone, J., Kim, P., Nuckolls, C. Covalently bridging gaps in single-walled carbon nanotubes with conducting molecules. *Science* **311**, 356 (2006).
- Guo, X., Whalley, A.C., Klare, J.E., Huang, L., O'Brien, S., Steigerwald, M.L., Nuckolls, C. Single-molecule devices as scaffolding for multicomponent nanostructure assembly. *Nano Lett.* **7**, 1119 (2007).
- Haick, H., Cahen, D. Making contact: Connecting molecules electrically to the macroscopic world. *Prog. Surf. Sci.* **83**, 217 (2008).
- He, J., Chen, B., Flatt, A.K., Stephenson, J.J., Doyle, C.D., Tour, J.M. Metal-free silicon-molecule-nanotube testbed and memory device. *Nature Mater.* **5**, 63 (2006).
- Herdt, G.C., Czanderna, A.W. Metal overlayer on organic functional groups of self-organized molecular assemblies. V. Ion scattering spectroscopy and X-ray photoelectron spectroscopy of Ag/COOH interfaces. *J. Vac. Sci. Technol. A* **13**, 1275 (1995).
- Holmlin, R.E., Haag, R., Chabiny, M.L., Ismagilov, R.F., Cohen, A.E., Terfort, A., Rampi, M.A., Whitesides, G.M. Electron transport through thin organic films in metal-insulator-metal junctions based on self-assembled monolayers. *J. Am. Chem. Soc.* **123**, 5075 (2001).
- Honciuc, A.M., R.M., Gong, A., Spangler, C.W. Elastic and inelastic tunnelling spectroscopy of a new rectifying monolayer. *J. Am. Chem. Soc.* **129**, 8310 (2007).
- Huang, Z., Xu, B., Chen, Y., Di Ventra, M., Tao, N.J. Measurement of current-induced local heating in a single molecule junction. *Nano Lett.* **6**, 1240 (2006).
- Joachim, C., Gimzewski, J.K., Aviram, A. Electronics using hybrid-molecular and monomolecular devices. *Nature* **408**, 541 (2000).
- Jung, D.R., Czanderna, A.W. Chemical and physical interactions at metal/self-assembled organic monolayer interfaces. *Crit. Rev. Solid State Mater. Sci.* **191**, 1 (1994).
- Jung, D.R., Czanderna, A.W., Herdt, G.C. Interactions and penetration at metal/self-assembled organic monolayer interfaces. *J. Vac. Sci. Technol. A* **14**, 1779 (1996).
- Kahn, A., Koch, N., Gao, W. Electronic structure and electrical properties of interfaces between metals and pi-conjugated molecular films. *J. Polym. Sci.: Part B: Polym. Phys.* **41**, 2529 (2003).
- Kergueris, C., Bourgoin, J.P., Palacin, S., Esteve, D., Urbina, C., Magoga, M., Joachim, C. Electron transport through a metal-molecule-metal junction. *Phys. Rev. B* **59**, 12505 (1999).
- Kervennic, Y.V., Van Der Zant, H.S., Morpurgo, A.F., Gurevitch, L., Kouwenhoven, L.P. Nanometer-spaced electrodes with calibrated separation. *Appl. Phys. Lett.* **80**, 321 (2002).
- Konstadinidis, K., Zhang, P., Opila, R.L., Allara, D.L. An in-situ X-ray photoelectron study of the interaction between vapor-deposited Ti atoms and functional groups at the surfaces of self-assembled monolayers. *Surf. Sci.* **338**, 300 (1995).
- Kornilovitch, P.E., Bratkovsky, A.M., Williams, R.S. Current rectification by molecules with asymmetric tunnelling barriers. *Phys. Rev. B* **66**, 165436 (2002).

- Kratochvilova, I., Kocirik, M., Zambova, A., Mbindyo, J., Mallouk, T.E., Mayer, T.S. Room temperature negative differential resistance in molecular nanowires. *J. Mater. Chem.* **12**, 2927 (2002).
- Krzeminski, C., Allan, G., Delerue, C., Vuillaume, D., Metzger, R.M. Theory of electrical rectification in a molecular monolayer. *Phys. Rev. B* **64**, 085405 (2001).
- Kubatkin, S., Danilov, A., Hjort, M., Cornil, J., Brédas, J.L., Stuhr-Hansen, N., Hedegard, P., Bjornholm, T. Single-electron transistor of a single organic molecule with access to several redox states. *Nature* **425**, 698 (2003).
- Kushmerick, J.G., Holt, D.B., Pollack, S.K., Ratner, M.A., Yang, J.C., Schull, T.L., Naciri, J., Moore, M.H., Shashidhar, R. Effect of bond-length alternation in molecular wires. *J. Am. Chem. Soc.* **124**, 10654 (2002a).
- Kushmerick, J.G., Holt, D.B., Yang, J.C., Naciri, J., Moore, M.H., Shashidhar, R. Metal-molecule contacts and charge transport across monomolecular layers: measurement and theory. *Phys. Rev. Lett.* **89**, 086802 (2002b).
- Kushmerick, J.G., Lazorcik, J., Patterson, C.H., Shashidhar, R., Seferos, D.S., Bazan, G. Vibronic contributions to charge transport across molecular junctions. *Nano Lett.* **4**, 643 (2004).
- Labonté, A.P., Tripp, S.L., Reifenberger, R., Wei, A. Scanning tunnelling spectroscopy of insulating self-assembled monolayers on Au(111). *J. Phys. Chem. B* **106**, 8721 (2002).
- Le, J.D., He, Y., Hoye, T.R., Mead, C.C., Kiehl, R.A. Negative differential resistance in a bilayer molecular junction. *Appl. Phys. Lett.* **83**, 5518 (2003).
- Lehn, J.-M. Supramolecular chemistry-scope and perspectives. Molecules, supermolecules and molecular devices (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **27**, 89 (1988).
- Lenfant, S. Monocouches organiques auto-assemblées pour la réalisation de diodes moléculaires. PhD, University of Lille (2001).
- Lenfant, S., Guerin, D., Tran Van, F., Chevrot, C., Palacin, S., Bourgoin, J.-P., Bouloussa, O., Rondelez, F., Vuillaume, D. Electron transport through rectifying self-assembled monolayer diodes on silicon: Fermi level pinning at the molecule - metal interface. *J. Phys. Chem. B* **110**, 13947 (2006).
- Lenfant, S., Krzeminski, C., Delerue, C., Allan, G., Vuillaume, D. Molecular rectifying diodes from self-assembly on silicon. *Nano Lett.* **3**, 741 (2003).
- Li, C., Fan, W., Lei, B., Zhang, D., Han, S., Tang, T., Liu, X., Liu, Z., Asano, S., Meyyappan, M., Han, J., Zhou, C. Multilevel memory based on molecular devices. *Appl. Phys. Lett.* **64**, 1949 (2004a).
- Li, C., Ly, J., Lei, B., Fan, W., Zhang, D., Han, J., Meyyappan, M., Thompson, M., Zhou, C. Data storage studies on nanowire transistors with self-assembled phorphyrin molecules. *J. Phys. Chem. B* **108**, 9646 (2004b).
- Li, C., Zhang, D., Liu, X., Han, S., Tang, T., Zhou, C., Fan, W., Koehne, J., Han, J., Meyyappan, M., Rawlett, A.M., Price, D.W., Tour, J.M. Fabrication approach for molecular memory arrays. *Appl. Phys. Lett.* **82**, 645 (2003).
- Li, C.Z., He, H.X., Tao, N.J. Quantized tunnelling current in the metallic nanogaps formed by electrodeposition and etching. *Appl. Phys. Lett.* **77**, 3995 (2000).
- Li, Q., Mathur, G., Homsi, M., Surthi, S., Misra, V., Malinovskii, V., Schweikart, K.-H., Yu, L., Lindsey, J.S., Liu, Z., Dabke, R.B., Yasseri, A.A., Bocian, D.F., Kuhr, W.G. Capacitance and conductance characterization of ferrocene-containing

- self-assembled monolayers on silicon surfaces for memory applications. *Appl. Phys. Lett.* **81**, 1494 (2002).
- Li, X., He, J., Hihath, J., Xu, B., Lindsay, S.M., Tao, N.J. Conductance of single alkanethiols: conduction mechanism and effect of molecule-electrode contacts. *J. Am. Chem. Soc.* **128**, 2135 (2006).
- Liang, W., Shores, M.P., Bockrath, M., Long, J.R., Park, H. Kondo effect in a single-molecule transistor. *Nature* **417**, 725 (2002).
- Likharev, K.K., Strukov, D.B. CMOL: Devices, circuits and architectures. In Cuniberti, G. (ed.) *Introduction to Molecular Electronics* (Springer, 2005).
- Liu, Z., Yasseri, A.A., Lindsey, J.S., Bocian, D.F. Molecular memories that survive silicon device processing and real-world operation. *Science* **302**, 1543 (2003).
- Long, D.P., Lazorcik, J.L., Mantooth, B.A., Moore, M.H., Ratner, M.A., Troisi, A., Yao, Y., Ciszek, J.W., Tour, J.M., Shashidhar, R. Effects of hydration on molecular junction transport. *Nature Mater.* **5**, 901 (2006).
- Long, D.P., Patterson, C.H., Moore, M.H., Seferos, D.S., Bazan, G., Kushmerick, J.G. Magnetic directed assembly of molecular junctions. *Appl. Phys. Lett.* **86**, 153105 (2005).
- Loo, Y.-L., Lang, D.V., Rogers, J.A., Hsu, J.W.P. Electrical contacts to molecular layers by nanotransfer printing. *Nano Lett.* **3**, 913 (2003).
- Loo, Y.-L., Willet, R.L., Baldwin, K.W., Rogers, J.A. *J. Am. Chem. Soc.* **124**, 7654 (2002).
- Loppacher, C., Guggiesberg, M., Pfeiffer, O., Meyer, E., Bammerlin, M., Lüthi, R., Schlitter, R.G., J.K., Tang, H., Joachim, C. Direct determination of the energy required to operate a single molecule switch. *Phys. Rev. Lett.* **90**, 066107 (2003).
- Mann, B., Kuhn, H. Tunnelling through fatty acid salt monolayers. *J. Appl. Phys.* **42**, 4398 (1971).
- Maoz, R., Sagiv, J. On the formation and structure of self-assembling monolayers. *J. Colloid Interf. Sci.* **100**, 465 (1984).
- Martin, A.S., Sambles, J.R., Ashwell, G.J. Molecular rectifier. *Phys. Rev. Lett.* **70**, 218 (1993).
- Metzger, R.M. The unimolecular rectifier: Unimolecular electronic devices are coming ... *J. Mater. Chem.* **9**, 2027 (1999).
- Metzger, R.M., Baldwin, J.W., Shumate, W.J., Peterson, I.R., Mani, P., Mankey, G.J., Morris, T., Szulczewski, G., Bosi, S., Prato, M., Comito, A., Rubin, Y. Electrical rectification in a Langmuir-Blodgett monolayer of dimethylanilinoazafullerene sandwiched between gold electrodes. *J. Phys. Chem. B* **107**, 1021 (2003).
- Metzger, R.M., Chen, B., Höpfner, U., Lakshminathan, M.V., Vuillaume, D., Kawai, T., Wu, X., Tachibana, H., Hughes, T.V., Sakurai, H., Baldwin, J.W., Hosch, C., Cava, M.P., Brehmer, L., Ashwell, G.J. Unimolecular electrical rectification in hexadecyquinolinium tricyanoquinodimethane. *J. Am. Chem. Soc.* **119**, 10455 (1997).
- Metzger, R.M., Panetta, C.A. The quest for unimolecular devices. *New, J. Chem.* **15**, 209 (1991).
- Metzger, R.M., Xu, T., Peterson, I.R. Electrical rectification by a monolayer of hexadecylquinolinium tricyanoquinodimethane measured between macroscopic gold electrodes. *J. Phys. Chem. B* **105**, 7280 (2001).

- Moresco, F., Meyer, G., Rieder, K.-H., Tang, H., Gourdon, A., Joachim, C. Conformational changes of single molecules induced by scanning tunnelling microscopy manipulation: A route to molecular switching. *Phys. Rev. Lett.* **86**, 672 (2001).
- Mottaghi, M., Lang, P., Rodriguez, F., Rumyantseva, A., Yassar, A., Horowitz, G., Lenfant, S., Tondelier, D., Vuillaume, D. Low-operating-voltage organic transistors made of bifunctional self-assembled monolayers. *Adv. Func. Mater.* **17**, 597 (2007).
- Ng, M.-K., Lee, D.-C., Yu, L. Molecular diodes based on conjugated diblock co-oligomers. *J. Am. Chem. Soc.* **124**, 11862 (2002).
- Nitzan, A. Electron transmission through molecules and molecular interfaces. *Annu. Rev. Phys. Chem.* **52**, 681 (2001).
- Nitzan, A., Ratner, M.A. Electron transport in molecular wire junctions. *Science* **300**, 1384 (2003).
- Okazaki, N., Sambles, J.R. New fabrication technique and current-voltage properties of a Au/LB/Au structure. International symposium on organic molecular electronics. Nagoya, Japan (2000).
- Parikh, A.N., Allara, D.L., Ben Azouz, I., Rondelez, F. An intrinsic relationship between molecular structure in self-assembled n-alkylsiloxane monolayers and deposition temperature. *J. Phys. Chem.* **98**, 7577 (1994).
- Park, H., Lim, A.K.L., Alivisatos, P.A., Park, J., McEuen, P.L. Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75**, 301 (1999).
- Park, H., Park, J., Lim, A.K.L., Anderson, E.H., Alivisatos, P.A., McEuen, P.L. Nanomechanical oscillations in a single-C<sub>60</sub> transistor. *Nature* **407**, 57 (2000).
- Park, J., Pasupathy, A.N., Goldsmith, J.I., Chang, C., Yaish, Y., Petta, J.R., Rinkoski, M., Sethna, J.P., Abrunas, H.D., McEuen, P.L., Ralph, D.C. Coulomb blockade and the Kondo effect in single-atom transistors. *Nature* **417**, 722 (2002).
- Patrone, L., Palacin, S., Bourgoin, J.-P. Direct comparison of the electronic coupling efficiency of sulfur and selenium alligator clips for molecules adsorbed onto gold electrodes. *Appl. Surf. Sci.* **212**, 446 (2003a).
- Patrone, L., Palacin, S., Bourgoin, J.-P., Lagoute, J., Zambelli, T., Gauthier, S. Direct comparison of the electronic coupling efficiency of sulfur and selenium anchoring groups for molecules adsorbed onto gold electrodes. *Chem. Phys.* **281**, 325 (2002).
- Patrone, L., Palacin, S., Charlier, J., Armand, F., Bourgoin, J.-P., Tang, H., Gauthier, S. Evidence of the key role of metal-molecule bonding in metal-molecule-metal transport experiments. *Phys. Rev. Lett.* **91**, 096802 (2003b).
- Pease, A.R., Jeppesen, J.O., Stoddart, J.F., Luo, Y., Collier, C.P., Heath, J.R. Switching devices based on interlocked molecules. *Acc. Chem. Res.* **34**, 433 (2001).
- Petit, C., Salace, G., Lenfant, S., Vuillaume, D. Inelastic tunnelling spectra of an alkly self-assembled monolayer using a MOS tunnel junction as a test-bed. *Microelectron. Eng.* **80**, 398 (2005).
- Petta, J.R., Slater, S.K., Ralph, D.C. Spin-dependent transport in molecular tunnel junctions. *Phys. Rev. Lett.* **93**, 136601 (2004).
- Pitters, J.L., Wolkow, R.A. Detailed studies of molecular conductance using atomic resolution scanning tunnelling microscopy. *Nano. Lett.* **6**, 390 (2006).
- Polymeropoulos, E.E. Electron tunnelling through fatty-acid monolayers. *J. Appl. Phys.* **48**, 2404 (1977).

- Polymeropoulos, E.E., Sagiv, J. Electrical conduction through adsorbed monolayers. *J. Chem. Phys.* **69**, 1836 (1978).
- Powell, R.J. Interface energy barrier determination from voltage dependence of photo-injected currents. *J. Appl. Phys.* **41**, 2424 (1970).
- Quek, S.Y., Neaton, J.B., Hybertsen, M.S., Kaxiras, E., Louie, S.G. Negative differential resistance in transport through organic molecules on silicon. *Phys. Rev. Lett.* **98**, 066807 (2007).
- Rakshit, T., Liang, G.-C., Ghosh, A.W., Datta, S. Silicon-based molecular electronics. *Nano Lett.* **4**, 1803 (2004).
- Ramachandran, G.K., Hopson, T.J., Rawlett, A.M., Nagahara, L.A., Primak, A., Lindsay, S.M. A bond-fluctuation mechanism for stochastic switching in wired molecules. *Science* **300**, 1413 (2003).
- Rampi, M.A., Schueller, O.J.A., Whitesides, G.M. Alkanethiol self-assembled monolayers as the dielectric of capacitors with nanoscale thickness. *Appl. Phys. Lett.* **72**, 1781 (1998).
- Rawlett, A.M., Hopson, T.J., Nagahara, L.A., Tsui, R.K., Ramachandran, G.K., Lindsay, S.M. Electrical measurements of a dithiolated electronic molecule via conducting atomic force microscopy. *Appl. Phys. Lett.* **81**, 3043 (2002).
- Reddy, P., Jang, S.-Y., Segalman, R.A., Majumdar, A. Thermoelectricity in molecular junction. *Science* **315**, 1568 (2007).
- Reed, M.A., Chen, J., Rawlett, A.M., Price, D.W., Tour, J.M. Molecular random access memory cell. *Appl. Phys. Lett.* **78**, 3735 (2001).
- Reed, M.A., Zhou, C., Muller, C.J., Burgin, T.P., Tour, J.M. Conductance of a molecular junction. *Science* **278**, 252 (1997).
- Reichert, J., Ochs, R., Beckmann, D., Weber, H.B., Mayor, M., Löhneysen, H.V. Driving current through single organic molecules. *Phys. Rev. Lett.* **88**, 176804 (2002).
- Reichert, J., Weber, H.B., Mayor, M., Löhneysen, H.V. Low-temperature conductance measurements on single molecules. *Appl. Phys. Lett.* **82**, 4137 (2003).
- Roth, K.M., Lindsey, J.S., Bocijan, D.F., Kuhr, W.G. Characterization of charge storage in redox-active self-assembled monolayers. *Langmuir* **18**, 4030 (2002).
- Roth, K.M., Yasseri, A.A., Liu, Z., Dabke, R.B., Malinovskii, V., Schweikart, K.-H., Yu, L., Tiznado, H., Zaera, F., Lindsey, J.S., Kuhr, W.G., Bocijan, D.F. Measurements of electron-transfer rates of charge-storage molecular monolayers on Si(100). Towards hybrid molecular/semiconductor information storage devices. *J. Am. Chem. Soc.* **125**, 505 (2003).
- Sakaguchi, H., Hirai, A., Iwata, F., Sasaki, A., Nagamura, T., Kawata, E., Nakabayashi, S. Determination of performance on tunnel conduction through molecular wire using a conductive atomic force microscope. *Appl. Phys. Lett.* **79**, 3708 (2001).
- Salomon, A., Boecking, M., Chan, C.K., Amy, F., Girshevitz, O., Cahen, D., Kahn, A. How do electronic carriers cross Si-bound alkyl monolayers? *Phys. Rev. Lett.* **95**, 266897 (2005).
- Salomon, A., Boecking, M., Seitz, O., Markus, T., Amy, F., Chan, C.K., Zhao, W., Cahen, D., Kahn, A. What is the barrier for tunnelling through alkyl monolayers? Results from n- and p-Si-alkyl/Hg junctions. *Adv. Mater.* **19**, 445 (2007).
- Salomon, A., Cahen, D., Lindsay, S.M., Tomfohr, J., Engelkes, V.B., Frisbie, C.D. Comparison of electronic transport measurements on organic molecules. *Adv. Mater.* **15**, 1881 (2003).

- Schreiber, F. Structure and growth of self-assembling monolayers. *Prog. Surf. Sci.* **65**, 151 (2000).
- Selzer, Y., Salomon, A., Cahen, D. Effect of molecule-metal electronic coupling on through-bond hole tunnelling across metal-organic monolayer-semiconductor junctions. *J. Am. Chem. Soc.* **124**, 2886 (2002a).
- Selzer, Y., Salomon, A., Cahen, D. The importance of chemical bonding to the contact for tunnelling through alkyl chains. *J. Phys. Chem. B* **106**, 10432 (2002b).
- Son, K.-A., Kim, H.I., Houston, J.E. Role of stress on charge transfer through self-assembled alkanethiol monolayers on Au. *Phys. Rev. Lett.* **86**, 5357 (2001).
- Star, A., Lu, Y., Bradley, K., Grüner, G. Nanotube optoelectronic memory devices. *Nano. Lett.* **4**, 1587 (2004).
- Stewart, D.R., Ohlberg, D.A.A., Beck, P.A., Chen, Y., Williams, R.S., Jeppesen, J.O., Nielsen, K.A., Stoddart, J.F. Molecule-independent electrical switching in Pt/organic monolayer/Ti devices. *Nano. Lett.* **4**, 133 (2004).
- Stokbro, K., Taylor, J., Brandbyge, M. Do Aviram-Ratner diodes rectify? *J. Am. Chem. Soc.* **125**, 3674 (2003).
- Szuchmacher Blum, A., Kushmerick, J.G., Long, D.P., Patterson, C.H., Yang, J.C., Henderson, J.C., Yao, Y., Tour, J.M., Shashidhar, R., Ratna, B.R. Molecularily inherent voltage-controlled conductance switching. *Nature Mater.* **4**, 167 (2005).
- Tao, N.J. Electron transport in molecular junctions. *Nature Nanotechnol.* **1**, 173 (2006).
- Taylor, J., Brandbyge, M., Stokbro, K. Theory of rectification in Tour wires: the role of electrode coupling. *Phys. Rev. Lett.* **89**, 138301 (2002).
- Tour, J.M., Van Zandt, W.L., Husband, C.P., Husband, S.M., Wilson, L.S., Franzon, P.D., Nackashi, D.P. Nanocell logic gates for molecular computing. *IEEE Trans. Nanotechnol.* **1**, 100 (2002).
- Tulevski, G.S., Miao, Q., Fukuto, M., Abram, R., Ocko, B.M., Pindak, R., Steigerwald, M.L., Kagan, C.R., Nuckolls, C. Attaching organic semiconductors to gate oxides: in situ assembly of monolayer field effect transistors. *J. Am. Chem. Soc.* **126**, 15048 (2004).
- Ulman, A. *An Introduction to Ultrathin Organic Films: From Langmuir-Blodgett to Self-Assembly* (Boston, Academic Press 1991).
- Ulrich, J., Esrail, D., Pontius, W., Venkataraman, L., Millar, D., Doerr, L.H. Variability of conductance in molecular junctions. *J. Phys. Chem. B* **110**, 2462 (2006).
- Van Der Wagt, J.P.A., Seabaugh, A.C., Beam, E.A. RTD/HFET low standby power SRAM gain cell. *IEEE Electron Dev. Lett.* **19**, 7 (1998).
- Venkataraman, L., Klare, J.E., Nuckolls, C., Hybertsen, M.S., Steigerwald, M.L. Dependence of single molecule junction conductance on molecular conformation. *Nature* **442**, 904 (2006a).
- Venkataraman, L., Klare, J.E., Tam, I.W., Nuckolls, C., Hybertsen, M.S., Steigerwald, M.L. Single-molecule circuits with well-defined molecular conductance. *Nano Lett.* **6**, 458 (2006b).
- Venkataraman, L., Park, Y.S., Whalley, A.C., Nuckolls, C., Hybertsen, M.S., Steigerwald, M.L. Electronics and chemistry: varying single-molecule junction conductance using chemical substituents. *Nano Lett.* **7**, 502 (2007).

- Vilan, A., Cahen, D. Soft contact deposition onto molecularly modified GaAs. Thin metal film flotation: principles and electrical effects. *Adv. Func. Mater.* **12**, 795 (2002).
- Vondrak, T., Cramer, C.J., Zhu, X.-Y. The nature of electronic contact in self-assembled monolayers for molecular electronics: evidence of strong coupling. *J. Phys. Chem. B* **103**, 8915 (1999).
- Vuillaume, D., Boulas, C., Collet, J., Allan, G., Delerue, C. Electronic structure of alkylsiloxane self-assembled monolayer-silicon heterostructure. *Phys. Rev. B* **58**, 16491 (1998).
- Vuillaume, D., Chen, B., Metzger, R.M. Electron transfer through a monolayer of hexadecylquinolinium tricyanoquinodimethanide. *Langmuir* **15**, 4011 (1999).
- Wang, W., Lee, T., Krestchmar, I., Reed, M.A. Inelastic electron tunnelling spectroscopy of an alkanedithiol self-assembled monolayer. *Nano. Lett.* **4**, 643 (2004).
- Wang, W., Lee, T., Reed, M.A. Mechanism of electron conduction in self-assembled alkanethiol monolayer devices. *Phys. Rev. B* **68**, 035416 (2003).
- Wang, W., Richter, C.A. Spin-polarized inelastic tunnelling spectroscopy of a molecular magnetic tunnel junction. *Appl. Phys. Lett.* **89**, 153105 (2006).
- Weber, H.B., Reichert, J., Weigend, F., Ochs, R., Beckmann, D., Mayor, M., Ahlrichs, R., Löheysen, H.V. Electronic transport through single conjugated molecules. *Chem. Phys.* **281**, 113 (2002).
- Wei, Z., Kondratenko, M., Dao, L.H., Perepichka, D.F. Rectifying diodes from asymmetrically functionalized single-wall carbon nanotubes. *J. Am. Chem. Soc.* **128**, 3134 (2006).
- Wold, D.J., Frisbie, C.D. Formation of metal-molecule-metal tunnel junctions: Micro-contacts to alkanethiol monolayers with a conducting AFM tip. *J. Am. Chem. Soc.* **122**, 2970 (2000).
- Wold, D.J., Frisbie, C.D. Fabrication and characterization of metal-molecule-metal junctions by conducting probe atomic force microscopy. *J. Am. Chem. Soc.* **123**, 5549 (2001).
- Wold, D.J., Haag, R., Rampi, M.A., Frisbie, C.D. Distance dependence of electron tunnelling through self-assembled monolayers measured by conducting probe atomic force microscopy: unsaturated versus saturated molecular junctions. *J. Phys. Chem. B* **106**, 23 (2002).
- Xia, Y., Whitesides, G.M. Soft lithography. *Angew. Chem. Int. Ed. Engl.* **37**, 550 (1998).
- Xu, B., Tao, N.J. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science* **301**, 1221 (2003).
- Xu, T., Peterson, I.R., Lakshminathan, M.V., Metzger, R.M. Rectification by a monolayer of hexadecylquinolinium tricyanoquinodimethanide between gold electrodes. *Angew. Chem. Int. Ed. Engl.* **40**, 1749 (2001).
- Yaliraki, S.N., Kemp, M., Ratner, M.A. Conductance of molecular wires: Influence of molecule-electrode binding. *J. Am. Chem. Soc.* **121**, 3428 (1999).

# **Self-assembly strategy of nanomanufacturing of hybrid devices**

*S. Hong, Y.-K. Kwon, J.S. Ha, N.-K. Lee, B. Kim, and M. Sung*

# 10

## **10.1 Introduction**

As the conventional device industry reaches its technological and economical limits, extensive efforts have been directed to explore new device architectures utilizing nanomaterials such as nanoparticles, organic molecules, carbon nanotubes (CNTs), and nanowires (NWs). These new devices are often called “hybrid devices” because they are comprised of new nanomaterials as well as conventional electronic elements. Examples include carbon-nanotube-based circuits, molecular electronics, etc. However, a major stumbling block holding back their practical applications can be a lack of a mass-production method for such devices. Since most new nanostructures are first synthesized in a solution or powder form, one has to pick up and assemble millions of individual nanostructures onto solid substrate to build integrated devices, which can be an extremely time-consuming task. On the other hand, biological systems have adapted self-assembly strategy to build macroscale systems from nanoscale molecular structures for billions of years. This chapter will discuss how the self-assembly strategy can be applied for nanomanufacturing of hybrid nanodevices.

## **10.2 Direct patterning of nanostructures**

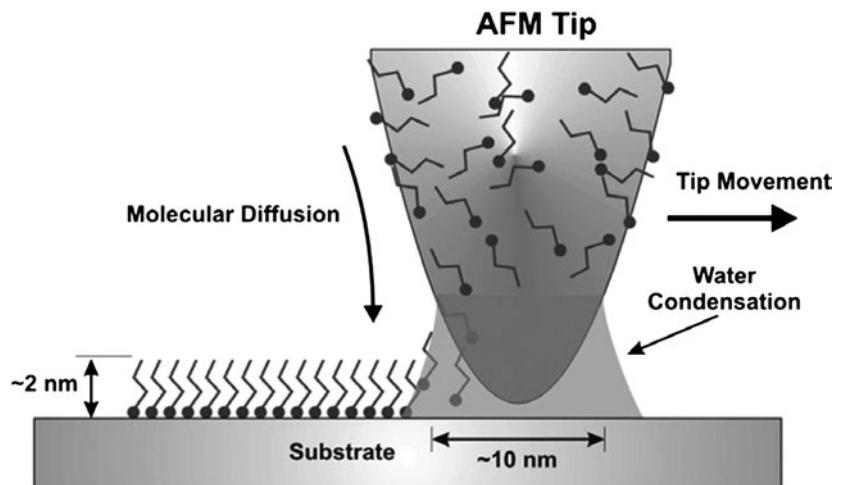
### **10.2.1 Dip-pen nanolithography**

#### **10.2.1.1 Basic concept of dip-pen nanolithography**

Dip-pen nanolithography (DPN) is a novel lithography technique that was first developed by Mirkin’s group at Northwestern University in 1998 (Piner *et al.* 1999). It is based on an atomic force microscope (AFM) and uses an AFM tip as a nanoscale pen, molecular substances as an ink, and solid substrates as a paper.

The DPN method has several advantages over other lithography techniques such as high resolution and registration, extremely versatile chemical and material flexibility, etc. In addition, there is no need to pre-modify the surface.

<b>10.1 Introduction</b>	343
<b>10.2 Direct patterning of nanostructures</b>	343
<b>10.3 Directed assembly of nanostructures</b>	357
<b>10.4 Characteristics of self-assembled hybrid nanodevices</b>	369
<b>10.5 Conclusion</b>	380
<b>Acknowledgments</b>	380
<b>References</b>	380



**Fig. 10.1** Basic mechanism of DPN.

These merits make DPN a unique and highly desirable tool for depositing biological and other soft materials on various surfaces.

Figure 10.1 shows the basic mechanism of DPN as well as transport scheme of a molecular “ink” from the tip to a substrate. When the molecule-coated tip is held in contact with the substrate, the molecules spread out by diffusion onto the substrate. Various-shape patterns are generated on the substrate by scanning the tip. Under ambient conditions, a water condensation forms between the AFM tip and the substrate. The size of the condensation is controlled by the relative humidity and affects the ink molecule transport rate, the effective tip–substrate contact area, and resolution.

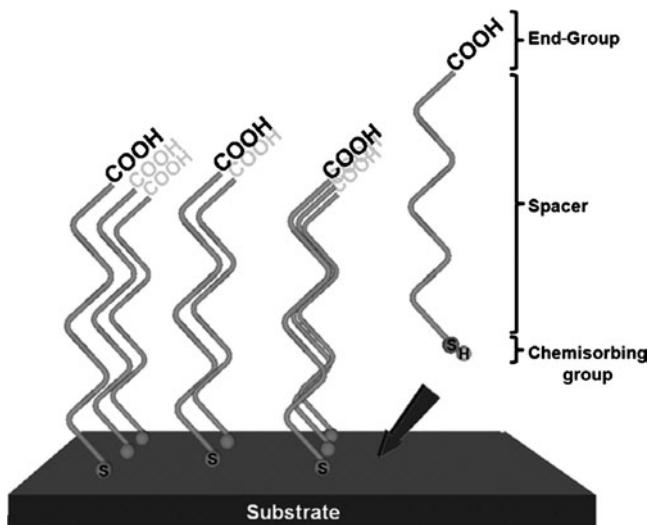
#### 10.2.1.2 *Self-assembled monolayer molecules*

Various molecules can be used as DPN ink. Figure 10.2 shows typical organic molecules forming self-assembled monolayers (SAMs) on substrates. Such molecules are composed of a chemisorbing group, spacer, and end-groups. If these molecules are placed onto the substrate, they are chemisorbed onto the substrate and generate a well-ordered stable crystalline monolayer film with a few nanometer thickness. Specific chemisorbing groups can be chosen depending on the substrate. By depositing these molecules onto the substrate, one can completely change the chemical properties of the surface to that of the end-group.

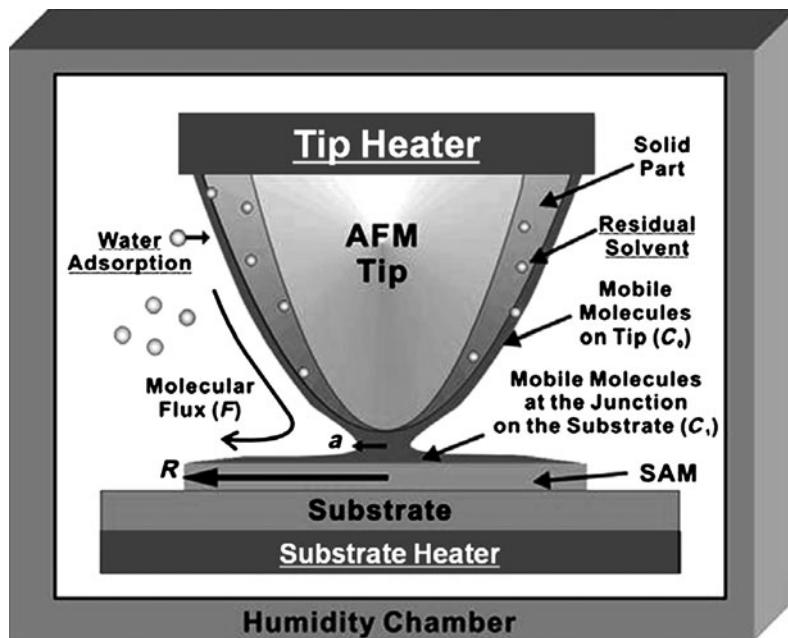
#### 10.2.1.3 *Diffusion theory*

When an AFM tip coated with ink molecules is held at a specific position on the solid substrate, the ink molecules diffuse out onto the substrate and form dot patterns of various sizes with respect to tip-dwelling time (Jang *et al.* 2001; Sheehan *et al.* 2002). To explain this phenomenon, the “double-molecular-layer” (DML) model was proposed, as shown in Figure 10.3 (Cho *et al.* 2006).

Since the DPN process usually utilizes a molecular species that has a higher melting point than room temperature and the tip is dried by N<sub>2</sub> blowing, most of the molecular layer on the tip can be assumed to



**Fig. 10.2** Self-assembled monolayer molecules.



**Fig. 10.3** Basic concepts of the DML model. Thermal energy, residual solvent and adsorbed water molecules affect molecular diffusion.

remain solid during the DPN process under ambient conditions. On the other hand, molecules on the surface of the molecular coating are likely to be much more mobile than those inside the molecular layer (Dash 1989). For this reason, the DML model assumes that the molecular ink forms a DML structure comprised of a bulk solid part and a thin mobile surface layer on the tip. The “solid part” continuously provides molecules to the mobile surface layer. Thus, the solid part plays the role of an ink reservoir by stabilizing the deposition rate in the DPN process. The DML

model suggests that the transport rate of molecules in DPN is mainly determined by the mobility and the number density of the mobile surface molecules.

The mobility of the surface molecules can be affected by the following factors; thermal energy, residual solvent and adsorbed water. Thermal energy increases the molecular diffusion constant and enhances the molecular flow. When the temperature goes up near the melting point, the diffusion constant of the surface molecules increases rapidly even for a small temperature variation. Therefore, if the melting point of the molecular ink becomes close to ambient, the deposition rate becomes strongly dependent on temperature. When the tip surface is coated with ink molecules, some residual solvent molecules should remain in the molecular coating, and they can enhance the surface molecular diffusion. When the DPN process is performed under ambient conditions, water from the air can be adsorbed onto the tip and substrate and then affect the mobility of the transporting molecules. In the case of hydrophilic molecular inks, the adsorbed water should improve the mobility.

The diffusion equation of DPN pattern generation can be solved using the DML model. The behavior of the ink molecules can be defined in the following four regions; tip region, nanoscale junction region, SAM region, and bare-substrate region.

In the AFM tip region, the surface density of mobile molecules on the tip  $C_0$  is assumed constant because the solid part of the molecular ink works as an ink reservoir. In the nanoscale junction region, assuming that the junction has circular cross-section with a radius  $a$  and that the net molecular transport is determined by the difference between flux transported toward the surface and that away from the surface, the molecular flux through the nanoscale junction can be described as  $\text{Flux} = 2\pi a (\alpha_0 C_0 - \alpha_1 C_1)$ , where  $\alpha_0$  and  $\alpha_1$  are the molecular transport rate coefficients from the tip to the substrate and that from the substrate to the tip, respectively.  $C_1$  is the density of mobile molecules on the substrate at the junction. It should be noted that  $\alpha_0$  and  $\alpha_1$  depend on various factors such as the water meniscus at the junction and attractive forces between the substrate and the molecules. However, when the substrate is covered by molecular layers, the diffusion coefficients are about the same on the tip and substrate, so  $\alpha_0 = \alpha_1 \equiv \alpha$  and  $\text{Flux} = 2\pi a \alpha (C_0 - C_1)$ .

On the substrate, the first-deposited molecules are assumed to chemically anchor to the bare substrate and form a uniform SAM. After this initial deposition, the molecules deposited on top of the SAM can diffuse out until they meet bare-substrate regions. In this case, the motion of the molecules on the SAM can be described by the diffusion equation,  $\frac{\partial C}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} (r D \frac{\partial C}{\partial r})$ , where  $C$  and  $D$  represent the surface density and diffusion constant of mobile molecules on the substrate, respectively. At the outer boundary of the SAM pattern, the moving boundary condition,  $\frac{\partial C}{\partial r}|_{r=R} = -C_R \frac{dR}{dt}$ , can be used to describe the increase in radius  $R$  of the SAM that occurs when molecules reaching the boundary become adhered to the substrate.  $C_R$  represents the density of molecules in the SAM. At the nanoscale junction, another boundary condition based on the molecular flux equation,  $\frac{\partial C}{\partial r}|_{r=a} = \alpha(C_0 - C_1)$ , is assumed. This means that molecules transported through the nanoscale junction diffuse onto

the substrate. With proper boundary condition, we can solve for  $dt/dR$  as,

$$\frac{dt}{dR} = \frac{1}{D} \left( \frac{C_R}{C_0} \right) \left\{ R \ln R + R \left( \frac{D}{a\alpha} - \ln a \right) \right\}, \quad (10.1)$$

(Crank 1993) Assuming initial deposition of  $R \rightarrow a$  as  $t \rightarrow 0$ , the solution to eqn (10.1) is

$$t = \frac{1}{4D} \left( \frac{C_R}{C_0} \right) \left\{ 2R^2 \ln(R/a) + (R^2 - a^2) \left( \frac{2D}{a\alpha} - 1 \right) \right\}. \quad (10.2)$$

In the *fast-diffusion limit case* (diffusion rate  $\gg$  deposition rate), eqn (10.2) becomes

$$t = \frac{R^2}{4D} \left( \frac{C_R}{C_0} \right) \left( \frac{2D}{a\alpha} \right) \left\{ 1 + \frac{2D}{a\alpha} \ln(R/a) \right\} \quad (R \gg a, D/a \gg 1). \quad (10.3)$$

From this, we obtain the power-law relation  $R \propto t^\nu$ , where

$$\nu \cong \frac{1}{2} \left( 1 - \frac{a\alpha}{2D} \right). \quad (10.4)$$

For the *general case* (deposition rate  $\sim$  diffusion rate), the general solution eqn (10.2) becomes

$$t \cong \frac{R^2}{4D} \left( \frac{C_R}{C_0} \right) [2 \ln(R/a)] \quad (D/a \sim 1, \ln(R/a) \gg [1 - (a^2/R^2)]). \quad (10.5)$$

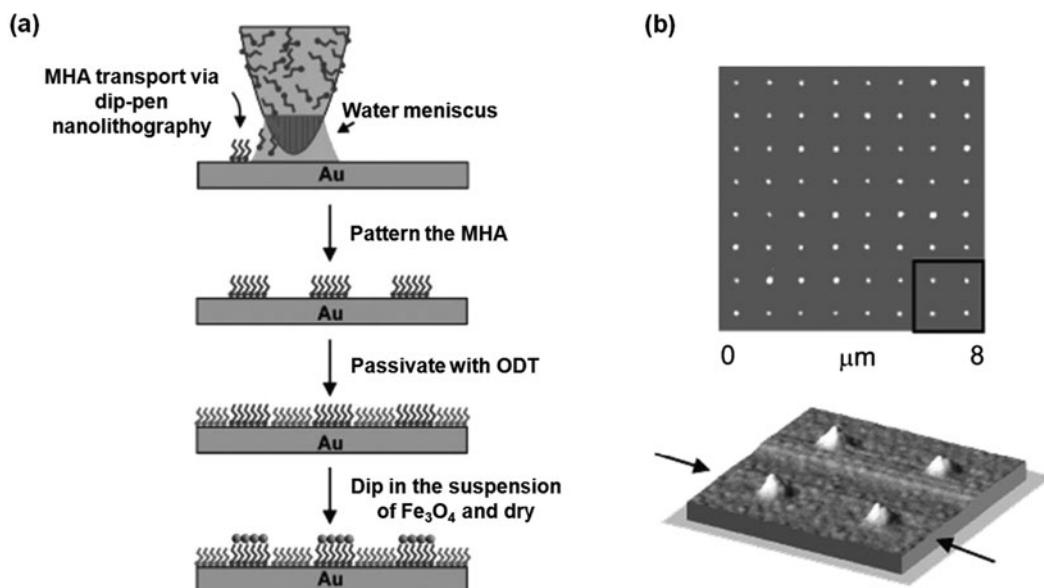
Since for  $R \gg a$ , we have  $1 < \ln(R/a) < R/a$ , we can write

$$\frac{1}{4D} \left( \frac{C_R}{C_0} \right) 2R^2 < t < \frac{1}{4D} \left( \frac{C_R}{C_0} \right) \frac{2R^3}{a}. \quad (10.6)$$

#### 10.2.1.4 Application

The DPN technique can generate patterns with various molecular species such as alkanethiols, biomolecules, salts, and conducting polymers, etc. Chemisorption of a molecular ink onto the substrate forms stable nanostructures, which can subsequently be used as templates to assemble different types of molecules or nanostructures of interest (Demers *et al.* 2001).

Figure 10.4(a) shows the typical procedure of pattern generation using the DPN method. The AFM tip is coated with 16-mercaptophexadecanoic acid (MHA) by dipping into a saturated solution of MHA in acetonitrile for 30 s. The coated tip is blown by compressed N<sub>2</sub> or other gases before being used. MHA patterns are generated by scanning the tip on Au substrate in pre-defined patterns. The patterned Au substrate is dipped into a saturated solution of 1-octadecanethiol (ODT) in acetonitrile for 30 s to form a passivating layer on the Au substrate around the MHA nanostructures. After rinsing with acetonitrile and drying, the samples are dipped into a suspension of the iron-oxide particles (magnetite, Fe<sub>3</sub>O<sub>4</sub>) for 30 s and then dried under a flow of nitrogen overnight (Fu *et al.* 2001). Figure 10.4(b) shows AFM topography image of iron-oxide dot array generated according to the scheme of Fig. 10.4(a) (Liu *et al.* 2002).



**Fig. 10.4** (a) Schematic representation of the procedure used to prepare magnetic nanostructures on a Au substrate. (b) AFM topography image of magnetic structures formed on MHA (from Liu, X., Fu, L., Hong, S., Dravid, V.P. and Mirkin, C.A. (2002) *Adv. Mater.* **14**, 231 ©2002 Copyright Wiley-VCH Verlag GmbH & Co. KGaA., reproduced with permission).

### 10.2.2 Microcontact printing of nanostructures

Microcontact printing (MCP) is a pattern-transfer technique where the patterned polymer stamp inked with molecules or nanomaterials is physically contacted with substrates and it has the advantage of easy transfer of micropatterns over a large area. This technique was originally adopted to transfer self-assembling molecular monolayers onto Au substrates and then either to etch Au by using such a formed self-assembled monolayer as a passivation layer (Kumar *et al.* 1993; Xia *et al.* 1996) or to fabricate chemical templates (Kumar *et al.* 1994a,b). During the last decade, MCP has been extended to patterning of polymer (Miller *et al.* 2002), metal film (Hur *et al.* 2004, 2005), nanomaterials (Yip *et al.* 2006; Sun *et al.* 2007), and biomaterials including DNA (Bernard *et al.* 2000; Tan *et al.* 2002; Csucs *et al.* 2003). MCP is done at relatively low temperatures, which enables its use on flexible plastic substrates (Yip *et al.* 2006). Owing to the above-mentioned advantages of MCP, there have been extensive efforts to replace the microfabrication process based on the photolithography technique, which is quite complicated and has high costs, with MCP (Xia *et al.* 1998; Love *et al.* 2005; Menard *et al.* 2007). In this section, we discuss several important points that should be understood well for the successful performance of traditional MCP in micropatterning of nanomaterials.

A general schematic of MCP is shown in Fig. 10.5. Since the pattern transfer occurs via physical contact, it is particularly important to have the inked stamp uniformly contacted with substrates and so to use the polymer stamp with elasticity and viscosity. As a stamp material, poly(dimethylsiloxane) (PDMS)

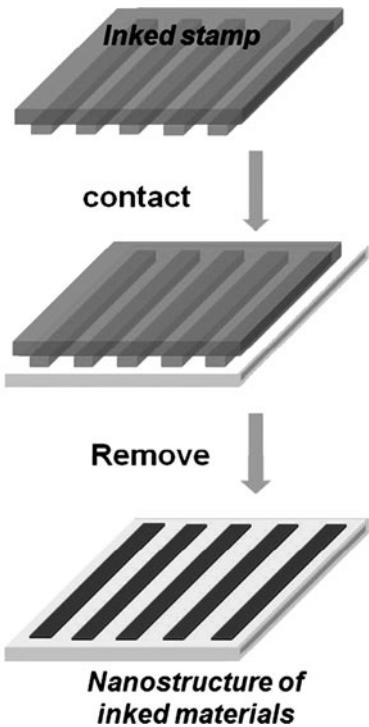
has been most widely used. PDMS is quite rigid to retain the micropatterns of the stamp and is also chemically inert not to be deformed or contaminated by chemicals during the patterning process (Xia *et al.* 1998; Menard *et al.* 2007). High-modulus PDMS or poly(urethaneacrylate) was also actively studied for the patterning of structures smaller than 100 nm (Schmid and Michel 2000) or for the patterning under easy deformation conditions of stamps (Yoo *et al.* 2004; Park *et al.* 2006b).

PDMS has a low surface energy ( $\sim 20 \text{ N/m}$ ) which enables easy transfer of ink materials onto substrate, but it is hydrophobic, which is not good for the uniform inking of polar materials. Materials dispersed in polar solvent forms a circular drop on the PDMS stamp surface with a high contact angle. Polar ink induces dewetting and it is not transferred onto PDMS stamp even under dry condition. Treatment of the hydrophobic stamps with oxygen plasma (Donzel *et al.* 2001; Delamarche *et al.* 2003), UV-ozone (Ouyang *et al.* 2000; Efimenko *et al.* 2002), and acid (Chaudhury *et al.* 1996) resulted in the hydrophilic PDMS stamp. However, such formed hydrophilic stamps recovered the intrinsic hydrophobic properties after a certain period of time (Donzel *et al.* 2001; Delamarche *et al.* 2003; Bodas and Malek 2007). In order to improve it, coating of the PDMS stamp surface via chemical reaction was devised (Donzel *et al.* 2001; Delamarche *et al.* 2003; Bodas and Malek 2007).

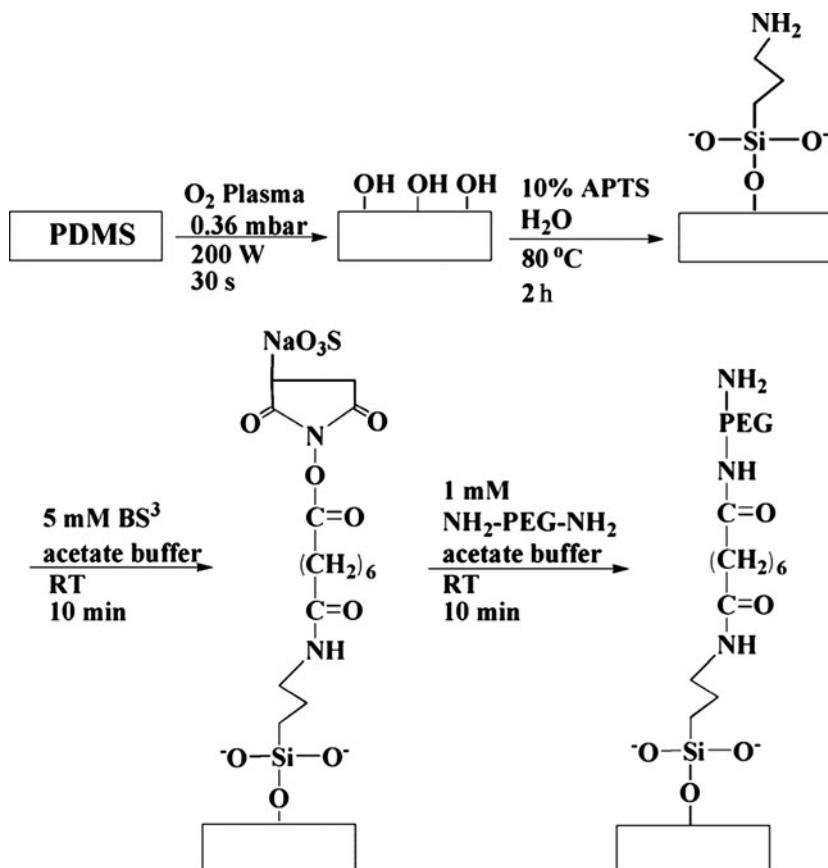
Figure 10.6 shows the grafting process of PEG molecules onto a PDMS surface (Park *et al.* 2006b). First, hydroxyl OH groups are formed on the surface via oxidation by  $\text{O}_2$  plasma treatment. Then, subsequent treatment with 3-aminopropyltriethoxysilane (APS), bis(sulfosuccinimidyl)-suberate)(BS<sup>3</sup>), and amino-terminated polyethylene glycol ( $\text{NH}_2\text{-PEG-NH}_2$ ) grafted the PEG molecules into PDMS, which prohibited the recovery of surface hydrophobicity. Instead of PDMS surface modification, use of hydrophilic poly(ether-block-ester) (Trimbach *et al.* 2007) and agarose (Weibel *et al.* 2005) as stamp materials could transfer various kinds of polar inks.

In traditional MCP, ink materials on the relief region of the stamp pattern are transferred onto the substrates as in general printing methods. However, it is also possible to transfer the ink in the recess region of the patterned stamp. In this case, confinement of ink inside the recess area is required and it can be easily done when the interaction between the stamp surface and the ink material is not attractive, which is called “discontinuous dewetting”. Figure 10.7 shows one example of patterning via discontinuous dewetting, where the  $\text{V}_2\text{O}_5$  nanowires in the recess area of the patterned PDMS stamp were transferred onto the substrates (Kim *et al.* 2006). If the hydrophobic stamps were slowly taken out from the  $\text{V}_2\text{O}_5$  nanowire solution after sufficient dipping time, polar nanowire solution would be stored inside the recess area without inking the relief area of the stamp. Then, evaporation of solvent in air or vacuum can make the concentrated nanowires spontaneously aggregate in the edge of the stamp. Therefore, aggregated nanowire patterns could be transferred via printing of the nanowire inks in the edges. In this way, submicrometer-sized patterns of nanowires could be obtained by using a stamp with micrometer-sized patterns.

On the contrary, use of hydrophilic stamps treated with UV-ozone could produce the transfer of nanowire patterns according to the stamp pattern, as expected. There were also reports on the patterning of Au nanoparticles and



**Fig. 10.5** Schematics of microcontact printing.

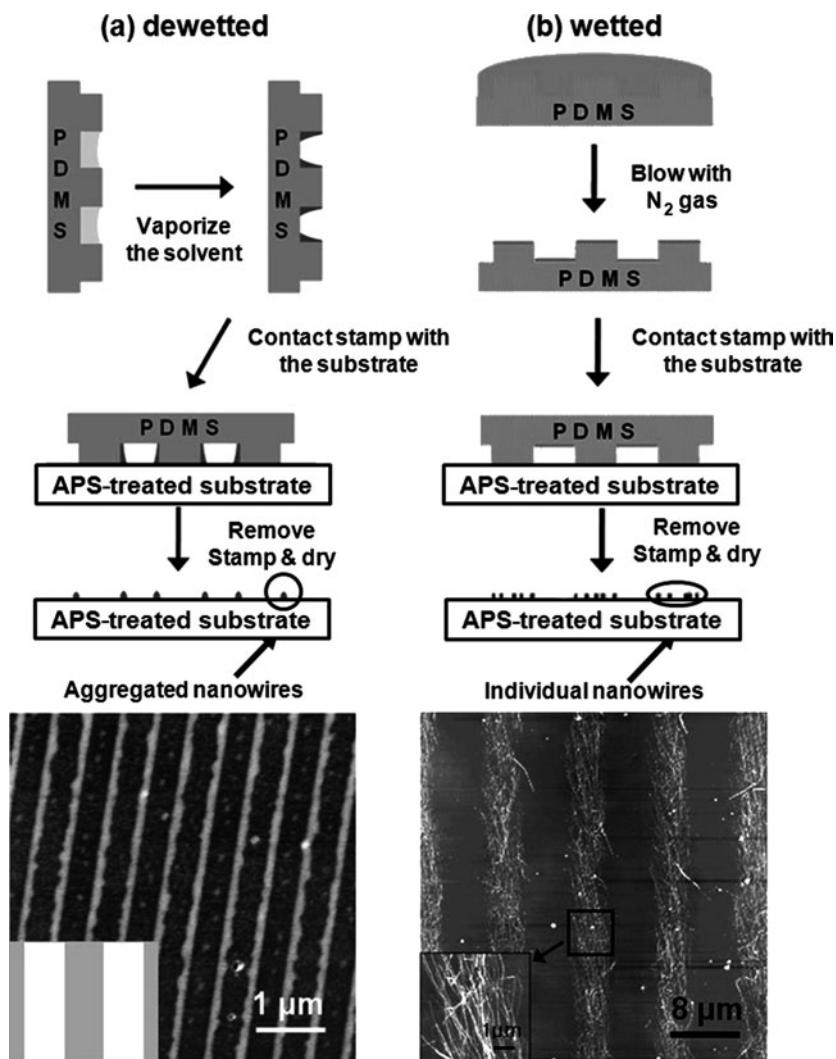


**Fig. 10.6** Preparation of hydrophilic PDMS surface by grafting (from Donzel, C., Geissler, M., Bernard, A., Wolf, H., Michael, B., Hilborn, J. and Delamarche, E. (2001) *Adv. Mater.* **13**, 1164 ©2001 Copyright Wiley-VCH Verlag GmbH & Co. KGaA, reprinted in part with permission).

alkanethiol molecules by use of edge patterns of the stamp (Cherniavskaya 2002).

In general, inking of nanoparticles larger than a few tens of nanometers or the nanowires longer than a few micrometers needs a special recipe, different from that of small charged particles or short-chain molecules. Figure 10.8 introduces the various interesting inking methods. Spin coating can form a uniform film in a short time. Inking of a PDMS stamp with single-wall carbon nanotubes (SWNTs) by spin coating is shown in Fig. 10.8(a) (Meitl *et al.* 2004). Coating of SWNT emulsion could not produce the high-density SWNTs on the stamp due to weak interaction between the PDMS stamp and SWNTs. However, addition of methanol could increase the density of SWNT on the stamp by enhancing the dissipation of more SWNTs via extracting the sodium dodecyl sulfate (SDS) from the SWNT emulsion. Therefore, SWNT patterns could be obtained by direct printing. Interestingly, alignment of SWNTs by centrifugal force during the spin coating was observed.

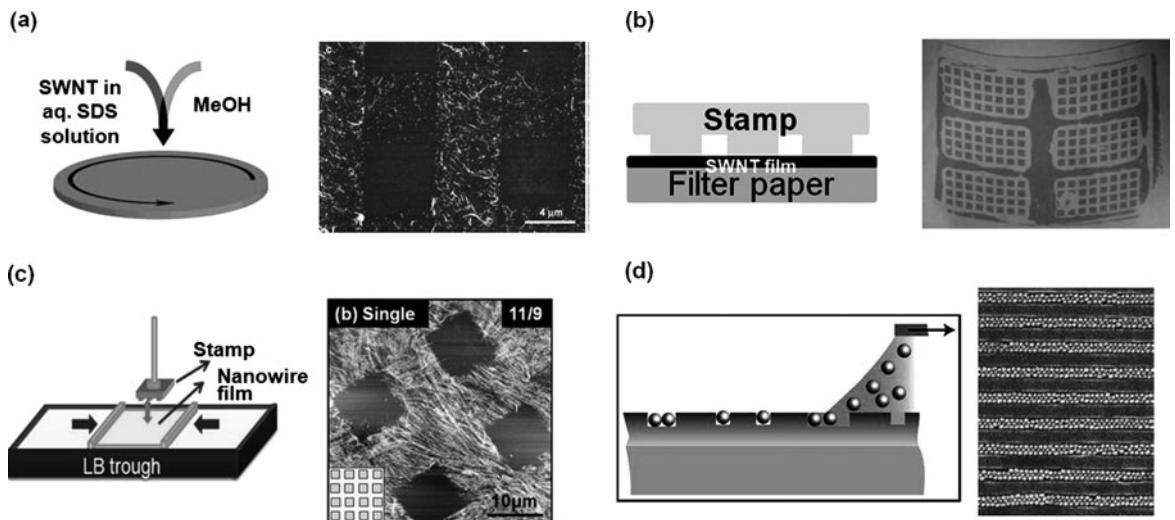
Via contact of PDMS stamp onto the uniform SWNT film prepared via vacuum filtration of SWNT emulsion dispersed in SDS solution by using an alumina filter, a transparent electrode pattern could be transferred, with a



**Fig. 10.7** MCP using a V<sub>2</sub>O<sub>5</sub> nanowire ink with (a) hydrophobic and (b) hydrophilic PDMS stamp (from Kim, Y.-K., Park, S.J., Koo, J.P., Oh, D.J., Kim, G.T., Hong, S. and Ha, J.S. (2006) *Nanotechnology* **17**, 1375 ©2006 Institute of Physics and IOP Publishing Limited, reprinted with permission).

transparency and a conductivity comparable to an ITO electrode (Fig. 10.8(b)) (Zhou *et al.* 2006; Hu *et al.* 2007). The Langmuir–Blodgett (LB) technique is very useful in the transfer of uniform molecular films formed via floating of amphiphilic molecules onto a water surface and controlling the surface area. Nanomaterial patterns can be also transferred via the LB technique. As shown in a Fig. 10.8(c), inking of patterned PDMS stamp with nanomaterial films on a water surface and subsequent direct printing onto the substrate could transfer the pattern (Santhanam *et al.* 2004; Park *et al.* 2007).

In the case of nanomaterials, which would not spontaneously float on a water surface to form a LB film, surface modification by surfactant (Kim *et al.* 2001; Tao *et al.* 2003; Whang *et al.* 2003) or formation of surfactant film to aggregate the nanomaterials (Yamaki *et al.* 2001; Park *et al.* 2007) could be done. Monolayer LB films of hexagonally ordered alkanethiol coated Au



**Fig. 10.8** Inking methods and the transferred patterns. (a) Spin coating (from Meitel, M.A., Zhou, Y., Gaur, A., Jeon, S., Usrey, M.L., Strano, M.S. and Rogers, J.A. (2004) *Nano Lett.* **4**, 1643 ©2004 American Chemical Society, reprinted in part with permission). (b) Direct transfer of filtered SWNT film (from Zhou, Y., Hu, L. and Grüner, G. (2006) *Appl. Phys. Lett.* **88**, 123109 ©2006 American Institute of Physics, reprinted in part with permission). (c) Langmuir–Schaefer technique (from Park, J., Shin, J. and Ha, J.S. (2007) *Nanotechnology* **18**, 405301 ©2007 Institute of Physics and IOP Publishing Limited, reprinted with permission). (d) Dewetting (from Kraus, T., Malaquin, L., Schumid, H., Riess, W., Spencer, N.D. and Wolf, H. (2007) *Nature Nanotech.* **2**, 570 ©2007 Nature Publishing Group, reprinted in part with permission).

nanoparticles were also transferred (Park *et al.* 2007). Transfer of the one-dimensionally aligned nanowire patterns without any deformation could be done in this way, too (Kim *et al.* 2001; Tao *et al.* 2003; Whang *et al.* 2003). Figure 10.8(d) shows the selective inking of Au nanoparticles onto the groove area of the stamp by hydrodynamic and surface forces (Kraus *et al.* 2007). Moving the meniscus of a colloidal Au suspension over the patterned PDMS stamp made the arrangement of dispersed Au particles inside the grooves. It is attributed to the dewetting properties of the stamp. Single-particle resolution was obtained by controlling the pattern size of the stamp.

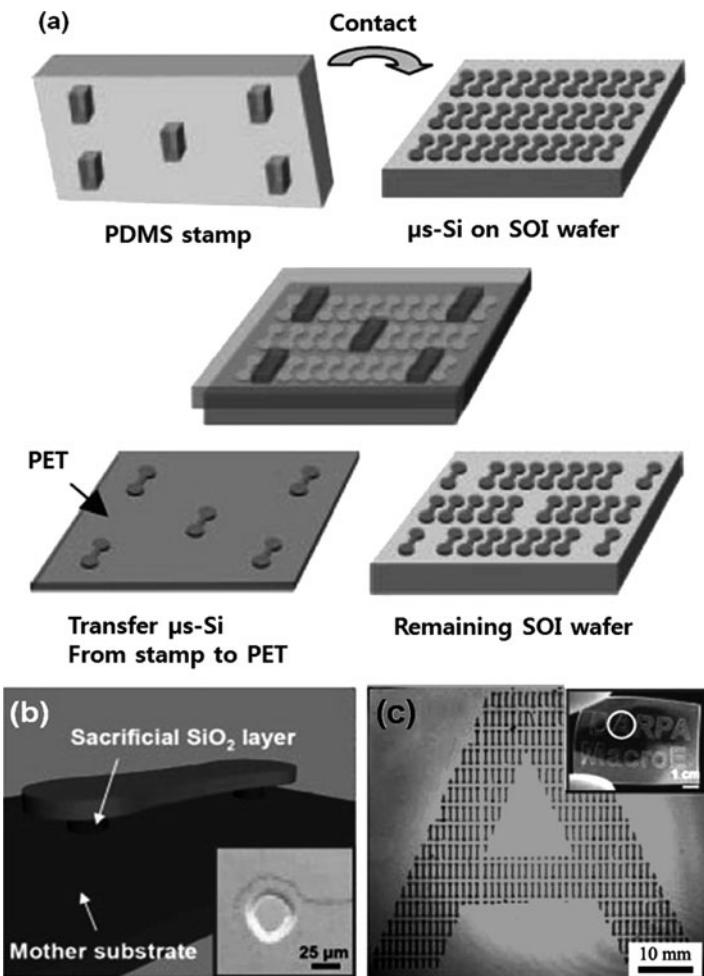
In MCP using a patterned stamp, transfer of the nanomaterials onto substrate as well as the inking of nanomaterials onto the stamp is very important. Attractive interaction between the ink materials and the stamp is very helpful in the inking process, but may be disadvantageous in the transfer of ink onto substrates. Except for the cases of chemical bonding and the transfer of charged materials, the transfer of the nanomaterials onto the substrates can be explained in terms of the work of adhesion calculated from the surface energy between stamp, nanomaterials, and the substrate (Adamson *et al.* 1997; Wang *et al.* 2003). When the force of adhesion between the nanomaterials and the substrate is stronger than that between the stamp and nanomaterials, nanomaterials can be spontaneously transferred onto the substrate. Individual surface energy can be easily calculated from the Young's equation and Wendt equation by measuring the contact angles (Adamson *et al.* 1997; Wang *et al.* 2003). This mechanism of pattern transfer could be applied to the selective removal of thin films by a stamp on top of the film as well as to the transfer printing of organic materials such as pentacene (Wang *et al.* 2003; Choi *et al.* 2005).

Surface modification by oxidation or coating of self-assembled monolayers is usually used to enhance the transfer rate by improving the adhesion between the nanomaterials and substrate, to enhance the resolution by hindering the diffusion of inks on the substrate, and to prohibit the aggregation of nanomaterials due to repulsive interaction between the substrate and the ink (Meitl *et al.* 2004; Hu *et al.* 2007; Kim *et al.* 2007). Amine-termination of SiO<sub>2</sub> surface could help the complete transfer of V<sub>2</sub>O<sub>5</sub> nanowires and enhance the resolution by increasing the interaction between V<sub>2</sub>O<sub>5</sub> nanowires and the substrate (Kim *et al.* 2007). However, pattern transfer is not simply explained just by the relative interaction between the stamp, nanomaterials, and the substrate. So, other factors such as viscosity and gravity should be considered. Recently, it was reported that ink materials could be either picked up by the stamp or transferred onto the substrate by change of the speed of peeling off the stamp during the printing process (Meitl *et al.* 2006; Feng *et al.* 2007).

### 10.2.3 Other printing methods of nanostructures

Semiconducting one-dimensional (1D) materials such as SWNT, Si, and metal-oxide nanowires have been extensively studied as promising component materials of future nanoelectronic devices (Li *et al.* 2006; Wang *et al.* 2007). 1D nanomaterials are expected to be applied to the future flexible devices due to their flexibility as well as their high electrical mobility (Hong *et al.* 2007; Ju *et al.* 2007). For the fabrication of actual devices, a patterning process of nanowires as well as the synthesis of high-quality nanowires should be established. Control of the position and alignment can enhance the device performance and the uniformity as well as the stability of the devices (Pimparkar *et al.* 2007). In this section, we will focus on the transfer-printing methods extended from the traditional MCP, where the materials printed from the stamp actually function in organic optoelectronics or other applications. Transfer-printing methods are simple, economical, processed under the mild conditions to be applied to the flexible substrates (Hong *et al.* 2007), and the resolution would exceed that of traditional MCP.

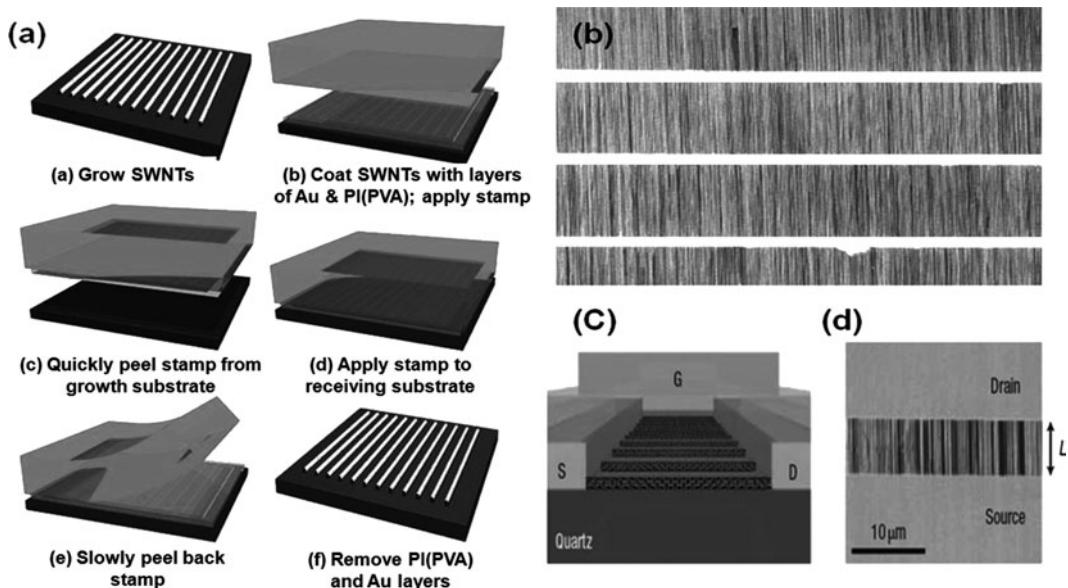
Via a transfer-printing method, nanowires and nanotubes grown or aligned on the mother substrates could be selectively transferred onto the substrate without change of properties and the alignment. Figure 10.9(a) shows the patterning process of microstructured single-crystalline Si ribbon, fabricated on the mother-substrate via a top-down method, onto the flexible PET substrate via a transfer-printing technique (Lee *et al.* 2005). Inking on the relief area of the stamp was performed by contacting the microstructures with a patterned PDMS stamp, and the transfer of the microstructured Si to PET was done by contacting the inked stamp on the PET substrate coated with UV-curable polyurethane (PU) and exposure to UV. This technique enabled the fabrication of high-performance devices based on high-quality microstructures from single-crystalline Si wafers (Menard *et al.* 2004; Lee *et al.* 2005; Sun *et al.* 2005; Baca *et al.* 2007), GaAs (Sun *et al.* 2004), and GaN (Lee *et al.* 2006a) and the SIO substrate via combination of top-down and bottom-up processes.



**Fig. 10.9** (a) Schematics of selective transfer of microstructured Si onto PET by using PDMS stamp. (b) Sacrificial  $\text{SiO}_2$  layer for pattern transfer. (c) Letter "A" composed of microstructured Si (from Lee, K.J., Motala, M.J., Meitl, M.A., Childs, W.R., Menard, E., Shim, A.K., Rogers, J.A. and Nuzzo, R.G. (2005) *Adv. Mater.* **17**, 2332 ©2005 Copyright Wiley-VCH Verlag GmbH & Co. KGaA, reproduced with permission).

Aligned nanowire films can be also obtained via printing of the inked stamp after transfer of the nanowires, grown on the mother substrate with an alignment, onto the stamp. That is a transfer technique of nanowires grown aligned on the mother substrate to the substrate for the nanodevice fabrication. Figure 10.10(a) shows transfer printing of CVD-grown perfectly aligned SWNTs onto the substrate (Kang *et al.* 2007a). Deposition of 100-nm thickness gold and coating of polyimide onto CVD-grown SWNTs was done. Au layers worked as an encapsulation of SWNTs for their successful transfer from the quartz substrate onto the stamp and the polyimide enhanced the transfer of SWNTs by improving the contact with PDMS. Printing of an inked stamp onto the substrate and sequential removal of polyimide and Au resulted in the aligned SWNT films without any deterioration form the mother substrate.

Various SWNT structures of cross and triangular junction could be obtained by repetition of the printing process on the same substrate and the fabrication of SWNT patterns in the desired position could also be done by additional

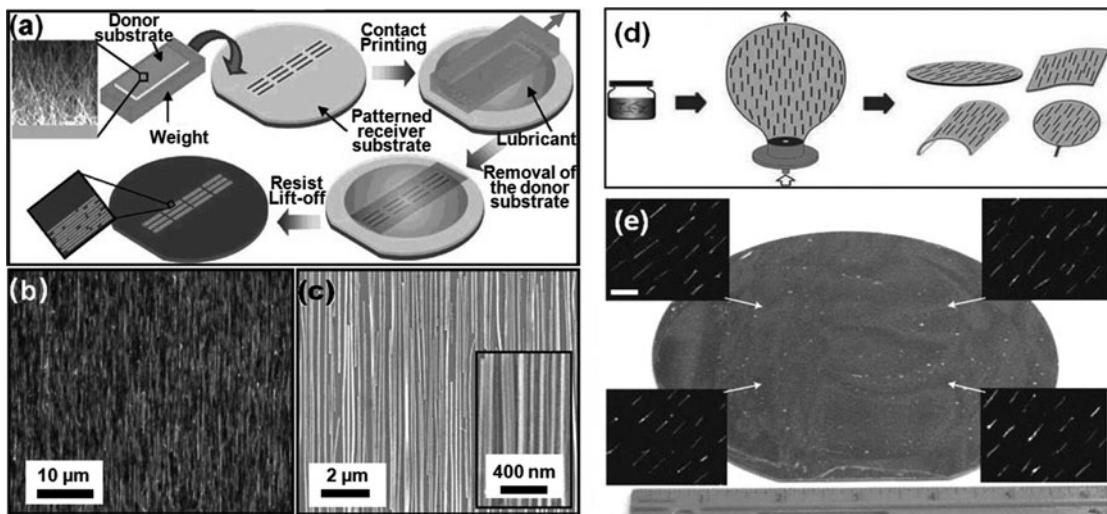


**Fig. 10.10** (a) Schematics of the transfer of grown SWNTs (from Kang, S.J., Kocabas, C., Kim, H.-S., Cao, Q., Meitl, M.A., Kang, D.-Y. and Rogers, J.A. (2007a) *Nano Lett.* **7**, 3343 ©2007 American Chemical Society, reprinted in part with permission). (b) SEM image of transferred SWNTs (c) Schematic of SWNT FET (d) SEM image of aligned SWNT channel region (from Kang, S.J., Kocabas, C., Ozel, T., Shim, M., Pimparkar, N., Alam, M.A., Rotkin, S.V. and Rogers, J.A. (2007b) *Nature Nanotech.* **2**, 230 ©2007 Nature Publishing Group, reprinted in part with permission).

photolithography and reactive ion etching processes. FET devices of aligned SWNTs showed a mobility of  $\sim 1000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . Such a transfer-printing technique is expected to be applied to the transfer of other materials onto the device substrates, regardless of their alignment.

Besides the fabrication of aligned nanowire patterns by using the originally aligned nanowires, unidirectional alignment of the nanowires from the randomly grown nanowires could be induced by intentional force during the contact-printing process. It is attributed to the anisotropic 1D structure of nanowires. As mentioned earlier, nanowires can be aligned perpendicular to the compression direction of the barrier in the LB trough (Kim *et al.* 2001; Tao *et al.* 2003; Whang *et al.* 2003) and they can be aligned along the flow of the carrier solvent inside the microfluidic channel (Messer *et al.* 2000; Huang *et al.* 2001; Park *et al.* 2006a). Figure 10.11(a) shows the patterning of nanowires via contact printing of the nanowire grown substrate onto the substrate (Javey *et al.* 2007; Fan *et al.* 2008).

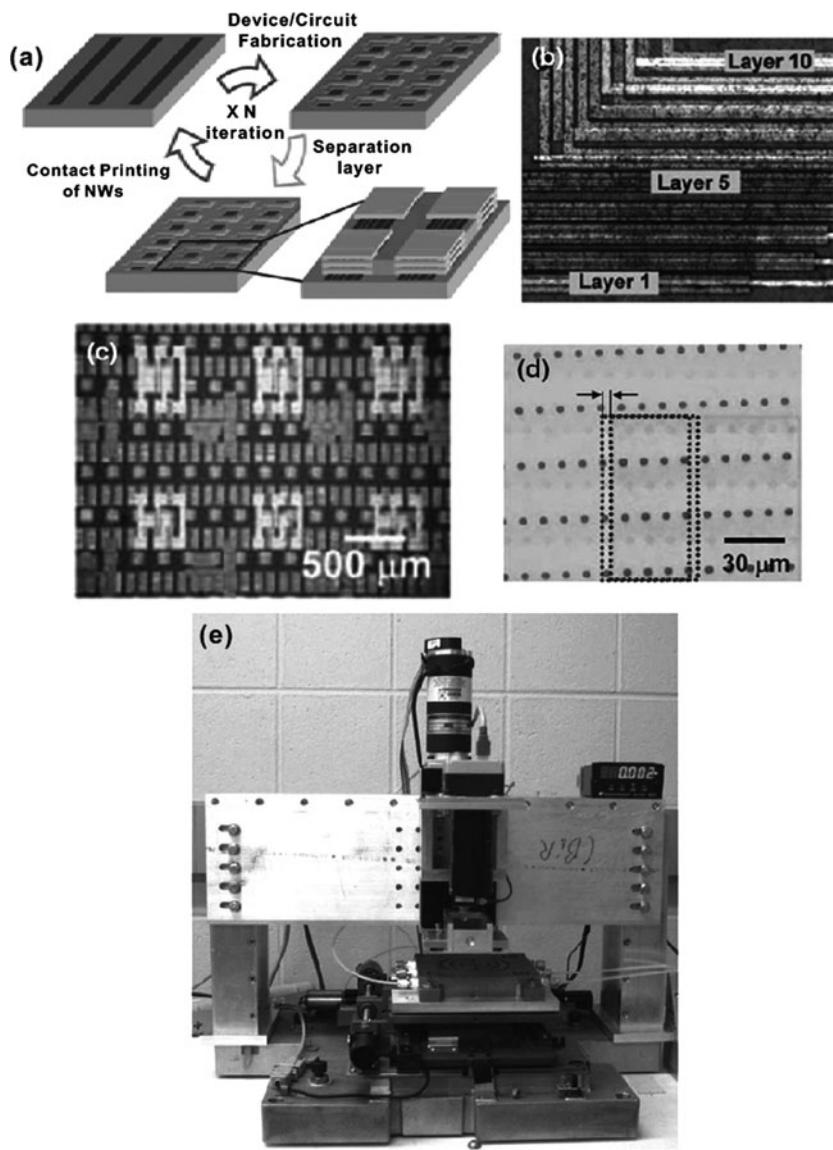
The donor substrate was contacted with the patterned substrate after the growth of the Ge/Si core/shell nanowires by CVD, and then it was pushed along the direction of the desired alignment of nanowires. During this process, nanowires are transferred onto the substrate and they are aligned by the sheer stress. Here, the cutting of nanowires due to the friction between two substrates could be prohibited by addition of a lubricant, a mixture of octane and mineral oil (Fan *et al.* 2008).



**Fig. 10.11** (a) Contact printing of nanowire arrays, (b) Optical image and (c) SEM image of Ge nanowires (from Fan, Z., Ho, J.C., Jacobson, Z.A., Yerushalmi, R., Alley, R.L., Razavi, H. and Javey, A. (2008) *Nano Lett.* **8**, 20 ©2008 Institute of Physics and IOP Publishing Limited, reprinted with permission). (d) Blowing-bubble process of SWNT alignment (e) Transferred aligned SWNTs on Si wafer (from Yu, G., Cao, A. and Lieber, C.M. (2007) *Nature Nanotech.* **2**, 372 ©2007 Nature Publishing Group, reprinted in part with permission).

Figure 10.11(b) shows the transfer of well-aligned Si nanowires onto 6" wafer substrate (Yu *et al.* 2007); alignment of the randomly dispersed emersion of Si nanowires could be done by extension stress during the blowing-bubble process and then the aligned Si nanowire patterns could be transferred by contact of the substrate onto the bubble. Such an alignment of randomly dispersed nanowires would be very useful in mass production.

Recently, three-dimensional multilayer stacks of nanowire arrays were fabricated (Ahn *et al.* 2006; Javey *et al.* 2007). Figure 10.12(a) shows the schematics of the fabrication of multistack structured devices. A maximum 9-layered device on the flexible substrate was fabricated via patterning of nanowires by transfer printing and repeated deposition of gate dielectric materials, insulating layers, or metal electrodes. The device performance on each level was uniform (Javey *et al.* 2007) and it was not deteriorated after repetition of bending. Heterogeneous three-level devices of SWNT, GaN, and Si nanoribbons were also fabricated (Ahn *et al.* 2006), where the interlayer alignment by ~2 μm was controlled by using the transfer-printing apparatus as shown in Fig. 10.12(e). Inverters by connecting the interlayer devices as well as the MOSFETs in a unilevel structure were fabricated in this way and showed a stable performance on bending. These printing-based patterning techniques can produce a wide range of multilayer and multifunctional 3D electronics in the future. Micro-contact printing, expected to replace the photolithography, is behind the other lithography techniques in the competition of minimum feature sizes, due to unexpected disadvantages. However, it will continue to play an important role in the fabrication of future electronic devices based on nanomaterials, owing to its simplicity, low cost, mild processing condition and wide applications to various nanomaterials.



**Fig. 10.12** (a) Schematic of 3D nanowire circuit fabrication, (b) 10 layers of Ge/Si nanowire FETs (from Javey, A., Nam, S.W., Friedman, R.S., Yan, H. and Lieber, C.M. (2007) *Nano Lett.* **7**, 773 ©2007 American Chemical Society, reprinted in part with permission). (c) 3D heterogeneously integrated electronic device, (d) Aligned 2-layer printed structure, (e) Automated stage for transfer printing (from Ahn, J.H., Kim, H.-S., Lee, K.J., Jeon, S., Kang, S.J., Sun, Y., Nuzzo, R.G. and Rogers, J.A. (2006) *Science* **314**, 1754 ©2006 American Association for the Advancement of Science, reprinted with permission).

## 10.3 Directed assembly of nanostructures

### 10.3.1 Directed assembly procedure

In this section, we describe a large-scale directed self-assembly process of nanostructures (Rao *et al.* 2003) without use of any external forces. Here, the substrate potential can be used to only further enhance the nanotubes (NTs) and nanowires (NWs) adsorption. This new assembly strategy is often called the “surface programmed assembly (SPA)” process (Hong *et al.* 2008). It is comprised of two simple steps: surface molecular patterning and

surface-directed assembly of nanostructures. Using this method, NTs and NWs can be assembled onto specific locations of solid substrates.

#### 10.3.1.1 *Carbon nanotubes*

Im *et al.* performed a SPA process to assemble single-walled carbon nanotubes (SWNTs) on various polar and non-polar self-assembled monolayers (SAMs) such as 2-mercaptopuridine (2MP), 4-mercaptopuridine (4MP), 2-mercaptopimidazole (2MI), cysteamine, 16-mercaptophexadecanoic acid (MHA) and 1-octadecanethiol (ODT) (Fig. 10.13(a)) (Im *et al.* 2006a,b).

The SAM was patterned onto the Au substrate to create surface regions with different polarities. At first, the first molecular species are patterned via DPN (Hong *et al.* 2000; Ivanisevic *et al.* 2001) or the microcontact printing (MCP) method (Kumar *et al.* 1993; Xia *et al.* 1998), and the remaining area is backfilled with the second molecular species by dipping the previously patterned substrate in the solution of the second molecules. Here, MHA is used for negatively charged areas and ODT for neutral regions. Cysteamine, 2MP, 4MP and 2MI are utilized to create positively charged surfaces. For SWNT assembly, the patterned substrate is placed in the SWNT suspension, usually for 10 s. Then, the substrate is rinsed thoroughly with 1,2-dichlorobenzene to remove extra SWNTs since the solution of SWNTs is dispersed in 1,2-dichlorobenzene.

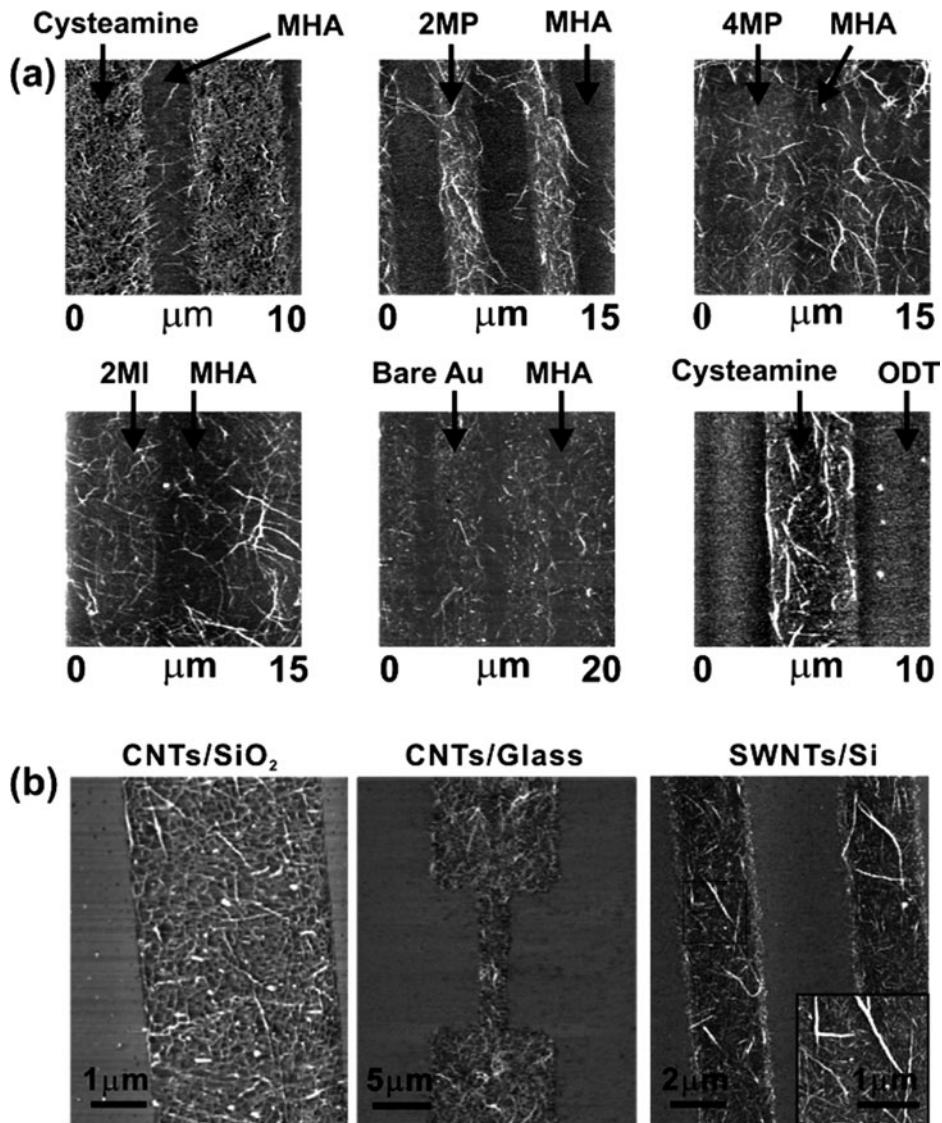
Lee *et al.* also reported the assembly of SWNTs onto various pristine substrates without any linker molecules using a similar method (Lee *et al.* 2006b). In this process, photoresist is first patterned onto solid substrates. And then, the substrate is rinsed with anhydrous hexane. After rinsing, the substrate is immediately dipped into the OTS solution for selective passivation. Finally, the photoresist is removed with acetone. Such a patterning of OTS layers results in some neutral regions and other bare surface regions on the substrates. The patterned substrate is then placed in a solution of SWNTs for selective assembly (Fig. 10.13(b)).

Furthermore, electric potential can be used to increase the adsorption of SWNTs (Fig. 10.14). However, the final assembled structures of adsorbed SWNTs are still completely determined by surface molecular patterns. Since electric potential is used only to control the amount of adsorbed SWNTs, precise alignment of SWNTs even with applied bias voltage can be possible. This SPA process can also be utilized to assemble double-walled carbon nanotubes (DWNT) and multiwalled carbon nanotubes (MWNT).

#### 10.3.1.2 *V<sub>2</sub>O<sub>5</sub> nanowire*

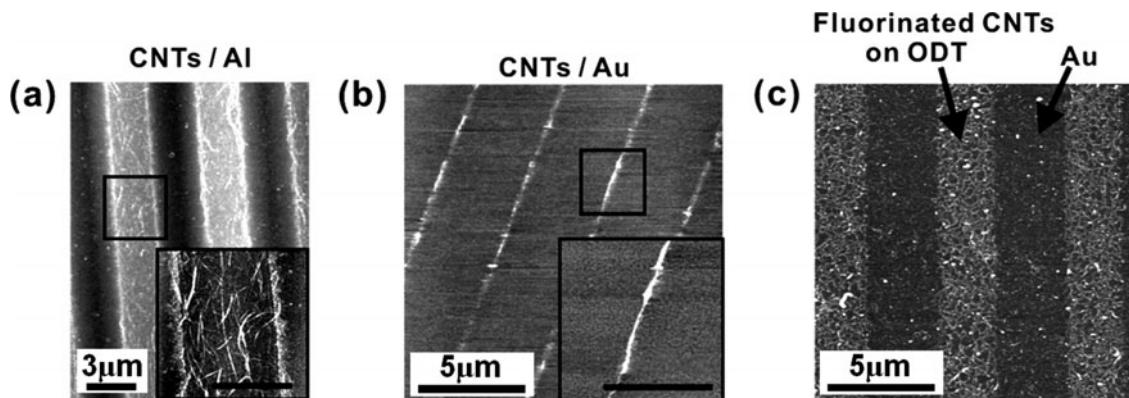
Myung *et al.* demonstrated precise assembly and alignment of V<sub>2</sub>O<sub>5</sub> nanowire arrays and nanowire-based devices over a large surface area ( $\sim 1\text{ cm} \times 1\text{ cm}$ ) (Myung *et al.* 2005). Specifically, positively charged surface molecular patterns are utilized to assemble and align V<sub>2</sub>O<sub>5</sub> NWs over a large area. The neutral molecular patterns are utilized to avoid the adsorption of nanowires (Figs. 10.15(a)–(d)).

A molecular patterning process is performed similarly by patterning the first molecular species via DPN, MCP or photolithography (Jaeger 2002; Busnaina 2007) as discussed before, and then the remaining area is backfilled with the second molecular species by a dipping process. The photolithography



**Fig. 10.13** (a) AFM topography images of SWNT patterns adsorbed onto SAM patterns comprised of: cysteamine and MHA, 2MP and MHA, 4MP and MHA, 2MI and MHA, 4MP and 2MI, 2MI and ODT. These patterns cover large surface area ( $\sim 1\text{ cm} \times 1\text{ cm}$ ) of the samples (from Im, J., Lee, M., Myung, S., Huang, L., Rao, S.G., Lee, D.G., Koh, J. and Hong, S. (2006) *Nanotechnology*, **17**, 3569, ©2006 Institute of Physics and IOP Publishing Limited, reproduced with permission). (b) SWNT networks formed directly on bare SiO<sub>2</sub> (left), glass (middle) and Si (right). OTS was used as the non-polar layer (from Lee, M., Im, J., Lee, B.Y., Myung, S., Kang, J., Huang, L., Kwon, Y.-K. and Hong, S. (2006) *Nature Nanotech.* **1**, 66 ©2006 Nature Publishing Group, reproduced in part with permission).

method is utilized to create large-scale molecular patterns comprised of 1-octadecyltrichlorosilane (OTS) for the neutral region and aminopropylmethoxysilane (APTES) for positively charged patterns on SiO<sub>2</sub>. In this case, photoresist is first patterned on SiO<sub>2</sub>, and the substrate is placed in the OTS solution (1:500 v/v in anhydrous hexane) for coating the bare SiO<sub>2</sub> regions.



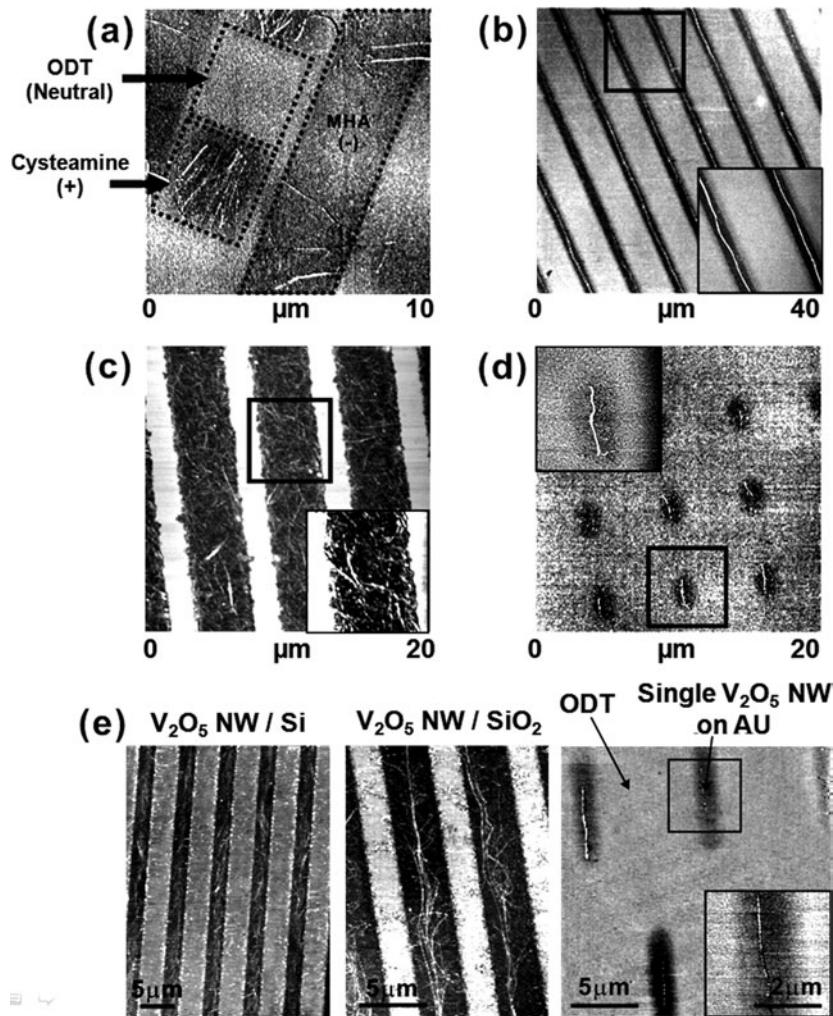
**Fig. 10.14** (a) SWNTs selectively adsorbed onto bare Al surfaces with a substrate bias of  $-1.5\text{ V}$ . OTS was used for passivation. (b) High-precision alignment of SWNTs onto Au surfaces with  $-1\text{ V}$  substrate bias. ODT was used as the non-polar layer. (c) Fluorinated SWNTs adsorbed onto ODT regions with  $-3\text{ V}$  substrate bias (from Lee, M., Im, J., Lee, B.Y., Myung, S., Kang, J., Huang, L., Kwon, Y.-K. and Hong, S. (2006) *Nature Nanotech.* **1**, 66 ©2006 Nature Publishing Group, reproduced in part with permission).

After the photoresist is removed with acetone, the substrate is dipped into APTES solution (1:500 v/v in ethanol) to backfill the remaining  $\text{SiO}_2$  area.

When the patterned substrates are placed in an aqueous solution of  $\text{V}_2\text{O}_5$  NWs for  $\sim 30\text{ s}$ , NWs are attracted toward the positively charged molecular regions, and aligned along the molecular patterns due to the negatively charged property of  $\text{V}_2\text{O}_5$  NWs (Fig. 10.15(c)). The substrate is then rinsed with thoroughly deionized water to remove any extra NWs. For Au substrate, SAM patterns are prepared using cysteamine as positively charged patterns and ODT for neutral regions, and successfully assembled a large number of pristine NWs over a  $1\text{ cm} \times 1\text{ cm}$  area (Figs. 10.15(a), (b), (d)). Since  $\text{V}_2\text{O}_5$  NWs do not adhere to bare Si or  $\text{SiO}_2$  surfaces without a bias voltage, electric potential can also be utilized to enhance the adsorption of NWs. As  $\text{V}_2\text{O}_5$  NWs are charged negatively in aqueous solution, we could assemble them by applying positive substrate bias voltage. Furthermore, even individual  $\text{V}_2\text{O}_5$  NWs can be assembled in small patterns on the substrates with the help of electric potential (Fig. 10.15(e)).

#### 10.3.1.3 *ZnO* nanowires

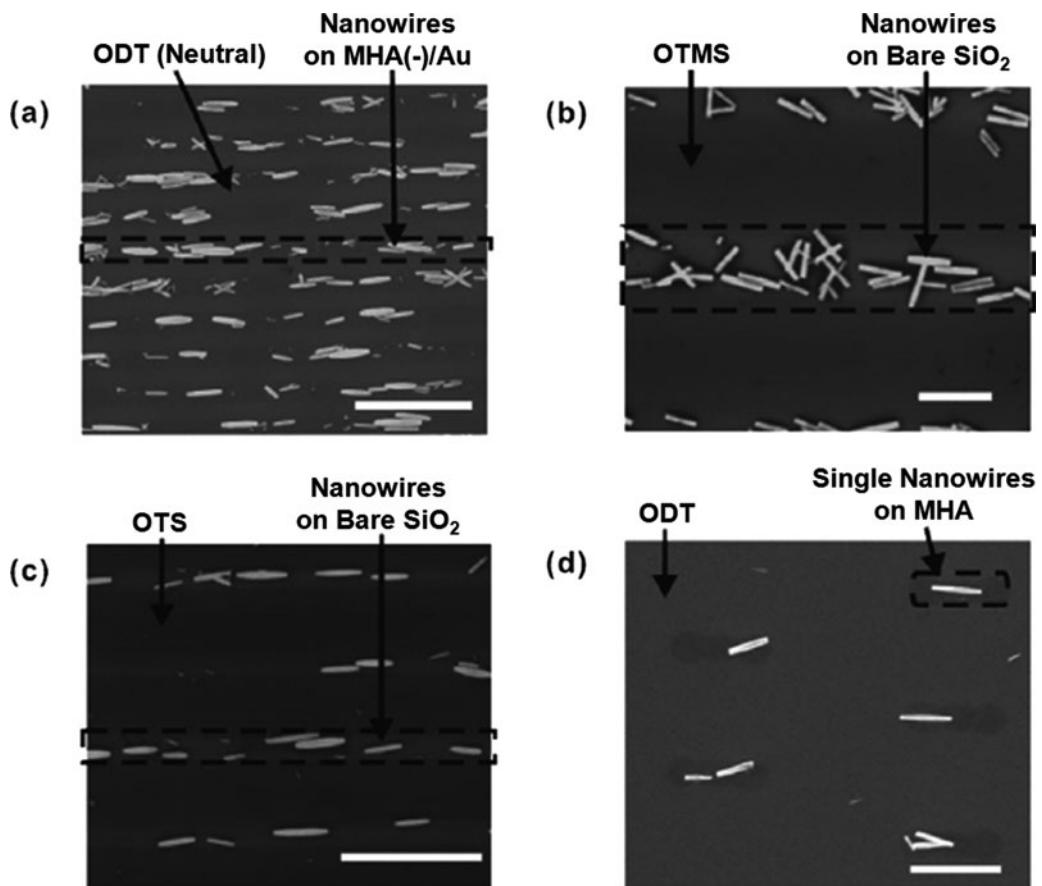
Kang *et al.* also reported ZnO NW adsorption and alignment using a similar method as described above (Kang *et al.* 2008). First, on the Au surfaces, MHA is used to create negatively charged molecular patterns, while ODT is used for neutral regions. On  $\text{SiO}_2$  substrates, the bare  $\text{SiO}_2$  surface is used as a negatively charged surface, while OTS or octadecyltrimethoxysilane (OTMS) are first patterned for neutral regions, while leaving some bare  $\text{SiO}_2$  regions. When the patterned substrate is placed in the ZnO NW solution for  $\sim 10\text{ min}$ , pristine ZnO NWs are attracted to the negatively charged region, and then assembled and aligned directly onto MHA region or the  $\text{SiO}_2$  surface (Fig. 10.16). The substrate is then rinsed with deionized water to remove any extra NWs. Moreover, the amount of adsorbed ZnO NWs could also be enhanced by applying a negative bias voltage to the substrates as in the previous NTs and NWs assembly.



**Fig. 10.15** (a) Atomic force microscopy (AFM) topography image of  $\text{V}_2\text{O}_5$  nanowires assembled onto SAM patterns comprised of ODT (neutral), MHA (negatively charged), and cysteamine (positively charged). (b) AFM topography image of  $\text{V}_2\text{O}_5$  nanowires (white lines) assembled on cysteamine patterns (dark areas) on Au. The remaining area was passivated with ODT (bright areas). (c) AFM topography image of  $\text{V}_2\text{O}_5$  nanowires assembled on APTES patterns on  $\text{SiO}_2$ . OTS was utilized for passivation. Insets show high-resolution AFM images of OTS (left) and APTES regions (right), respectively. (d) AFM topography image of individual  $\text{V}_2\text{O}_5$  nanowires (white lines) assembled on cysteamine patterns (dark areas) on Au. ODT (bright areas) was utilized for passivation (from Myung, S., Lee, M., Kim, G.T., Ha, J.S. and Hong, S. (2005) *Adv. Mater.* **17**, 2361 ©2005 John Wiley & Sons Limited, reproduced with permission). (e)  $\text{V}_2\text{O}_5$  NWs selectively adsorbed onto bare Si (left) and  $\text{SiO}_2$  (middle) surfaces using a substrate bias 3 V and 6 V, respectively. Individual  $\text{V}_2\text{O}_5$  NWs directly adsorbed onto bare Au surfaces using 3 V substrate bias using ODT as the non-polar layer (from Lee, M., Im, J., Lee, B.Y., Myung, S., Kang, J., Huang, L., Kwon, Y.-K. and Hong, S. (2006) *Nature Nanotech.* **1**, 66 ©2006 Nature Publishing Group, reproduced in part with permission).

### 10.3.2 Mechanism of various assembly processes of nanostructures

In this section, we discuss a theoretical model describing the collective behavior of molecules in direct deposition processes such as dip-pen nanolithography and other directed assembly procedures. Interplay between intermolecular interactions and substrate–molecule interactions may result in diverse shapes of molecular patterns. Computer simulations based on a simple model reveal circular and star-like patterns at low and intermediate densities of preferentially attractive surface sites, respectively. At a large density of such surface sites, the molecules form a two-dimensional invasion percolation cluster. We also discuss the self-assembly process of a long rod-like molecule on patterned substrate immersed in a two-dimensional sliding chamber.



**Fig. 10.16** Scanning electron microscopy (SEM) images and analysis results of ZnO NWs assembled on the various substrates via the solution-phase method. (a) ZnO NWs assembled on MHA line patterns on a Au substrate without a substrate bias voltage. The remaining area was passivated by ODT to prevent non-specific adsorption of ZnO NWs. The scale bar represents 20  $\mu\text{m}$ . (b) ZnO NWs assembled directly onto bare SiO<sub>2</sub> surfaces without substrate bias voltage. OTMS SAM was utilized for passivation. The scale bar represents 5  $\mu\text{m}$ . (c) ZnO NWs assembled directly onto bare SiO<sub>2</sub> surfaces without a substrate bias voltage. OTS SAM was utilized for passivation. The scale bar represents 20  $\mu\text{m}$ . (d) Individual ZnO NWs assembled onto MHA patterns on Au without a substrate bias voltage. ODT was utilized for passivation. The scale bar represents 5  $\mu\text{m}$  (from Kang, J., Myung, S., Kim, B., Oh, D., Kim, G.T. and Hong, S. (2008) *Nanotechnology* **19**, 095303 ©2008 Institute of Physics and IOP Publishing Limited, reprinted with permission).

### 10.3.2.1 Modelling of collective behavior

In a common DPN process, a molecule-coated AFM tip is held at a fixed position in contact with the substrate so that molecules flow onto the substrate to form various-size “dot” patterns. Various direct deposition experiments show that if deposited molecules strongly bind to the substrates, the pattern always grows in a circular shape (Piner *et al.* 1999; Hong *et al.* 2000; Rozhok *et al.* 2003; Ginger *et al.* 2004; Peterson *et al.* 2004). Examples include alkanethiol molecules on gold. On the other hand, when the binding to the substrates is relatively weak compared to the attractive molecule–molecule interactions, non-circular patterns are often observed. Most representative patterns are fractal-like patterns with the fractal dimension of  $\sim 1.51$  when 1-dodecylamine (DDA) molecules are deposited onto mica (Manandhar *et al.* 2003). Other

examples include proteins and conducting polymers (Wilson *et al.* 2001; Lim *et al.* 2002).

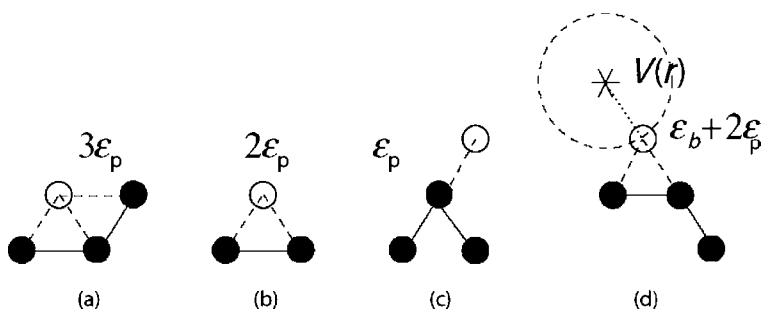
Several theoretical models (Jang *et al.* 2001; Sheehan *et al.* 2002; Weeks *et al.* 2002) describe the motion of deposited molecules as Fickian diffusion on top of other molecular layers until they covalently anchor to bare substrates. The application of this model seems to be limited to the circular pattern formation of strong surface-binding molecules. Obviously, the diffusion model is insufficient to explain non-circular pattern formation, which is common in weak surface-binding cases.

Lee and Hong (2006) proposed a theoretical model that successfully accounts for diverse shapes of molecular patterns in dip-pen nanolithography of weak-binding molecules. In this model, the focus has been placed on the role of intermolecular interactions and inherent non-uniform substrate–molecule interactions. In contrast to previous models where deposited molecules diffuse independently, here the motion of molecules is identified as collective rearrangements of a cluster whose periphery adapts its shape to form the minimum energy state for all. The deposited molecules bind weakly onto most of the substrate regions except for several impurity sites on substrates. Furthermore, attractive intermolecular interactions combine all molecules together. The interactions with strong binding sites result in non-trivial pattern formation when molecules aggregate via attractive intermolecular interactions. Lee and Hong have pointed out that for charged molecules such as DDA, randomly distributed surface charges on mica play roles as strong binding sites. For DDA, molecules are attracted to each other due to hydrophobic interactions. As the molecules are preferentially attracted to certain surface sites, the tension can be released by making a channel toward the strong binding sites where molecules can flow in. The frontier moves by invading local areas in bursts. The frontier will become circular again unless it encounters any other attractive surface sites. When the frontier encounters other attractive sites, a new branch appears on the top of the growing branch.

Computer simulations based on a simple model taking into account the above feature demonstrate various self-assembly patterns (Lee and Hong 2006). The molecules live on two-dimensional hexagonal lattices. Each lattice site can be denoted using the index  $(i, j)$ . The growth starts from the center by occupying neighboring sites that provide the maximum energy gain. When there are multiple minimal energy sites, the closest to the center becomes occupied first. The Hamiltonian of a cluster with  $M$  deposited molecules is

$$\frac{H}{k_B T} = - \sum_{\{i,j\}}^M \varepsilon_p s_{i,j} (s_{i,j+1} + s_{i,j-1} + s_{i-1,j+1} + s_{i+1,j-1} + s_{i+1,j} + s_{i-1,j}) + \sum_n V(|R_{i,j} - r_n|),$$

where  $\varepsilon_p$  is the energy gain by pairwise interactions between molecules, and it contributes to the total energy gain only when two molecules are next to each other. The  $s_{ij}$  is an Ising-type variable indicating the occupancy of the corresponding site.  $V(r)$  is the potential energy of a molecule at distance  $r$  apart from the strong surface binding site.  $r_n$  and  $R_{i,j}$  are the position of

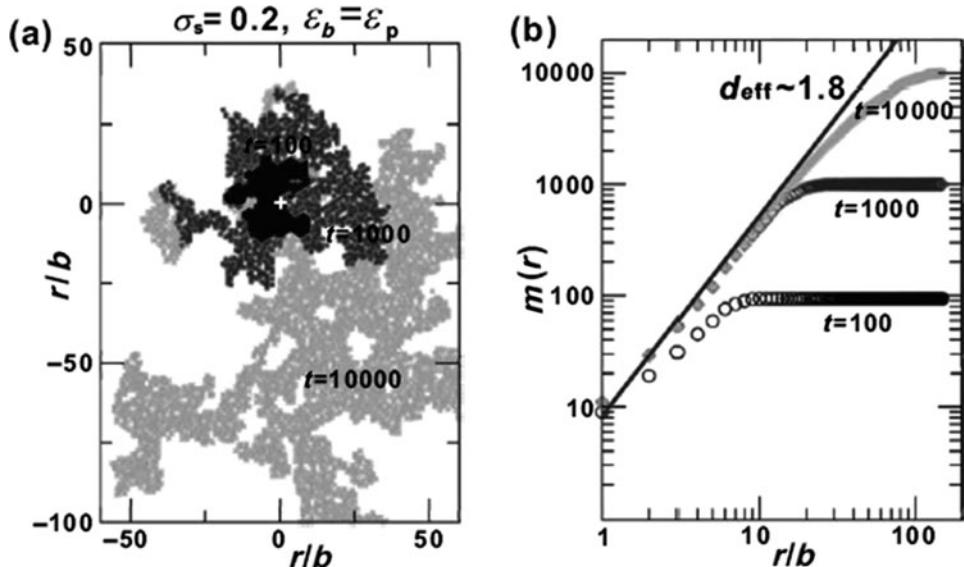


**Fig. 10.17** Schematic representations of a cluster growth on hexagonal lattices. A deposited molecule may occupy a new sites (o) by creating (a) 3 pairs, (b) 2 pairs, (c) 1 pair of intermolecular contacts and (d) by advancing toward a strong binding site (\*).

the  $n$ th strong-binding site and that of the molecule at the lattice site  $(i, j)$ , respectively.

Figure 10.17 demonstrates the advance of the cluster frontiers. The energy gain by advancing from a concave boundary (Fig. 10.17(a)), a flat boundary (Fig. 10.17(b)), and a convex site (Fig. 10.17(c)) are  $3\epsilon_p$ ,  $2\epsilon_p$ , and  $\epsilon_p$ , respectively. Thus, the growing frontier has a tendency to fill the concave site first, leading to a circular-shape frontier. When a molecule is within the interaction range  $r_b (\geq b)$  with a strong binding site, the energy gain by growing into the direction of the strong-binding site is  $\epsilon_b + n \epsilon_p$  where  $n$  is the number of intermolecular interaction pairs by making a bridge (Fig. 10.17(d)). Here,  $\epsilon_b = \delta V$  indicates the potential-energy gain at the periphery of the interaction range. If the substrate–molecule interaction wins over the line tension of the cluster boundary ( $\epsilon_b > \epsilon_p$ ), the molecules can advance in the direction of the strong-binding site rather than filling an available concave site. The major parameters controlling the shape of the cluster growth are: (a) the number density of strong-binding sites and (b) the strength and range of the substrate–molecule and intermolecular interactions.

This simple but efficient framework can be applied for various DPN patterning experiments simply by plugging in proper parameters. As an example, we consider DDA molecules on mica, which can be a representative case of weak surface-binding molecules. DDA molecules have hydrophobic backbones with amine ( $-\text{NH}_2$ ) end-groups. When deposited on mica, they aggregate among themselves via hydrophobic interactions and are bound rather weakly to the surface sites. Randomly distributed surface charges on mica act as strong binding sites and attract positively charged DDA molecules. The hydrophobic interaction is  $\sim 1$  kcal/mol per methylene group. Thus,  $\epsilon_p \approx 10$  kcal/mol for a DDA pair consisting of 12 methylene groups provided that molecules are well aligned. Since mica surfaces are normally covered by a thin water layer, the Bjerrum length  $l_B \equiv e^2/k_B T \approx 7$  Å in water sets the interaction range where the electrostatic interactions are active. The diameter of a deposited DDA molecule sets the lattice constant of the hexagonal lattices to be  $b = 3$  Å. Thus, the “bridging” interaction range  $r_b$  is  $r_b = 2b$  on the lattices. Considering that the typical strength of the electrostatic interaction at an atomic distance is 40–50 kcal/mol, we estimate the interaction strength  $\epsilon_b = \delta V$  at a Bjerrum length away to be about  $2\epsilon_p$ . The reported surface charge density  $-0.34$  C/m<sup>2</sup> (Pashley 1981) of mica corresponds to the number density  $\sigma_s = 0.13$  on this

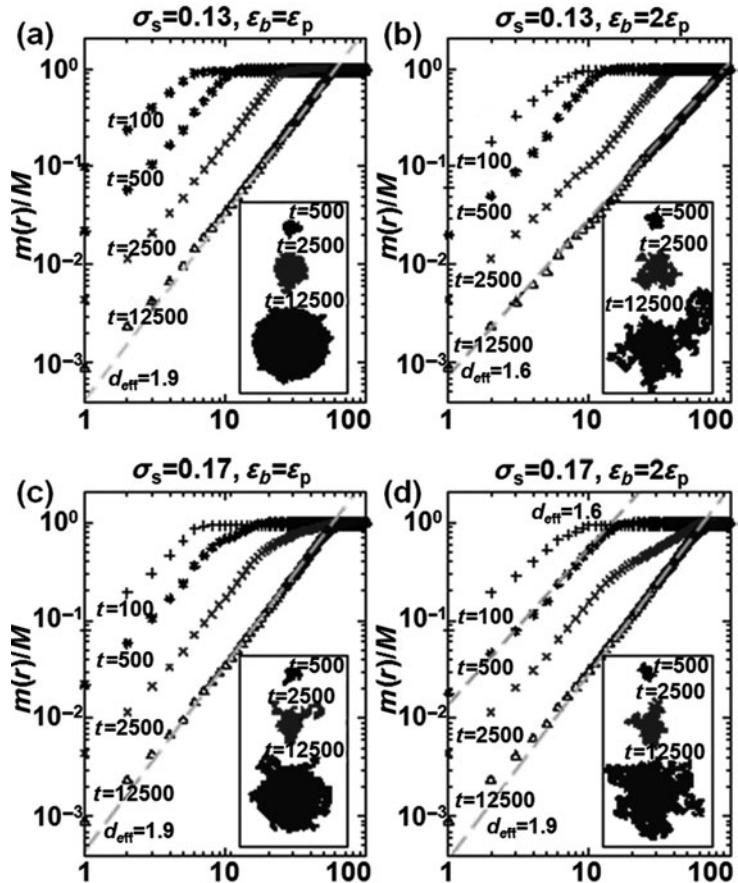


**Fig. 10.18** (a) Cluster growth in invasion percolation regime. (b) The number of deposited molecules  $m(r)$  inside of a circle of radius  $r$ .

hexagonal lattices ( $b = 3 \text{ \AA}$ ). The simulation results using these parameters predict the growth of irregular-shape clusters similar to the reported DPN experiments using DDA molecules (Manandhar *et al.* 2003).

The parameters for simulations are extracted from various types of molecules. Figure 10.18 demonstrates cluster formation with  $\sigma_s = 0.2$  and  $\varepsilon_p = \varepsilon_b$ . At this specific setup, the cluster growth resembles that of the *invasion percolation* process in fluid dynamics (Wilkinson *et al.* 1983). For instance, when water is injected slowly into a porous medium filled with oil, a water cluster grows with a fractal dimension of 1.82 (Furuberg *et al.* 1988). Here, the dynamics are determined by the capillary forces and the pore size. We measured the *effective* fractal dimension  $d_{\text{eff}}$  of each pattern by counting the number of molecules  $M$  at given  $R \sim M^v$ ,  $v = 1/d_{\text{eff}}$ . Slopes in log-log plots depict the effective fractal dimension at given times. The measured fractal dimension of  $\sim 1.8$  for times  $t \geq 10000$  is in agreement with the known fractal dimension of invasion percolation clusters. The growth direction is chosen by a few branches at the moment of a burst. The merging of these branches will destroy the hierarchy of fractal structures at later times. The invasion percolation cluster shape appears if the strong-binding site density is larger than a certain threshold  $\sigma_c$  so that the occupancy of a strong-binding site always guarantees filling neighboring binding sites before all concave sites around it.

The cluster shapes at various values of  $\varepsilon_p$ ,  $\varepsilon_b$  and  $\sigma_c$  are investigated. In the limit of strong intermolecule interaction ( $\varepsilon_p \gg \varepsilon_b$ ), the molecules always aggregate into a trivial circular form. In the other limit ( $\varepsilon_b \gg \varepsilon_p$ ), the cluster shape can be fractal-like patterns at lower density of strong-binding sites. Figure 10.18 demonstrates some representative cluster growth. For  $\varepsilon_b \leq \varepsilon_p$  with small  $\sigma_s = 0.13 < \sigma_c$ , the overall shape is still circular (Fig. 10.19(a)). Molecules with strong intermolecular interactions fall into this category. At

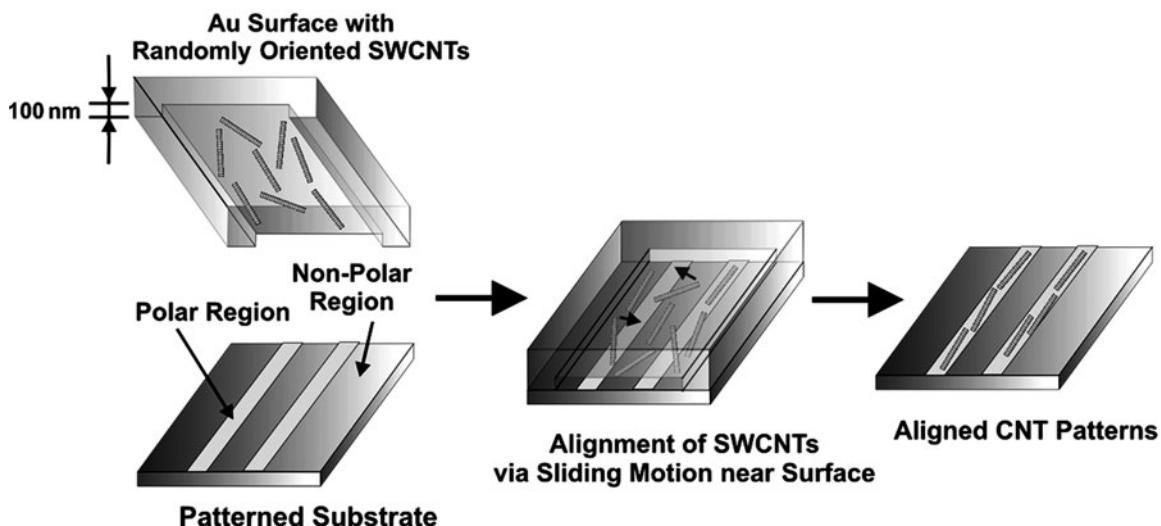


**Fig. 10.19** Clusters (inserts) simulated with various surface charge densities, hydrophobic interactions, surface–molecule interactions. The cluster size  $m(r)$  is the number of molecules with the range of  $r$  from the center. The slopes in the graph indicate the effective fractal dimension  $d_{\text{eff}}$ .

a larger substrate–molecule interaction  $\epsilon_b = 2\epsilon_p$ , we observe mixed patterns (Fig. 10.19(b)). At small times, branches and the central part are occupied with similar probabilities. At longer times, the outer branches continue developing but the central region no longer grows.

This is because, after the cluster size reaches a certain value, the lowest-energy sites are always found at the newly explored area. The measured fractal dimension is 1.6, which is similar to the reported value for DDA patterns on mica (Manandhar *et al.* 2003). It should be noted that the model parameters used in this simulation are extracted from the real interactions of DDA molecules.

For larger molecules (e.g. globular proteins), we can expect a different lattice constant. Consider molecules whose diameter is about 10% larger. Accordingly, the number density of strong-binding sites is scaled to  $\sigma_s = 0.17$  (Figs. 10.19(c) and (d)). If  $\epsilon_b = \epsilon_p$ , the pattern shows the circular core and small branches. Boundary tension still suppresses the development of fractal-like structures. For  $\epsilon_b = 2\epsilon_p$ , branching is more favorable than occupying concave sites (Fig. 10.19(d)). These branches create new branches in a recursive way similar to the case  $\sigma_s = 0.13$ ,  $\epsilon_b = 2\epsilon_p$  (Fig. 10.19(b)), but they start



**Fig. 10.20** Sliding reaction chamber experiment (from Im, J., Huang, L., Kang, J., Lee, M., Lee, D.J., Rao, S.G., Lee, N.-K. and Hong, S. (2006) *J. Chem. Phys.* **124**, 224707 ©2006 American Institute of Physics, reprinted in part with permission).

overlapping almost immediately. At earlier times ( $t < 2500$ ), the growth can be anisotropic and shows the effective fractal dimension of  $\sim 1.6$ . It is not yet clear whether the cluster shape at this regime is fractal. At later times ( $t > 2500$ ), the merged branches cover the substrates isotropically, resulting in the effective fractal dimension of  $\sim 2$ . A series of simulations on rectangular lattices reports similar behavior.

This model can also be generalized to strong surface-binding cases such as alkylthiolates on Au. In this case, all lattice sites are equally attractive to molecules. The simulation results show trivial circular patterns. One interesting situation is when strong-binding molecules can slide on the substrates due to the translation invariance. This effect can be taken into account by introducing a certain probability with which molecules can escape from the central aggregates. The escaped molecules may still diffuse on the substrate. The simulation results confirm a circular pattern with a blurred boundary, as observed in some DPN experiments (Sheehan *et al.* 2002).

#### 10.3.2.2 Self-assembly of long molecules

We now discuss the adsorption of stiff rod-like molecules on patterned surfaces. Since long molecules have internal degrees of freedom that relax only at large timescales, the detailed description of kinetics is required to characterize the adsorption behaviors.

It is suggested that the selective adsorption of long polymer molecules can be used to create nanoscale devices. In particular, such molecules like CNT, DNA are extremely useful from the technological point of view. Various adsorption experiments of long molecules on a patterned surface have been carried out. In order to direct CNT or NWs, electric fields or external liquid flow have been utilized (Huang *et al.* 2001; Krupke *et al.* 2003). It is also shown that (Liu *et al.* 2002; Rao *et al.* 2003; Nuraje *et al.* 2004; Tsukruk *et al.* 2004; Myung *et al.* 2005) the SAM (self-assembled monolayer) patterns can

be utilized to direct the assembly of nanoscale wires such as SWNT (Iijima 1991) onto specific locations of the substrates. The experimental images of Im *et al.* (2006c) suggest that adsorption of these molecules is irreversible to some extent, i.e. in timescales of relaxation time of CNT,  $\tau \propto L^3$ . This also suggests irreversible adsorption kinetics plays an important role to determine the adsorption patterns.

The adsorption of long molecules can be considered in two stages: (a) Landing from the bulk to attractive surface sites; (b) Relaxation of the internal degree of freedom of a molecule. The first step, the adsorption (landing) of SWNTs onto relatively large polar SAM patterns, can be described by the standard Langmuir isotherm (Karpovich and Blanchard, 1994). In the Langmuir isotherm, the adsorption rate is proportional to the bulk concentration and the number of the vacant sites, and the desorption rate is proportional to the number of adsorbed molecules. As more and more molecules are adsorbed, the number of available adsorption sites becomes smaller, until the equilibrium coverage is reached. In aqueous solution, the attraction between SWNT is estimated to be about  $40k_B T/\text{nm}$ . This means typical micrometer-long CNTs are strongly attractive to each other resulting in heavy aggregation of SWNTs in aqueous solution. In dichlorobenzene solvent, however, SWNTs disperse very well, implying that the attractive forces between CNTs are yet much weaker than that in water. Hence, the progress of SWNT adsorption onto polar SAM favors less and less incoming flux. If we neglect CNT–CNT interaction ( $\varepsilon_{\text{CNT}} \ll \varepsilon_{\text{ads}}$ ), the equilibrium surface coverage is obtained when  $\varepsilon_{\text{ads}} + k_B T \log(\theta b^2) = k_B T \log(Cb^3)$  with  $\theta$  and  $C$  being the surface and bulk concentration of CNTs, respectively, and  $b$  is the length unit.

Let us consider a SWNT that is confined in a two-dimensional plane, a so-called “sliding chamber”, in order to investigate the relaxation of the internal degree of freedom as shown in Fig. 10.2 (Im *et al.* 2006a).

For a long molecule to adsorb, the initial segments landed with a certain angle with patterned attractive sites adsorb first. Rotational diffusion allows exploring around the fixed strand. If only a fraction of a SWNT falls on the attractive region, the remaining fluctuating part (landed in a non-polar part) diffuses and eventually adsorbs on the polar region, trading off the bending energy cost. This kinetics is similar to Kramer’s reaction theory (reaction through the energy barrier).

The rotational and translational mobility of SWNT depends on the surface conditions. The reason for small mobility on polar SAMs is attributed to strong adhesion forces and more importantly “chemical roughness”, which often appears in experimental situations and breaks translational invariance. Experimentally, the “sliding” mobility to SWNTs in a sliding chamber is provided by thermal energy or ultrasonic vibration until the SWNTs spread uniformly on the surface. Without sonication, SWNTs start landing. In the chamber, the segments of the SWNTs perform mainly diffusive motions until they are captured by the interaction range of surface binding sites. The aforementioned segments, lying initially on the polar region, are immobile at least on the timescale of interest during which segments on the non-polar region move and accumulate at the polar boundary. If the whole SWNT falls on the non-polar region, it will diffuse until some part of the CNT is captured by the polar region

and aligns along the polar–non-polar boundary. If some fraction of a SWNT enters the interaction ranges of a polar SAM, the segment further moves into the pattern until it reaches the other boundary. This process is deterministic. The parts of a SWNT in the non-polar region diffuse until they encounter the attractive polar region (Fig. 10.20). Subsequent adsorption of the remaining segment results in bending toward the polar region with accumulation of bending energy. The strong CNT–polar SAM interactions entail progressive zipping of CNT segments on the polar–non-polar boundary.

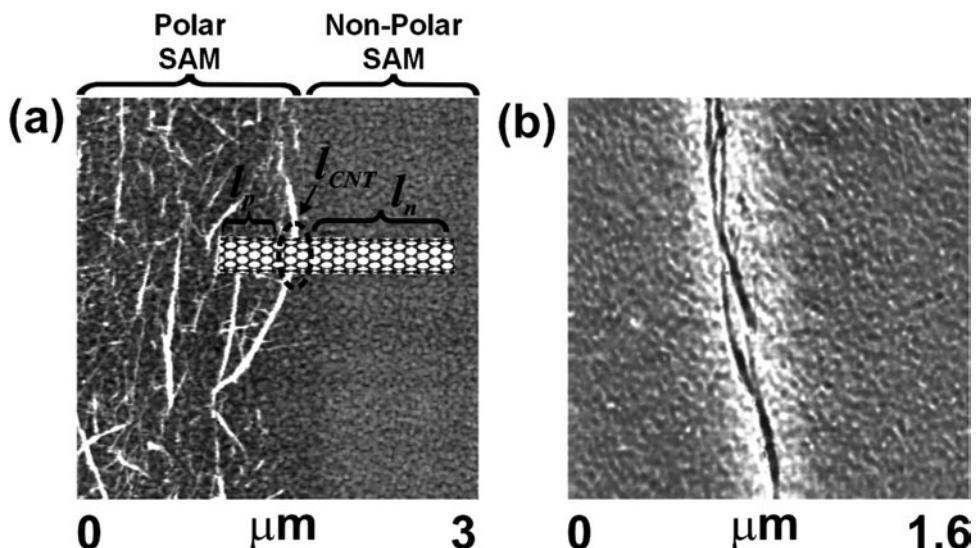
The total energy of a SWNT whose fraction is adsorbed on the polar region consists of the adsorption energy of CNT on the polar substrates and bending energy  $E_{\text{elastic}}$ :

$$E_{\text{tot}} = \int_0^{l_p} \gamma_p ds + \frac{\kappa}{2} \int_0^L \left( \frac{d^2 r}{ds^2} \right)^2 ds,$$

where  $\gamma_p (< 0)$  is the interface energy per unit length for the SWNT segments on a polar SAM and  $\kappa/k_B T$  is the so-called persistence length that represents the bending stiffness of SWNTs.  $L, l_p$  stands for the total SWNT length and the segment length landed on the polar region. The integration is performed following the contour of the SWNT segments. An adsorption experiment (Im *et al.* 2006a) indicates that the attractive energy of SWNT segments on a non-polar SAM or other SWNTs is negligible, and the interface energy terms of those segments are omitted. We also neglect the entropic contribution, which is of the order of  $k_B T$ . The equilibrium conformation will be obtained by minimizing the free-energy condition. The aggregated CNTs expected to align straight and accumulate in polar SAM regions (Fig. 10.21(b)). We found that the SWNTs are often trapped in local energy-minimum states. In order to escape from local minima, a large activation energy is required. We estimate the energy barrier from the curved conformation of CNT bundles at the polar boundary (Fig. 10.21(a)). The curved structure lasts as long as the accumulated bending energy is smaller than the energy barrier. When the total stored energy exceeds the energy barrier, the overall SWNT shape straightens, as is shown in Fig. 10.21(b). From the measured value of the typical radius of curvature  $r \sim 1.34 \mu\text{m}$  of a micrometer-long SWNT segment, we estimate that the restoring force is about 0.056 pN and the typical energy barrier associated with the roughness for a micrometer-long SWNT is about  $13k_B T$  per  $1 \mu\text{m}$ .

## 10.4 Characteristics of self-assembled hybrid nanodevices

The feature size of silicon-based transistors has been reduced exponentially as Moore's law indicates. Nowadays, current semiconducting industries have, however, confronted a fundamental obstacle preventing further size reduction by top-down approaches. Over the past decade, various research efforts in nanoscience and nanotechnology have shown that nanostructured materials including carbon nanotubes and nanowires can be used as various electronic



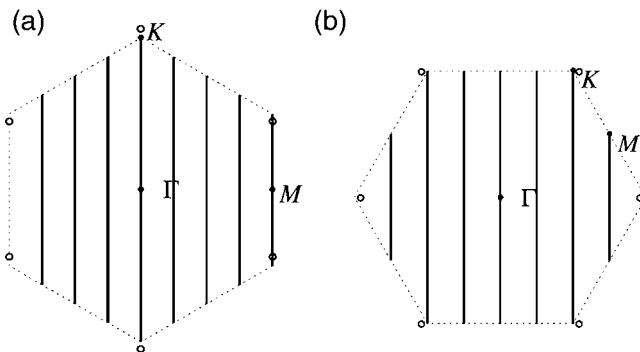
**Fig. 10.21** (a) Schematic diagram showing a SWNT assembled on pre-existing SWNT patterns and AFM topography image of SWNTs assembled onto the patterns on polar and non-polar regions. (b) High-resolution lateral force image of SWNTs aligned on polar SAM patterns (from Im, J., Huang, L., Kang, J., Lee, M., Lee, D.J., Rao, S.G., Lee, N.-K. and Hong, S. (2006) *J. Chem. Phys.* **124**, 224707 ©2006 American Institute of Physics, reprinted in part with permission).

devices or device elements. In this section, we will describe device characteristics of a wide range of such new devices from a single carbon-nanotube device to self-assembled network devices.

#### 10.4.1 Electronic and transport properties of carbon nanotubes

Depending on their chirality and diameter, single-wall nanotubes can be either metallic or semiconducting (Hamada *et al.* 1992; Mintmire *et al.* 1992; Saito *et al.* 1992). In this section, we review this unique electronic structure and electrical transport properties of single-wall carbon nanotubes.

The electronic band structure of a nanotube somewhat resembles that of a graphene sheet. The difference between the two cases comes from the periodic boundary conditions. A graphene sheet is regarded as an infinitely extended two-dimensional plane, whereas a nanotube is a one-dimensional structure that is infinitely long along the tube axis. However, the periodic boundary condition,  $\mathbf{C} \cdot \mathbf{k} = 2\pi l$ , is imposed for a finite period along the circumference. Here,  $\mathbf{C}$  is a chiral vector defining the chiral index  $(n, m)$  of a nanotube;  $\mathbf{k}$  is an allowed wave vector; and  $l$  is an integer. This results in Bloch wavefunctions with discretely selected wave vectors, which are shown by the straight lines in Fig. 10.22, in the first Brillouin zone of a graphene sheet. Note that the bonding and antibonding  $\pi$  bands, which are located near the Fermi level, are degenerate at the  $K$  point in the Brillouin zone of a graphene sheet, while  $sp^2 \sigma$  bands are located far from the Fermi level. Therefore, if at least one of the selected wave vectors crosses the  $K$  point, the nanotube becomes metallic.



**Fig. 10.22** Due to the periodic boundary condition along the circumference, only discrete wave vectors are allowed on the Brillouin zone of a graphene sheet. The allowed wave vectors are, for example, shown in (a) for a (4, 4) armchair nanotube and in (b) for a (6, 0) zigzag nanotube. Open circles show the degenerate point shifted from the K point.

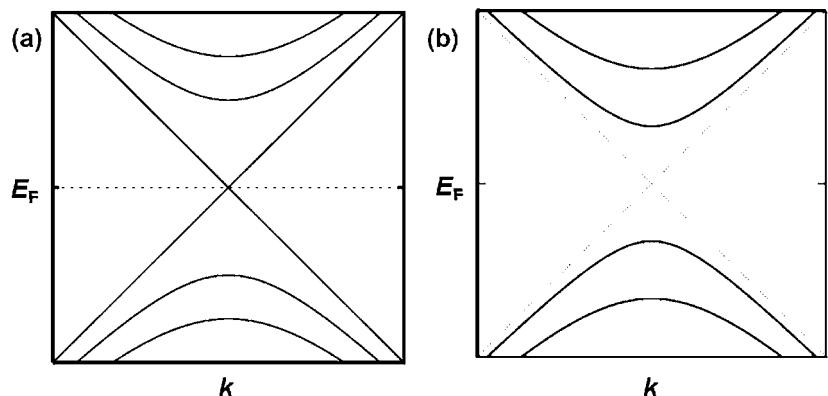
The degenerate point in the nanotube, however, does not exactly correspond to the  $K$  point due to its curvature causing  $p_z$  atomic orbitals, which point in the radial direction, not to be exactly parallel to each other. Thus, the  $\sigma$ -bond components are incorporated and thus an additional electron transfer is produced between those orbitals. Hence, metallic nanotubes classified in this way become small-bandgap semiconductors with an exception of  $(n, n)$  armchair nanotubes, which are truly metallic regardless of their curvature, as discussed below.

Figure 10.22 explains this situation for achiral nanotubes. The vertical lines in Fig. 10.22(a) show the set of allowed wave vectors of a (4, 4) armchair nanotube on the first Brillouin zone of a graphene sheet, which is surrounded by the dashed line in the figure. Although the degenerate point is shifted from the  $K$  point to the positions shown by open circles, by the enhancement of the electron transfer along the tube circumference, it is always on the  $\Gamma - K$  line, which exactly overlaps with one of the allowed wave vectors. Therefore, the nanotube becomes truly metallic. This feature is essential for any  $(n, n)$  armchair nanotubes. The set of allowed wave vectors of a (6, 0) zigzag nanotube is shown in Fig. 10.22(b). Although one of the allowed wave vectors crosses the  $K$  point, the shifted degenerate point shown by the open circles is not crossed by any of the allowed wave vectors. Therefore, this tube is not metallic, but narrow-bandgap semiconducting. For general  $(n, 0)$  zigzag nanotubes, the allowed wave vectors cross the points that divide the doubled  $\Gamma - M$  line into  $n$  parts. If  $n$  is a multiple of 3, one of the allowed wave vectors crosses the  $K$  point, which is close to the degenerate point, like in a (6, 0) nanotube and thus the  $(n, 0)$  tube will be a narrow-bandgap semiconductor. If  $n$  is not a multiple of 3, then the tube will be a moderate- or wide-bandgap semiconductor.

By using periodic boundary conditions, the condition for general  $(n, m)$  nanotubes to be either metallic or semiconducting can be easily derived. The condition is

$$n - m \begin{cases} = 0, & \text{for metallic nanotubes} \\ = 3q, & \text{for narrow-bandgap semiconducting nanotubes} \\ \neq 13q, & \text{for wide-bandgap semiconducting nanotubes,} \end{cases}$$

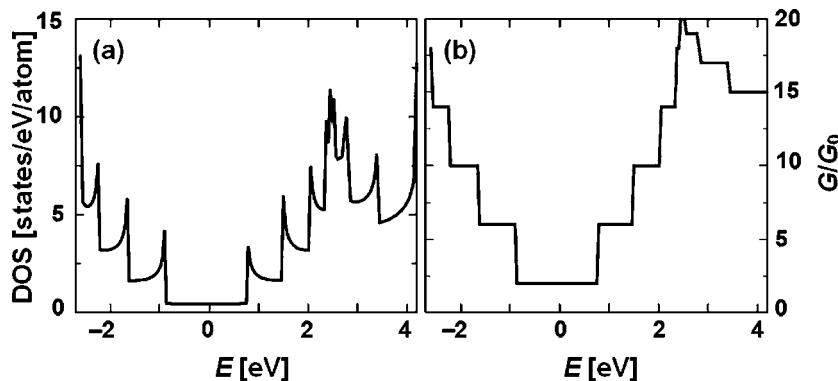
where  $q$  is any non-zero integer.



**Fig. 10.23** Band structures of (a) metallic armchair and (b) semiconducting carbon nanotubes near the Fermi level along the tube axis. These band diagrams were plotted near the  $K$  point in the Brillouin zone of graphene.

Electronic band structures of carbon nanotubes near one of the  $K$  points of graphene and near the Fermi level are shown in Fig. 10.23, where (a) is for a metallic armchair ( $n, n$ ) nanotube and (b) is for semiconducting one. The electronic band structure of single-wall (Hamada *et al.* 1992; Mintmire *et al.* 1992; Saito *et al.* 1992) and multiwall carbon nanotubes (Saito *et al.* 1993; Lambin *et al.* 1994; Kwon *et al.* 1998b), as well as single-wall nanotube ropes (Delaney *et al.* 1998; Kwon *et al.* 1998a) is now well documented. It has been shown that interwall coupling leads to the formation of pseudogaps near the Fermi level in multiwall nanotubes (Kwon *et al.* 1998b) and single-wall nanotube ropes (Delaney *et al.* 1998; Kwon *et al.* 1998a), the latter of which were later confirmed by experiments (Ouyang *et al.* 2001). Note that these different electronic structures are determined only by the geometry of the nanotubes depending on how they have been rolled from the planar graphene structure and their relative orientation with respect to neighboring hexagonal configurations. This fact implies that nanotube samples should be a mixture of metallic and semiconducting nanotubes unless they are separated in controlled ways. Experimentally synthesized samples are indeed a mixture of different chiral nanotubes and one of most serious issues in various nanotube research fields.

According to the scattering theory (Landauer 1970), the conductance is expected to be  $2G_0$ , where  $G_0 = 2e^2/h \approx (12.9\text{ k}\Omega)^{-1}$  is the quantum conductance (Tian and Datta 1994; Lin and Shung 1995; Chico *et al.* 1996). Direct evidence of the delocalization of the wavefunction along the tube axis has been already shown (Bockrath *et al.* 1997; Tans *et al.* 1997), while a direct measurement of the conductance quantization for single-wall nanotubes was recently done (Purewal *et al.* 2007). Electron transport in perfect single nanotubes is believed to be ballistic in nature, implying the absence of inelastic scattering. Conductance measurements of multiwall carbon nanotubes raised a significant controversy due to the observation of unexpected conductance values, in apparent disagreement with theoretical predictions (Frank *et al.* 1998). In these experiments, multiwall carbon nanotubes, when brought into contact with liquid mercury, exhibit not only even, but also odd multiples of the conductance quantum  $G_0$ . An even bigger surprise was the observation of



**Fig. 10.24** Electronic density of states (DOS) (a) and conductance  $G$  (b) of an isolated single-wall (10, 10) carbon nanotube. The DOS is given in arbitrary units, and  $G$  is given in units of the conductance quantum  $G_0 = 2e^2/h$  (from Sanvito, S., Kwon, Y.-K., Tománek, D. and Lambert, C.J. (2000) *Phys. Rev. Lett.* **84**, 1974 ©2000 The American Physical Society, reprinted in part with permission).

non-integer quantum conductance values, such as  $G \approx 0.5 G_0$ , since conductance is believed to be quantized in units of  $G_0$  (Landauer 1970).

The unexpected conductance behavior can arise from the interwall interaction in multiwall or in bundled nanotubes (Sanvito *et al.* 2000). This interaction may not only block some of the quantum conductance channels, but also redistribute the current non-uniformly over the individual tubes. Under the experimental conditions described in Frank *et al.* (1998), this effect may reduce the conductance of the whole system to well below the expected value of  $G_0$ .

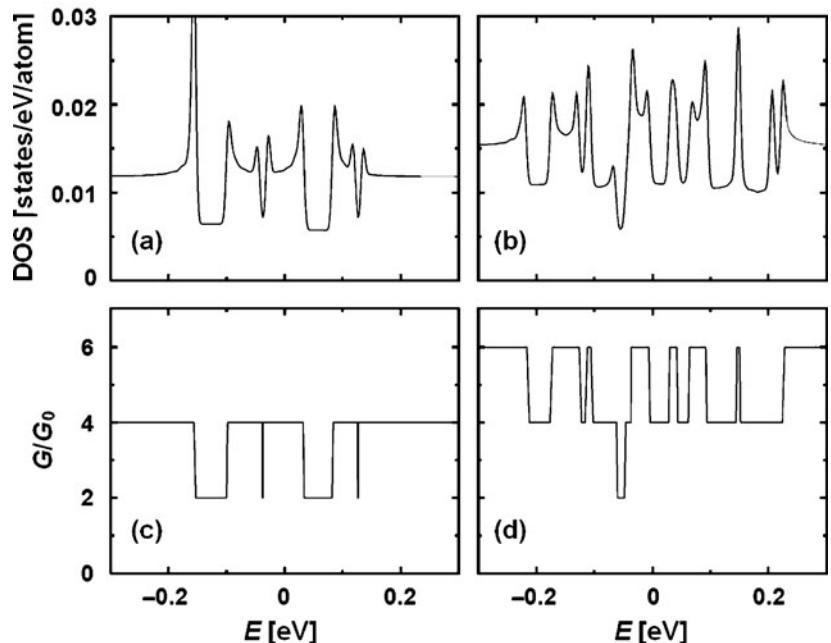
The scattering technique determines the quantum-mechanical scattering matrix  $S$  of a phase-coherent “defective” region that is connected to “ideal” external reservoirs (Sanvito *et al.* 1999). At zero temperature, the energy-dependent electrical conductance  $G(E)$  is given by the Landauer–Büttiker formula (Büttiker *et al.* 1985)

$$G(E) = \frac{2e^2}{h} T(E),$$

where  $T(E)$  is the total transmission coefficient evaluated at the energy  $E$ , which, in the case of small bias, is the Fermi energy  $E_F$ .

For a homogeneous system,  $T(E)$  assumes integer values corresponding to the total number of open scattering channels at the energy  $E$ . For individual  $(n, n)$  armchair tubes, this integer is further predicted to be even (Tian and Datta 1994; Lin and Shung 1995; Chico *et al.* 1996), with a conductance  $G = 2G_0$  near the Fermi level. As a reference to previous results, the density of states and the calculated conductance of an isolated (10, 10) nanotube is shown in Fig. 10.24.

The corresponding results for the (10, 10)@(15, 15) double-wall nanotube (Kwon *et al.* 1998b) and the (5, 5)@(10, 10)@(15, 15) triple-wall nanotube, where the interwall interaction significantly modifies the electronic states near the Fermi level, are shown in Figs. 10.25. The density of states of the double- and the triple-wall nanotubes are shown in Figs. 10.25(a) and (b), respectively. The corresponding results for the total conductance are given in Figs. 10.25(c) and (d), respectively. The conductance results suggest that some of the conduction channels have been blocked close to  $E_F$ . The interwall interaction, which is responsible for this behavior, also leads to a redistribution of the total conduction current over the individual tube walls. The partial conductances of



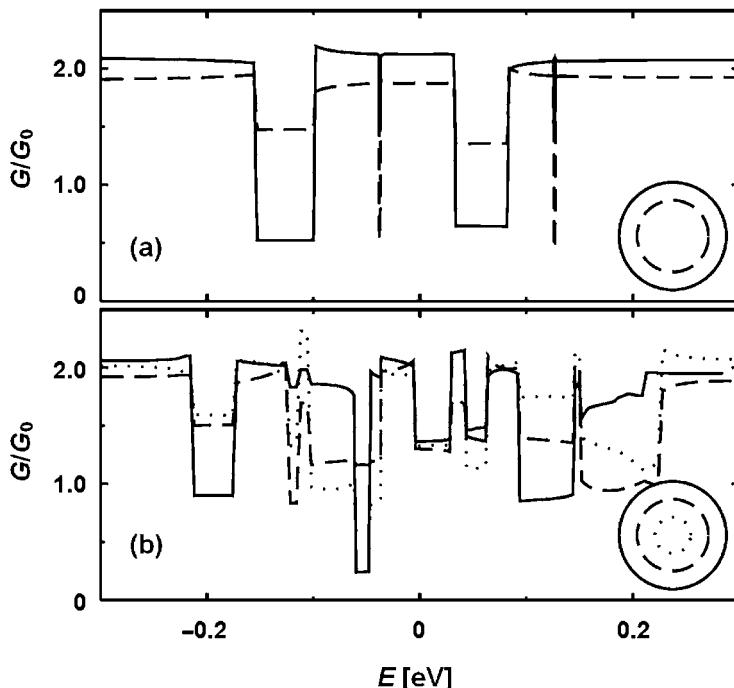
**Fig. 10.25** Electronic density of states and conductance of a double-wall  $(10, 10)@(15, 15)$  nanotube [(a) and (c), respectively] and a triple-wall  $(5, 5)@(10, 10)@(15, 15)$  nanotube [(b) and (d), respectively] (from Sambri, S., Kwon, Y.-K., Tománek, D. and Lambert, C.J. (2000) *Phys. Rev. Lett.* **84**, 1974 ©2000 The American Physical Society, reprinted in part with permission).

the tube walls are defined accordingly as projections of the total conductance and shown in Fig. 10.26. It is noticed that the partial conductance is strongly non-uniform within the pseudogaps, where the effects of intertube interactions are stronger.

The interwall interaction in multiwall nanotubes not only blocks certain conduction channels, but also redistributes the current non-uniformly across the walls. The interwall interaction will play a similar role in a bundle of single-wall nanotubes, or a rope. Such interwall interaction becomes important when new nanoscale devices are formed from nanotube networks. We will discuss the electrical characteristics of network-based devices in the next section.

#### 10.4.2 Electronic and transport properties of network devices

In this section, we will describe electronic and transport properties of hybrid nanodevices based on mixed-chirality nanotube networks, which are readily available to industry. No theoretical foundation for nanocomposite network thin-film transistors comparable to Shockley's theory of classical transistors currently exists. We have developed a methodology for modelling the electrical properties of nanotube composite thin films, and optimizing the production specifications for yielding the optimal performance, while minimizing the number of defective components. Our computational approach is to model the device as a random dispersion of nanotubes. The dispersion can be modelled with fractal diffusion taking into account the lens effect of the substrate template that imposes alignment of the nanotubes at the channel edges. The dispersion is also represented as a grid of resistors and nodes. Using

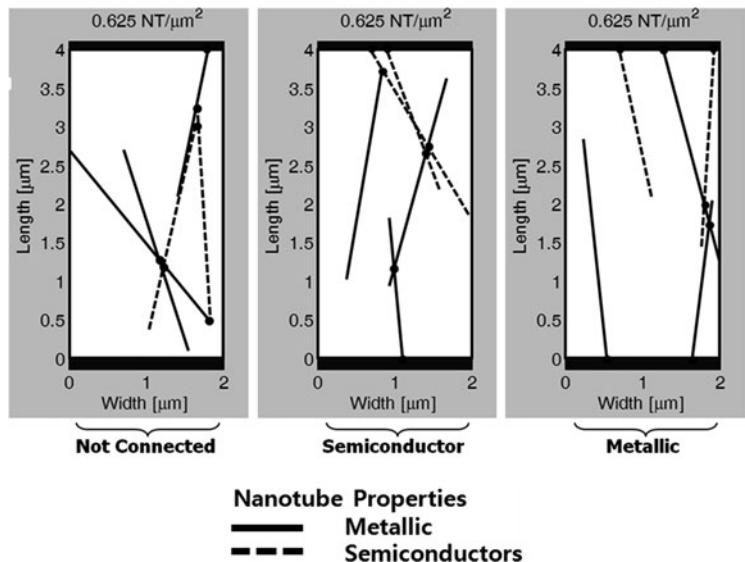


**Fig. 10.26** Partial conductance of the constituent tubes of (a) a double-wall (10, 10)@(15, 15) nanotube and (b) a triple-wall (5, 5)@(10, 10)@(15, 15) nanotube. Values for the outermost (15, 15) tube are given by the solid line, for the (10, 10) tube by the dashed line, and for the innermost (5, 5) tube by the dotted line (from Sanvito, S., Kwon, Y.-K., Tománek, D. and Lambert, C.J. (2000) *Phys. Rev. Lett.* **84**, 1974 ©2000 The American Physical Society, reprinted in part with permission).

Kirchhoff's laws and measured and modelled nanotube conductance and nanotube junction conductance, the electronic properties of individual model channels can be estimated. A Monte Carlo approach to statistically characterize outcomes shows the optimal channel dimensions and nanotube density resulting in the most semiconducting devices; and models ensemble statistics of the electrical properties of these thin films.

#### 10.4.2.1 Model description

Figure 10.27 shows a birds eye view of a mixed-chirality network. We can regard it as a system of percolating sticks. Here, the solid lines represent metallic nanotubes and the dashed lines semiconducting nanotubes. The overlay shows possible subregions that would result from such dispersion on a template. We know from fractal or percolation theories that an infinitely large network of sticks will percolate at a threshold density of  $\rho_{\text{Th}} = 4.26^2/\pi L_S^2$  or higher. This value is about 0.64 nanotubes in  $1 \mu\text{m}^2$  for a random dispersion of 3- $\mu\text{m}$  nanotubes. Channels are not infinite in extent, and local regions have three possible connectivity classes as shown in Fig. 10.27. The first possibility is that connectivity is broken, so that the device lacks a conducting path from the one electrode to the other. If all the tubes are semiconducting, high densities would solve this problem. But with mixed chirality, a high density in which the fraction of metallic tubes will result in a device with metallic properties. For the infinite extent channel, when the density of metallic CNTs becomes  $0.64 \text{ NTs}/\mu\text{m}^2$ , the device will be metallic. The second possible outcome is a channel that has a conducting path from one electrode to the other exclusively with metallic tubes. The broken connectivity and the metallic



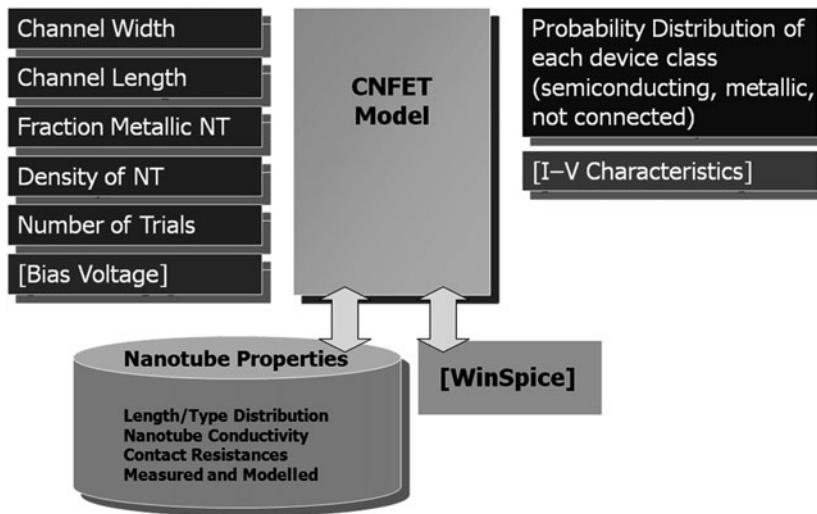
**Fig. 10.27** Classification of the different chirality of assembled SWNTs in channels.

connectivity are undesired outcomes. What industry seeks are techniques for creating semiconducting channels of nanotube networks. The third possible outcome is a channel with at least one semiconducting nanotube along each path of nanotubes connecting the two electrodes. Thus, there are three possible classes of devices: not connected, metallic, and semiconducting. Our model can be used to predict the likelihood of each class for a given set of processing parameters.

A conceptual block diagram of our model is provided in Fig. 10.28. The inputs to the model are the channel dimensions, fraction of metallic nanotubes, and nanotube density. These are the parameters used to describe a single channel or a single device. The number of trials for the Monte Carlo statistics is also an input and specifies the number of such devices that will be modelled in the simulation. As the model runs it accesses a database of nanotube properties. The model outputs the probability distribution functions for each class. A bias voltage can be specified to model  $I-V_g$  curves for individual channels.

The program is modular and the database is flexible and extensible. Currently, there are nominal properties for nanotube conductivity operating at room temperatures, in an ambient environment, and dispersed on a generic substrate. The program is designed to allow the database to grow and include nanotube properties including alternative algorithms for the dispersion statistics (the position and orientation in the channel) that simulate bundling or other dynamics; and nanotube properties that are environmentally active to address questions of using the thin film as a sensor and modelling the sensor performance or determining when a device might fail; and compare the performance of different chirality mixtures on different substrates.

The central algorithm is shown in Fig. 10.29. First, a channel is modelled by populating it with nanotubes. Internally, the orientation and classification (defining properties) of each nanotube is maintained in a list. Next, the



**Fig. 10.28** Conceptual block diagram describing our model.

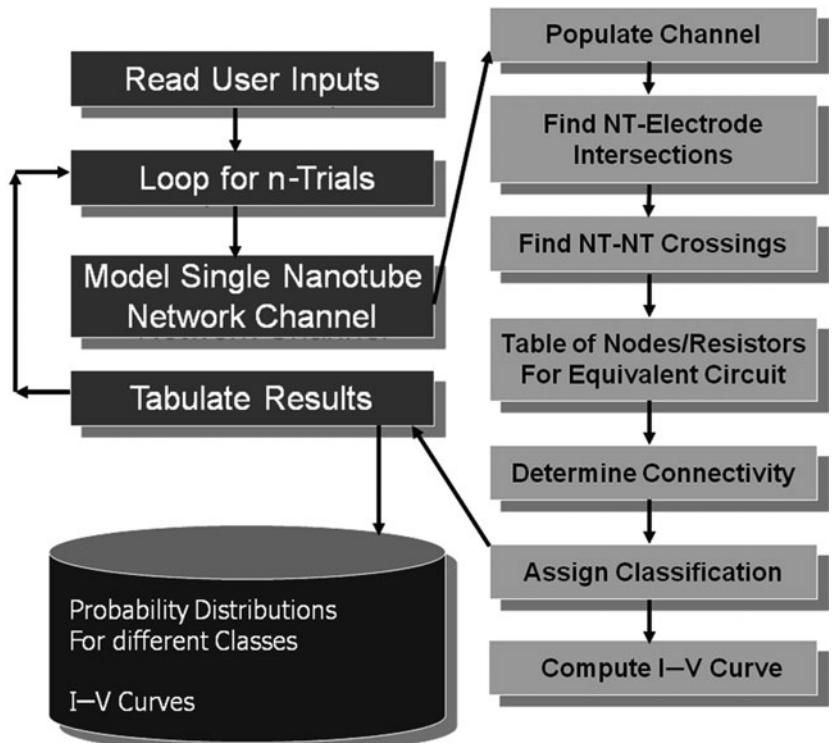
locations where nanotubes contact the electrodes and the nanotube junctions in the channel are found. These locations are used to segment the nanotubes into resistors and nodes. This equivalent circuit is analyzed to determine path connectivity and assign a classification: broken, metallic, or semiconducting, to the device channel. The electronic properties are determined by solving the equivalent circuit using WinSpice. Performing this sequence many times to acquire statistics for the probability distribution functions of each class as a function of density or other manufacturing variable.

#### 10.4.2.2 Monte carlo results

The probability distributions as functions of nanotube density are shown in Fig. 10.30. The dashed lines represent the percentage of semiconducting nanotube networks as a function of CNT densities. Manufacturers want a high probability of semiconducting nanotubes, and processing parameters that are not too sensitive to variation. For example, the statistics ideally won't change significantly as the width, length or density changes. Optimally, the high turnout of semiconducting networks will occur for a wide range of densities, not within a narrow peak.

The probability distribution of not-connected networks resembles the Fermi-Dirac distribution function starting with 100% not connecting, having a nearly linearly slope near the 50% not connecting point, and ending with 100% connecting, shown with solid lines in Fig. 10.30. In percolation theory, the percolation threshold characterizes the criteria at which a system will percolate. But this applies generally to systems that are infinite in extent. It is inconvenient to use a percolation threshold to characterize systems of percolating sticks that are directional and finite in extent. A more convenient metric for characterizing connectivity is the reference point at which 50% of the channels percolate. When the channel width is held constant the  $\rho_0$  point occurs at higher densities, as the channel length is increased.

When the channel length is held constant, the  $\rho_0$  point occurs at lower density as the width increases. This result is counterintuitive at first glance, when



**Fig. 10.29** Central algorithm of the model.

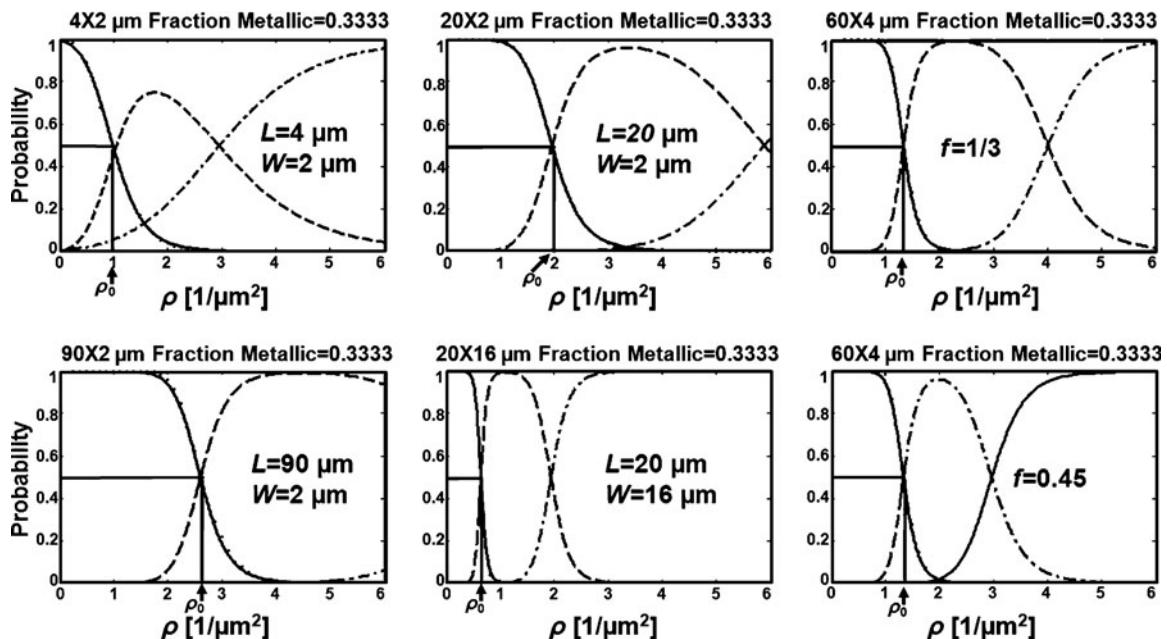
considering percolation as a function of area. But the channel connectivity imposes a directional connectivity. Consider two channels side by side. The probability for both being unconnected follows the product rule, thus 75% will have at least one of the two channels connecting at the single channel  $\rho_0$  density point. Thus, doubling the channel width substantially increases the number of connecting devices. The final set in Fig. 10.30 illustrates the sensitivity to small changes in chirality mix. Even a small change in the fraction of metallic content can result in significantly more or fewer metallic devices.

#### 10.4.2.3 Nanotube network electronic properties

Our database currently focuses on four resistances for modelling the equivalent circuit. First, measurements of metallic SWNT are length dependent and about  $4 \text{ k}\Omega/\mu\text{m}$  (McEuen and Park 2004). The conductivity of the semiconducting nanotubes is reported to be length dependent and dependent on the gate voltage (Burke 2002; McEuen and Park 2004; Llani *et al.* 2006). Nanotubes have been found to have conductances that vary inversely as length and are proportional to applied voltage (McEuen and Park 2004):

$$G = \mu C_g (V_g - V_{g0}) / L,$$

where  $C_g$  is the capacitance per unit length of the tube,  $V_{g0}$  is the threshold voltage,  $\mu$  is the mobility, and  $L$  is the tube length. Here, an array of

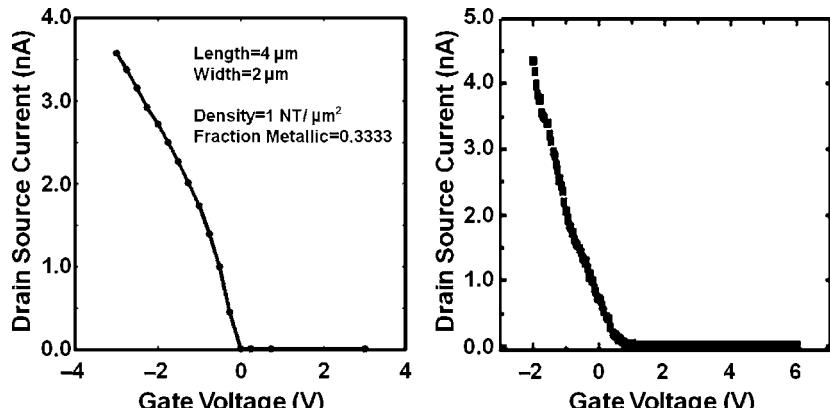


**Fig. 10.30** Probability distributions as functions of nanotube density with selected manufacturing parameters. The solid lines represent “not connected” devices; the dashed lines show the probability of semiconducting devices at a given density, whereas the dash-dotted lines do this for metallic devices.

values is generated to create an  $I$ – $V$  curve, though the factor of capacitance and mobility can vary significantly. Nanotube mobilities have been measured from 1000–10 000  $\text{cm}^2/\text{V s}$  for CVD tubes, and recently a value as high as 200 000  $\text{cm}^2/\text{V s}$  has been reported for graphene (Bolotin *et al.* 2008). The electrical measurements were done on various junctions of crossed nanotubes (Fuhrer *et al.* 2000) and modelled computationally (Yoon *et al.* 2001). We used a nominal value of 40  $\text{k}\Omega$  but these are highly sensitive to pressure, so that the network might be useful as an electromechanical sensor. Finally, we set the contact resistance between a nanotube and an electrode to be 50  $\text{k}\Omega$ .

The shape of the resulting  $I$ – $V$  curves resembles those found experimentally as shown in Fig. 10.31. We find similar orders of magnitude difference for semiconducting and metallic networks, but the capacitance/mobility factor could change the outcome significantly.

We have shown that nanotube networks exhibit excellent semiconducting properties; and it is possible to produce hybrid nanodevices from mixed-chirality nanotubes. We have developed a model for estimating the probability of successful and defective semiconductors resulting from a set of manufacturing parameters. For example, over 90% of devices with  $40 \times 2 \mu\text{m}^2$  channel dimension, of 1/3 metallic content, and  $\rho = 4 \text{ NTs}/\mu\text{m}^2$  will have semiconducting properties.



**Fig. 10.31**  $I-V_g$  characteristic calculated (left) and experimental measured one (right).

## 10.5 Conclusion

The self-assembly process has been a key strategy in building biological systems for billions of years. This process is now being used to assemble inorganic nanostructures for large-scale functional devices. The self-assembly strategy has been used for various device fabrication methods such as pen-writing, printing, and directed assembly. The advanced fabrication strategy allows us to pattern quite versatile materials such as organic molecules, nanoparticles and nanowires, and it enabled the fabrication of functional devices such as carbon-nanotube network-based transistors. Considering that most new nanostructures are prepared in solution or powder form, the self-assembly strategy is expected to play a key role in nanomanufacturing of hybrid devices based on nanomaterials in the future.

## Acknowledgments

S.H. acknowledges the financial support from NRL (R0A-2004-000-10438-0), NSI-NCRC, and TND programs from KOSEF. Y.-K.K. acknowledges the support from NSF (NSF-0425806 and CMMI-0708541) and NERSC (DOE DE-AC02-05CH1 1231).

## References

- Adamson, A.W., Gast, A.P. *Physical Chemistry of Surfaces*, 6th edn (John Wiley & Sons Inc., New York, 1997).
- Ahn, J., Kim, H., Lee, K.J., Jeon, S., Kang, S.J., Sun, Y., Nuzzo, R.G., Rogers, J.A. *Science* **314**, 1754 (2006).
- Baca, A.J., Meitl, M.A., Ko, H.C., Mack, S., Kim, H., Dong, J., Ferreira, P.M., Rogers, J.A. *Adv. Funct. Mater.* **17**, 3051 (2007).
- Bernard, A., Renault, J.P., Micheal, B., Bosshard, H.R., Delamarche, E. *Adv. Mater.* **12**, 1067 (2000).

- Bockrath, M., Cobden, D.H., McEuen, P.L., Chopra, N.G., Zettl, A., Thess, A., Smalley, R.E. *Science* **275**, 1922 (1997).
- Bodas, D., Malek, C.K. *Sens. Actuators B* **123**, 368 (2007).
- Bolotin, K.I., Sikes, K.J., Jiang, Z., Klima, M., Gudenberg, G., Hone, J., Kim, P., Stormer, H.L. *Solid State Commun.* **146**, 351 (2008).
- Busnaina, A. *Nanomanufacturing Handbook* (CRC Press, Boston, 2007) p. 34–37.
- Burke, P.J. *Proc. IEEE-NANO*, **2002**, 393 (2002).
- Büttiker, M., Imry, Y., Landauer, R., Pinhas, S. *Phys. Rev. B* **31**, 6207 (1985).
- Chaudhury, M.K., Weaver, T. *J. Appl. Phys.* **80**, 30 (1996).
- Cherniavskaya, O., Adzic, A., Knutson, C., Gross, B.J., Zang, L., Liu, R., Adams, D.M. *Langmuir* **18**, 7029 (2002).
- Chico, L., Benedict, L.X., Louie, S.G., Cohen, M.L. *Phys. Rev. B* **54**, 2600 (1996).
- Cho, N., Ryu, S., Kim, B., Schatz, G., Hong, S. *J. Chem. Phys.* **124**, 024714 (2006).
- Choi, J., Kim, D., Yoo, P.J., Lee, H.H. *Adv. Mater.* **17**, 166 (2005).
- Crank, J. *The Mathematics of Diffusion* (Clarendon, Oxford, 1993).
- Csues, G., Michel, R., Lussi, J.W., Textor, M., Danuser, G. *Biomaterials* **24**, 1713 (2003).
- Dash, J.G. *Contemp. Phys.* **30**, 2 (1989).
- Delamarche, E., Donzel, C., Kamounah, F.S., Wolf, H., Geissler, M., Stutz, R., Winkel, P.S., Michel, B., Mathieu, H.J., Schaumburg, K. *Langmuir* **19**, 8749 (2003).
- Delaney, P., Choi, H.J., Ihm, J., Louie, S.G., Cohen, M.L. *Nature* **391**, 466 (1998).
- Demers, L.D., Park, S.-J., Taton, T.A., Li, Z., Mirkin, C.A. *Angew. Chem. Int. Ed.* **40**, 3071 (2001).
- Donzel, C., Geissler, M., Bernard, A., Wolf, H., Michel, B., Hilborn, J., Delamarche, E. *Adv. Mater.* **13**, 1164 (2001).
- Efimenko, K., Wallace, W.E., Genzer, J. *J. Colloid Interf. Sci.* **254**, 306 (2002).
- Fan, Z., Ho, J.C., Jacobson, Z.A., Yerushalmi, R., Alley, R.L., Razavi, H., Javey, A. *Nano Lett.* **8**, 20 (2008).
- Feng, X., Meitl, M.A., Bowen, A.M., Huang, Y., Nuzzo, R.G., Rogers, J.A. *Langmuir* **23**, 12555 (2007).
- Frank, S., Poncharal, P., Wang, Z.L., de Heer, W.A. *Science* **280**, 1744 (1998).
- Fu, L., Dravid, V.P., Johnson, D.L. *Appl. Surf. Sci.* **181**, 173 (2001).
- Führer, M.S., Nygard, J., Shih, L., Forero, M., Yoon, Y.-G., Mazzoni, M.S.C., Choi, H.J., Ihm, J., Louie, S.G., Zettl, A., McEuen, P.L. *Science* **288**, 494 (2000).
- Furuberg, L., Feder, J., Aharony, A., Jossang, T. *Phys. Rev. Lett.* **61**, 2117 (1988).
- Ginger, D.S., Zhang, H., Mirkin, C.A. *Angew. Chem. Int. Ed.* **43**, 30 (2004).
- Hamada, N., Sawada, S., Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579 (1992).
- Hong, S., Mirkin, C.A. *Science* **288**, 1808 (2000).
- Hong, S., Myung, S. *Nature Nanotech.* **2**, 207 (2007).
- Hong, S., Kim, T.H., Lee, J., Byun, K., Koh, J., Kim, T., Myung, S. *NANO* **2**, 333 (2008).
- Huang, Y., Duan, X., Wei, Q., Lieber, C.M. *Science* **291**, 630 (2001).
- Hu, L., Gruner, G. *J. Appl. Phys.* **101**, 016102 (2007).

- Hur, S., Khang, D. *Appl. Phys. Lett.* **85**, 5732 (2004).
- Hur, S., Kocabas, C., Gaur, A., Park, O.Ok., Shim, M., Rogers, J.A. *J. Appl. Phys.* **98**, 114302 (2005).
- Iijima, S. *Nature* **354**, 56 (1991).
- Im, J., Huang, L., Kang, J., Lee, M., Rao, S.G., Lee, N.-K., Hong, S. *Nanotechnology* **124**, 224707 (2006a).
- Im, J., Kang, J., Lee, M., Kim, B., Hong, S. *J. Phys. Chem. B* **110**, 12839 (2006b).
- Im, J., Lee, M., Myung, S., Huang, L., Rao, S.G., Lee, D.J., Koh, J., Hong, S. *Nanotechnology* **17**, 3569 (2006c).
- Ivanisevic, A., Mirkin, C.A. *J. Am. Chem. Soc.* **123**, 7887 (2001).
- Jaeger, R.C. *Introduction to Microelectronic Fabrication*, 2nd edn., vol. V, Chap. 2 (Prentice Hall, New Jersey, 2002).
- Jang, J., Hong, S., Schatz, G.C., Ratner, M.A. *J. Chem. Phys.* **115**, 2721 (2001).
- Javey, A., Nam, S., Friedman, R.S., Yan, H., Lieber, C.M. *Nano Lett.* **7**, 773 (2007).
- Ju, S., Facchetti, A., Xuan, Y., Liu, J., Ishikawa, F., Ye, P., Zhou, C., Marks, T.J., Janes, D.B. *Nature Nanotech.* **2**, 378 (2007).
- Kang, J., Myung, S., Kim, B., Oh, D., Kim, G.T., Hong, S. *Nanotechnology* **19**, 095303 (2008).
- Kang, S.J., Kocabas, C., Kim, H., Cao, Q., Meitl, M.A., Khang, D., Rogers, J.A. *Nano Lett.* **7**, 3343 (2007a).
- Kang, S.J., Kocabas, C., Ozel, T., Shim, M., Pimparkar, N., Alam, M.A., Rotkin, S.V., Rogers, J.A. *Nature Nanotech.* **2**, 230 (2007b).
- Karpovich, D.S., Blanchard, G.J. *Langmuir* **10**, 3315 (1994).
- Kim, F., Kwan, S., Akana, J., Yang, P. *J. Am. Chem. Soc.* **123**, 4360 (2001).
- Kim, Y., Park, S.J., Koo, J.P., Oh, D.J., Kim, G.T., Hong, S., Ha, J.S. *Nanotechnology* **17**, 1375 (2006).
- Kim, Y., Park, S.J., Koo, J.P., Kim, G.T., Hong, S., Ha, J.S. *Nanotechnology* **18**, 015304 (2007).
- Kraus, T., Malaquin, L., Schmid, H., Riess, W., Spencer, N.D., Wolf, H. *Nature Nanotech.* **2**, 570 (2007).
- Krupke, R., Hennrich, F., von Lohneysen, H., Kappes, M.M. *Science* **301**, 344 (2003).
- Kumar, A., Whitesides, G.M. *Appl. Phys. Lett.* **63**, 2002 (1993).
- Kumar, A., Biebuyck, H.A., Whitesides, G.M. *Langmuir* **10**, 1498 (1994a).
- Kumar, A., Whitesides, G.M. *Science* **263**, 60 (1994b).
- Kwon, Y.-K., Saito, S., Tománek, D. *Phys. Rev. B* **58**, R16001 (1998a).
- Kwon, Y.-K., Tománek, D. *Phys. Rev. B* **58**, R13314 (1998b).
- Lambin, Ph., Philippe, L., Charlier, J.C., Michenaud, J.P. *Comput. Mater. Sci.* **2**, 350 (1994).
- Landauer, R. *Philos. Mag.* **21**, 863 (1970).
- Lee, K.J., Motala, M.J., Meitl, M.A., Childs, W.R., Menard, E., Shim, A.K., Rogers, J.A., Nuzzo, R.G. *Adv. Mater.* **17**, 2332 (2005).

- Lee, K.J., Meitl, M.A., Ahn, J., Rogers, J.A., Nuzzo, R.G. *J. Appl. Phys.* **100**, 124507 (2006a).
- Lee, M., Im, J., Lee, B.Y., Myung, S., Kang, J., Huang, L., Kwon, Y.K., Hong, S. *Nature Nanotech.* **1**, 66 (2006b).
- Lee, N.-K., Hong, S. *J. Chem. Phys.* **124**, 114711 (2006).
- Lee, T., Mitrofanov, O., Hsu, J.W.P. *Adv. Funct. Mater.* **15**, 1683 (2005).
- Li, Y., Qian, F., Xiang, J., Lieber, C.M. *Mater. Today* **9**, 18 (2006).
- Lim, J.H., Mirkin, C.A. *Adv. Mater.* **14**, 1474 (2002).
- Lin, M.F., Shung, K.W.-K. *Phys. Rev. B* **51**, 7592 (1995).
- Liu, X., Fu, L., Hong, S., Dravid, V.P., Mirkin, C.A. *Adv. Mater.* **14**, 231 (2002).
- Llani, S., Donev, L.A.K., Kindermann, M., McEuen, P.L. *Nature Phys.* **2**, 687 (2006).
- Love, J.C., Estroff, L.A., Kriebel, J.K., Nuzzo, R.G., Whitesides, G.M. *Chem. Rev.* **105**, 1103 (2005).
- Manandhar, P., Jang, J., Schatz, G.C., Ratner, M.A., Hong, S. *Phys. Rev. Lett.* **90**, 115505 (2003).
- McEuen, P.L., Park, J.-Y. *MRS Bull.* **29**, 272 (2004).
- Meitl, M.A., Zhou, Y., Gaur, A., Jeon, S., Usrey, M.L., Strano, M.S., Rogers, J.A. *Nano Lett.* **4**, 1643 (2004).
- Meitl, M.A., Zhu, Z., Kumar, V., Lee, K.J., Feng, X., Huang, Y., Adesida, I., Nuzzo, R.G., Rogers, J.A. *Nature Mater.* **5**, 33 (2006).
- Menard, E., Lee, K.J., Khang, D., Nuzzo, R.G., Rogers, J.A. *Appl. Phys. Lett.* **84**, 5398 (2004).
- Menard, E., Meitl, M.A., Sun, Y., Park, J., Shir, D.J., Nam, Y., Jeon, S., Rogers, J.A. *Chem. Rev.* **107**, 1117 (2007).
- Messer, B., Song, J.H., Yang, P. *J. Am. Chem. Soc.* **122**, 10232 (2000).
- Miller, S.M., Troian, S.M. *J. Vac. Sci. Technol. B* **20**(6), 2320 (2002).
- Mintmire, J.W., Dunlap, B.I., White, C.T. *Phys. Rev. Lett.* **68**, 631 (1992).
- Myung, S., Lee, M., Kim, G.T., Ha, J.S., Hong, S. *Adv. Mater.* **17**, 2361 (2005).
- Nuraje, N., Banerjee, I.A., MacCuspie, R.I., Yu, L., Matsui, H. *J. Am. Chem. Soc.* **126**, 8088 (2004).
- Ouyang, M., Yuan, C., Muisener, R.J., Boulares, A., Koberstein, J.T. *Chem. Mater.* **12**, 1591 (2000).
- Ouyang, M., Huang, J.-L., Cheung, C.L., Lieber, C.M. *Science* **292**, 702 (2001).
- Park, J., Meitl, M.A., Hur, S., Usrey, M.L., Strano, M.S., Kenis, P.J.A., Rogers, J.A. *Angew. Chem. Int. Ed.* **45**, 581 (2006a).
- Park, J., Shin, G., Ha, J.S. *Nanotechnology* **18**, 405301 (2007).
- Park, S.Y., Kwon, T., Lee, H.H. *Adv. Mater.* **18**, 1861 (2006b).
- Pashley, R.M. *J. Colloid. Interf. Sci.* **83**, 531 (1981).
- Peterson, E.J., Weeks, B.L., De Yoreo, J.J., Schwarz, P.V. *J. Phys. Chem. B* **108**, 15206 (2004).
- Pimparkar, N., Kocabas, C., Kang, S.J., Rogers, J., Alam, M.A. *IEEE Electron Dev. Lett.* **28**, 593 (2007).
- Piner, R., Zhu, J., Xu, F., Hong, S., Mirkin, C.A. *Science* **283**, 661 (1999).

- Purewal, M.S., Hong, B.H., Ravi, A., Chandra, B., Hone, J., Kim, P. *Phys. Rev. Lett.* **98**, 186808 (2007).
- Rao, S.G., Huang, L., Setyawan, W., Hong, S. *Nature* **425**, 36 (2003).
- Rozhok, S., Piner, R., Mirkin, C.A. *J. Phys. Chem. B* **107**, 751 (2003).
- Saito, R., Fujita, M., Dresselhaus, G., Dresselhaus, M.S. *Appl. Phys. Lett.* **60**, 2204 (1992).
- Saito, R., Dresselhaus, G., Dresselhaus, M.S. *J. Appl. Phys.* **73**, 494 (1993).
- Santhanam, V., Anders, R.P. *Nano Lett.* **4**, 41 (2004).
- Sanvito, S., Lambert, C.J., Jefferson, J.H., Bratkovsky, A.M. *Phys. Rev. B* **59**, 11936 (1999).
- Sanvito, S., Kwon, Y.-K., Tománek, D., Lambert, C.J. *Phys. Rev. Lett.* **84**, 1974 (2000).
- Schmid, H., Michel, B. *Macromolecules* **33**, 3042 (2000).
- Sheehan, P.E., Whitman, L.J. *Phys. Rev. Lett.* **88**, 156104 (2002).
- Smith, R.K., Lewis, P.A., Wiess, P.S. *Progr. Surf. Sci.* **75** (2004).
- Sun, Y., Rogers, J.A. *Nano Lett.* **4**, 1953 (2004).
- Sun, Y., Khang, D., Hua, F., Hurley, K., Nuzzo, R.G., Rogers, J.A. *Adv. Funct. Mater.* **15**, 30 (2005).
- Sun, Y., Rogers, J.A. *Adv. Mater.* **19**, 1897 (2007).
- Tan, J.L., Tien, J., Chen, C.S. *Langmuir* **18**, 519 (2002).
- Tans, S.J., Devoret, M.H., Dai, H., Thess, A., Smalley, R.E., Geerligs, L.J., Dekker, C. *Nature* **386**, 474 (1997).
- Tao, A., Kim, F., Hess, C., Goldberger, J., He, R., Sun, Y., Xia, Y., Yang, P. *Nano Lett.* **3**, 1229 (2003).
- Tian, W., Datta, S. *Phys. Rev. B* **49**, 5097 (1994).
- Trimbach, D.C., Stapert, H., Orselen, J.V., Jandt, K.D., Bastiaansen, C.W.M., Broer, D.J. *Adv. Eng. Mater.* **9**, 1123 (2007).
- Tsukruk, V.V., Ko, H., Peleshanko, S. *Phys. Rev. Lett.* **92**, 065502 (2004).
- Wang, Z., Zhang, J., Xing, R., Yuan, J., Yan, D., Han, Y. *J. Am. Chem. Soc.* **125**, 15278 (2003).
- Wang, Z. L., Lieber, C.M. *MRS Bull.* **32**, 99 (2007).
- Weeks, B.L., Noy, A., Miller, A.E., De Yoreo, J.J. *Phys. Rev. Lett.* **88**, 255505 (2002).
- Weibel, D.B., Lee, A., Mayer, M., Brady, S.F., Bruzewicz, D., Yang, J., DiLuzio, W.R., Clardy, J., Whitesides, G.M. *Langmuir* **21**, 6436 (2005).
- Whang, D., Jin, S., Wu, Y., Lieber, C.M. *Nano Lett.* **3**, 1255 (2003).
- Wilkinson, D., Willemesen, J.F. *J. Phys. A* **16**, 3365 (1983).
- Wilson, D.L., Martin, R., Hong, S., Cronin-Golomb, M., Mirkin, C.A., Kaplan, D.L. *Proc. Natl. Acad. Sci.* **98**, 13660 (2001).
- Xia, Y., Zhao, X.M., Whitesides, G.M. *Microelectron. Eng.* **32**, 255 (1996).
- Xia, Y., Whitesides, G.M. *Angew. Chem. Int. Ed.* **37**, 550 (1998).
- Yamaki, T., Asai, K. *Langmuir* **17**, 2564 (2001).
- Yip, H., Zou, J., Ma, H., Tian, Y., Tucker, N.M., Jen, A.K. *J. Am. Chem. Soc.* **128**, 13042 (2006).

Yoo, P.J., Choi, S., Kim, J.H., Suh, D., Baek, S.J., Kim, T.W., Lee, H.H. *Chem. Mater.* **16**, 5000 (2004).

Yoon, Y.-G., Mazzoni, M.S.C., Choi, H.J., Ihm, J., Louie, S.G. *Phys. Rev. Lett.* **86**, 688 (2001).

Yu, G., Cao, A., Lieber, C.M. *Nature Nanotech.* **2**, 372 (2007).

Zhou, Y., Hu, L., Grüner, G. *Appl. Phys.* **88**, 123109 (2006).

# 11

# Templated carbon nanotubes and the use of their cavities for nanomaterial synthesis

11.1 Introduction	386
11.2 Synthesis of carbon nanotubes and carbon nano-test-tubes	387
11.3 Controlled filling of magnetic materials into carbon nano-test-tubes	392
11.4 Synthesis of water-dispersible and magnetically responsive carbon nano-test-tubes	397
11.5 Carbon nanotube cavities as a reaction field of hydrothermal synthesis	403
11.6 Conclusions	412
References	413

## 11.1 Introduction

The versatility in carbon-atom hybridization gives rise to a rich diversity of structural forms of solid carbon. Most of the materials dealt with in carbon science and industry are considered to be composed of mainly large polycyclic aromatic molecules. The differences in the shape of such macromolecules and the way the molecules assemble (how they stack and how they are connected to one another) again lead to an immense variety of possibilities. If one could control the shape of the macromolecules and the state of their assemblage at the nanometer level, it would be possible to prepare carbon materials with a unique nanostructure, thereby expecting novel and useful characteristics from the structure. Such control is, however, a very difficult task, because the nanostructure of carbon materials is not uniform by nature, except for perfect single-walled carbon nanotubes, fullerene, graphite and diamond.

One of the most powerful and promising ways to allow for such precise control of a carbon nanostructure is the template carbonization method. This method consists of the carbonization of an organic compound in nanospace of a template inorganic substance and the liberation of the resulting carbon from the template. Various types of unique carbon materials have been synthesized using this method so far. For example, ultrathin graphite film has been prepared from the carbonization of organic polymer in the two-dimensional opening between the lamellae of layered clay such as montmorillonite and taeniolite (Kyotani *et al.* 1988, 1994; Sonobe *et al.* 1991). This research revealed that even a typical non-graphitizable carbon precursor like polyfurfuryl alcohol can be graphitized very well by the template method using the layered clays. This finding was beyond the bounds of the conventional common knowledge of carbon science, where it was said that the final structure of a carbon material strongly depends on the nature of the original precursor rather than

its nurture (the conditions of the carbonization process). Besides this two-dimensional approach, the template technique allows one to prepare one- and three-dimensional carbons such as carbon nanotubes and nanoporous carbons. Carbon nanotubes are nanometer-wide needle-like cylindrical tubes of concentric graphitic carbon. They have attracted much attention due to their potential applications as well as a fundamental interest in their properties.

The present chapter introduces the template synthesis of uniform carbon nanotubes. In particular, we would like to show how effectively the carbon nanotubes can be functionalized by the template carbonization technique. Moreover, we demonstrate that the cavities of such templated nanotubes can be utilized as a reaction field for obtaining one-dimensional nanomaterials.

## 11.2 Synthesis of carbon nanotubes and carbon nano-test-tubes

### 11.2.1 Carbon nanotubes

Using uniform and straight nanochannels of an anodic aluminum oxide (AAO) film as a template, carbon nanotubes can be prepared by pyrolytic carbon deposition on an AAO film (Kyotani *et al.* 1995, 1996). Briefly, an AAO film was subjected to carbon deposition from the pyrolytic decomposition of propylene at 800 °C, which resulted in a uniform pyrolytic carbon coating on the inner wall of the template nanochannels. Then, the AAO template was removed with acid or alkaline washing, and only carbon was left as an insoluble fraction. The formation process of carbon tubes using this chemical vapor deposition (CVD) technique is illustrated in Fig. 11.1.

From this process, it is easy to understand that the diameter and the thickness of the resulting nanotubes can be controlled by simply changing the inner

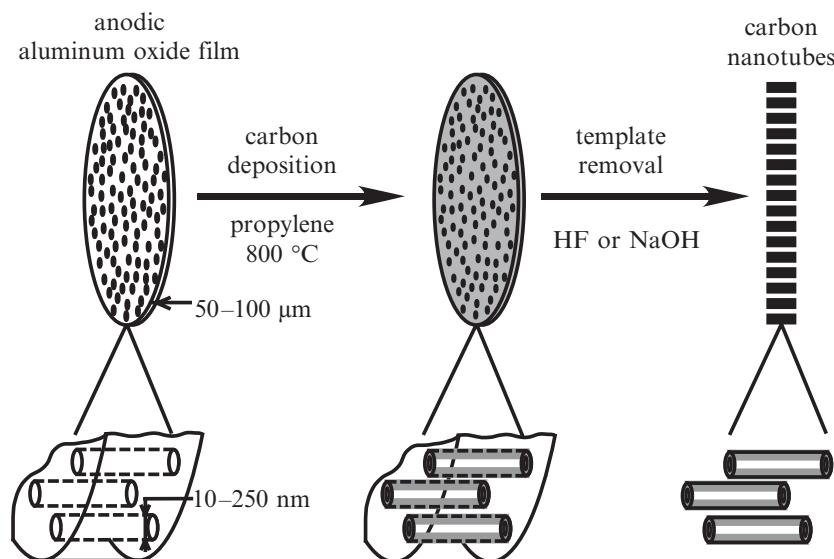
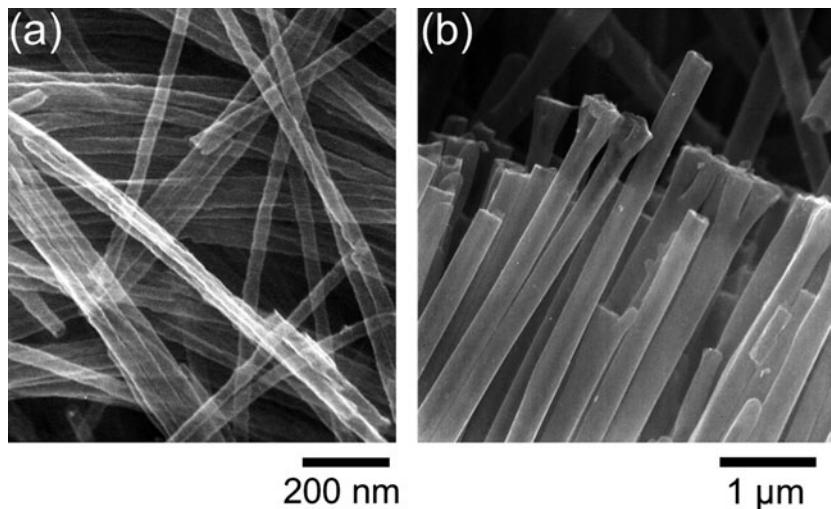


Fig. 11.1 Schematic drawing of the template synthesis of carbon nanotubes.



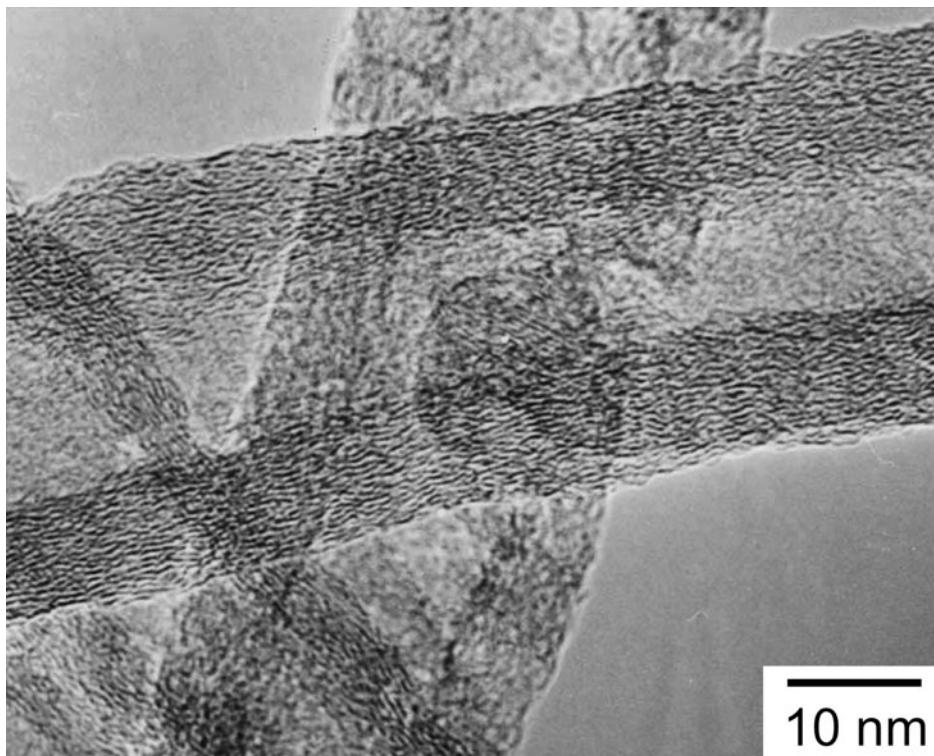
**Fig. 11.2** SEM photographs of the carbon tubes prepared by carbon deposition of propylene: (a) a carbon deposition period of 1 h on an anodic aluminum oxide film with 30-nm channels; (b) a period of 12 h on the film with 230-nm channels.

diameter and the length of the nanochannels of an AAO film, respectively. Moreover, this process does not produce any other forms of carbons only the uniform carbon nanotubes.

Figure 11.2 shows scanning electron microscope (SEM) photographs of the carbon nanotubes prepared using two types of AAO films with different channel diameters (30 and 230 nm). These photographs clearly indicate that in both cases the samples consist of only cylindrical tubes and their outer diameter is the same as the channel diameter of the corresponding AAO film. Furthermore, the length of each type of carbon nanotubes was exactly the same as the thickness of the corresponding AAO film (not shown here). Carbon nanotubes with such uniform diameters and lengths cannot be synthesized by conventional arc-evaporation and catalyst CVD techniques, which generally produce tubes of different sizes together with many types of impurities including metal particles. A high-resolution transmission electron microscope (HRTEM) image (Fig. 11.3) for the carbon tubes thus prepared reveals that the size of most graphene layers in carbon walls is less than 10 nm and they wrinkle to a great extent. This structure is far from graphite, but all the layers are orientated toward the direction of the carbon tube axis and these can easily be graphitized by further thermal treatment at as high a temperature as more than 2400 °C (Kyotani *et al.* 1996; Delpeux-Ouldriane *et al.* 2006). Another important feature of this template-synthesized carbon tube is that the tubes are not capped at both ends, unlike conventional carbon nanotubes synthesized by the arc-discharge and the catalytic CVD methods. In conclusion, the template technique allows one to prepare multiwalled carbon tubes of uniform diameter and length without any metal catalyst.

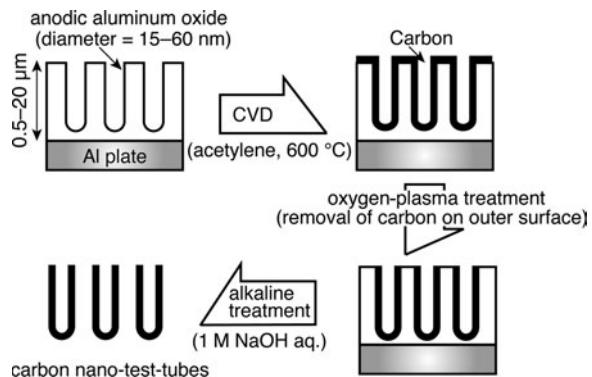
### 11.2.2 Carbon nano-test-tubes

As described before, carbon nanotubes hold great potential for a variety of industrial applications, but the extremely poor solubility of carbon nanotubes



**Fig. 11.3** High-resolution TEM image of the carbon nanotubes from an anodic aluminum oxide film with 30-nm channels under a carbon deposition period of 6 h.

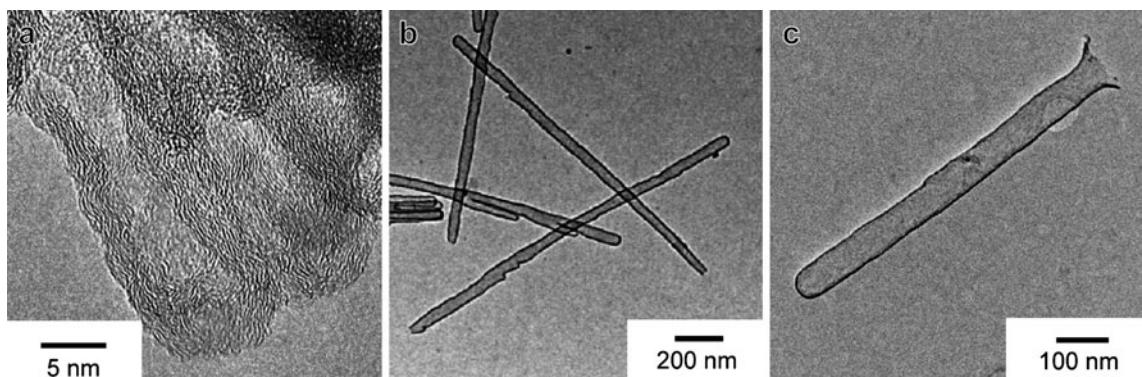
in solvents hampers their practical use in several applications. To illustrate, the application of nanotubes in the field of biotechnology, which has recently started to emerge with great hopes, is based on the premise that nanotubes are dispersible in water (Sun *et al.* 2002; Pantarotto *et al.* 2003). The most typical way for the solubilization is to produce carboxyl groups on the surface of nanotubes by strong acid treatment and then functionalize the nanotubes with large molecules through the resulting carboxyl groups (Chen *et al.* 2001; Kam *et al.* 2004). Alternatively, some solubilizing agents such as polymers (O'Connell *et al.* 2001) and DNA (Zheng *et al.* 2003) have been utilized as additives. In any case, an additional treatment is always necessary for the solubilization of nanotubes after they are synthesized. As already explained, the template method using straight nanochannels of an AAO film is a useful technique to obtain uniform carbon nanotubes of a desired size (in both diameter and length). Furthermore, using this method we can obtain nanotubes with both ends open, or with one end open but the other end closed. The latter type of carbon can be termed a carbon “nano-test-tube” (CNTT). Therefore, the CNTT synthesized by this method, if they were water dispersible, would be best suited for applications of biotechnology, especially as a capsule for a drug-delivery system. In this section, we introduce the template synthesis of water-dispersible and uniform CNTTs (Orikasa *et al.* 2006).



**Fig. 11.4** Synthesis process of carbon "nano-test-tubes" by the template method.

Figure 11.4 shows an overview of the synthesis process of CNTTs. One side of an aluminum plate was anodically oxidized until the depth of straight nanochannels reached a desired length, and pyrolytic carbon was then deposited by the CVD method at 600 °C using acetylene gas. Note that the CVD was conducted over the AAO film supported on the Al plate, and this is different from the case of Fig. 11.1, where the CVD was done over the whole AAO film (without any Al plate). The CVD temperature should therefore be less than the melting temperature of Al metal (660 °C) and this CVD temperature (600 °C) is lower than the one (800 °C) in Fig. 11.1. The carbon-coated AAO film was subjected to oxygen-plasma treatment to remove only the carbon layer deposited on the outer surface, followed by the template dissolution in alkali solution. The liberated tubes in the alkaline suspension were thoroughly washed with copious amounts of deionized water. As illustrated in Fig. 11.4, each nanochannel of the template is always closed at their bottom. Thus, it is easily understood that one end of each tube is always open, while the other end is closed.

The microscopic features of the tubes were examined with a transmission electron microscope (TEM), and its images of the tubes prepared using AAO films with different channel sizes (15, 35 and 60 nm) are shown in Fig. 11.5.



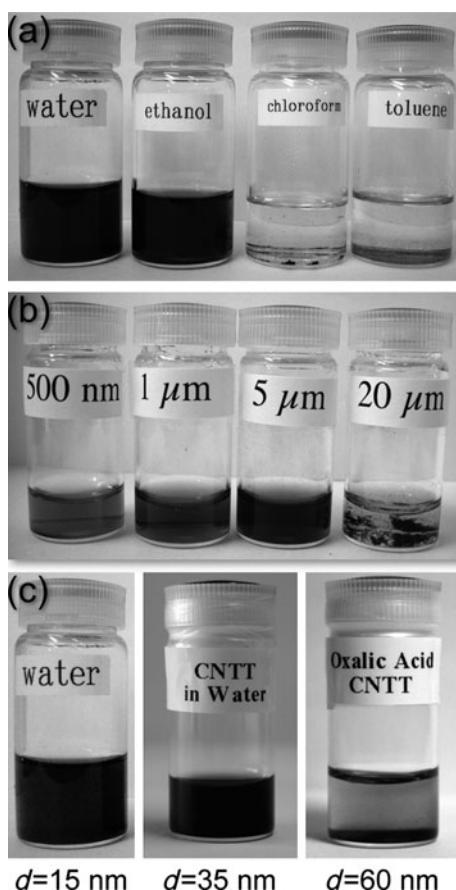
**Fig. 11.5** TEM images of uniform carbon "nano-test-tubes" having various diameters. The tube diameters in the three images are (a) 15 nm (some tubes are overlapped), (b) 35 nm and (c) 60 nm, respectively.

It is evident from these images that the bottoms of all tubes are closed and their shape looks like a test tube.

In Fig. 11.5(a), many short and wrinkled lines that correspond to graphene layers are observed in the walls, indicating the ill-crystallized nature of the present tubes. The crystallinity looks much lower than that of the carbon nanotubes prepared from the 800 °C CVD (Fig. 11.3). This is probably due to the lower CVD temperature (600 °C) in the present case.

To examine the solubility of the present CNTTs in various solvents, about 0.1 mg of the test tubes having a length of 1  $\mu\text{m}$  and a diameter of 15 nm were mixed with 7 ml of each solvent. Four different solvents (water, ethanol, chloroform, and toluene) were used and the resulting suspensions were ultrasonicated for several tens of seconds. Figure 11.6(a) shows the dispersion state of the test tubes in the various solvents after two weeks. Surprisingly, the color of both the water and ethanol suspensions was dense black and this state was kept for several months after the suspensions were prepared. On the other hand, in both toluene and chloroform, the CNTTs were agglomerated and precipitated within several minutes.

This finding clearly indicates the hydrophilic nature of the present tubes, even though carbon is hydrophobic by nature. Since the length of CNTTs



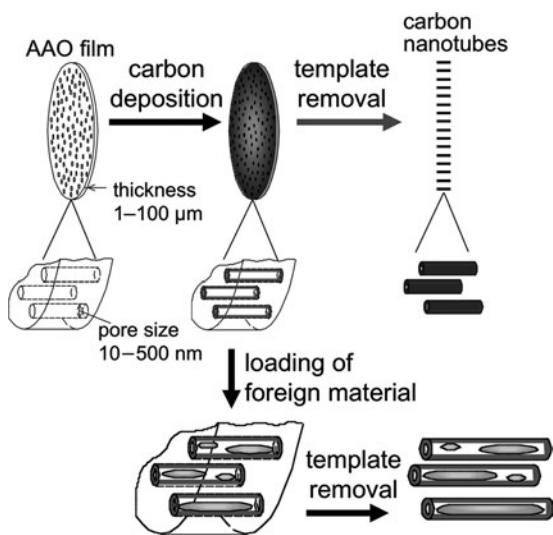
**Fig. 11.6** Dispersion state of the carbon “nano-test-tubes” in solvents. (a) Mixtures of the 1- $\mu\text{m}$  long tubes with four different solvents (water, ethanol, chloroform, and toluene) and (b) water mixtures of the tubes with different lengths of 0.5, 1, 5, and 20  $\mu\text{m}$ . (c) Water dispersions of CNTTs with different diameters (15 nm, 35 nm, and 60 nm).

is fully controllable with the template method, we were able to examine the effect of the tube length on their solubility. CNTTs with different lengths of 0.5, 1, 5, and 20  $\mu\text{m}$  (but the same diameter of 15 nm) were prepared using AAO films of the corresponding lengths, respectively, and then CNTTs of each length (0.07 mg) were mixed with 2 ml of deionized water. The dispersion state after 3 days is shown in Fig. 11.6(b), where CNTTs with lengths of 0.5, 1, and 5  $\mu\text{m}$  were dispersed in water, but the longest ones (20  $\mu\text{m}$ ) were not dispersed (they were precipitated in just 3 days). Moreover, the use of CNTTs with the different diameters (15, 35 and 60 nm) reveals the effect of tube diameter on water dispersibility. The two types of the thicker CNTTs (about 0.3 mg) was, respectively, mixed with 3 ml of deionized water and their dispersion state is shown in Fig. 11.6(c), together with the result of CNTTs with a diameter of 15 nm. The lengths of all these tubes are much less than 20  $\mu\text{m}$ , so there is probably no effect of tube length. This figure clearly indicates good water dispersibility for the 15- and 35-nm tubes, but the thicker ones (60 nm) were almost precipitated in 3 days. At the present moment, we don't have a clear explanation why such a difference in tube size should cause the observed difference in water dispersibility, but we believe that this is the first report on the size effect on the dispersibility of carbon nanotubes in water.

It was found from  $\zeta$ -potential measurement of all types of CNTTs that they were negatively charged in water. This confirms the presence of an electric double layer (EDL) on the surface of each tube and the observed solubility in water can be ascribed to the presence of the EDL. The analyses using X-ray photoelectron and Fourier transform infrared spectrometers revealed the presence of acidic oxygen-containing surface groups on the tubes. Probably such a presence gives rise to the negative charge on the carbon surface. Taking the less-crystallized nature of the tubes into consideration, we can judge that there may be many defects such as dangling bonds on the outer surface that were completely protected by the template. When these reactive sites were exposed to the NaOH solution for the tube liberation, these sites would easily be reacted to form oxygen functional groups on the whole outer surface of CNTTs.

### 11.3 Controlled filling of magnetic materials into carbon nano-test-tubes

Since carbon nanotubes are hollow, it is therefore possible, at least in principal, to introduce "foreign materials" into the tube cavities. Thus, there have been many attempts to insert foreign materials into the tubes. Details of the methods were well reviewed by Terrones (2004). The template method also produces the filled uniform carbon nanotubes. Furthermore, the presence of the template materials makes this method fundamentally different from the other ones. The schematic of the filling process is illustrated in Fig. 11.7, where the precursor of "foreign material" was loaded into the carbon-deposited AAO film before the dissolution of alumina by alkaline treatment using concentrated NaOH aqueous solution. Since all the openings are placed on one side of

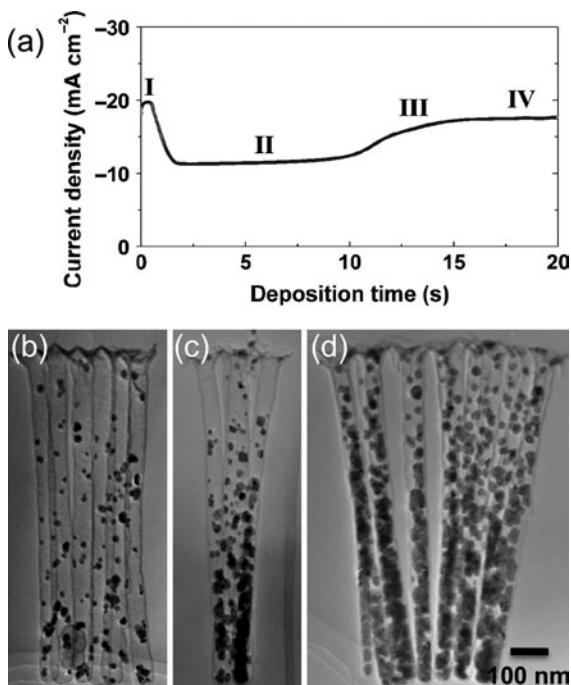


**Fig. 11.7** Schematic of the filling process into templated carbon nanotubes.

the flat AAO film (Fig. 11.7), the filling of foreign materials becomes very easy as long as it is performed at the stage of the carbon/AAO composite film. Using this encapsulation method, so far we have introduced many types of foreign materials such as Pt, Ag, NiO and  $\text{Fe}_3\text{O}_4$  in the forms of nanoparticles, nanowires and/or nanorods. It should be noted that for all cases these materials are present only inside the cavities of the tubes, which are completely free from the metals and/or the metal oxides on the outer surface (Kyotani *et al.* 1997, 1999; Pradhan *et al.* 1998, 1999; Matsui *et al.* 2001). Among foreign materials, special attention should be paid to the introduction of magnetic substances, because such introduction endows carbon nanotubes with magnetism and, as a result, the carbon nanotubes would have a variety of industrial applications such as magnetic storage media, and capsules for magnetic drug-delivery systems. In the present section, we present our attempt to introduce Ni-Fe alloy into the cavities of uniform CNTTs by an electrochemical route and demonstrate that the degree of such filling can be controlled simply by changing the electrochemical deposition conditions (Wang *et al.* 2007).

### 11.3.1 Electrochemical deposition of Ni-Fe alloy

The template synthesis of CNTTs was conducted by chemical vapor deposition at 600 °C using acetylene as a carbon source. Since the as-prepared carbon-coated AAO film is hydrophobic, a process of pre-treatment to give some hydrophilicity is indispensable, which allows aqueous plating solution to enter the tube cavities and deposition can therefore occur therein. In a typical pre-treatment, the as-prepared carbon-coated AAO film supported on the Al substrate was soaked in dilute  $\text{H}_2\text{O}_2$  solution (10 wt%), immersed in hot water and dried. After the pre-treatment, the carbon-coated AAO film became somewhat hydrophilic. Ni-Fe alloy deposition into the carbon tubes array was conducted in a home-made two-electrode plating cell in which the



**Fig. 11.8** Visualization of the filling process into CNTs during the constant-voltage ECD. (a) Current transient recorded during ECD. (b–d) TEM images of CNTs filled with Ni-Fe alloy at 24.0 V for 5 s (b); 10 s (c); and 15 s (d). The open ends of CNTTs connected by outer carbon are pointing upward.

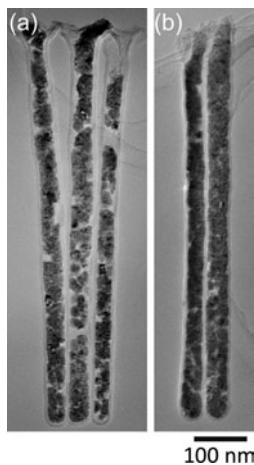
carbon-coated AAO film supported on the Al substrate and a nickel plate were used as cathode and anode, respectively. Only the carbon side of the cathode was exposed to the plating electrolyte. The constituents of the plating electrolyte were the same as in the literature (Girard 1967) but with addition of methanol (the volume ratio of methanol to the original plating electrolyte is 1:4). We found that the methanol addition efficiently suppressed hydrogen generation during the electrochemical deposition (ECD), which is very helpful for the filling of the alloy into the tube cavities at a high filling ratio. Three routes, constant-voltage ECD, voltage-stepped ECD and pulsed ECD, were studied to fill NiFe into CNTTs. After the filling, NiFe-filled carbon tubes were liberated from the AAO film in 1.0 M NaOH solution at room temperature. Note that at this stage each CNTT is connected to one another at the open end, because the external carbon layer deposited on the AAO surface still remains even upon the removal of the template. In this section, we unveil the ECD behavior of Ni-Fe alloy into the cavities of such a connected CNTT array.

At first, the filling of NiFe was carried out by the constant-voltage ECD at  $-4.0\text{ V}$ . During the ECD at this voltage, the current transient at the beginning exhibits four distinct stages (Fig. 11.8(a)). This feature is similar to that for the electrochemical formation of Ni and Co nanowires in nanochannels of an insulating polycarbonate membrane, where a thin copper film was sputter deposited on one side of the membrane, which facilitates the initial nucleation of Ni and Co only at the bottom of channels (Whitney *et al.* 1993). Whereas in the present work, the AAO film is fully covered with an electronically conductive carbon film on both the outer surface and the inner walls of channels of the AAO

film. Such a conductive carbon film can fully facilitate the initial nucleation of Ni-Fe alloy on both the outer surface and the inner walls. Thanks to CNTTs, each scene of the ECD could be kept as intact as possible, which allows us to visualize the time course of the deposition process for the first time with the assistance of TEM. After ECD for 5 s, some nanoparticles were distributed randomly on the inner wall of each tube but with some enrichment at the bottom (Fig. 11.8(b)). As ECD proceeded (10 s and 15 s), the nanoparticles at the bottom preferentially grew larger, so that the CNTTs were filled upwards from the bottom (Figs. 11.8(c) and (d)).

Chemical analyses with energy-dispersive spectroscopy demonstrate that the particles consist of Ni and Fe elements. Close examination on the nanoparticles deposited in the tubes with the assistance of selected-area electron diffraction (SAED) indicates that the lattice parameters calculated from the diffused rings correspond to Ni-Fe alloy having a face-centred cubic structure. It should be noted that the appearance of the external surface for each carbon-coated AAO film after the deposition for 5, 10 and 15 s does not show metallic luster, while it is of metallic luster for the sample by ECD for 20 s. This evidently indicates that NiFe has deposited on the outer surface of the sample by ECD for 20 s. In this way, the preferred deposition into the tube cavities has been confirmed in combination with the direct observation with TEM. Moreover, these results clearly indicate that the metal filling started from the closed end of each CNTT in spite of the conductive carbon film covering the outer surface of the AAO film. This unexpected bottom-up filling makes it possible to achieve controlled filling of materials of interest into uniform CNTTs in arrays, as demonstrated in the following paragraphs.

There are a few voids in the tubes (Fig. 11.8(d)) filled with the constant-voltage ECD for 15 s. The presence of voids is unfavorable for the application of high-density magnetic recording media (Kawai *et al.* 1976). Such voids may be attributed to the blocking of metal ion migration through the tube cavities by already-deposited NiFe particles inside the tubes. To reduce the voids, we proposed a voltage-stepped method. The key idea of the voltage-stepped method is that smaller nuclei with a larger population density are created first at a high voltage, followed by a low voltage that gives a low current density. In a practical case of voltage-stepped ECD, a high voltage of  $-4.0\text{ V}$  was applied for 5 s in the first step, followed by a low voltage at  $-1.5\text{ V}$  for 10 min. The voltage applied at  $-1.5\text{ V}$  corresponds to a current density of about  $-0.75\text{ mA/cm}^2$ . This current density is much lower than that at  $-4.0\text{ V}$  (Fig. 11.8(a)), so that the particle growth rate should be reduced significantly and the blocking effect should be alleviated as a result. Figure 11.9(a) shows a TEM image of the CNTTs filled by the voltage-stepped method. It is obvious from the TEM image that the metal filling was improved when compared with the constant-voltage ECD. We should point out that there are still some voids in the tubes filled by the voltage-stepped ECD, suggesting that the number of deposition centers is not sufficient. To further increase such deposition centers, higher overpotential or higher voltage corresponding to higher current density are needed. We then applied pulsed ECD, which was adopted by Nielsch *et al.* (2000) for the filling of Ni into the nanochannels of a bare AAO film.

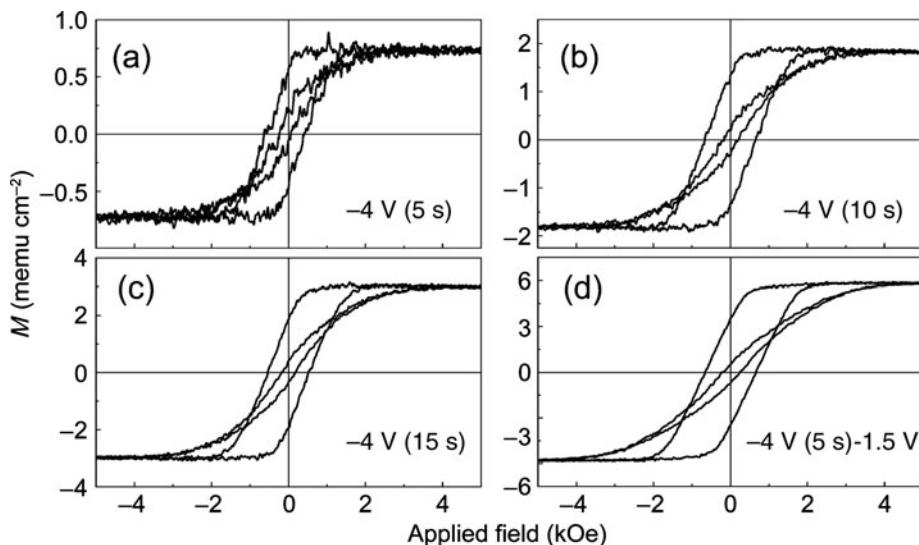


**Fig. 11.9** TEM images of NiFe-filled CNTTs by the voltage-stepped ECD. (a) The ECD was conducted at 24.0 V for 5 s, followed by 21.5 V for 10 min. (b) The filling was performed by the pulsed ECD.

In one cycle of a typical pulsed ECD, a high current density ( $-50\text{ mA/cm}^2$ ) was applied for 40 ms, followed by a low current density ( $-10\text{ mA/cm}^2$ ) for 200 ms. A relaxation period during which the current density was zero, 1 s, was inserted between each cycle in order to allow the concentration of metal ions in the tube cavities to recover. The first current density applied in the pulsed ECD is much higher than those in the constant-voltage ECD and in the voltage-stepped ECD, thus the nucleus density is also much higher. As a result, the number of deposition centers was increased. Figure 11.9(b) presents a TEM image of CNTTs filled by the pulsed ECD. It can be seen that the cavities are almost completely filled. With the pulsed ECD, the metal-filling ratio was successfully improved.

### 11.3.2 Magnetic properties of NiFe-filled CNTTs

The magnetic properties of the NiFe-filled CNTT array embedded in the AAO film were examined at room temperature. The external magnetic field was applied parallel and perpendicular to the tube axis. Figure 11.10 shows hysteresis loops corresponding to the three specimens prepared by constant-voltage ECD for 5, 10 and 15 s and that prepared by the voltage-stepped ECD. As a matter of course, the saturated magnetization increases with the amount of metal deposited into tubes, in the following order: constant-voltage ECD (5 s), (10 s), (15 s), voltage-stepped ECD. With the applied magnetic field along the tube axis, the filled tube array exhibits relatively square loops with high coercivity (520–660 Oe) and high remanence (0.60–0.80), while the loops are sheared with low coercivity in the range of 150–220 Oe and low remanence



**Fig. 11.10** Magnetic hysteresis loops recorded at room temperature. (a–c) NiFe-filled CNT array by the constant-voltage ECD at 24.0 V for 5 s; 10 s (b); 15 s (c). (d) NiFe-filled CNT array by the voltage-stepped ECD. The external magnetic field was applied parallel (large loops) and perpendicular (small loops) to the CNT axis. Note that the measurement was performed for the CNTTs embedded in the AAO template. (Wang et al. *J. Mater. Chem.* **17**, 986–991 (2007), Reproduced by permission of The Royal Society of Chemistry)

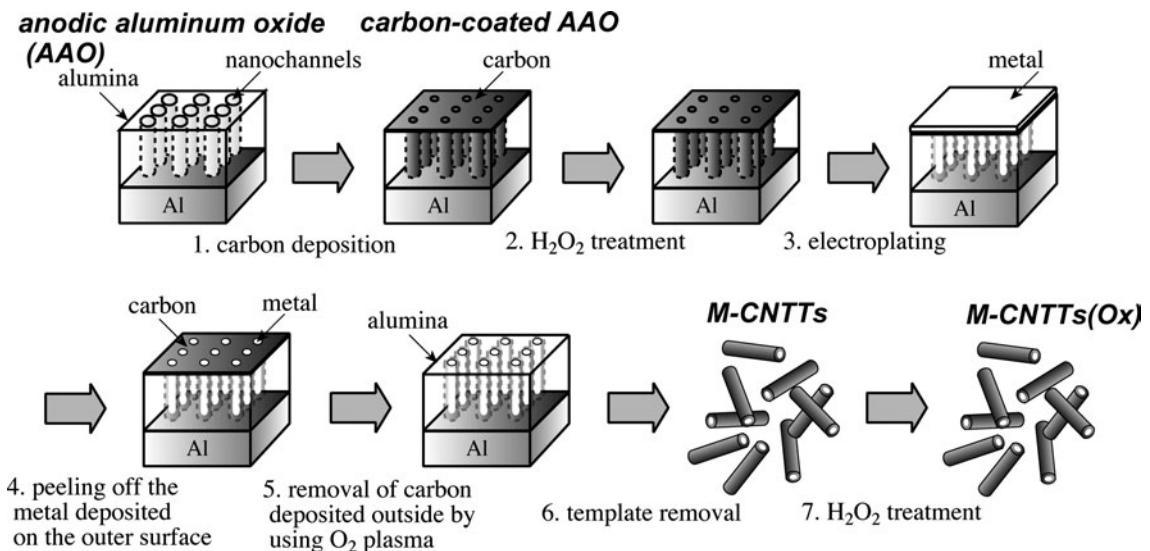
in the range of 0.10–0.19 when the field was applied perpendicular to the tube axis. This feature demonstrates that all NiFe-filled tube arrays exhibit magnetic anisotropy with an easy magnetization orientation along the tube axis. The previous extensive investigations have shown that ferromagnetic nanowires often exhibit magnetic anisotropy with a preferential magnetic orientation along the nanowires or nanorods (Kawai *et al.* 1976; AlMawlawi *et al.* 1991; Fert *et al.* 1999; Nielsch *et al.* 2001; Sellmyer *et al.* 2001). In this work, surprisingly, once NiFe nanoparticles instead of nanowires are arranged on the inner walls of the vertically aligned CNTT array (Figs. 11.8(b)–(d)), they exhibit an analogous magnetization behavior with an easy magnetization direction along the tube axis to the nanowires or nanorods. To the best of our knowledge, this is the first report on anisotropic magnetization behavior for nanoparticles arranged in vertically aligned tubes. The mean distance between two nanoparticles in each CNTT is much shorter than that in different CNTTs and the tube cavities are of high aspect ratio. As a result, the entity of the nanoparticles in each high aspect ratio CNTT can be regarded as a pseudonanowire. The magnetization behavior of such pseudonanowires in an external magnetic field follows the principle of the nanowires.

## 11.4 Synthesis of water-dispersible and magnetically responsive carbon nano-test-tubes

As introduced in Section 2, by using the nanochannels of an AAO film as a template, we have succeeded in the synthesis of water-dispersible CNTTs with tuneable size both in diameter and length. If magnetic materials are loaded into the CNTTs and the CNTTs keep their water dispersibility even after the loading, they are useful in many fields of applications. For example, when drugs or biomacromolecules are introduced into such magnetoresponsive CNTTs, they are applicable as carriers for magnetic delivery systems, where the carriers are guided to a target tissue by an external applied magnetic field. So far, many types of magnetically responsive nanocarbons have been prepared for this purpose (Correa-Duarte *et al.* 2005; Georgakilas *et al.* 2005; Korneva *et al.* 2005; Sharma *et al.* 2006; Utsumi *et al.* 2006), but none of them are completely free from serious problems such as the lack of water dispersibility and size controllability. Needless to say, both of these properties are indispensable for such biomedical applications. In this section, we introduce the synthesis of size-controlled, water-dispersible and magnetoresponsive CNTTs (Orikasa *et al.* 2008). Moreover, we try to estimate the interaction potentials that work among the metal-loaded CNTTs and understand the effect of the magnetic interaction on the water dispersibility.

### 11.4.1 Synthesis

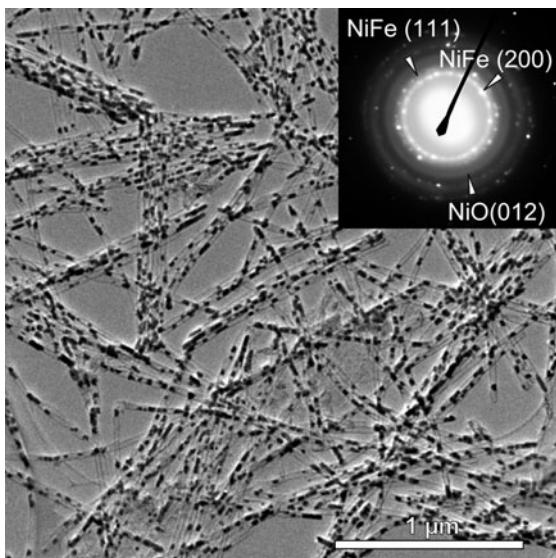
Except for one process, a synthesis process of the magnetoresponsive CNTTs (Fig. 11.11) is almost the same as that for the Ni-Fe alloy filling in the previous



**Fig. 11.11** Synthesis process of the magnetic metal loaded CNTTs using the template technique.

section. In the present process, we added an oxygen-plasma treatment for the removal of the carbon layer deposited on the external AAO layer just after the electroplating process, because, in this case, we needed individual and separated CNTTs rather than the CNTT array. Using the same plating electrolyte and cell as in the previous section, the ECD of Ni-Fe alloy was performed at a constant voltage of  $-4.0\text{ V}$  for a period as long as 30 min. As a result, the metal deposited not only inside the tube cavities but also on the top surface of the AAO substrate, but the latter thick metal layer was easily peeled off by tweezers. After the metal-loaded and individual CNTTs were liberated from the AAO template by alkaline treatment, some of them were subjected to the surface oxidation treatment with  $H_2O_2$  to improve the hydrophilicity of the CNTT outer walls.

A TEM image of the CNTTs synthesized is shown in Fig. 11.12, where CNTTs with a diameter of about 35 nm are observed and they contain many small particles inside the tube cavities. A SAED pattern from several metal-loaded CNTTs is shown in the inset of Fig. 11.12. Many bright diffraction spots and two evident rings constituted of bright spots were observed, all of which can be assigned to the diffraction from Ni-Fe alloy crystals. In addition, diffused rings that can be ascribed to NiO are present, indicating that the NiFe is partially oxidized to NiO. Since the diffused rings from NiO were observed even in the sample before the  $H_2O_2$  treatment, NiO may be formed during air exposure and/or the alkaline treatment for the AAO removal. The magnetization curve was measured for the metal-loaded CNTTs liberated from the template and it showed a clear hysteresis loop, proving the sample to be ferromagnetic. Moreover, the CNTTs hold a significantly large coercivity over 1000 Oe, although the coercivity of Ni-Fe alloy is generally less than 130 Oe despite its composition (Herzer *et al.* 1995). Such large coercivity would originate from the exchange anisotropy induced



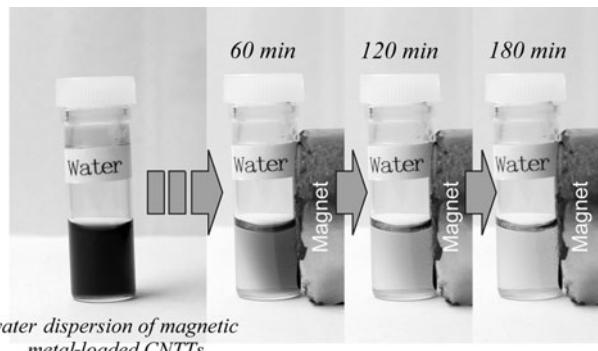
**Fig. 11.12** A TEM image and the corresponding SAED pattern of metal-loaded CNTTs.

by interfacial exchange coupling between the ferromagnetic NiFe and the antiferromagnetic NiO.

### 11.4.2 Water dispersibility

As we have already reported in Section 2, empty CNTTs were dispersible in both ethanol and water without any post-treatment. However, for the metal-loaded CNTTs, they were dispersible in ethanol but not in water. The surface oxidation by using  $H_2O_2$  was necessary to make the metal-loaded CNTTs water dispersible. The achieved dispersed state in water could be kept for at least a month. When a magnet (neodymium magnet, surface inductive flux = 4500 G) was put near a vial of the water dispersion, the CNTTs were slowly attracted by the magnet, and the color of the dispersion liquid changed over a period of hours from dense black to transparent and colorless (Fig. 11.13). After the removal of the magnet from the vial, the agglomerated CNTTs could be redispersed by mild agitation for several seconds. This observed magnetoresponsive property shows promise as carriers for the delivery of biomacromolecules.

The improvement of the water dispersibility upon the  $H_2O_2$  treatment would be attributed to an increase in the amount of acidic oxygen-containing functional groups on the carbon surface. Indeed, X-ray photoelectron spectroscopy analysis indicated that the O/C atomic ratio on the surface was increased by the  $H_2O_2$  treatment from 0.14 to 0.21. Furthermore, the  $\zeta$ -potential measured in neutral water was negatively increased by the treatment (Table 11.1). Such an increase in  $\zeta$ -potential fostered the repulsive force among the CNTTs due to their EDL and the dispersibility was improved, as a result. However, the increase in the  $\zeta$ -potential can not fully explain the change in the water dispersibility. To evaluate the water dispersibility more quantitatively, “salting-out” of the CNTTs was performed by the addition of KCl aqueous



**Fig. 11.13** Magnetoresponsive behavior of metal-loaded CNTTs dispersed in water. When a magnet was put near a vial, the CNTTs were slowly attracted by the magnet.

solution into the water suspension (0.02 mg metal-loaded CNTTs in 1 mL of water). The results of the salting-out are summarized in the third column of Table 11.1. For empty CNTTs, irrespective of the  $\text{H}_2\text{O}_2$  treatment, they are water dispersible and keep the dispersed state until the KCl concentration reaches  $0.040 \text{ mol L}^{-1}$  (treated) and  $0.015 \text{ mol L}^{-1}$  (non-treated). However, for the metal-loaded CNTTs, the non-treated ones are not water dispersible, and even for the treated ones, they precipitated when the KCl concentration was  $0.004 \text{ mol L}^{-1}$  or higher. Obviously, the metal-loaded CNTTs are less dispersible than both types of empty CNTTs, suggesting that an attractive force that works among the magnetic metal particles present inside the cavities of the CNTTs affects the water dispersibility.

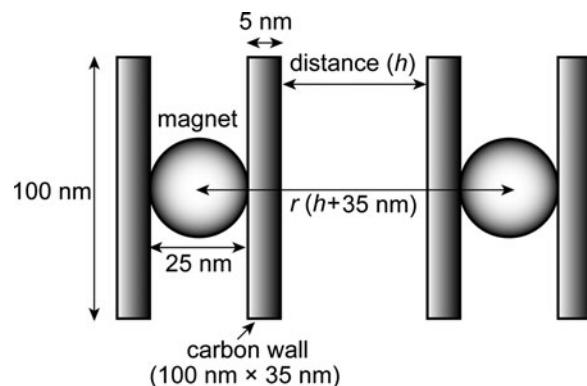
#### 11.4.3 Effect of magnetic interaction

To understand the effect of the magnetic interaction on the water dispersibility, the interaction between two metal-loaded CNTTs was roughly estimated by using a simple model (Fig. 11.14), where the carbon wall of a part of one CNTT is simplified to a flat graphite plate (100 nm in length, 35 nm in width and 5 nm in thickness) and a spherical NiFe particle with a diameter of 25 nm is placed in contact with the graphite plate. The area of the graphite plate ( $100 \text{ nm} \times 35 \text{ nm}$ ) corresponds to the projected area of one CNTT with a length of 100 nm. Since we assume in this model that the 25-nm diameter particle is present in

**Table 11.1**  $\zeta$ -potential and critical aggregation concentration of KCl for various types of CNTTs.

Sample	$\zeta$ -potential/mV	Critical aggregation concentration of KCl/mol L <sup>-1</sup>
empty CNTTs without $\text{H}_2\text{O}_2$ treatment	-29	0.015
empty CNTTs with $\text{H}_2\text{O}_2$ treatment	-39	0.040
metal-loaded CNTTs without $\text{H}_2\text{O}_2$ treatment	-25	n.a.*
metal-loaded CNTTs with $\text{H}_2\text{O}_2$ treatment	-37	0.004

\*not dispersible in water.



**Fig. 11.14** A simple model of magnetic-metal-loaded CNTTs.

the 100-nm length CNTT, the metal-filling ratio corresponds to about 25% and this value is almost the same as the actual filling ratio determined from the CNTTs in the TEM image (Fig. 11.12). The overall potential curve ( $U$ ) for this simple model can be calculated as the sum of the repulsive potential due to the presence of EDL ( $E$ ), the attractive potential from van der Waals force between the two graphite plates ( $V$ ) and the magnetic dipole–dipole interaction between the two metal particles ( $D$ ). The repulsive interaction by EDL at 25 °C is expressed by the following equations:

$$E = \frac{1.58 \times 10^8 c}{\kappa} \gamma^2 \exp(-\kappa h) \times S$$

$$\kappa = 3.3 \times 10^9 \sqrt{c}$$

$$\gamma = \tanh \left( \frac{e\phi_0}{4kT} \right),$$

where  $c$ ,  $k$ ,  $T$  (25 °C) and  $\phi_0$  are the ionic strength, Boltzmann constant, temperature and the surface potential (assumed to be a little lower than the measured  $\zeta$ -potential: -50 mV), respectively. The van der Waals potential ( $V$ ) between the two facing flat carbon plates was estimated from the following equation.

$$V = -\frac{A}{12\pi h^2} \times S,$$

where  $A$  is the Hamaker constant (assumed to be  $1 \times 10^{-19}$  J) (Dagastine *et al.* 2002; Li *et al.* 2005),  $h$  is a distance between the two plates and  $S$  is an area of the carbon plates (35 nm × 100 nm in the model). Since the carbon plates (the wall of CNTTs) have a thickness of 5 nm, “ $r$ ” is always larger than “ $h$ ” by 35 nm (Fig. 11.14). The potential of the dipole–dipole interaction ( $D_{ij}$ ) between two magnetic moments ( $m_i$  and  $m_j$ ) with a distance of “ $r$ ” is expressed as follows:

$$D_{ij} = \frac{1}{r^3} \left\{ m_i \cdot m_j - \frac{3}{r^2} (m_i \cdot r_{ij})(m_j \cdot r_{ij}) \right\},$$

where  $r_{ij}$  is a vector parallel to the line joining the centers of the two dipoles. The  $D_{ij}$  become a minimum when  $m_i$  and  $m_j$  are parallel. Since the

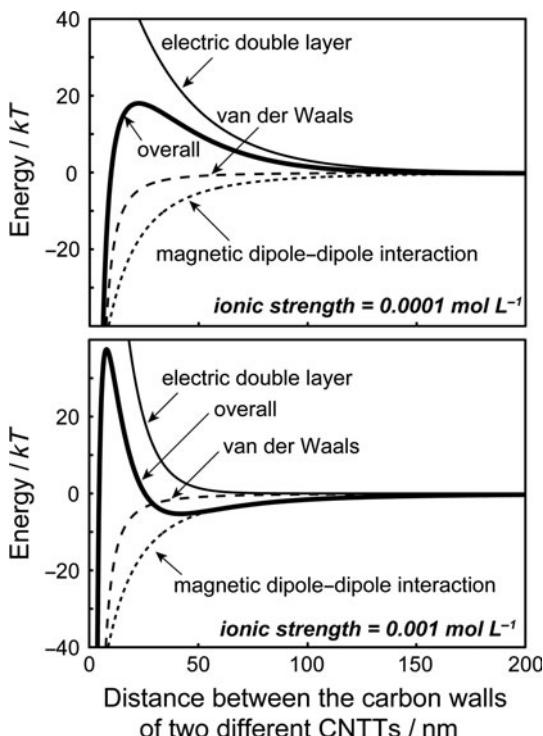
magnitudes of  $m_i$  and  $m_j$  are identical for the model, the minimum of  $D_{ij}$  ( $D$ ) is defined as follows:

$$D = -2m^2/r^3$$

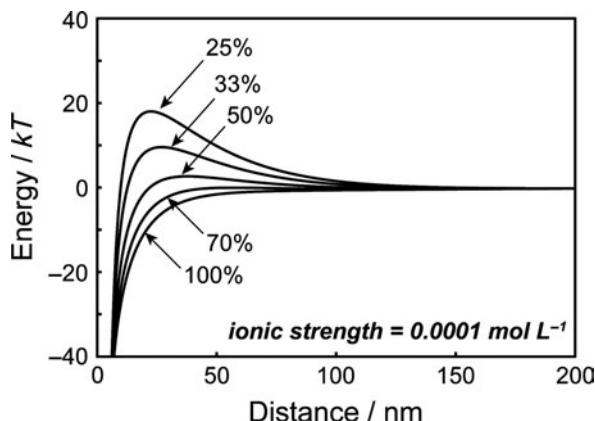
$$m = Mv \times \pi d^3/6,$$

where  $m$  is the magnitude of  $m_i$  and  $m_j$ ,  $d$  is a diameter of the particles and  $M_v$  is a volumetric magnetization of the magnetic metal (assumed to be 1000 emu/cm<sup>3</sup>).

The resulting three potential curves are plotted in units of kinetic energy ( $kT$ ) under two different ionic strengths (0.0001 mol L<sup>-1</sup> and 0.001 mol L<sup>-1</sup>) at 25 °C in Fig. 11.15, where the sum of the three curves, i.e. an overall potential curve is also drawn. In both ionic strengths, the magnitude of the magnetic interaction is comparable with that of the repulsive interaction of EDL. In other words, the magnetic interaction certainly affects the water dispersibility. When the ionic strength is low (0.0001 mol L<sup>-1</sup>), there is an energy barrier high enough to overcome the energy of Brownian motion ( $kT$ ) at around the distance of 20 nm. The CNTTs can thus keep their dispersed state. On the other hand, under the higher ionic strength (0.001 mol L<sup>-1</sup>), a valley of the potential was observed at around the distance of 50 nm. Since the bottom of the valley is deep enough to be metastable, the metal-loaded CNTTs should form an agglomerate. Indeed, in the salting-out experiment, the metal-loaded CNTTs precipitated at 0.004 mol L<sup>-1</sup> (Table 11.1) and this



**Fig. 11.15** Potential curves that work between two metal-loaded CNTTs.



**Fig. 11.16** Overall potential curves for the CNTTS with different metal-filling ratios (25, 33, 50, 70 and 100 vol.%).

value is not far from the value estimated based on the present simple model ( $\sim 0.001 \text{ mol L}^{-1}$ ).

The model in Fig. 11.14 represents the CNTTs with a metal filling ratio of 25%. (The 25-nm diameter particle is present in the 100-nm length CNTT.) When the filling ratio of the metal is increased, the overall potential curve changes significantly. Figure 11.16 shows the potential curves with different filling ratios (25, 33, 50, 70 and 100 vol.%). The ratio is changed by controlling the length of the graphite plate in this model so that the 25, 33, 50, 70 and 100 vol.% corresponds to the length of 100, 75, 50, 35 and 25 nm, respectively. Figure 11.16 clearly indicates that the potential barrier that hampers the agglomeration of the metal-loaded CNTTs is lowered when the filling ratio is increased and the barrier becomes comparable with the thermal energy ( $kT$ ) when the filling ratio is over 50%. In other words, the metal-loaded CNTTs with a filling ratio higher than 50% are not water dispersible even after the  $\text{H}_2\text{O}_2$  treatment. To improve the water dispersibility of CNTTs with a higher metal-filling ratio, further surface modification might be effective. Alternatively, to employ CNTTs with thicker carbon walls would be useful. When the carbon wall is thicker, the magnetic interaction is reduced relative to the repulsive interaction. From our rough estimation, even for fully metal-filled CNTTs, they would be water dispersible when the thickness of the carbon wall is over 15 nm.

## 11.5 Carbon nanotube cavities as a reaction field of hydrothermal synthesis

The straight nanochannels of an AAO film are quite useful as a template for fabricating one-dimensional nanosized materials. This template material is, however, unable to be used under severe reaction conditions such as in strong acid or alkaline solutions. Carbon materials are, therefore, very attractive as a template, because they are thermally stable and chemically inert. Some researchers have synthesized one-dimensional zeolite (Pham-Huu *et al.* 2004),  $\text{NbS}_2$  (Zhu *et al.* 2001), GaN (Han *et al.* 1997) and  $\text{CoFe}_2\text{O}_4$  (Keller *et al.* 2004) by using multiwalled carbon nanotubes as a template. Madsen *et al.*

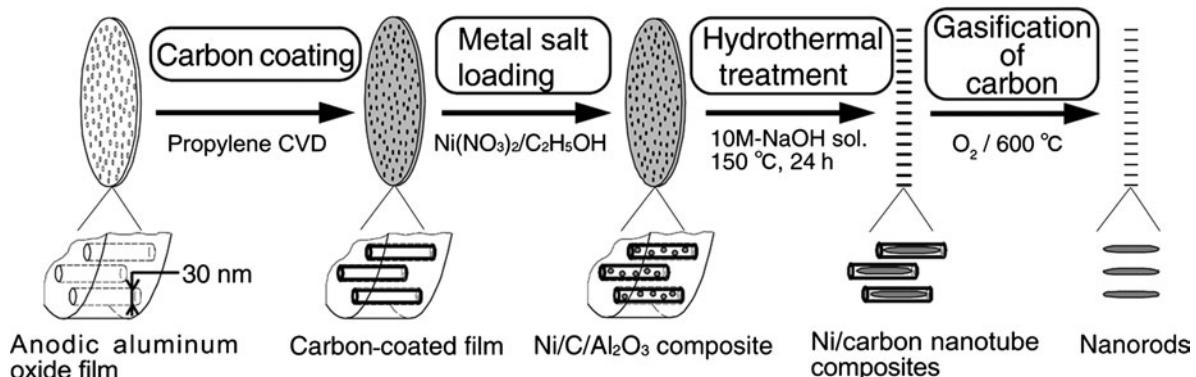
(1999) have used interparticle voids of carbon blacks as the reaction field for the hydrothermal synthesis of nanosized zeolite crystals. However, the size control of cavities in the carbon templates is not easy, as is evident in conventional carbon nanotubes and carbon black.

As described before, an AAO film was coated with carbon by the CVD technique, where the carbon deposition was surprisingly uniform and fully covered the whole surface of the film including the inner walls of the nanochannels. Such uniform carbon coating must improve the chemical stability of the films, because the carbon deposit will function as a protective layer. The carbon-coated nanochannels in the film, therefore, could be used as a template in hydrothermal treatment. Highly crystallized nickel hydroxide, which is usually prepared in a hydrothermal treatment, has widely been used as a positive-electrode active material in nickel batteries. The electrochemical properties and practical capacity are directly affected by the morphology and size of  $\text{Ni(OH)}_2$  crystals. In particular, reducing their size to the nanoscale is effective for the improvement of the electrochemical performance (Oliva *et al.* 1982; Reisner *et al.* 1997; Cai *et al.* 2004). Such nanoscale morphology and size control is possible, but not easy in usual hydrothermal methods, which always require very careful consideration in the reaction conditions and the selection of starting materials (Yang *et al.* 2004; Cheng *et al.* 2005; Jiao *et al.* 2007). In marked contrast, the template method using the carbon-coated AAO film looks quite straightforward to control both the size and shape, because both the size and shape of the resulting  $\text{Ni(OH)}_2$  crystals directly reflect those of the nanochannels of the AAO film.

In the present section, we describe our attempts at the hydrothermal synthesis of single-crystal  $\beta\text{-Ni(OH)}_2$ , and NiO nanorods in the cavity of the carbon-coated nanochannels (Matsui *et al.* 2002; Orikasa *et al.* 2007) and demonstrate for the first time that the hydrothermal synthesis method can be applied to template-based preparation of nanosized materials. Moreover, we examine how  $\text{Ni(OH)}_2$  crystals are formed and grow during the hydrothermal process in the carbon-coated nanochannels. In particular, we try here to follow the time course of crystal formation and growth in the nanochannels.

### 11.5.1 Hydrothermal synthesis of single-crystal $\text{Ni(OH)}_2$ , and NiO nanorods

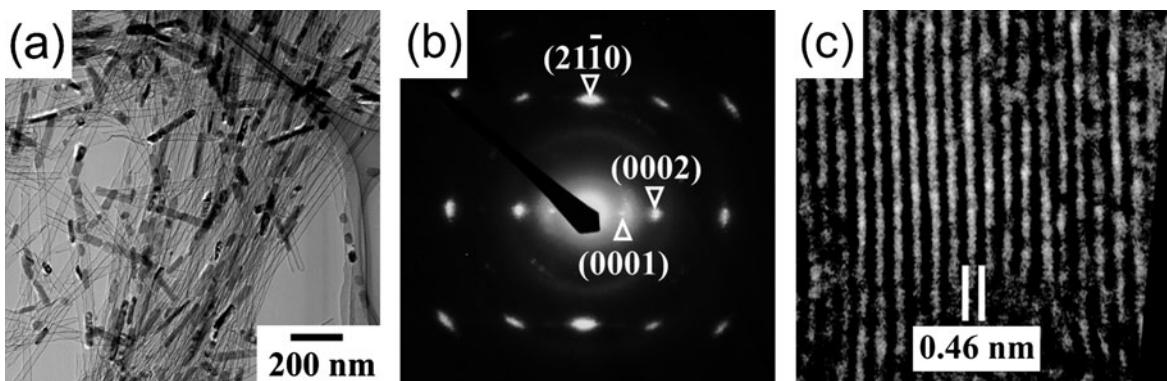
After an AAO film of  $75\ \mu\text{m}$  thickness was prepared on an aluminum plate, the film was separated from the plate and then an impervious layer was etched by immersing the film in sulfuric acid. As a result, a closed end of each AAO nanochannel became open and its diameter was increased to  $30\ \text{nm}$ . The AAO film with both-end-opened channels was subjected to propylene CVD at  $800\ ^\circ\text{C}$  for 2 h. The resulting carbon-coated film was impregnated with ethanol solution of  $\text{Ni(NO}_3)_2 \cdot 6\text{H}_2\text{O}$  at room temperature and dried under vacuum at  $80\ ^\circ\text{C}$ . The film loaded with the nickel salt was subjected to hydrothermal treatment with  $\text{NaOH}$  ( $10\text{ M}$ ) solution at  $150\ ^\circ\text{C}$  for 24 h in



**Fig. 11.17** Schematic of the synthesis of nanorods by hydrothermal treatment in the template.

an autoclave. Thereby, a Ni/carbon nanotube composite was obtained as an insoluble fraction. The composite was then heat treated in  $\text{O}_2$  at  $600^\circ\text{C}$  to burn out carbon nanotubes. The synthesis process is illustrated in Fig. 11.17. Note that in this case both ends of each channel in the AAO film are open to the outside, forming a contrast with the case of the CNTT formation, where one of the channel ends is closed.

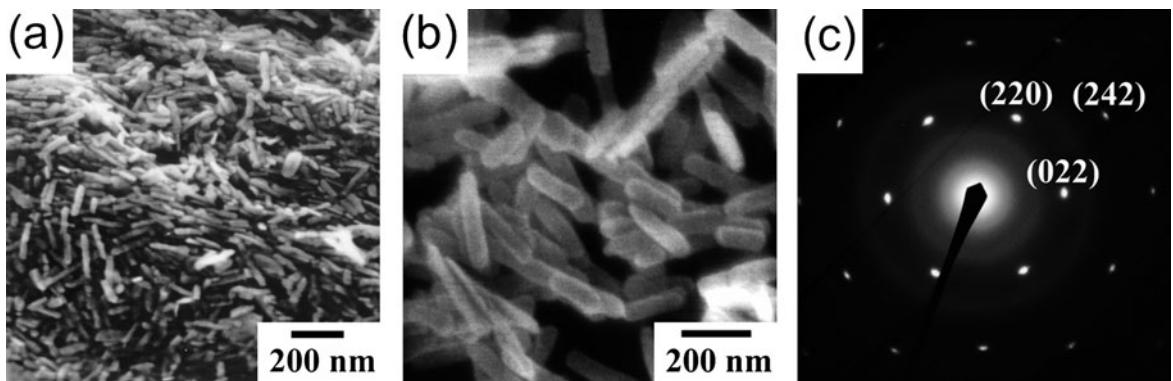
Figure 11.18 shows the electron microscopic features of nickel/carbon nanotube composite obtained as an insoluble fraction in the hydrothermal treatment. The AAO film was completely dissolved into NaOH solution under the present hydrothermal treatment conditions, and the carbon remains as carbon nanotubes. In the cavity of carbon nanotubes with an inner diameter of 25 nm, many nanorods were observed as dark images (Fig. 11.18(a)). Although some of the carbon nanotubes were empty, the others were filled with many nanorods. The diameter of the nanorods was around 25 nm, which is the same as the inner diameter of the carbon nanotubes. The length of the nanorods varies from 50 to a few hundred nm. The nanorods were identified as  $\text{Ni}(\text{OH})_2$  from their X-ray diffraction peaks.



**Fig. 11.18** (a) Low-magnification TEM image of nickel/carbon nanotube composites obtained as an insoluble fraction from the hydrothermal treatment. (b) SAED pattern of a  $\text{Ni}(\text{OH})_2$  nanorod in the cavity of carbon nanotube, showing the  $[120]$  zone axis of  $\text{Ni}(\text{OH})_2$  single crystal. (c) the corresponding HRTEM image of the  $\text{Ni}(\text{OH})_2$  nanorod.

For further understanding of the crystallographic features of the present nanorods, many individual Ni(OH)<sub>2</sub> nanorods were examined with the SAED and HRTEM techniques. A SAED pattern from a single nanorod is shown in Fig. 11.18(b), in which we observe many spots and all of them were identified as the diffraction from hexagonal  $\beta$ -Ni(OH)<sub>2</sub>. The appearance of the diffraction not as rings, but as spots, indicates high crystallinity of the Ni(OH)<sub>2</sub> nanorods. Each of the spots can be assigned the Miller indices as in Fig. 11.18(b). From the examination of the spatial arrangement of the spots in the SAED patterns, it was found that the set of lattice planes giving rise to all these spots derives from a single hexagonal crystal with its [120] direction being oriented toward the direction of the electron beam. In other words, the SAED pattern of Fig. 11.18(b) corresponds to the diffraction pattern with [120] zone axis. From the position of (001) spots in the SAED pattern, the (001) planes of Ni(OH)<sub>2</sub> should run parallel to the longitudinal axis of the nanorod. This is confirmed by an HRTEM image (Fig. 11.18(c)), where the lattice fringes in the nanorod are clearly observed and the lattice planes are parallel to the nanorod axis. The regular spacing of the observed lattice planes was 0.46 nm, which is consistent with the *d* spacing of (001) planes of hexagonal Ni(OH)<sub>2</sub>. For the other nanorods, we performed a similar TEM analysis and found that most of the nanorods could be regarded as a single Ni(OH)<sub>2</sub> crystal. No polycrystalline nanorod was observed as far as our examination is concerned. Furthermore, the (001) planes of the single-crystal nanorods were always parallel to the Ni(OH)<sub>2</sub> nanorod axis.

After the combustion of carbon nanotubes, the remaining substances were only nanorods, indicating that the nanorods could be liberated completely from carbon nanotubes by the combustion. Figures 11.19(a) and (b) show SEM images of the nanorods obtained upon the combustion. Numerous numbers of nanorods seem to aggregate with each other. The size of individual nanorods seen in the SEM images was almost the same as those observed in the carbon nanotubes before the combustion process. More detailed features of the nanorods were studied with a TEM. In the TEM observation, only a few nanorods were observed as a disengaged rod from the aggregates and its



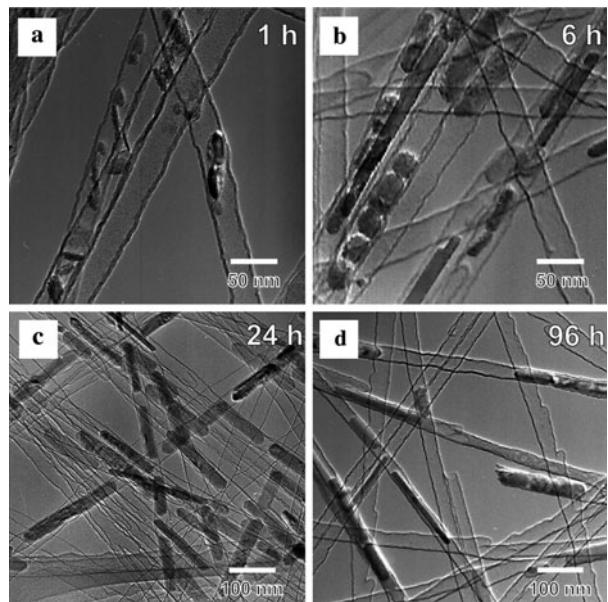
**Fig. 11.19** (a, b) SEM images of NiO nanorods liberated from carbon nanotubes by combustion. (c) SAED pattern of a NiO nanorod with <111> zone axis of single-crystal NiO.

resulting SAED pattern (Fig. 11.19(c)) presents an hexagonal arrangement of the {220} spots that corresponds to the diffraction pattern with the  $<111>$  zone axis of NiO crystal. These results indicate that the nanorod observed here is a single crystal of NiO, and that Ni(OH)<sub>2</sub> nanorods in the carbon nanotubes were converted to NiO during the combustion with little change in their shape and size. The SAED pattern (Fig. 11.19(c)) reveals that the (111) planes of this NiO nanorod are parallel to its longitudinal axis, implying that the (111) planes used to be (001) planes of the original Ni(OH)<sub>2</sub> nanorod. Such little change in crystal size and smooth conversion of (001) planes in hexagonal Ni(OH)<sub>2</sub> to (111) planes in fcc NiO by thermal decomposition are well-known and well-understood phenomena for large crystal particles (Cronan *et al.* 1976; Durand-Keklikian *et al.* 1994; Carpenter *et al.* 1999). A similar mechanism can therefore be applied to the case of the present nanorods.

### 11.5.2 Crystal formation and growth during hydrothermal synthesis

In this way, single-crystal Ni(OH)<sub>2</sub> nanorods were synthesized by the hydrothermal reaction in the cavity of a carbon-coated AAO template. The size of the nanorods is the same as the inner diameter and the length is around a few hundred nm. Using exactly the same approach but with a different type of metal salt as a precursor, we can produce many other types of single-crystal nanorod. To illustrate, single-crystal nanorods of  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> were obtained when iron nitrate was used as a precursor (Matsui *et al.* 2002). The formation of such nanorods suggests that the nanochannels in an AAO film can be used as a reaction space for the hydrothermal synthesis as long as they were uniformly coated with carbon. However, the overall reaction behavior in such a nanospace as the AAO channels has not experimentally been examined and there has been no systematic understanding of the crystal formation and growth in nanospace. A deep insight into the reaction behavior is quite essential for the precise control of the morphology and shape of Ni(OH)<sub>2</sub> crystals. In this section, we examine how single Ni(OH)<sub>2</sub> crystals are formed and grow during the hydrothermal process in the carbon-coated nanochannels.

Three types of AAO films with different channel diameters were used; two of them were prepared by anodic oxidation in different anodizing baths (sulfuric acid and oxalic acid) and the other is a commercially available AAO film (Whatman Ltd., Anodisc 25). The sulfuric and oxalic acid-anodized AAO films possess straight nanochannels of 30 and 100 nm diameter, respectively. The commercially available AAO film is characterized by its large channels (300 nm in diameter). The thickness of all the AAO films is in the range of 60–80  $\mu\text{m}$  and both ends of each nanochannel in these films are open to the outside. The carbon deposition over the AAO films and the subsequent loading of Ni(NO<sub>3</sub>)<sub>2</sub> were performed in the same manner as above, but the resulting carbon-coated and Ni(NO<sub>3</sub>)<sub>2</sub>-loaded AAO films were subjected to hydrothermal treatment for different periods (1, 6, 24 and 96 h). The use of the three types of AAO films allows one to examine the effect of the channel

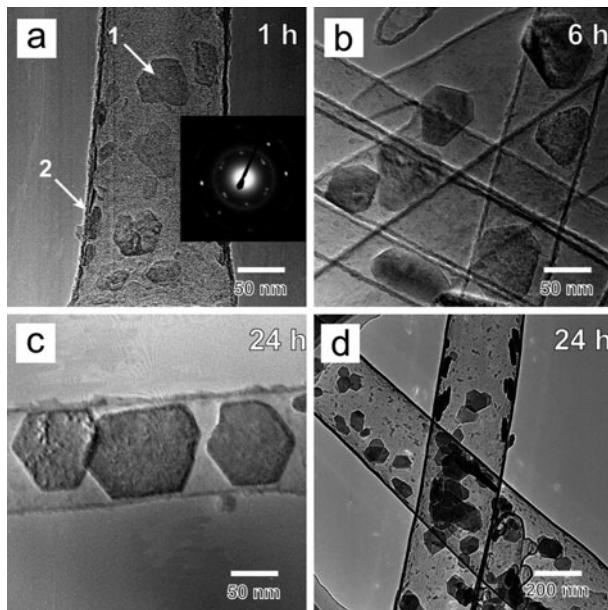


**Fig. 11.20** TEM images of  $\text{Ni}(\text{OH})_2$  crystals synthesized for different hydrothermal periods in the carbon-coated nanochannels with a diameter of 25 nm: (a) 1 h; (b) 6 h; (c) 24 h; and (d) 96 h.

size on the crystal growth of  $\text{Ni}(\text{OH})_2$ . Moreover, the hydrothermal treatment for different periods reveals the time course of the crystal formation and growth.

Figure 11.20 shows TEM images of the carbon/nickel composites synthesized for the different hydrothermal periods using the carbon-coated films with a channel diameter of 25 nm (it was 30 nm before the carbon deposition). In the composites obtained by the reaction for 1 h (Fig. 11.20(a)), flake-shaped inserts with a size less than 50 nm are observed. Upon reaction for 6 h (Fig. 11.20(b)), such flaky inserts seem to disappear, but short nanorod-like materials with a size as large as the tube inner diameter appear instead. For most of the nanorods, their length is no longer than 150 nm. Further hydrothermal treatment (the reactions for 24 and 96 h) makes the length noticeably longer, as shown in Figs. 11.20(c) and (d), where nanorods with a length over 150 nm are often found. For some of these nanorods, SAED analysis was performed. In any case, clear diffraction spots were observed and all of them can be identified as the diffraction from an hexagonal crystal of  $\beta\text{-Ni}(\text{OH})_2$ . We examined several tens of individual nanorods, and confirmed that all of them are a single crystal.

To examine the effect of the cavity size on the  $\text{Ni}(\text{OH})_2$  crystal formation, the carbon-coated films with different channel diameters (100 and 300 nm) were employed for the hydrothermal synthesis. Figures 11.21 (a)–(c) show typical TEM images of the composites synthesized by using the films with a channel diameter of 100 nm for different treatment periods (1, 6 and 24 h). At the reaction for 1 h (Fig. 11.21(a)), some hexagonal crystals with sizes less than 50 nm were found in the cavity (as indicated by arrow 1). As well as the hexagonal crystals, some tiny and thin crystals were placed around the sidewall of the carbon tube (as indicated by arrow 2). By the reaction for 6 h, the size of such hexagonal crystals was increased. In addition, some nanorod-like



**Fig. 11.21** TEM images and a SAED pattern of  $\text{Ni}(\text{OH})_2$  crystals synthesized for different hydrothermal periods in the nanochannels with diameters of 100 and 300 nm. The crystals in the images (a, b, and c) were prepared for different periods of 1, 6, and 24 h in the 100-nm nanochannels, respectively. The inset in (a) corresponds to a SAED pattern from a crystal indicated by arrow 1 in this TEM image. The TEM image (d) shows crystals prepared in the 300-nm nanochannels for the hydrothermal period of 24 h.

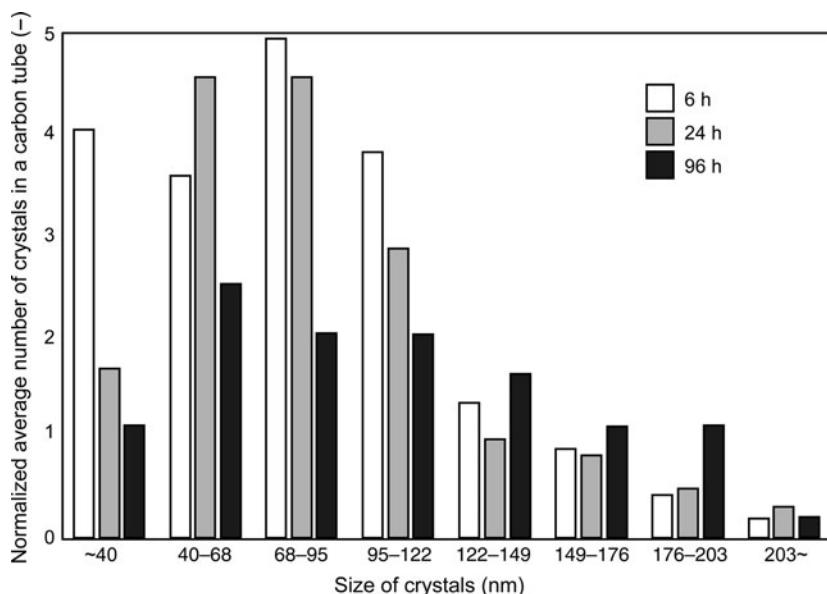
inserts are observed, but they are not major ones (Fig. 11.21(b)). When the reaction time was extended to 24 h, the size of most of the hexagonal platelets almost reaches the inner diameter of the carbon tubes (Fig. 11.21(c)). A SAED analysis was conducted for the two hexagonal flat crystals lying horizontally in the left half of Fig. 11.21(c). Again, all of the spots are attributable to the diffraction from  $\beta\text{-Ni}(\text{OH})_2$  and the analysis of the spatial arrangement of the spots reveals that the (001) planes of  $\text{Ni}(\text{OH})_2$  run parallel to the flat surface of the platelets. This structure is exactly the same as that of usual hexagonal  $\text{Ni}(\text{OH})_2$  crystals hydrothermally prepared without any spatial restriction (Fievet and Figlarz 1975; Yang and Gao 2006).

When the reaction was continued up to 96 h, many nanorod-like crystals were formed together with hexagonal crystals (not shown here). In Fig. 11.21(d), the sample obtained by using the template with nanochannels having a diameter of 300 nm (reaction period, 24 h) is shown. Only platelet crystals were formed and their sizes (50–100 nm) are similar to those of the crystals shown in Fig. 11.21(c), where the reaction time was the same but the narrower cavities (100 nm in diameter) were used.

It is obvious from these TEM results that the template effect becomes more pronounced when a narrower cavity is employed. For example, in the case of the 25-nm carbon-coated channels, the reaction for 6 h is enough to obtain short nanorods and their length was increased by further treatment (Fig. 11.20). On the other hand, it took 96 h to obtain nanorods as a major product for the 100-nm channels and nanorods were not obtained within 96 h for the 300-nm channels. The hexagonal crystals synthesized by the reaction for 24 h in the 100-nm channels range in size from 50 to 100 nm, the range of which is similar to that of the crystals in the 300-nm channels (Fig. 11.21(d)). Furthermore, a similar size range (35–140 nm) was reported in Liu *et al.* (2005), where hexagonal  $\text{Ni}(\text{OH})_2$  crystals were prepared under almost the

same hydrothermal conditions (at 160 °C for 24 h using 10 M NaOH) but no template was used. These findings suggest that the nucleation and subsequent crystal growth of  $\beta\text{-Ni(OH)}_2$  in the one-dimensional channels took place just in the same manner as the case without spatial restriction unless the crystal size becomes larger than the channel diameter.

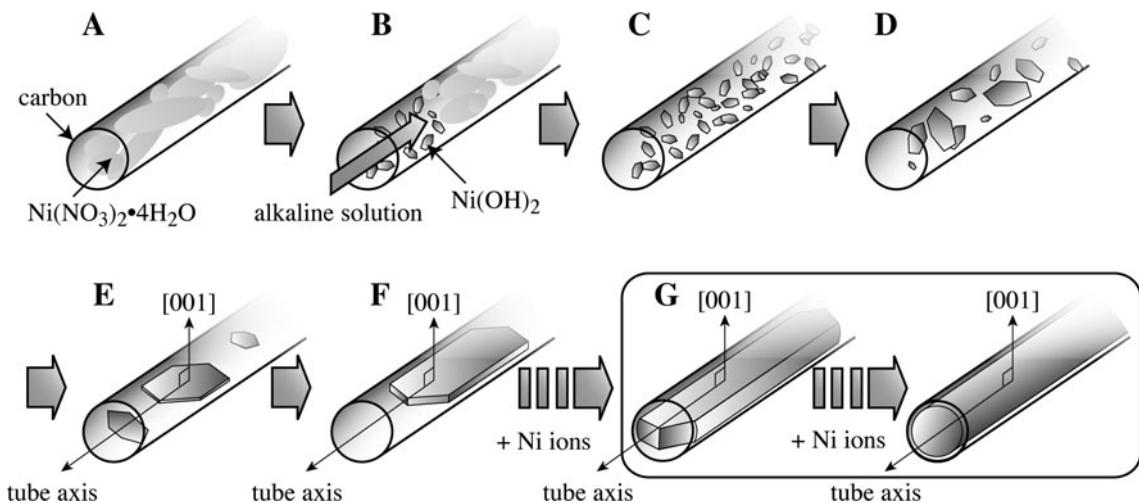
The amount of  $\text{Ni(OH)}_2$  in the carbon nanotube/nickel composites was thermogravimetrically analyzed and, as a result, it was found that the amount did not change with the hydrothermal reaction time. Moreover, this analysis suggested that the Ni atoms loaded in the 25-nm nanochannels as  $\text{Ni}(\text{NO}_3)_2$  did not go out of both open ends of each carbon nanotube through the whole hydrothermal process and the atoms remained as  $\text{Ni(OH)}_2$  crystals inside the nanotubes. This was the case also in the hydrothermal synthesis using the 100-nm nanochannels. The TEM observation for  $\text{Ni(OH)}_2$  crystals synthesized in the 25-nm channels reveals that the number of small platelet crystals decreased and the nanorod length increased with the hydrothermal treatment time. To obtain more quantitative information from the TEM images with different treatment times, we counted the number of all the crystals and measured their maximum caliper size. Then, we normalized the number by dividing the total length of all the carbon nanotubes observed in the corresponding TEM image. Such analysis was not carried out for the 1-h-treated sample, because crystals in this sample are hardly discrete. There are 338, 391 and 375 crystals inside the cavities of 56, 76 and 102 carbon nanotubes in the 6-, 24- and 96-h-treated samples and the average sizes of these crystals are calculated to be 88, 93 and 108 nm, respectively. The change in the average size looks small, but a striking difference among the three samples was found when the distribution of crystal size (Fig. 11.22) was taken into consideration. For the sample synthesized for 6 h, the number of crystals with sizes less than 40 nm



**Fig. 11.22** Crystal-size distribution obtained from the TEM images of  $\text{Ni(OH)}_2$  synthesized for different hydrothermal periods (6, 24 and 96 h) in the carbon-coated nanochannels with a diameter of 25 nm. The average numbers of crystals in these size ranges were normalized in terms of a unit carbon-nanotube length (10  $\mu\text{m}$ ). Mechanism of  $\text{Ni(OH)}_2$  nanorod formation in nanochannels. (Orikasa *et al.* *Dalton Transactions* **34**, 3757–62 (2007), reproduced by permission of The Royal Society of Chemistry.)

is much greater than those in the samples synthesized for 24 and 96 h. Upon the reaction for 24 h, the number of these small crystals is decreased significantly, but the number of crystals having sizes above 40 nm is not changed so much. When the reaction time is extended to 96 h, the size distribution is changed to a great extent. The crystals with sizes less than 122 nm are greatly decreased, but the crystals more than 122 nm are increased instead. As well as the changes in the size distribution, the number density of  $\text{Ni(OH)}_2$  crystals in the nanotubes is decreased with the reaction time; the numbers of crystals in a nanotube length of  $10\ \mu\text{m}$  are, respectively, 19.4, 16.5 and 12.0 in the 6-, 24- and 96-h-treated samples, while the total amounts of  $\text{Ni(OH)}_2$  are kept constant. These findings strongly suggest that the crystal growth of  $\text{Ni(OH)}_2$  in the nanochannels is governed by Ostwald ripening, i.e. many small crystals (less than 40 nm, in the present case) form in a system initially, but slowly disappear except for a few that grow larger at the expense of the small crystals.

Based on the results obtained in the above analyses, we propose a crystal-growth mechanism in the 25-nm nanochannels, as illustrated in Fig. 11.23. From the amount of  $\text{Ni(NO}_3)_2 \cdot 4\text{H}_2\text{O}$  loaded in the channels, it can be estimated that 75% of the whole cavity in each channel of the carbon-coated AAO film was filled with the solid nickel nitrate (Fig. 11.23(A)). When the alkaline solution gains entrance into the channels, the solution is, at first, allowed to react with the nitrate present near the open mouth to immediately form small flaky  $\text{Ni(OH)}_2$  crystals, which opens the way to the channel inside because the molar volume of the resulting hydroxide ( $22\ \text{cm}^3$ ) is much smaller than that of the original nitrate ( $190\ \text{cm}^3$ ) (Fig. 11.23(B)). As a result, the reaction proceeds



**Fig. 11.23** Crystal formation and growth process of  $\beta\text{-Ni(OH)}_2$  in the cavity of a carbon nanotube. (A)  $\text{Ni(NO}_3)_2 \cdot 4\text{H}_2\text{O}$  is loaded in the cavities of carbon-coated nanochannels. (B) Small hexagonal platelet  $\text{Ni(OH)}_2$  crystals are formed immediately after the nitrate is exposed to  $\text{NaOH}$  solution. (C) The nitrate is completely converted to the hydroxide. (D) Some crystals grow at the expense of smaller crystals by Ostwald ripening. (E) Further crystal growth makes the crystal size as large as a tube inner diameter. (F) Crystals grow further only toward the tube axis by the supply of Ni ions from other smaller crystals (not shown) in the cavity. (G) Nanorod crystals are formed by the crystal growth along the thickness direction. (Orikasa *et al.*, *Dalton Transactions*, **34**, 3757–62 (2007), reproduced by permission of The Royal Society of Chemistry.)

deep into the center of the channels and the remaining nitrate is completely exhausted within 1 h (Fig. 11.23(C)), as this was evident from the XRD pattern of the 1-h sample. Thus, there is almost no chance for the nickel to get out of the channels, as we already confirmed that almost all the nickel from the loaded nitrate was converted to  $\text{Ni(OH)}_2$  crystals.

In the early stage of the hydrothermal reaction, small flaky  $\text{Ni(OH)}_2$  crystals are formed and they are randomly oriented in the cavity of each carbon nanotube (Fig. 11.23(C)). According to the Ostwald-ripening mechanism, some of the flaky  $\text{Ni(OH)}_2$  crystals grow at the expense of other smaller ones (Figs. 11.23(D), (E)). If there were no steric hindrance imposed by the carbon nanotube, platelet hexagonal crystals with a size of several hundreds of nanometers would be formed, as observed in the case of the 300-nm channels (Fig. 11.21(d)). However, in the present case, when the size of platelet hexagonal crystals reaches the tube inner diameter, their (001) planes become parallel to the tube axis (Fig. 11.23(E)). Then, the crystals grow toward the axis and ribbon-like crystals are formed as a result (Fig. 11.23(F)). With further crystal growth not only toward the long axis but also along the thickness direction, single crystals of  $\text{Ni(OH)}_2$  nanorods are finally formed (Fig. 11.23(G)). When the nanorods were formed, the exposed surface area of nanorods is almost uniform, i.e. the solubility would be almost the same for all nanorods (note that the sidewalls of the nanorods are completely covered with the carbon wall of the CNT). Thus, Ostwald ripening would stop after the nanorods formation. Indeed, further crystal growth of large nanorods (over 203 nm) was not observed even when the reaction time was extended from 6 h to 96 h (Fig. 11.22).

## 11.6 Conclusions

This chapter can be summarized as follows. The template technique using the straight nanochannels in an AAO film allows the precise control of the length and diameter of carbon nanotubes. Moreover, this method enables one to control the status (open or closed) of both ends of each nanotubes. To illustrate, it is easy to produce one-end-opened carbon nanotubes (referred as to carbon nano-test-tubes (CNTTs) in this chapter). The CNTTs thus prepared could be used in various fields, especially as a capsule for a drug-delivery system since they have many advantages such as excellent size uniformity and controllability, solubility in water, and the presence of open ends. The encapsulation of a drug into CNTTs becomes very easy if it is performed at the stage of the carbon-coated AAO film, because all the openings are placed on one side of the flat AAO film (Fig. 11.7). This advantage is actually demonstrated when we tried to insert “foreign materials” into the cavities of CNTTs. Using the electrochemical deposition technique, Ni-Fe alloy was successfully introduced into CNTTs. Despite the electrically conductive nature of CNTTs, NiFe filling started from the pore bottoms of CNTTs. Our unexpected finding of such bottom-up filling phenomenon opens a way for the fabrication of CNTTs filled in a controlled manner with materials of interest. By the further improvement of hydrophilicity, the NiFe-loaded CNTTs also become water dispersible.

Due to their magnetically responsive property, they are promising as a capsule for magnetic delivery systems. In addition, their water dispersibility was analyzed in a semi-quantitative manner, and the analysis revealed that the magnetic interaction substantially affects the water dispersibility. Apart from the CNTTs, the uniform cavities of the carbon-coated nanochannels in an AAO film can be used as a reaction space for the synthesis of one-dimensional nano-materials under severe reaction conditions such as hydrothermal treatment, as we exemplified by the hydrothermal synthesis of single-crystal Ni(OH)<sub>2</sub> nanorods. From the TEM observation of the time course of the hydrothermal reaction, we found that the crystal growth of Ni(OH)<sub>2</sub> is governed by the Ostwald ripening even in such one-dimensional nanospace as the AAO channels.

As demonstrated in this chapter, the template technique using an AAO film is a very powerful method not only for the control of tube size (diameter and length) but also for the production of various types of carbon nanotube composites. In other words, not only an AAO film plays a role of a simple template to transfer its uniform nanochannel shape to the resulting carbon materials, but also the presence of the template material makes the synthesis method unique, because its presence facilitates many processes such as the occurrence of hydrophilicity on the outer surface of the nanotubes, the insertion of “foreign materials” into the cavities, and the use as a reaction field of hydrothermal treatment. In conclusion, the template method is quite attractive in many respects. One may consider the use of this method to prepare carbon nanotubes and/or their composites that can not be obtained by conventional synthesis methods.

## References

- AlMawlawi, D., Coombs, N., Moskovits, M. *J. Appl. Phys.* **70**, 4421 (1991).
- Cai, F.-S., Zhang, G.-Y., Chen, J., Gou, X.-L., Liu, H.-K., Dou, S.-X. *Angew. Chem. Int. Ed.* **43**, 4212 (2004).
- Carpenter, C.J.C., Wronski, Z.S. *Nanostruct. Mater.* **11**, 67 (1999).
- Chen, J., Rao, A.M., Lyuksyutov, S., Itkis, M.E., Hamon, M.A., Hu, H., Cohn, R.W., Eklund, P.C., Colbert, D.T., Smalley, R.E., Haddon, R.C. *J. Phys. Chem. B* **105**, 2525 (2001).
- Cheng, D., Gao, L. *Chem. Phys. Lett.* **405**, 159 (2005).
- Correa-Duarte, M.A., Grzelczak, M., Salgueirño-Maceira, V., Giersig, M., Liz-Marzán, L.M., Farle, M., Sieradzki, K., Diaz, R. *J. Phys. Chem. B* **109**, 19060 (2005).
- Cronan, C.L., Micale, F.J., Topi, M., Leidheiser, H., Zettlemoyer, Jr. A.C. *J. Colloid Interface Sci.* **55**, 546 (1976).
- Dagastine, R.R., Prieve, D.C., White, L.R. *J. Colloid Interface Sci.* **249**, 78 (2002).
- Delpoux-Ouldriane, S., Szostak, K., Frackowiak, E., Béguin, F. *Carbon* **44**, 799 (2006).
- Durand-Keklikian, L., Haq, I., Matijevic, E. *Colloids Surfaces, A* **92**, 267 (1994).
- Fert, A. Piraux, L. *J. Magn. Magn. Mater.* **200**, 338 (1999).
- Fievet, F., Figlarz, M. *J. Catal.* **39**, 350 (1975).

- Georgakilas, V., Tzitzios, V., Gournis, D., Petridis, D. *Chem. Mater.* **17**, 1613 (2005).
- Girard, R. *J. Appl. Phys.* **38**, 1423 (1967).
- Han, W., Fan, S., Li, Q., Hu, Y. *Science* **277**, 1287 (1997).
- Herzer, G. *Scripta Metall. Mater.* **33**, 1741 (1995).
- Jiao, Q.-Z., Tian, Z.-L., Zhao, Y. *J. Nanopart. Res.* **9**, 519–522 (2007).
- Kam, N.W.S., Jessop, T.C., Wender, P.A., Dai, H. *J. Am. Chem. Soc.* **126**, 6850 (2004).
- Kawai, S., Ishiguro, I. *J. Electrochem. Soc.* **123**, 1047 (1976).
- Keller, N., C. Pham-Huu, C., Estournés, C., Grenéche, J.-M., Ehret, G., Ledoux, M. *J. Carbon* **42**, 1395 (2004).
- Korneva, G., Ye, H., Gogotsi, Y., Halverson, D., Friedman, G., Bradley, J.-C., Kornev, K.G. *Nano Lett.* **5**, 879 (2005).
- Kyotani, T., Tsai, L.-F., Tomita, A. *Chem. Commun.* 701 (1997).
- Kyotani, T., Mori, T., Tomita, A. *Chem. Mater.* **6**, 2138 (1994).
- Kyotani, T., Pradhan, B.K., Tomita, A. *Bull. Chem. Soc. Jpn.* **72**, 1957 (1999).
- Kyotani, T., Sonobe, N., Tomita, A. *Nature* **331**, 331 (1988).
- Kyotani, T., Tsai, L.F., Tomita, A. *Chem. Mater.* **7**, 1427–28 (1995).
- Kyotani, T., Tsai, L.F., Tomita, A. *Chem. Mater.* **8**, 2109 (1996).
- Li, J.-L., Chun, J., Wingreen, N.S., Car, R., Aksay, I.A., Saville D.A. *Phys. Rev. B* **71**, 235412 (2005).
- Liu, X., Qiu, G., Wang, Z., Li, X. *Nanotechnology* **16**, 1400 (2005).
- Madsen, C., Jacobsen, C.J.H. *Chem. Commun.* 673 (1999).
- Matsui, K., Kyotani, T., Tomita, A. *Mol. Cryst. Liq. Cryst.* **387**, 225 (2002).
- Matsui, K., Kyotani, T., Tomita, A. *Adv. Mater.* **14**, 1216 (2002).
- Matsui, K., Pradhan, B.K., Kyotani, T., Tomita, A. *J. Phys. Chem. B* **105**, 5682 (2001).
- Nielsch, K., Muller, F., Li, A.P., Gosele, U. *Adv. Mater.* **12**, 582 (2000).
- Nielsch, K., Wehrspohn, R.B., Barthel, J., Kirschner, J., Gosele, U., Fischer, S.F., Kronmuller, H. *Appl. Phys. Lett.* **79**, 1360 (2001).
- Oliva, P., Leonardi, J., Laurent, J.F., Delmas, C., Braconnier, J. J., Figlarz, M., Fievet, F., de Guibert, A. *J. Power Sources* **8**, 229 (1982).
- Orikasa, H., Inokuma, N., Itisanronnachai, S., Wang, X.-H., Kitakami, O., Kyotani, T. *Chem. Commun.* DOI: 10.1039/b800475g (2008).
- Orikasa, H., Inokuma, N., Okubo, S., Kitakami, O., Kyotani, T. *Chem. Mater.* **18**, 1036 (2006).
- Orikasa, H., Karoji, J., Matsui, K., Kyotani, T. *Dalton Trans.* **34**, 3757 (2007).
- O'Connell, M.J., Boul, P., Ericson, L.M., Huffman, C., Wang, Y., Haroz, E., Kuper, C., Tour, J., Ausman, K.D., Smalley, R.E. *Chem. Phys. Lett.* **342**, 265 (2001).
- Pantarotto, D., Partidos, C.D., Graff, R., Hoebeke, J., Briand J.-P., Prato, M., Bianco, A. *J. Am. Chem. Soc.* **125**, 6160 (2003).
- Pham-Huu, C., Winé, G., Tessonniere, J.-P., Ledoux, M.-J., Rigolet, S., Marichal, C. *Carbon* **42**, 1941 (2004).
- Pradhan, B.K., Kyotani, T., Tomita, A. *Chem. Commun.* 1317 (1999).
- Pradhan, B.K., Toba, T., Kyotani, T., Tomita, A. *Chem. Mater.* **10**, 2510 (1998).
- Reisner, D.E., Salkind, A.J., Strutt, P.R., Xiao, T.D. *J. Power Sources* **65**, 231 (1997).

- Sellmyer, D.J., Zheng, M., Skomski, R. *J. Phys.: Condens. Matter* **13**, R433 (2001).
- Sharma, A., Nakagawa, H., Miura, K. *Carbon* **44**, 2090 (2006).
- Sonobe, N., Kyotani, T., Tomita, A. *Carbon* **29**, 61 (1991).
- Sun, Y.-P., Fu, K., Huang, W. *Acc. Chem. Res.* **35**, 1096 (2002).
- Terrones, M. *Int. Mater. Rev.* **49**, 325 (2004).
- Utsumi, S., Urita, K., Kanoh, H., Yudasaka, M., Suenaga, K., Iijima, S., Kaneko, K. *J. Phys. Chem. B* **110**, 7165 (2006).
- Wang, X.H., Orikasa, H., Inokuma, N., Yang, Q.-H., Hou, P.-X., Oshima, H., Itoh, K., Kyotani, T. *J. Mater. Chem.* **17**, 986 (2007).
- Whitney, T.M., Jiang, J.S., Searson, P.C., Chien, C.L. *Science* **261**, 1316 (1993).
- Yang, D., Yang, R., Zhang, J., Liu, Z. *J. Phys. Chem. B* **108**, 7531 (2004).
- Yang, R., Gao, L. *J. Colloid Interface Sci.* **297**, 134 (2006).
- Zheng, M., Jagota, A., Semke, E.D., Diner, B.A., Mclean, R.S., Lustig, S.R., Richardson, R.E., Tassi, N.G. *Nature Mater.* **2**, 338 (2003).
- Zhu, Y.Q., Hsu, W.K., Kroto, H.W., Walton, D.R.M. *Chem. Commun.* 2184 (2001).

# 12

# Nanocatalysis

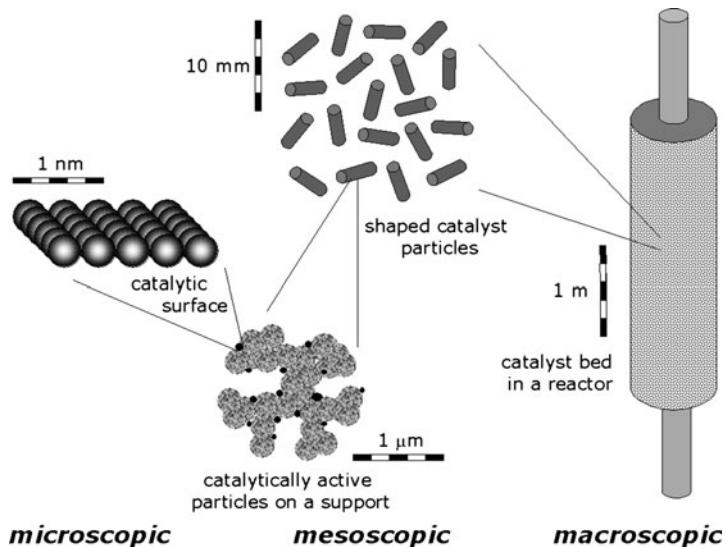
*R.T. Vang, S. Wendt, and F. Besenbacher*

12.1 Introduction	416
12.2 Surface characterization	419
12.3 Single-crystal surfaces	427
12.4 Changing the reactivity at the atomic scale: Design of new catalysts from first principles	433
12.5 Nanoparticles	437
12.6 TEM studies of nanoclusters on high surface area supports	458
12.7 Conclusions and outlook	464
References	465

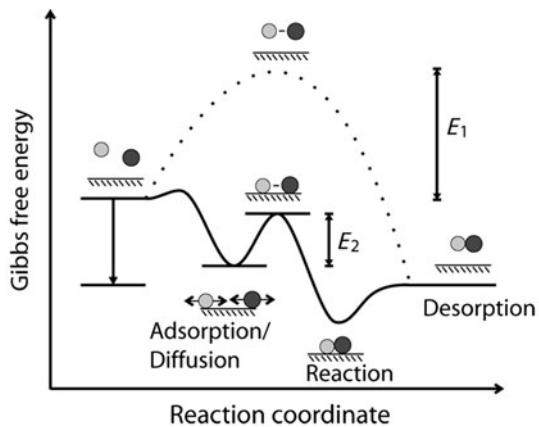
## 12.1 Introduction

A catalyst is a substance that accelerates the rate of a chemical reaction or changes its product distribution without being part of its final products. Catalysis constitutes a cornerstone of life as we know it; it provides the key technology for large-scale production of basic chemicals, from fuels, fertilizers, plastics to pharmaceuticals, and is of utmost importance in many biological processes. Catalysis is essential to a healthy economy in our society, and it has been estimated that approximately 25% of the production in the industrialized world is dependent on catalysis directly or indirectly. Also in living organisms, biochemical processes are governed by enzymatic catalysts, which have been created and refined through evolution. Today, there is an increasing awareness worldwide that the way we produce energy may jeopardize the climate and the environment of the Earth and it has been realized that novel catalysts will be mandatory for new vital environmental protection, for a better and more economic exploitation of the scarce fossil-fuel reserves (oil refining) or by facilitating a transition from carbon-based energy sources to a more sustainable energy production based on, for example, hydrogen or light-harvesting photocatalysts.

Real industrial heterogeneous catalysts, such as for example the steam reforming catalyst converting natural gas into synthesis gas ( $H_2$  and  $CO$ ), are immensely complex structures typically consisting of 1–10-nm large metallic nanoparticles supported on a sponge-like porous high surface area ceramic support (see Fig. 12.1) and the processes involved are often highly influenced by a wide range of external parameters (temperature, pressure, etc.). As a consequence of this complexity and despite the great importance and increasing application of catalysis in society, the detailed atomic-scale understanding and principles governing the catalyzed chemical transformations involved in even simple reactions are in general not established. In heterogeneous catalysis, the reactants are temporarily bound to the solid surface of the catalyst, which provides an alternative reaction route with lower energy barriers, thereby enabling the reaction to proceed at a much higher rate (Fig. 12.2). A catalyzed chemical reaction is therefore the outcome of a complex interplay between a number of sequential elementary processes that typically involve adsorption of reactants, diffusion to the active site, docking to this site followed by bond breaking and formation of new bonds and a final desorption step of the



**Fig. 12.1** The different length scales involved in heterogeneous catalysis. Reproduced from Chorkendorff and Niemantsverdriet (2003) Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.



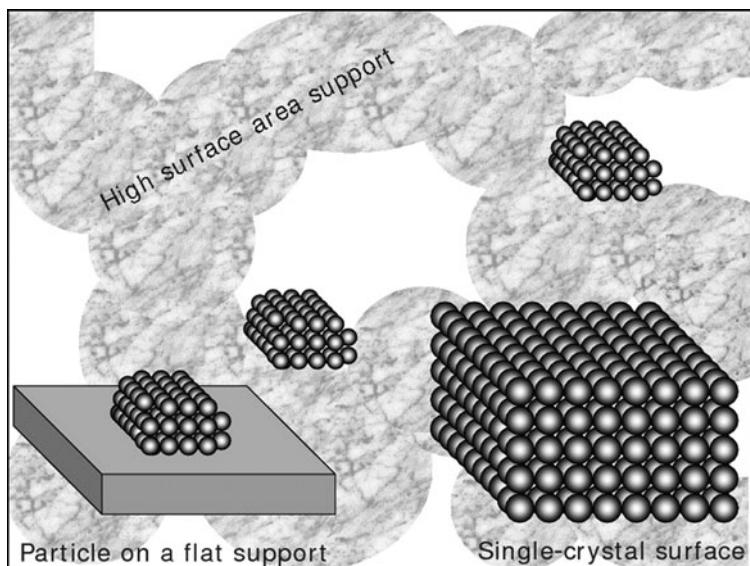
**Fig. 12.2** Potential-energy diagram for a surface-catalyzed reaction. The energy barrier for the surface-catalyzed reaction ( $E_2$ ) is reduced compared to the barrier for the gas-phase reaction ( $E_1$ ).

products. Being inherently atomic-scale processes, fundamental understanding of catalytic reactions thus requires the ability to characterize surface structures, and in particular the so-called catalytically active sites, at the atomic level, and thus ideally requires studies of the catalyst and single reactants with high spatial and temporal resolution.

The activity of a real high surface area catalyst scales directly with the number of exposed active sites on the surface, and the activity is optimized by dispersing the active material as nanometer-sized particles onto highly porous supports, e.g. alumina or silica, with surface areas often in excess of  $100 \text{ m}^2/\text{g}$ . The properties of the nanosized particles are in general not scalable from those of the bulk materials, and the saying “small is different” within the area of nanoscience, is very much relevant for catalysts, one of the most important nanomaterials. Numerous examples that

show that nanoclusters reveal new physical and chemical properties, and thus unique catalytic properties as their dimensions are reduced down to the nanoscale exist. One of the most distinct examples is the unique catalytic properties of Au nanoparticles with a size of 1–3 nm, which have a high activity for oxidation of, e.g. CO, although bulk Au is known to be a very noble metal with very low chemical reactivity (Haruta 2004; Chen and Goodman 2006; Hashmi and Hutchings 2006).

The structural complexity of the materials, combined with the high temperatures and pressures of catalysis, may limit the possibilities for detailed structural characterization of real catalysts. This is why in the past catalysts have often been developed mainly by trial-and-error methods based on intuition, but the need for better and more efficient heterogeneous catalysts has recently led to the development of a plethora of specialized experimental nanoscience techniques that provide detailed structural and chemical insight into complex catalysts. To gain such detailed fundamental insight into the elementary steps involved in catalytic reactions many researches have adopted the so-called “surface-science approach” (Ertl 2008). To simplify the considerations of reactions on surfaces, model systems consisting of either flat single-crystal surfaces or well-defined nanoclusters on surfaces are studied under clean and well-controlled, often ultrahigh-vacuum (UHV), conditions (Fig. 12.3). Within the surface-science approach a variety of different surface-science techniques have been developed and employed. These traditional surface-science techniques have given an impressive amount of very detailed insight into a number of fundamental processes on surfaces, but the studies have typically been based on diffraction or spectroscopy signals averaged over macroscopic areas of the surface and in general do not provide direct information about individual atoms and particles. The area of surface science and thus the area of catalysis was,



**Fig. 12.3** Schematic drawing of a high surface area supported catalyst and model systems consisting of a single-crystal surface or nanoparticles on a flat support. Reproduced from Chorkendorff and Niemantsverdriet (2003) Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

however, revolutionized by the development of scanning probe microscopes in the 1980s. These scanning probe microscope (SPM) techniques are a family of direct, real-space, local probe techniques that are capable of exploring the atomic-scale realm of surfaces with atomic resolution. With the best and most stable SPM microscopes it has become possible to use SPM as an outstanding tool for monitoring atomic and molecular dynamics on surfaces, the adsorption, diffusion and reaction of molecules that are the basic steps involved in reactions on surfaces with direct relevance to heterogeneous catalysis. Furthermore, the ultimate resolving power of the SPMs has led to unprecedented new atomic-scale insight into catalytic active sites, and thereby emphasized the catalytic important edges, kinks, atom vacancies and other defect sites, which are often extremely difficult to detect when using more traditional averaging techniques. The knowledge base thereby obtained has provided valuable information about a number of catalytic reactions, and several examples where this basic insight has been used to design new high surface area catalysts now exist.

The simplified surface-science approach, however, introduces the so-called “materials gap” and the “pressure gap” in catalysis, referring to the gap between UHV-based experiments conducted on clean and well-defined model surfaces and the complex structure of a real catalyst operating at high pressures and temperatures. Similarly, novel nanoscience synthesis techniques that allow for the synthesis of nanoparticles with well-controlled size and chemical composition have recently been developed. A detailed understanding of the interrelation between the structure, composition and properties of catalysts will open up the possibility to design novel catalysts with tailored catalytic properties for specific reactions.

## 12.2 Surface characterization

Within the area of surface science and catalysis a variety of different surface-science techniques have been developed and employed. These traditional surface-science techniques, many of which were dependent on the development within the area of ultrahigh-vacuum (UHV) technology have given an impressive amount of very detailed insight into a number of fundamental processes on surfaces, but the studies have typically been based on diffraction or spectroscopy signals averaged over macroscopic areas of the surface and in general do not provide direct information about individual atoms and particles. It is beyond the scope of this review chapter to provide a complete and detailed discussion of the large number of available techniques and their influence on surface science and catalysis, and for a more thorough review we refer to the literature, e.g. Somorjai (1994). The area of surface science and thus the area of catalysis were, however, revolutionized by the development of scanning probe microscopes, the scanning tunnelling microscope (STM) or the atomic force microscope (AFM) in the 1980s. These scanning probe microscopy techniques are, on the contrary, a family of direct, real-space, local probe techniques that are capable of exploring the atomic-scale realm of surfaces with atomic resolution. Most of the examples presented throughout this chapter

will be based on STM or transition electron microscopy (TEM) studies. This bias towards STM and TEM studies reflects the experience and methods of choice of the authors and is not an attempt to neglect the impact of other techniques.

### 12.2.1 UHV techniques

The wide range of surface-characterization techniques can be divided into three main areas: spectroscopy, diffraction, and microscopy. Spectroscopic techniques exploit the response of the sample to an external excitation caused by, e.g. electrons, photons, ions, neutral atoms, or heat. The group of surface spectroscopy techniques include X-ray photoemission spectroscopy (XPS) (photons in—electrons out), infra-red spectroscopy (photons in—photons out), Auger electron spectroscopy (AES) (electrons in—electrons out), secondary ion mass spectroscopy (SIMS) (ions in—ions out), and temperature-programmed desorption (heat in—atoms/molecules out). These spectroscopy techniques all provide information about the chemical and electronic structure of the surface but no or only indirect information about the surface structure. The surface spectroscopy techniques are thus ideal for the characterization of the chemical properties of a model catalyst surface. To obtain surface structural information, however, diffraction techniques are more appropriate. The two most widely used diffraction techniques are based on surface diffraction of photons or electrons in, e.g. (surface) X-ray diffraction ((S)XRD) and low-energy electron diffraction (LEED).

Although spectroscopy and diffraction provide information about the atomic structure of the surface, these techniques are averaging techniques in the sense that they typically probe a large area of the surface compared to the interatomic distances, which implies that it is very difficult with these techniques to obtain information about local surface phenomena such as step edges and kinks, point defects, or single adsorbates. Truly local information can be obtained by means of advanced microscopy techniques such as TEM and STM. Since these microscopy techniques will be used throughout this chapter to illustrate important aspects of nanocatalysis, these techniques will be discussed in some detail.

The fundamental principle of STM is conceptually rather simple; an atomically sharp metal tip, typically made of W or PtIr, is brought into such close proximity (0.3–0.5 nm) to a sample surface that an overlap occurs between the tip and the surface electronic wavefunctions, which decay exponentially in the junction gap. If a small bias voltage ( $V_t$ ) is applied to the sample, electrons can tunnel elastically from filled tip states into sample-tip states or vice versa, depending on the polarity of  $V_t$ . This vacuum tunnelling, lending its name to the microscope, establishes a small tunnel current ( $I_t$ ) within the nanoampere range.

In the usual mode of operation, the STM tip is raster scanned across the surface at a fixed bias voltage with a piezo-scanner used to control the  $x$ - $y$ - $z$  motion of the tip. Because the tunnel current depends exponentially on the distance  $z$  between the tip and surface, the individual atoms on the

surface will give rise to current variations as the tip is scanned across the corrugated surface; that is, the tunnelling current will increase (or decrease) as the separation ( $z$ ) between the tip and sample decreases (or increases). As the tip sweeps over surface structures, a feedback circuit regulates the tip–sample  $z$  separation in such a manner that the tunnelling current is kept at the constant preset value  $I_0$ , and the  $z$  position of the tip, or rather the feedback signal, is recorded to produce a topographic map of the surface.

The mechanical design of an STM instrument has the inherent complication that the tip-to-sample distance has to be stabilized on a picometer scale by construction elements with dimensions in the centimeter range (Wiesendanger 1994; Chen and Friend 1997). This mismatch of more than nine orders of magnitude is a major difficulty, because construction elements, such as tip, scan unit, sample, sample holder, manipulator, and interconnecting base systems, are affected by vibrations and thermal drift. Today, the technical difficulties have largely been overcome, and STM instruments that routinely provide atomic resolution on surfaces have been built.

Modern state-of-the-art transmission electron microscopes are very powerful tools within the area of catalysis since they combine imaging and electron diffraction with local elemental analysis and electronic structure information obtained from areas ranging from the  $\mu\text{m}$ -level down to the atomic scale (Datye *et al.* 2008). Images can be recorded by use of either a parallel electron beam (in TEM) or a focused probe that is scanned across the specimen (in STEM). In either case, a spatial resolution down to about 0.1 nm that allows atomic columns to be resolved in images of crystalline materials can be obtained (Datye *et al.* 2008). The transmission geometry provides 2D images representing projections of the 3D atomic arrangement in the sample. TEM images revealing the crystal lattice can therefore be obtained if the crystalline structures are aligned with a zone axis close to the direction of the electron beam. Moreover, atomic-scale information about the exposed surfaces and interface can be obtained when supported nanoparticles are imaged in profile. In addition to the atom-resolved images, the local chemical composition from areas  $\sim 1 \text{ nm}$  or less in diameter can be analyzed by energy-dispersive X-ray spectroscopy (EDS) or electron energy-loss spectroscopy (EELS). Such spectroscopic information may be presented in 2D maps showing the spatial element distribution in the specimen. Finally, information about the local density of unoccupied electron states of a specific element can be obtained from simultaneously recorded EELS data and used to estimate the oxidation state and the local coordination geometry of the excited atoms.

### 12.2.2 *In-situ* techniques

In recent years, one of the most important developments within surface-science-based catalysis research is the large number of *in-situ* techniques that have emerged. These techniques all aim to study both real high surface area catalysts and model catalysts as close to their operating conditions as possible, i.e. at high pressure (HP) and at high temperature. One of the most challenging

problems to overcome when working at high gas pressure is the low mean-free path of electrons, ions, and neutrals, which are used as probing particles in many techniques. For high-energy photons, however, the interaction with the gas-phase particles is negligible, and X-rays are thus an excellent probe for surface characterization at HP, and several X-ray based techniques have been developed for *in-situ* studies. For “X-ray in—X-ray out” techniques such as SXRD and X-ray absorption fine-structure spectroscopy (XAFS) the extension from UHV to HP is mainly a matter of constructing suitable sample chambers, and setups where these techniques are used for *in-situ* studies at high pressures have been developed (Evans *et al.* 2007; Ferrer *et al.* 2007). SXRD is the method of choice for the determination of surface structures, and it is thus extremely useful to extend the use of this technique to reaction conditions. HP SXRD has, e.g. been used to study the adsorption of CO and the CO-induced surface reconstruction on Ni surfaces at HP of CO. In this case it was shown that above 430 K CO induces a strong restructuring of the Ni(110), resulting in (111) microfaceting (Peters *et al.* 2001), whereas no strong restructuring of the Ni(111) was observed at CO pressures up to 1.2 bar at room temperature (Quiros *et al.* 2003). When the SXRD studies are carried out in a gas mixture resembling the reactants of a catalytic reaction, SXRD can also provide information about the surface structure *during* reaction. As an example, Ackerman *et al.* used a combination of SXRD and mass spectrometry to study the surface structure of Ni(111) in a mixture of CO and H<sub>2</sub> during methane production (CO + 3H<sub>2</sub> → CH<sub>4</sub> + H<sub>2</sub>O). The authors showed that although a surface carbide phase develops between 100 and 200 °C the surface structure during methane production is identical or very close to that of clean metallic Ni(111) (Ackermann *et al.* 2004). This finding was in contrast to previous speculations that the surface of the catalyst was partly covered by a carbide phase during the methanation reaction. This example clearly demonstrates the importance of *in-situ* surface-characterization techniques. Another very important application of the HP SXRD technique has been the characterization of a series of thin surface oxides on transition metals, which are often found to be the active phase of the catalyst surface during oxidation reactions. This work includes studies on thin surface oxides on Rh(111) (Gustafson *et al.* 2004), Pd(100) (Lundgren *et al.* 2004) and Ag(111) (Schmid *et al.* 2006).

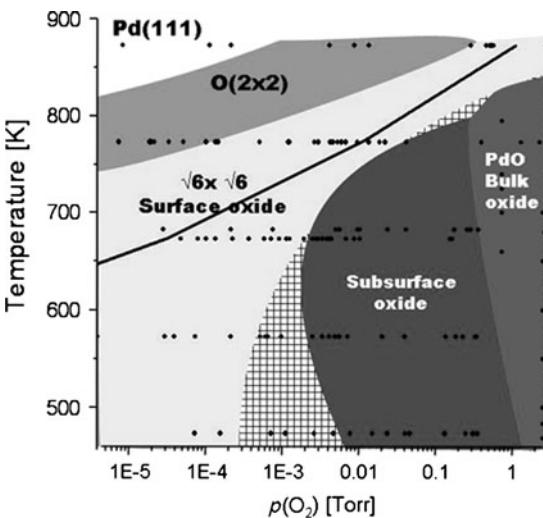
XAFS is a versatile technique that is complementary to the SXRD technique in several ways. XAFS probes the electronic structure of metal centers rather than their surface structure, and, furthermore, XAFS is applicable to non-crystalline materials, and thus XAFS is one of few surface-science techniques suited for studies of high surface area catalysts (Evans *et al.* 2007). The time resolution in XAFS measurements can be reduced to ~50 ms, which enables dynamic real-time studies of catalysts during reaction or during the initial activation (Clausen *et al.* 1998). The information obtained by the XAFS technique is averaged over macroscopic areas and provides an average of the different structures within the extent of the X-ray beam. For this reason it is desirable to use XAFS in combination with other techniques, and Clausen *et al.* successfully combined XAFS with XRD and online gas analysis during activation and reaction of high surface area catalysts. With such a combined setup

structure–reactivity relationships can be determined in great detail as Clausen *et al.* demonstrated for the activation of CuO/ZnO and CuO/ZnO/Al<sub>2</sub>O<sub>3</sub> methanol synthesis catalysts. During activation the formation of small metallic clusters was observed, and during reaction a morphological transition from disk-like particles under dry conditions to spherical particles under wet conditions was revealed (Grunwaldt and Clausen 2002) (the CuO/ZnO catalyst will be discussed in further detail towards the end of this chapter).

The adaptation of XPS to high pressures is very challenging because of the need to detect photoemitted electrons with a low mean-free path ( $\sim 1$  mm at 1 mbar for typical kinetic energies encountered in XPS). Furthermore, the design of a high-pressure (HP) XPS system is complicated by the need to operate the electron detector at high vacuum. This challenge is accomplished by using several differential pumping stages between the sample and the detector. Recently, further progress has been made by combining the differential pumping with electrostatic lenses, which focus the incoming electron beam, and thereby the apertures between the pumping stages can be minimized without severely reducing the intensity of the electron beam. With this state-of-the-art HP-XPS setup it is now possible to perform XPS studies at pressures up to 10 mbar, and with further refinement of this technique one should be able to record data even at pressures up to  $\sim 100$  mbar (Ogletree *et al.* 2002; Salmeron and Schlögl 2008).

*In-situ* XPS measurements are particularly useful to determine the oxidation state of a surface at or close to reaction conditions. The understanding of the oxidation of metal surfaces is indeed a very important problem in catalysis since many surfaces oxidize at reaction conditions. Salmeron's group used HP XPS to scrutinize the mechanism of oxidation of Pd(111) and determined the  $p$ ,  $T$  phase diagram for the O/Pd(111) system depicted in Fig. 12.4 (Ketteler *et al.* 2005). Each point in this phase diagram represents a single measurement, and the different points were obtained by increasing the O<sub>2</sub> pressure while keeping the sample at a constant temperature. From the experiments three different phases besides chemisorbed O were identified: a surface oxide with  $\sqrt{6} \times \sqrt{6}$  periodicity, a subsurface oxide, and bulk PdO oxide. The phase diagram represents the measurements and is not identical to the thermodynamic phase diagram, i.e. the hatched region in the phase diagram represents  $p$ ,  $T$  conditions at which bulk PdO is stable when the O<sub>2</sub> pressure is reduced at constant temperature. From the measurements, a two-step bulk oxidation mechanism that starts with the formation of the metastable subsurface oxide followed by a transformation into bulk PdO could be established, and also the activation barriers for forming both the subsurface oxide and the bulk oxide were determined (Ketteler *et al.* 2005).

Vibrational spectroscopy is a valuable technique to identify surface adsorbates including reactants, products, precursors, intermediates, or spectator species. Since such information is complementary to the insight gained from the X-ray-based techniques, vibrational spectroscopy operating at elevated pressures is also very important for *in-situ* studies of model catalysts. The standard technique for probing the vibrational properties of adsorbates is infrared spectroscopy. However, infrared light interacts not only with the



**Fig. 12.4**  $p, T$  phase diagram for Pd(111)/O determined from *in-situ* XPS measurements. All points represent  $p, T$  conditions where an XPS spectrum was acquired. The solid indicates the phase transition from bulk Pd to bulk PdO calculated by tabulated values of the enthalpies and heat capacities. The hatched region marks the PdO bulk stability region when the pressure is reduced from 1 Torr at fixed temperature. Reproduced with permission from Ketteler *et al.* (2005). Copyright 2005 Am. Chem. Soc.

adsorbed species but also with the gas-phase atoms/molecules, and surface analysis in a gas atmosphere can therefore only be obtained by means of a technique that is either intrinsically surface sensitive or one where the gas-phase contribution can be reduced or averaged out. The most widely used vibrational spectroscopy techniques for *in-situ* studies are sum frequency generation (SFG) and polarization modulation infrared reflectance absorption spectroscopy (PM-IRAS) (Rupprechter and Weilach 2008). SFG probes the surface with two laser pulses (one infrared and one visible) and detects the outgoing photons with a frequency corresponding to the sum of the two impinging photons. These “sum frequency photons” are generated in a second-order process that is only allowed at the 2D interface, where the inversion symmetry is broken, and hence this technique is intrinsically surface sensitive (Rupprechter and Bruce 2007). With respect to the PM-IRAS technique the contribution to the absorption signal from the surface adsorbates is singled out since the absorption by surface adsorbates is not allowed for s-polarized light, and the infrared absorption spectra acquired with s-polarized light is thus identical to the gas-phase absorption spectra. Absorption spectra obtained with p-polarized light have, on the other hand, a contribution from the surface species, and thus the surface contribution can be determined by subtracting the s-polarized absorption spectrum from the p-polarized spectrum. *In-situ* SFG and PM-IRAS are versatile techniques that provide unique information on, e.g. the nature and reactivity of specific adsorption sites, carbonaceous species, surface oxides and determine the local adsorption environment on bimetallic surfaces (Rupprechter and Bruce 2007). SFG and PM-IRAS are thus very suitable *in-situ* techniques for the study of, e.g. high-symmetry adsorbate structures (Su *et al.* 1996; Rupprechter *et al.* 2001) and dynamic processes occurring during activation, reaction or regeneration of the catalyst under reaction conditions (Ozenoy *et al.* 2004; Ozenoy and Goodman 2004; Baumer *et al.* 2007).

As already discussed, true local insight at the atomic level into the mechanism of a catalytic surface process can only be obtained by means of high-resolution microscopy, and it is thus important that, e.g. TEM and STM also can be implemented at elevated pressures. Two approaches have recently been pursued to enable TEM imaging at elevated pressures, where the major challenge is the confinement of gas (or liquid) environments close to the catalyst specimen within the vacuum column of the TEM (Gai *et al.* 2007). One method uses several differential pumping stages (Boyes and Gai 1997) in a similar fashion to that used for the HP XPS technique, whereas the other method is based on gas-tight electron-transparent thin films to separate the sample holder cell from the vacuum column (Giorgio *et al.* 2006). These capabilities for studying catalysts *in situ* during exposure to elevated gas pressures and temperatures are often referred to as environmental TEM (ETEM) and the different approaches are capable of providing ETEM images with a resolution of 0.2–0.3 nm at gas pressures of  $\sim$ 10 mbar. Improved confinement of the gas phase by means of a MEMS device was recently demonstrated to allow atomic-resolution ETEM imaging at 1 atm. pressure (Creemer *et al.* 2008), which will facilitate a more direct comparison of the atomic-scale observations with information from other *in-situ* techniques, surface-science and reactivity measurements.

Unlike most other electron-based techniques the performance of an STM is not limited by the presence of a gas phase since the electrons only have to tunnel a very short ( $\sim$ 1 nm) distance between the sample surface and the tip apex atom, and as a consequence STM can be performed in the entire pressure range from UHV to atmospheric or even higher pressures (Rasmussen *et al.* 1998; Jensen *et al.* 1999; Rössler *et al.* 2005; Vang *et al.* 2007). HP STM studies were pioneered by Salmeron and Somorjai, who designed a dedicated HP STM system and used a vacuum transfer cell for the sample transfer between the UHV chamber and the HP cell (McIntyre *et al.* 1993). Later, this design was improved so that they could perform an *in-situ* transfer without breaking the vacuum, and the upper temperature for the HP STM studies was increased to 675 K (Jensen *et al.* 1999). The group of Frenken integrated their HP STM into a microreactor (Rasmussen *et al.* 1998), and with this setup they have been able to image, e.g. a Pt(110) surface during the CO oxidation reaction (Hendriksen and Frenken 2002).

The techniques discussed here are only part of the growing number of new tools for surface and catalyst characterization that are adapted to or specifically developed for *in-situ* studies, and this development is advancing the field further towards the ultimate goal of actually “seeing” catalysis at the atomic level.

### 12.2.3 Theory

Within the last decade the development of computer technology along with the development of efficient computational algorithms have led to a major breakthrough in the theoretical understanding of catalysts and catalytic processes. First-principles calculations on large and complex systems have

become possible, and theory has reached a point where it can account for the energetics of catalytic model systems in qualitative or even quantitative agreement with experimentally determined parameters, such as the binding energies of atoms and molecules or the activation barriers for elementary processes on metal surfaces (Hammer and Nørskov 2000; Greeley *et al.* 2002).

Theory has thus developed to a stage where it can predict binding energies and activation energies (Christoffersen *et al.* 2001; Greeley and Mavrikakis 2004) and has as such become a very valuable complement to experimental surface-science techniques, and it can often provide insight into atomistic details that are difficult, or even impossible, to probe with experimental techniques. For an accurate and computationally affordable theoretical treatment of adsorption properties, reaction pathways, and activation energies for surface chemical reactions, density-functional theory (DFT) has emerged as the method of choice (Hammer and Nørskov 2000; Greeley *et al.* 2002). Today, numerous examples exist that show that the interplay between DFT and state-of-the-art experimental studies has been one of the main driving forces in the advancement of the atomic-scale understanding of heterogeneous catalysis over the last ten years.

The results obtained from the DFT calculations represent the total energies at zero temperature and zero pressure and thus do not directly relate to catalytically relevant conditions at HP and high temperature. Recently, however, Reuter *et al.* have demonstrated how DFT in combination with statistical-mechanics calculations can be used to construct phase diagrams of surface structures as functions of partial gas pressures and temperatures (Reuter and Scheffler 2002).

One of the main goals of surface-science studies of catalyst model systems is to use the detailed experimental and/or theoretical insight gained from the fundamental studies of elementary processes and surface structures on model systems to predict the catalytic activity of a real high surface area catalyst under realistic working conditions at high pressure and high temperatures. One very successful approach has been to use the theoretical *microkinetic modelling* (Stoltze 2000) method, in which one assumes a reaction mechanism and treats the individual steps as being either in thermodynamic equilibrium or as kinetically hindered one-way reactions. In the simplest case only one of the individual reaction steps is not in thermodynamic equilibrium, in which case this will be the rate-limiting step for the overall reaction. This relatively simple approach, however, requires a high degree of detailed knowledge about the reaction mechanism and about a number of physical properties that determine the equilibrium constants and rate constants of the individual reaction steps. With great success Stoltze and Nørskov performed microkinetic modelling of the ammonia synthesis reaction based to a high degree on the fundamental knowledge gained by Ertl and coworkers using the surface-science approach to study the details of the reaction mechanisms on Fe single-crystal model surfaces. The microkinetic model revealed a very convincing agreement with reactivity measurements on supported Fe catalysts, and the rate-limiting step was confirmed to be the dissociation of N<sub>2</sub> molecules on the Fe surface (Stoltze and Nørskov 1985).

Although successful, the microkinetic method used by Stoltze and Nørskov neglects the adsorbate–adsorbate interactions and dynamic effects such as adsorbate-induced surface restructuring, which may be important parameters in certain cases. It has been demonstrated by Reuter *et al.* that such dynamic effects can be taken into account in the kinetic Monte Carlo simulations, exemplified for the CO oxidation on RuO<sub>2</sub>. In this case it was revealed that such dynamic surface reconstruction is indeed an important factor since the high rate of CO oxidation is found along the boundary between different surface structures in the temperature–pressure phase diagram (Reuter *et al.* 2004). Honkala and colleagues included adsorbate–adsorbate interactions in a microkinetic model for the ammonia synthesis over Ru catalysts by calculating the energetics of the rate-limiting activation of N<sub>2</sub> for all possible ensembles of adsorbates at the nearest-neighbor positions (Honkala *et al.* 2005). In both of these studies all parameters entering the kinetic Monte Carlo simulations were determined from first-principles calculations and in both cases the macroscopic reaction rates were found to be in good agreement with the experimentally measured reaction rates. These findings prove that theory has reached a level of accuracy where it can describe the macroscopic kinetics of the working catalyst. This underlines the important role of theory in the understanding of heterogeneous catalysis.

## 12.3 Single-crystal surfaces

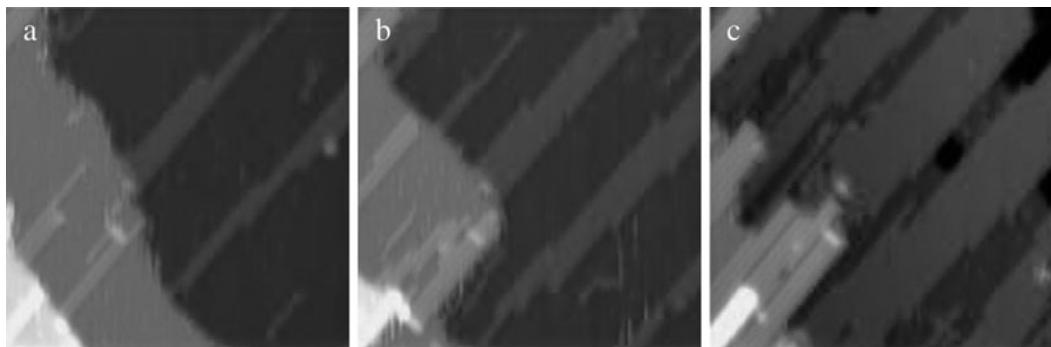
Within the surface-science approach the simplest possible scheme to model a heterogeneous catalyst is to use a single-crystal surface of the active catalyst material as the model system. Despite the great simplicity of this approach, it has proven to be extremely successful in describing the fundamental surface properties, the rate-limiting steps, and the active sites relevant to a catalytic reaction occurring on a high-surface-area technical catalyst. In certain cases, one of the most notable examples being the ammonia production over a Fe catalyst, it has been possible to describe the reaction on the macroscopic catalyst in great detail solely on the basis of microscopic fundamental studies (Stoltze and Nørskov 1985). In this connection it is worth mentioning that the 2007 Nobel Prize in chemistry was awarded to Prof. Gerhard Ertl not least for his pioneering studies of ammonia synthesis, where he very elegantly exploited the capability of the surface-science approach, i.e. studied model systems consisting of either flat single-crystal surfaces or well-defined nanoclusters on surfaces under clean and well-controlled, often ultrahigh-vacuum (UHV), conditions and thereby revealed the microscopic processes underlying this extremely important ammonia process (Ertl 2008).

The amount of work carried out on single-crystal model catalysts is overwhelming, and it is beyond the scope of this chapter to make a comprehensive overview of this entire exciting surface-science research field. Rather, we will, by discussing a few examples, illustrate how detailed atomic-scale insight into elementary surface processes relevant to catalysis can be obtained mainly by means of high-resolution STM studies on single-crystal surfaces.

### 12.3.1 Surface structure

Since the pioneering studies of Langmuir a simplified approach has been to consider the surface of a catalyst as a static “checkerboard” of adsorption sites formed by a rigid, undistorted substrate lattice onto which gas-phase molecules can adsorb, dissociate, react and form reactants that desorb without affecting the substrate template (Zangwill 1988). Modern surface science has, however, shown that this static view of the lattice represents a highly idealized picture that is the exception rather than the rule for many adsorbates. Instead, the surface has to be considered as a dynamic medium, the structure of which changes in response to a changing chemical environment induced by the adsorbates (Besenbacher *et al.* 1994).

As an example of an adsorbate-induced reconstruction of a metal surface we will briefly discuss the  $(2 \times 1)$  surface reconstruction of Cu(110) induced by dissociative adsorption of O<sub>2</sub> as initially revealed by the pioneering LEED study by Ertl in 1967 (Ertl 1967). Despite numerous experimental studies, the exact structure of the  $(2 \times 1)$  reconstruction and how this reconstruction of copper nucleates and grows was not settled until 1990 when Coulman *et al.* (1990) and Jensen *et al.* (1990) established the now widely accepted “added row” structure model. In the “added row” model adsorbed O atoms combine with Cu ad-atoms to form low-coordinated [001]-directed –Cu-O- added rows nucleating on top of the Cu(110) terrace (Besenbacher *et al.* 1994). The nucleation and growth mode of this  $(2 \times 1)$  reconstruction can be visualized by time-lapsed dynamic STM movies, and in Fig. 12.5 a series of snapshots from such a movie is depicted. From the STM movie/images it is revealed how Cu atoms detach from the step edges and diffuse around the surface until they combine with an adsorbed O atom and nucleate into [001] directed –Cu-O-Cu- added rows that nucleate and grow from the step edges. The dynamics of the reconstruction is inherently linked to the final structure, and based on the unique insight provided by dynamic STM, the competing “buckled-row” structural model could be rejected (Kuk *et al.* 1990) because it does not involve the long-range mass transport directly revealed from the STM movies. Once completed and fully developed, the “added row” structure is



**Fig. 12.5** Snapshots from an STM movie ( $23.5 \times 25.6 \text{ nm}^2$ ) showing a Cu(110) surface during exposure to  $\sim 10^{-8}$  torr O<sub>2</sub> at room temperature. The removal of Cu atoms from step edges is accompanied by the nucleation of added rows in the [001] direction. Reproduced from Besenbacher *et al.* (1994) with kind permission from Springer Science + Business Media.

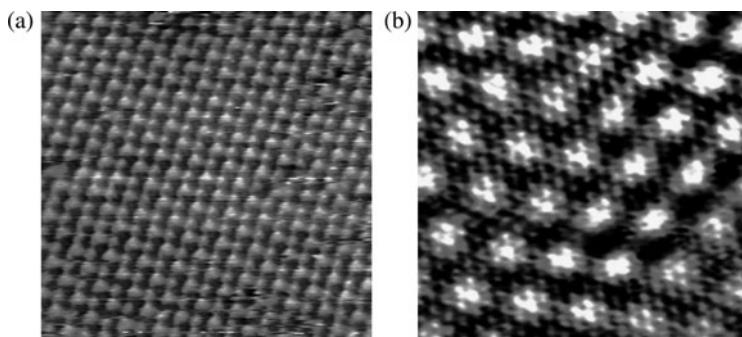
identical to a “missing row” structure, but in terms of mass transport these two models differ significantly. A “missing row” model would lead to the growth of the step edges due to Cu atoms squeezed out from the terrace as opposed to the “added row” model, where Cu is etched away and transported to the terrace.

This example shows the close linkage between the dynamic surface mass-transport process and the final surface reconstruction, and it also stresses the fact that surfaces in the presence of strongly interacting adsorbates in general cannot simply be modelled as a static checkerboard. Instead, models of adsorption and catalytic reactions must include the coupling to the substrate distortions as a pre-requisite for understanding the trends of catalytic activities.

### 12.3.2 Adsorption

Before any catalytic surface reaction can take place, reactants must adsorb on the surface of the catalyst, and as such adsorption is one of the most fundamental properties to study in relation to catalysis. In the following we will, as an illustrative example, discuss one of the most well-studied adsorbate-metal surface systems, i.e. CO on Pt(111). The great interest in this particular adsorption system is not least due to the importance of CO and Pt in numerous catalytic processes, such as CO oxidation in car catalysts, or the CO poisoning of Pt anode catalysts in low-temperature fuel cells. As for many other adsorbate systems the majority of the surface-science studies are performed at low pressure ( $<\sim 10^{-6}$  Torr), and indeed a very comprehensive and detailed understanding of the CO adsorption structures formed at low pressures exists (McEwen *et al.* 2003). However, with respect to the relevance for catalytic reactions it is important to know the adsorption structure under reaction conditions, i.e. at high pressure and high temperature.

The high-pressure (HP) adsorption structures of CO on Pt(111) have been studied by both spectroscopic (SFG) (Su *et al.* 1996; Rupprechter *et al.* 2001) and microscopic (STM) (Vestergaard *et al.* 2002; Davies *et al.* 2004; Longwitz *et al.* 2004) techniques. By means of atomically resolved HP STM, Longwitz *et al.* studied the adsorption structure of CO on Pt(111) at room temperature over the entire pressure range from  $10^{-6} - 720$  Torr (Longwitz *et al.* 2004). Figure 12.6 depicts both the well-known  $c(4 \times 2)$ -2CO structure formed at room temperature and low pressure and the Moiré-type superstructure formed



**Fig. 12.6** STM images ( $7.5 \times 7.5 \text{ nm}^2$ ) of CO adsorption structures. (a)  $c(4 \times 2)$  structure formed at  $10^{-7}$  Torr. (b) Moiré superstructure formed at 720 Torr. Reproduced from Davies *et al.* (2004) Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

at high CO pressure still at room temperature. The Moiré pattern is caused by the interference of a hexagonally close-packed CO adsorbate layer formed on top of the underlying hexagonally close-packed Pt(111) surface. Moiré superstructures with varying periodicity and direction relative to the substrate were observed in the entire pressure range from  $10^{-6}$  Torr up to 720 Torr, and the variation in the Moiré superstructure was found to be caused by a reversible expansion/contraction of the CO overlayer with increasing/decreasing CO pressure.

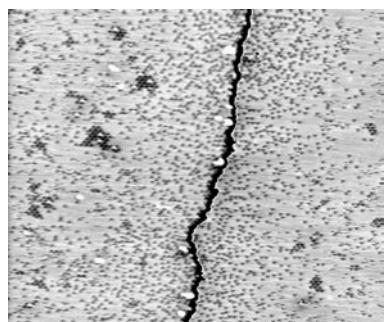
At low pressures, low temperatures and high CO coverage a Moiré superstructure identical to that observed at high pressure and room temperature was revealed (Vestergaard *et al.* 2002). For the CO/Pt(111) adsorption system it therefore seems that raising the pressure is equivalent to lowering the temperature as long as the thermodynamic equilibrium structure remains kinetically accessible, as has also been observed for other adsorbate systems (Österlund *et al.* 2001; Thostrup *et al.* 2003; Vang *et al.* 2005b). The reversible contraction and expansion of the CO overlayer can, however, only be observed at high pressure and high temperature where the adsorbed CO is in equilibrium with the gas phase. The finding that the adsorbed CO is in equilibrium with the gas phase at RT is indeed relevant to the understanding of anode catalysts used in low-temperature fuel cells. The anode catalysts typically consists of Pt, which is very effective for H<sub>2</sub> dissociation but prone to poisoning (site blocking) by trace amounts of CO in the H<sub>2</sub> feed gas, which are difficult to avoid since the majority of H<sub>2</sub> is produced from reforming of fossil fuels. It has been found that the addition of Ru to the Pt catalyst increases the CO tolerance of the catalyst considerably (Oetjen *et al.* 1996), and two mechanisms have been proposed to be responsible for this effect. In the *bifunctional mechanism* the Ru atoms act as active sites for CO oxidation and thereby remove the CO molecules from the surface (Gasteiger *et al.* 1994). The *ligand effect*, on the other hand, ascribes the increased CO tolerance to a change in the electronic structure of Pt, which weakens the interaction with CO and reduces the CO coverage on the catalyst surface (Igarashi *et al.* 2001). It has been found both experimentally (Buatier de Mongeot *et al.* 1998) and theoretically (Christoffersen *et al.* 2001) that Pt overlayers on a Ru(0001) substrate indeed have a lower CO binding energy than pure Pt. This will, however, only have an effect on the CO coverage if the adsorbed CO is in equilibrium with the gas phase under the conditions of the fuel cell. The Co-Pt(111) results presented above show that the adsorbed CO is indeed in equilibrium with the gas phase at room temperature and therefore also at higher temperatures, which gives strong support to the ligand effect: Of course the bifunctional mechanism cannot not be discarded, and it is indeed very likely that the enhanced CO tolerance of Pt-Ru is due to a combination of the two effects.

### 12.3.3 Dissociation and diffusion

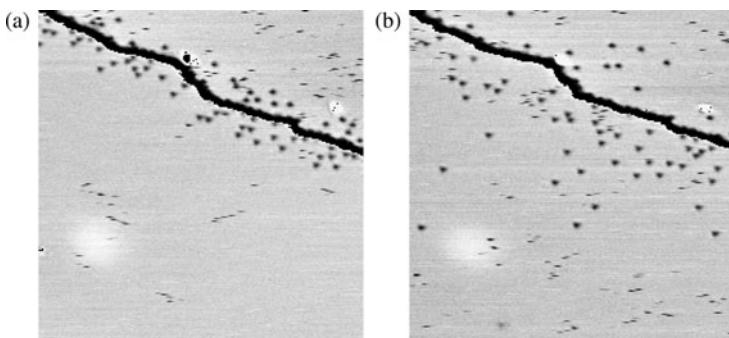
A surface reaction is inherently a dynamic process involving diffusion and/or dissociation of atoms and molecules adsorbed on the surfaces. An example

illustrating both these important phenomena at the atomic scale is the dissociative adsorption of NO molecules and the subsequent diffusion of N atoms on the Ru(0001) surface (Zambelli *et al.* 1996a). Figure 12.7 shows an STM image of the Ru(0001) surface after the surface is exposed to 0.3 Langmuir NO at 315 K. From similar adsorption experiments with N and O adsorbates, respectively, it could be concluded that the dark spots close to the step edge can be identified as N atoms. From the measured diffusion profile of the N atoms it can be concluded that the NO molecules dissociate at the step edge (the diffusion of the O atoms happens so fast that O islands have already been formed away from the step edge at the time of the STM image (Wintterlin *et al.* 1997a)). Such STM experiments provide not only direct evidence of the dissociative adsorption of NO molecules, they, furthermore, directly reveal that the step edges on the Ru(0001) surface act as the active sites for the dissociation of NO molecules. The concept of special active sites on the catalyst surface dates back to the pioneering work of Taylor from 1925 and the STM studies of NO dissociation on Ru(0001) represents the first direct atomic-scale identification of the active site for a surface-catalyzed reaction (Zambelli *et al.* 1996b).

The fact that NO molecules dissociate exclusively at the step edges of Ru(0001) was later exploited by Zambelli *et al.* in a series of very elegant studies to derive the barrier for N diffusion on Ru(0001) both from a (microscopic) profile analysis and from the tracking of individual single jump events for N adatom diffusion on the Ru(0001) surface (Zambelli *et al.* 1996a). The dissociation of NO at the step edges creates a highly localized initial coverage of N ad-atoms at the step edges, and the subsequent temporal evolution of the N concentration profile could be monitored. In Fig. 12.8 the Ru(0001) surface is shown 6 min (a) and 2 h (b) after initial NO adsorption at room temperature, and it is clearly observed that in time the N atoms diffuse away from the step edges. From the analysis of the temporal broadening of the concentration profile at variable temperatures, the activation energy for N diffusion was determined from a simple Arrhenius analysis to be  $E_a = 0.94$  eV. In another series of STM experiments the diffusion of individual N atoms led to an almost identical diffusion constant (Zambelli *et al.* 1996a). Normally, the diffusion barriers determined from a temporal analysis of an initially localized profile do not show such good agreement with the diffusion barriers measured by directly monitoring the single jumping events of the individual adsorbates,



**Fig. 12.7** STM image ( $38 \times 33 \text{ nm}^2$ ) of a Ru(0001) surface after the dissociative adsorption of 0.3 Langmuir NO at 315 K. The dark spots near the step edge are individual N atoms, and the small islands made up of dark spots are O islands. Reproduced from Zambelli *et al.* (1996b) reprinted with permission from AAAS.



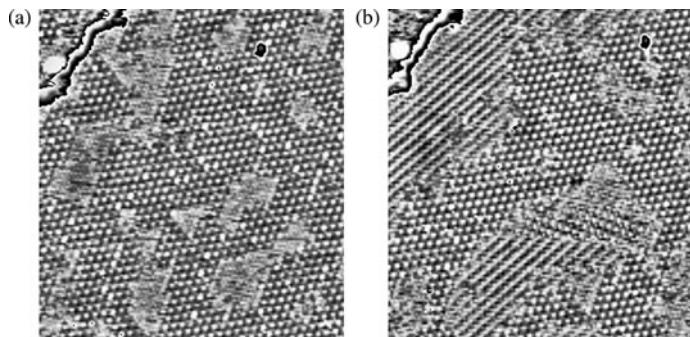
**Fig. 12.8** A Ru(0001) surface imaged by STM ( $18 \times 20 \text{ nm}^2$ ) 6 min (a) and 2 h (b) after adsorption of 0.1 Langmuir NO. The large dark depressions are single N atoms. The streaky dark features are O atoms, which diffuse on the same timescale as the line-scanning rate. Reproduced with permission from Zambelli *et al.* (1996a), copyright 1996 by the American Physical Society.

due to the fact that a profile analysis is usually carried out at a macroscopic level where steps and other defects cannot be neglected. Temporally resolved atomic-scale studies are thus important for the determination of diffusion barriers that are in turn essential to the microscopic description of the surface reactivity.

### 12.3.4 Surface chemical reactions

The reaction between adsorbed species is, needless to say, of utmost fundamental importance in surface reactions proceeding according to the Langmuir–Hinshelwood mechanism, in which it is assumed that all species are adsorbed and accommodated in thermal equilibrium with the surface before they take part in any reactions. In general it is, however, extremely difficult to observe directly a chemical reaction with the sufficient high spatial and temporal resolution.

In a few cases it is possible to observe and quantify the rate-limiting reaction step at the atomic scale in an indirect manner by observing the temporal evolution of the adsorbates. For the CO oxidation reaction on Pt(111) Wintterlin *et al.* performed titration experiments in which an O pre-covered Pt(111) surface was exposed to CO, and the reaction was monitored directly by STM (Wintterlin *et al.* 1997b). Figure 12.9(a) shows an image of such an O pre-covered Pt(111) surface after exposure to  $5 \times 10^{-8}$  mbar CO for 140 s at 247 K. At this point no reaction has taken place, and the O coverage is identical to the initial O coverage before CO exposure. The adsorbed CO has, however, led the adsorbed O to form large  $(2 \times 2)$  islands due to the CO–O repulsion (Campbell *et al.* 1980). After an additional exposure of 140 s the areas of the  $(2 \times 2)$  O islands start to shrink, and simultaneously, islands with a c( $4 \times 2$ ) structure, characteristic for CO on Pt(111), start to nucleate and grow on the Pt(111) surface (Fig. 12.9(b)). This reaction was interpreted as CO reacting with the pre-adsorbed O to form  $\text{CO}_2$ , which desorbs immediately, leaving behind vacant adsorption sites on the Pt(111) surface for CO molecules to adsorb on from the gas phase. In this way it was thus possible to follow both reactants on the surface during the reaction, and it is clearly revealed that the reactants are not randomly distributed as assumed in the Langmuir–Hinshelwood model, but rather two different phases exist on the surface, and



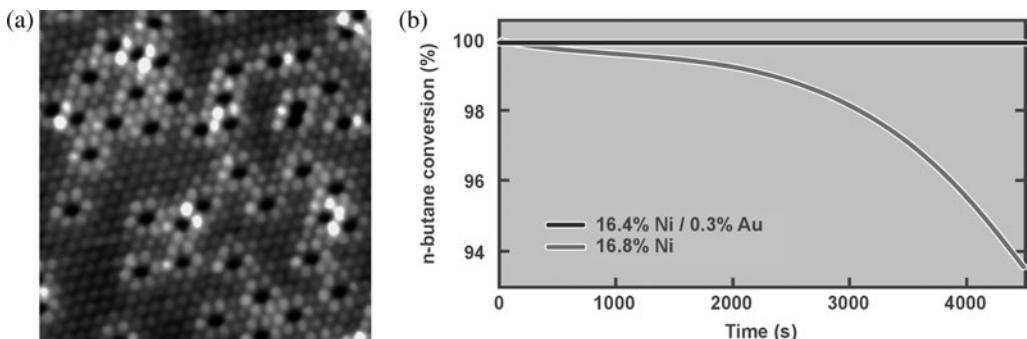
**Fig. 12.9** O pre-covered Pt(111) surface during reaction with CO at 247 K imaged by STM ( $18 \times 17 \text{ nm}^2$ ) after 140 s (a) and 600 s (b) exposure to  $5 \times 10^{-8}$  mbar CO. The  $(2 \times 2)$  O structure is decreasing, and the c( $4 \times 2$ ) CO structure is growing with time. Reproduced from Wintterlin *et al.* (1997b) reprinted with permission from AAAS.

the reaction only occurs at the boundary between these two phases. This point was further quantified by obtaining the microscopic reaction rate from the sequence of STM images, and indeed this reaction rate was found to be proportional to the edge length along the phase boundary, rather than to the product of the two reactant surface coverages ( $\Theta_{\text{CO}} \cdot \Theta_{\text{O}}$ ) as one would have expected from a Langmuir–Hinshelwood model. Based on STM experiments carried out at different temperatures, the authors were also able to determine the activation energy of the reaction to be 0.49 eV, in excellent agreement with a previous molecular beam study that reported a value of 0.51 eV (Campbell *et al.* 1980). The results of Winterlin *et al.* thus serve as an excellent example that a macroscopic quantity can be rationalized in terms of a detailed atomic-scale understanding.

## 12.4 Changing the reactivity at the atomic scale: Design of new catalysts from first principles

One of the Holy Grails of the surface-science approach to catalysis is to reach a point where the fundamental insight gained from studies into surface structure and reactivity on well-controlled model systems can be applied directly to the design of new catalysts with tailored properties from first principles, or phrased differently to advance from a purely descriptive to a predictive scientific approach. Given the complexity of a catalytic reaction and of the structure of a catalyst this is indeed an extremely ambitious target that is yet far from being reached in general. The advancement of experimental and theoretical tools and the hereby gained understanding of surface-science processes in general have, however, progressed considerably within the last decade, and examples where the fundamental insight gained from model studies has been used to guide the design of new improved catalysts are indeed appearing. In this section we will discuss a few such examples.

Besenbacher *et al.* used a combination of STM experiments with DFT calculations to design a Au/Ni alloy catalyst for the steam reforming reaction ( $\text{CH}_4 + \text{H}_2\text{O} \rightarrow 3 \text{ H}_2 + \text{CO}$ ) (Besenbacher *et al.* 1998). The steam reforming of natural gas (mainly methane,  $\text{CH}_4$ ) is of utmost importance for hydrogen production, and the catalyst most widely used has Ni nanoclusters as the active material (Rostrup-Nielsen and Rostrup-Nielsen 2002; Rostrup-Nielsen *et al.* 2002). One of the main shortcomings with the use of Ni catalysts for steam reforming is that Ni also catalyzes the formation of carbon and subsequently the growth of carbon filaments (to be discussed below), which eventually may lead to the breakdown of the catalyst and an expensive reactor shutdown to replace the worn-out catalyst. Besenbacher *et al.* used STM to show that Au and Ni form stable 2D surface alloys despite the fact that the two metals are bulk-immiscible (Nielsen *et al.* 1993). When Au is deposited on a Ni (111) surface, the Au atoms squeeze out Ni atoms and are substituted into Ni atom lattice positions, and thereby a surface alloy is formed, as depicted from the atom-resolved STM image in Fig. 12.10(a). Although the Au atoms are larger than the Ni atoms and therefore protrude geometrically from the surface, the Au atoms are imaged as depressions in the STM images. This reflects the

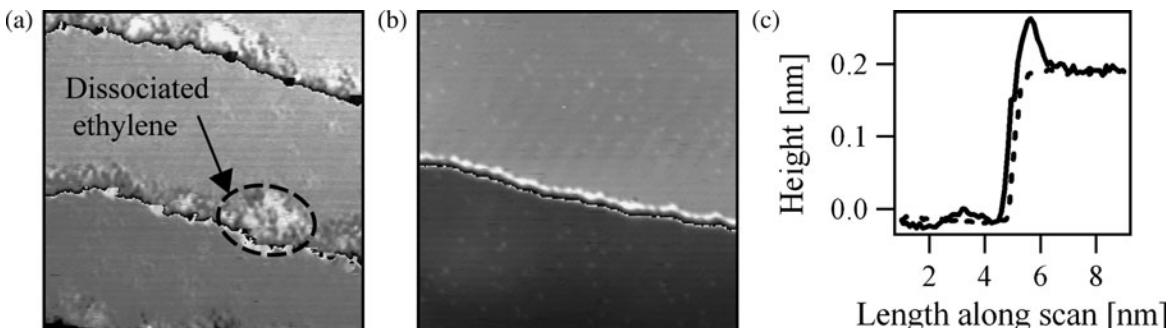


**Fig. 12.10** (a) STM image of Au/Ni(111) surface alloy with ~7% Au. The Au atoms are imaged as dark depressions, and the neighboring Ni atoms appear brighter in the STM image. (b) Conversion of n-butane as a function of time for a Au/Ni alloy catalyst and a pure Ni catalyst. Reproduced from Besenbacher *et al.* (1998) with permission from AAAS.

fact that STM images in general represent a rather complicated convolution of the geometric and electronic structure of the surface and from a stricter theoretical point of view should be interpreted as contours of constant local density of states at the Fermi level at the position of the tip rather than as simple topographic maps (Tersoff and Hamann 1983; 1985).

In the STM images, the Ni atoms with a neighboring Au atom are imaged brighter than the Ni surface atoms with only Ni neighbors, and this effect is even more pronounced for those Ni atoms with two Au neighbors. DFT calculations confirmed that this effect cannot be explained by an outwards relaxation of the Ni atoms, and it was concluded that the electronic structure was significantly perturbed for those Ni atoms that had neighboring Au atoms alloyed into the Ni surface. The DFT calculations further revealed that the tendency of the surface to bind carbon and form graphite is strongly impeded by the presence of the Au atoms alloyed into the topmost layer of the Ni (111) surface. The Ni atoms with Au nearest neighbors are shown to have a somewhat higher barrier for activation of hydrocarbon molecules, such as CH<sub>4</sub>, and the overall effect of the Au atoms is thus to increase the selectivity at the expense of a slightly reduced activity.

Motivated by these fundamental surface-science findings a high surface area, MgAl<sub>2</sub>O<sub>4</sub>-supported Ni catalyst (with 16.5 wt.% Ni) modified with 0.3 wt.% Au was synthesized (Besenbacher *et al.* 1998). By means of XAFS, it was verified that Au atoms are indeed alloyed into the first layer of the Ni nanoclusters of the catalyst. The activity of the steam reforming reaction of n-butane (see Fig. 12.10(b)) on this high surface area Au-Ni catalyst was compared to the activity of the pure Ni catalyst. Steam reforming of n-butane was used as a model reaction to test the steam reforming activity since n-butane is known to give rise to the most severe graphite-formation problems. As depicted in Fig. 12.10(b) the conventional Ni catalyst is deactivated rapidly, whereas the conversion factor for the Au-Ni catalyst remains almost constant, implying that for the new Au-Ni surface alloy catalyst, the graphite formation is significantly reduced. These results show that by combining several experimental surface-science techniques with theory and insight into synthesis and *in-situ* characterization of high surface area catalysts, it has been possible to



**Fig. 12.11** (a) STM image ( $20 \times 20 \text{ nm}^2$ ) of a Ni(111) surface after exposure to 1 Langmuir ethylene at room temperature. Reaction products have nucleated along the upper step edges. (b) STM image ( $40 \times 40 \text{ nm}^2$ ) of a Ni(111) surface after evaporation of Ag and postannealing at 800 K. Ag atoms are decorating the step edge. (c) Line scans perpendicular to a step edge on pure Ni(111) (dotted line) and on a Ni(111) surface with Ag at the step edges (full line). Reproduced by permission from Macmillan Publishers Ltd: (Vang *et al.* 2005a), copyright 2005.

convert the fundamental understanding of the atomic processes involved in the catalytic reaction into the design of an improved catalyst for the steam reforming reaction.

The reactivity of catalytic surfaces is often dominated by very reactive low-coordinated atoms, such as step edge sites (Dahl *et al.* 1999; Hammer 1999; Gambardella *et al.* 2001; Ciobica and Santen 2003; Liu and Hu 2003), as it was, e.g., illustrated for the dissociation of NO on Ru(0001) (Zambelli *et al.* 1996b). These very reactive sites on the catalyst surface may also be decisive for the *selectivity* of a given catalytic reaction, which is an equally important parameter as the *activity*. Vang *et al.* demonstrated how the very reactive step edges on a Ni surface can be selectively blocked, whereby the selectivity is changed for certain reactions (Vang *et al.* 2005a). The authors used STM to show that ethylene ( $\text{C}_2\text{H}_4$ ) decomposes with a much higher rate at the step edges on Ni(111) as compared to the higher-coordinated terrace sites. The STM image in Fig. 12.11(a) shows the Ni(111) surface after exposure to 1 Langmuir ethylene at room temperature, where a brim of hydrocarbon reaction products has formed along the step edges. From an interplay with DFT calculations this finding was confirmed, and it was, furthermore, found that this step-edge effect is considerably more pronounced for the initial C–C bond breaking than for the initial C–H bond breaking, and thus steps play an important role in the bond-breaking selectivity of ethylene on Ni(111). Vang *et al.* also demonstrated how the number of reactive step sites can be controlled on the atomic scale by blocking the steps with Ag. The selective blocking of the step edge sites was achieved by the evaporation of Ag and subsequent annealing to 800 K. This surface preparation leads to a Ni(111) surface with Ag atoms decorating all the step edges, as depicted in Fig. 12.11(b). Exposure of the Ag/Ni(111) surface to ethylene at room temperature revealed that the Ag atoms indeed lowered the reactivity of the step edges since ethylene did not decompose at the Ag-decorated step edges (or anywhere else on the surface).

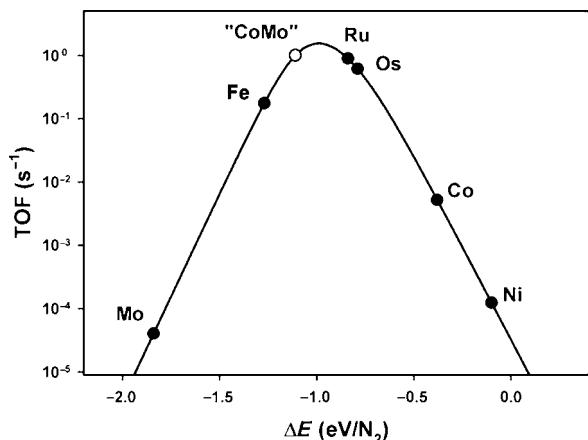
To bridge the gap between such fundamental STM and DFT surface-science studies and applied catalysis, a high surface area oxide-supported Ag–Ni catalyst was synthesized and tested for the hydrogenolysis of ethane

( $\text{C}_2\text{H}_6 + \text{H}_2 \rightarrow 2\text{CH}_4$ ), which is the simplest possible reaction used to probe the activity for C–C bond breaking. The rate constant for ethane hydrogenolysis was found to be approximately an order of magnitude lower for the Ag–Ni catalyst than for the conventional Ni catalyst. The rate did not drop to zero because not all step edges were covered with Ag on the highly dispersed oxide-supported catalyst.

These studies have thus demonstrated how the insight gained through fundamental surface-science studies has formed the basis for the nanoscale design of a new high surface area catalyst with improved bond-breaking selectivity between dissociation and primary dehydrogenation. The studies also show that one cannot, in general, simply view the step atoms as more reactive sites but also has to consider the relative increase in reactivity between different reaction pathways. The step atoms may thus, to a high degree, control the selectivity in a given catalytic process. Such knowledge gained from fundamental surface-science experiments could become a powerful tool in the nanoscale design of new and improved catalysts.

Another example of catalyst design from first principles was published by Jacobsen and colleagues, who used a combination of *ab initio* DFT calculations, detailed experiments on model systems, and synthesis and *in-situ* characterization of nanostructured catalysts to design a new improved catalyst for the ammonia synthesis (Jacobsen *et al.* 2001). By studying the trends in ammonia activity versus the nitrogen binding energy for certain elements in the periodic system, a typical volcano curve was revealed. This volcano plot establishes that the nitrogen adsorption energy is a good descriptor for the activity of a metal for ammonia synthesis. The optimal nitrogen binding energy is found to lie in a “gap” between Fe and Ru, i.e. none of the pure elements have the optimal binding energy (Logadottir *et al.* 2001), and hence new possibilities seem to exist if an alloy is formed with a nitrogen adsorption energy closer to the optimum value. If the adsorption properties of an alloy are somehow an average of those of the constituent elements, one should be able to identify alloys with the optimum N adsorption energy by alloying metals to the left with those to the right of the optimum in Fig. 12.12. This approach was used to guide the design of a new “CoMo” alloy catalyst for ammonia synthesis consisting of Mo with too high a binding energy, and Co with too low a binding energy. An alloy catalyst with the composition of  $\text{Co}_3\text{Mo}_3\text{N}$  was synthesized. This alloy catalyst indeed turned out to have a higher activity than that of a state-of-the-art promoted ruthenium and iron catalyst (Jacobsen *et al.* 2001). It was also established that the adsorption properties of the CoMo alloy are indeed intermediate between those of pure Co and Mo.

Recently, Studt *et al.* used a similar approach where they found the binding energy of methyl to be a good descriptor for the selectivity for the hydrogenation of acetylene in a background of ethylene (Studt *et al.* 2008). Rather than simply choosing two metals with too low and too high binding energy, respectively, the authors performed an extensive theoretical screening of a large series of metals and binary alloys. Besides the presently used Ag–Pd catalysts this analysis identified several different promising catalysts. Given the fact that the price of the precious metals is a severe drawback with the Ag–Pd catalyst, the authors chose the cheapest of the identified candidates, a NiZn



**Fig. 12.12** Calculated turnover frequencies for ammonia synthesis as a function of the adsorption energy of nitrogen. None of the pure elements have the optimum N binding energy. Reproduced with permission from Jacobsen *et al.* (2001). Copyright 2001 Am. Chem. Soc.

alloy, for further experimental testing. A high surface area Ni-Zn alloy catalyst with varying Zn content on  $MgAl_2O_4$  spinel supports was synthesized, and the experimental results confirmed the high selectivity predicted by DFT. The catalyst with highest Zn content was even better than the best of the Ag-Pd catalysts also tested in this series of experiments.

The above examples demonstrate that we are approaching a new era in catalysis research where the development of heterogeneous catalysts is no longer a “black art” based on “trial-and-error” methods. The emerging molecular level understanding of catalysis is the result of a rapid development in experimental fundamental surface-science methods and theoretical first-principles calculations over that last decade or two. However, it should be stressed that we are still at the beginning of this development, and the examples to date all rely on either modifications of known catalysts as the AuNi and AgNi catalysts or reactions where simple one-parameter models can be constructed such as the ammonia synthesis and selective hydrogenation of acetylene.

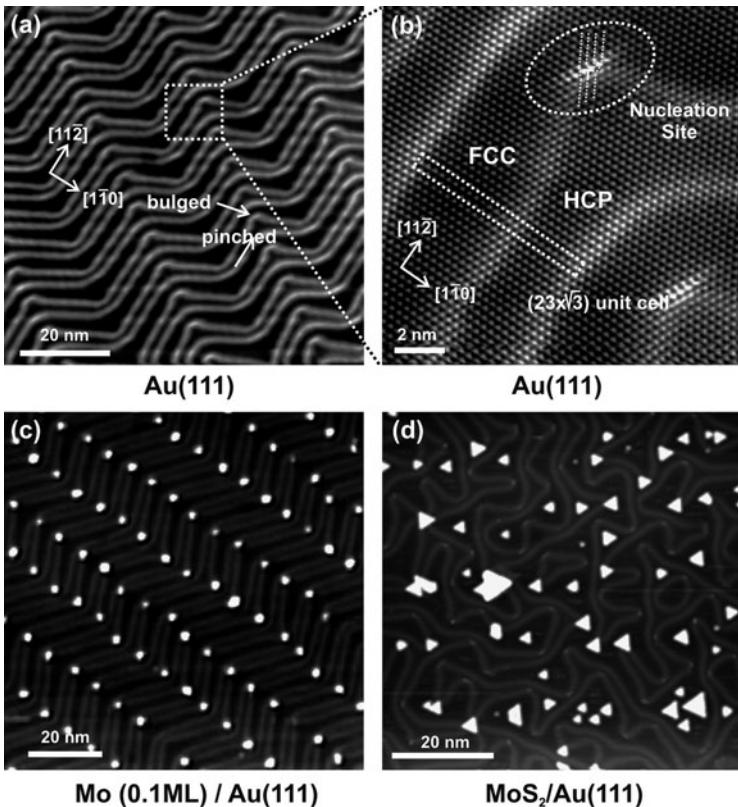
## 12.5 Nanoparticles

The use of single-crystal surfaces as model catalysts is undisputedly extremely useful for many catalytic processes. However, examples also exist where this is not the case, i.e. a single-crystal surface of the active material is a very poor model for the real catalyst, since the catalytic reactivity appears to be strongly correlated to the detailed nanoscale structure of the catalyst particles. In such cases, surface-science studies have to be carried out on more elaborate model systems consisting, for example, of nanoparticles deposited onto a substrate or on high surface area supported nanoclusters, i.e. real catalysts. In this section we discuss two examples where nanoparticles are used as model catalysts. The first example is the  $MoS_2$ -based hydrodesulphurization catalyst used for the pre-treatment of crude oil, and the second example is catalysis by Au nanoparticles.

### 12.5.1 Hydrodesulfurization of MoS<sub>2</sub>-based model catalysts

Hydrotreating catalysts are used worldwide in oil refineries for upgrading and purifying crude oil fractions by hydrogenation reactions (HYD) and removing sulfur (hydrodesulfurization, HDS) or nitrogen (hydrodenitrogenation, HDN). In addition to the motivation to limit the emission of harmful gaseous sulfur and nitrogen species from fuels in a well-controlled way, there is also the important technological aspect that even trace levels of sulfur will poison other catalysts (e.g. Ni, Cu, Pd or Pt) which are used in other catalytic processes (hydrocracking, reforming, automotive catalysis, fuel cells, etc.). Furthermore, there is in general a huge and urgent demand for ultralow-sulfur fuels and a better exploitation of heavy-fuel reserves. Consequently, better hydrodesulfurization processes and catalysts are of utmost importance, but in the past progress has been hampered by a lack of fundamental insight into the atomic-scale structure and exact working principle of the catalyst. The most common hydrotreating catalyst consists typically of a few-nanometer wide single-layer MoS<sub>2</sub> nanoparticles mainly promoted with Co or Ni and supported on a porous alumina carrier (Prins *et al.* 1989). For many years a considerable effort has been devoted to relate, e.g. the HDS activity and selectivity to microscopic properties such as catalyst composition, geometric and electronic structure of the MoS<sub>2</sub> nanoclusters (Topsøe *et al.* 1996; Prins 1997; Gosselink 1998; Whitehurst *et al.* 1998; Kabe *et al.* 1999). Molybdenite (2H-MoS<sub>2</sub>) is a layered compound consisting of S-Mo-S layers held together by weak van der Waals forces. Despite the impressive number of studies of HDS catalysts, a series of fundamental questions has remained unanswered. Much of the controversy is directly related to the fact that the traditional spectroscopy-based techniques are not able to unequivocally map the real-space atomic structure of the few-nanometer wide CoMoS nanoclusters. It was recognized early that only very weak physisorption occurs on the (0001) sulfur-terminated surface of MoS<sub>2</sub>. Thus, the basal plane of MoS<sub>2</sub> crystallites is very inactive in catalytic reactions and mainly coordinatively unsaturated sites (CUS) associated with sulfur vacancies (or Mo<sup>δ+</sup> centres) on the edges of the S-Mo-S layers in MoS<sub>2</sub> nanoclusters have been considered to be catalytically active (Topsøe *et al.* 1996). By the use of probe molecules such as CO, NO or O<sub>2</sub> it has been possible to quantify their numbers and correlate them with the catalytic activity (Shuxian *et al.* 1986; Topsøe and Topsøe 1986; Topsøe *et al.* 1996). However, in order to pinpoint and understand more precisely the active sites, it is necessary to obtain more detailed information about the atomic-scale structure of MoS<sub>2</sub> nanoclusters. For instance, what is the preferential shape of the MoS<sub>2</sub> nanoclusters, what are the active edge structures, and what is the role of the promoters in CoMoS?

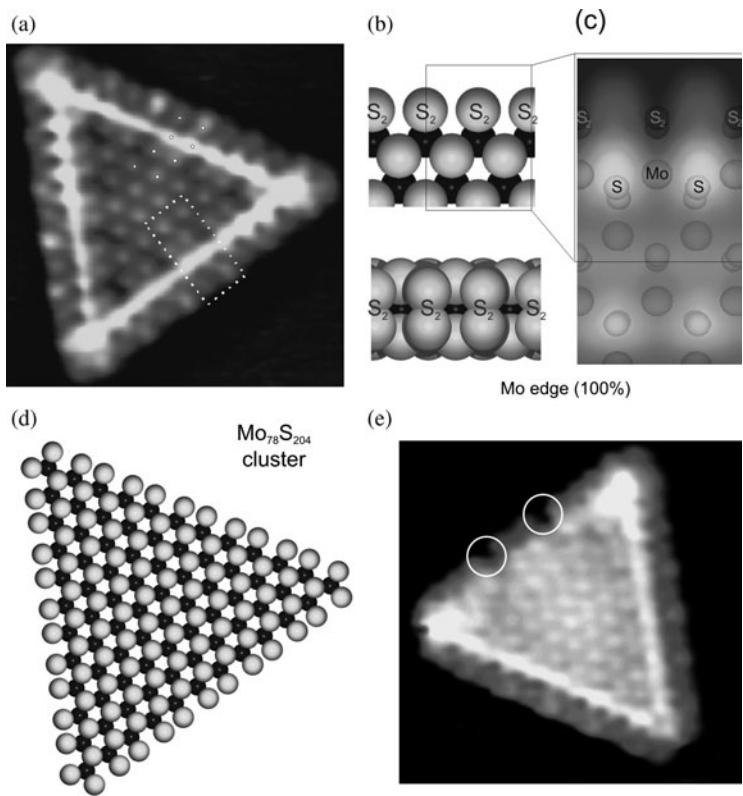
Recently, using the surface-science approach it was demonstrated that high-resolution STM was capable of addressing the atomic-scale structure of a HDS model system consisting of MoS<sub>2</sub> nanoclusters grown on a single-crystal gold substrate (Helveg *et al.* 2000; Lauritsen *et al.* 2001) and thus for the first time many of the questions discussed above with respect to the catalytic properties of these nanoclusters were successfully elucidated



**Fig. 12.13** (a) STM image of clean Au(111) surface ( $80 \times 80 \text{ nm}^2$ ). (b) Atom-resolved STM image showing the atomic structure of the herringbone reconstruction on Au(111) ( $14 \times 14 \text{ nm}^2$ ). (c) Mo nanoclusters ( $\sim 0.1 \text{ ML}$ ) on Au(111) ( $100 \times 100 \text{ nm}^2$ ). (d) MoS<sub>2</sub> nanoclusters on Au(111) ( $75 \times 75 \text{ nm}^2$ ). Reproduced from Lauritsen and Besenbacher (2006), copyright 2006, with permission from Elsevier.

(Helveg *et al.* 2000; Lauritsen *et al.* 2004a, 2007; Lauritsen and Besenbacher 2006). The Au(111) surface was selected as a model substrate for the synthesis of MoS<sub>2</sub> nanoclusters since clean Au(111) exhibits the so-called herringbone reconstruction with a periodic array of surface dislocations (the elbows in the STM image Fig. 12.13(a) and the atom-resolved STM image in Fig. 12.13(b)) (Barth *et al.* 1990). This characteristic template pattern is known to have a high affinity for nucleation and growth of highly dispersed metallic (Mo) nanoclusters, as depicted in Fig. 12.13(c) (Helveg *et al.* 2000). In a subsequent step, these Mo nanoclusters were sulfided at  $10^{-6}$  mbar H<sub>2</sub>S pressure and 673 K, which results in the formation of a well-dispersed ensemble of single-layer MoS<sub>2</sub> nanoclusters oriented with the (0001) basal plane in parallel with the Au substrate (Fig. 12.13(d)). A similar approach has subsequently been used for synthesis of Co and Ni promoted metal-sulfide nanoclusters such as Co-Mo-S and Ni-Mo-S (Lauritsen *et al.* 2007).

Under sulfiding conditions, it was found from the STM images that the equilibrium morphology of the MoS<sub>2</sub> nanocluster was always triangular independent of the cluster size (Fig. 12.13(d)), and in Fig. 12.14(a) the atomic-scale structure of a typical MoS<sub>2</sub> nanocluster is depicted. The protrusions on the basal plane of the nanoclusters reflect the hexagonally arranged sulfur atoms in the topmost layer of MoS<sub>2</sub>. Surprisingly, a characteristic and pronounced bright *brim* was revealed to extend all the way around the edges of the

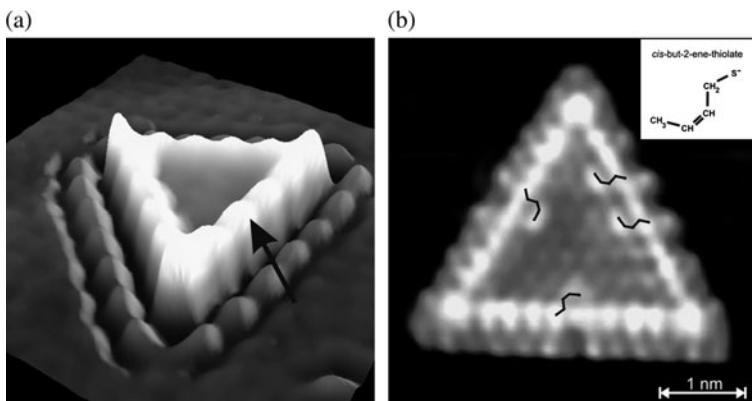


**Fig. 12.14** (a) Atom-resolved STM image of a single-layer MoS<sub>2</sub> nanocluster ( $4 \times 4 \text{ nm}^2$ ). (b) Ball model of the fully sulfided MoS<sub>2</sub>(1010) Mo edge terminating the clusters (Mo blue; S: yellow). (c) STM simulation of an edge section corresponding to the ball model. (d) Top-view ball model of the MoS<sub>2</sub> triangle. (e) Atom-resolved STM image of a triangular MoS<sub>2</sub> nanocluster showing the formation of S vacancies after exposure to pre-dissociated H (white circles). (a) and (e) reproduced with permission from Helveg *et al.* (2000), copyright 2000 by the American Physical Society. (c) reproduced with permission from Bollinger *et al.* (2001), copyright 2001 by the American Physical Society.

nanocluster, and the edge protrusions are imaged out of registry with the S protrusions at the basal plane (see superimposed dots). Both the bright brim and the apparent shifted registry of the edge protrusions visible in the STM images of the MoS<sub>2</sub> triangles in Fig. 12.14(a) can be traced back to a modified electronic structure at the edges, reflected in variations of the LDOS as recorded by the STM. From an interplay with DFT calculations and a detailed analysis of the electronic structure the existence of two distinct one-dimensional electronic *edge states* on the fully sulfided Mo edges was revealed (Byskov *et al.* 1999; Cristol *et al.* 2000; Raybaud *et al.* 2000b; Bollinger *et al.* 2001). The so-called (1010) Mo edges with a full coverage of S (forming S<sub>2</sub> dimers on the edge as shown in Fig. 12.14(b)) were found to be the energetically stable edge structure and a theoretically calculated STM image using the simple Tersoff–Hamann model was in excellent agreement with experimentally recorded STM images (Fig. 12.14(c)), reproducing the appearance of the metallic brim state and the shifted geometry of the protrusion on the edges, reflecting the enhanced LDOS in between the S<sub>2</sub> dimers.

From a simple coordination chemistry point-of-view, the fully sulfur-saturated Mo edges of the MoS<sub>2</sub> nanoclusters in Fig. 12.14(b) are normally not considered reactive. It was found that sulfur vacancies preferentially form at the nanocluster edges by exposing the clusters to atomic (pre-dissociated) hydrogen at a sample temperature of 673 K, and as illustrated in the STM

image in Fig. 12.14(e) (Lauritsen *et al.* 2003, 2004b). However, the distinct and characteristic metallic “brim” states on the MoS<sub>2</sub> nanocluster edges were observed to result in another hitherto unknown type of active sites with a reactivity that is different from the one expected from sulfur vacancies. The nature of these active brim sites on the cluster edges was revealed by selectively adsorbing thiophene (C<sub>4</sub>H<sub>4</sub>S) molecules in combination with hydrogen on the MoS<sub>2</sub> clusters (Lauritsen *et al.* 2003, 2004b). Exposing the fully sulfided nanoclusters to thiophene alone at room temperature we observed no adsorbed thiophene molecules, but by reducing the temperature of the sample during exposure and imaging at 200 K we observed how the metallic-like sites on top of the bright brim gradually became populated with thiophene molecules adsorbed with the aromatic-like ring system in parallel with the cluster plane (Fig. 12.15(a)). These thiophene molecules were only weakly physisorbed and found to desorb easily when the crystal was heated slightly. If, however, the MoS<sub>2</sub> nanoclusters were first exposed to atomic hydrogen and subsequently to thiophene at the elevated temperature of 500 K, other, significantly more strongly adsorbed thiophene-related species were revealed at the position of the bright metallic brim. Atom-resolved STM images (see, e.g. Fig. 12.15(b)) reveal several “bean-like” protrusions at positions adjacent to the edges, and these species were observed in time-lapsed STM movies to be very mobile on the cluster edges. From an interplay with DFT calculations and STM simulations of various possible adsorption configurations of thiophene and thiophene derivatives (Lauritsen *et al.* 2004b), the “bean-like” features were determined to be thiophene-related reaction intermediates, *cis*-but-2-ene-thiolates (C<sub>4</sub>H<sub>7</sub>S-), coordinated through the terminal sulfur atom to the metallic brim, resulting from a partial hydrogenation reaction occurring on the metallic brim states. The thiolates are formed by a sequential hydrogenation of one of the double bonds in thiophene by hydrogen adsorbed on the edges (from the S–H groups) followed by C–S bond cleavage. The hydrogen driving this reaction originates from pre-dissociated H atoms adsorbed on the terminal S atoms on the edges in the form of S–H groups (Lauritsen *et al.* 2004b). The combination of hydrogen atoms adsorbed on the edges and the unusual metallic brim sites at the edges on the MoS<sub>2</sub> nanoclusters presents a favorable active-site situation for a hydrogenation reaction. The observed species in

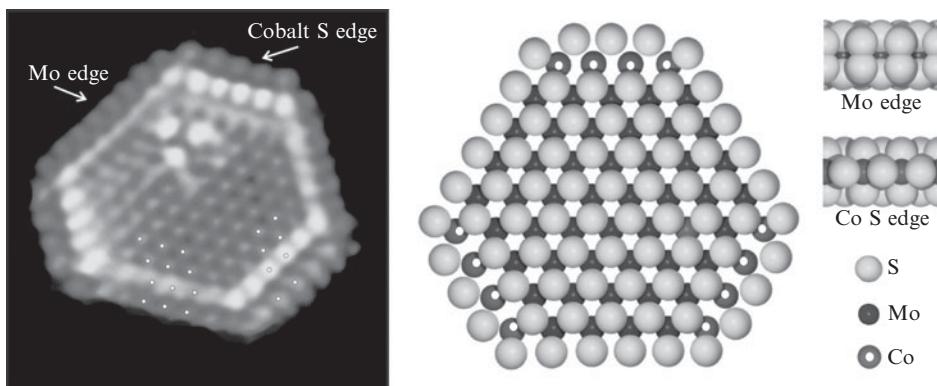


**Fig. 12.15** (a) 3D STM representation of thiophene molecules adsorbed on the metallic brim. Thiophene also adsorbs next to the cluster on the Au substrate. Reproduced from Lauritsen *et al.* (2004a), copyright 2004, with permission from Elsevier. (b) Atom-resolved STM image ( $5 \times 5 \text{ nm}^2$ ) of the MoS<sub>2</sub> nanocluster after exposure to atomic hydrogen and thiophene, revealing adsorbed *cis*-but-2-ene-thiolates (C<sub>4</sub>H<sub>7</sub>S-) species as determined in DFT-based simulations. Reproduced from Lauritsen *et al.* (2003) with permission from IOPP.

Fig. 12.15(b) resulting from the hydrogenation and C–S bond breaking are associated with a simple thiol in which the S is known to be much easier to remove, and the final extrusion of this S may thus proceed on sulfur vacancies. STM studies have showed that the metallic brim sites are not inhibited by H<sub>2</sub>S adsorption, which explains the retained hydrogenation activity of the MoS<sub>2</sub>-based catalyst even under highly sulfiding conditions (Logadottir *et al.* 2006). Finally, recent studies have indeed shown that these metallic sites are important also for hydrogenation of both large S-containing molecules and N-substituted aromatic compounds (Logadottir *et al.* 2006; Moses *et al.* 2007)

Interestingly and surprisingly, using the surface-science approach, the interplay between STM and DFT studies has thus revealed a new route for an initial activation of a relatively inert S-containing molecules such as thiophene, and these processes are found to take place on the active metallic brim edge states of the fully saturated Mo edges, which have the ability to accept or donate electrons and thus act as catalytic sites, just like ordinary reactive metal surfaces. Thus, the studies on HDS model catalysts have obtained the first atomic-scale view and revealed the quite surprising bifunctional nature of the active sites and their selectivity in HYD and HDS processes.

Cobalt is known to act as a promoter for the HDS catalysts, and from STM studies it was revealed that also the edges of promoted CoMoS nanoclusters possess metallic edge states. It was therefore proposed that metallic brim states also play a role for the catalytic properties of the promoted phase. Figure 12.16 shows an atomically resolved STM image of a single-layer CoMoS cluster formed by codeposition of Mo and Co onto the Au substrate during exposure to an H<sub>2</sub>S atmosphere and subsequent annealing. The main new finding is that the CoMoS nanoclusters adopt an hexagonal shape as opposed to the triangular morphology of unpromoted MoS<sub>2</sub>. This change in the equilibrium shape was attributed to the incorporation of cobalt into the MoS<sub>2</sub> structure, i.e. the formation of the CoMoS nanoclusters. The pre-dominant hexagonal morphology implies that both fundamental types of low-indexed edge



**Fig. 12.16** (a) Atom-resolved STM image of a CoMoS nanocluster. Size is  $5.1 \times 5.2 \text{ nm}^2$ . Notice the very intense brim associated with the Co-substituted S edge (shorter edges). (b) Ball model of the proposed CoMoS structure. The CoMoS cluster is shown in the top view exposing the unpromoted Mo edge and a Co-promoted S edge (Mo: dark, S: bright, Co: dark with spot). Also shown on the basal plane is a single Co inclusion. The Mo edge appears to be unaffected by Co and is shown in a side-view ball model. The Co-substituted S edge with a tetrahedral coordination of each edge Co is also shown. Reproduced from Lauritsen *et al.* (2006) with permission from Elsevier.

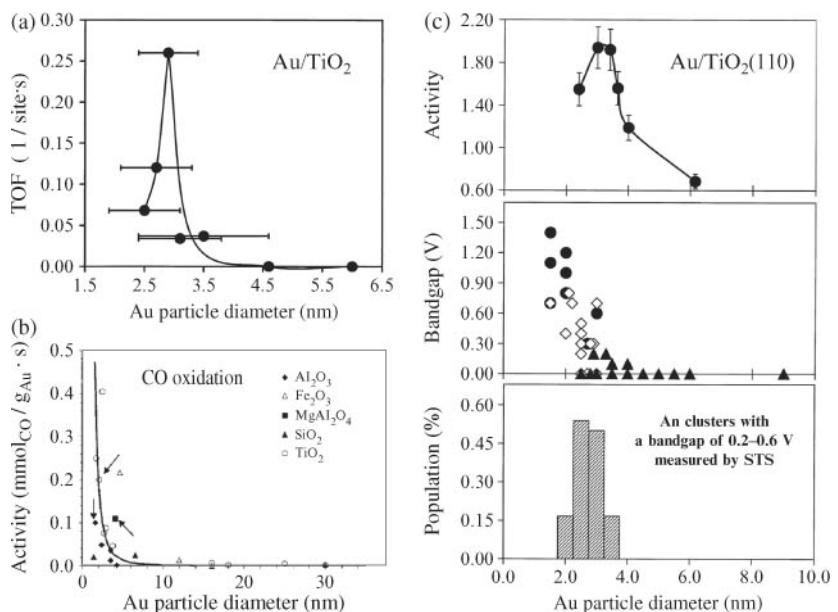
terminations of MoS<sub>2</sub> must be present, i.e. the Mo edge and the S edge. One edge type in the CoMoS was found to be similar to that observed for the MoS<sub>2</sub> triangles, with the edge protrusions clearly imaged out of registry with the lattice of S atoms on the basal plane and a bright brim along the edge. These edges were therefore identified as Mo edges, fully sulfided with two S dimers per Mo edge atom as in the triangles in Fig. 12.14(a). From the symmetry of MoS<sub>2</sub> nanoclusters the other, shorter edges must consequently be attributed to S-type edges. From the detailed atomic-scale information provided by the STM images, a structural model of the CoMoS nanoclusters in which Co atoms have substituted Mo atoms along the S edges of hexagonally truncated nanoclusters was determined. As depicted in the ball model in Fig. 12.16, a tetrahedral environment of the Co atoms is produced if the outermost protrusions are assumed to be S monomers, which agrees well with previously published XAFS results on supported CoMoS catalysts (Clausen *et al.* 1981; Niemann *et al.* 1990; Bouwens *et al.* 1991; Louwers and Prins 1992) and previous DFT studies (Byskov *et al.* 1999; Raybaud *et al.* 2000a). Interestingly, in the STM images the promoted edges are seen to exhibit an even brighter brim. This finding suggests that metallic brim states also exist in the promoted CoMoS structures, and in view of the result on the unpromoted clusters it was proposed that the promoted brim states may be connected to the higher catalytic activity of the Co-promoted phase.

Based on the new fundamental insight obtained from the surface-science studies on the model systems described above and in particular the fact that the existence of a reactive metallic site near the edges of the MoS<sub>2</sub> nanoclusters, it was recently possible for Haldor Topsøe Research Laboratories to synthesize and manufacture a novel generation of hydrotreating catalysts that optimize the number of brim sites and enhance the hydrogenation properties (Zeuthen and Skyum 2004; Hinnemann *et al.* 2005; Topsøe *et al.* 2005). Again, this shows that information gained in fundamental studies of idealized model systems can generate the knowledge required to develop new and better industrial catalysts. It is doubtful whether any experimental technique other than STM could unequivocally point to the presence of electronic states on the edge of highly dispersed MoS<sub>2</sub> nanoclusters of this size and, furthermore, directly identify their role in the adsorption of relevant molecules.

### 12.5.2 Novel catalytic properties of supported gold nanoclusters

Although bulk gold (Au) is known to be the noblest metal (Hammer and Nørskov 1995), Haruta and colleagues reported in the late 1980s that Au particles finely dispersed on high surface area oxide supports are catalytically very active, even at low temperature (Haruta *et al.* 1987). It is indeed so that in many cases the catalytic activity and selectivity of dispersed Au nanoparticles exceed those of the commonly used transition-metal catalysts such as Pt, Rh and Pd. The unique catalytic properties of Au nanoclusters have since spurred intense research activities both in the field of applied catalysis as well as within the more fundamental surface-science approach (Haruta *et al.* 1989; Bocuzzi and

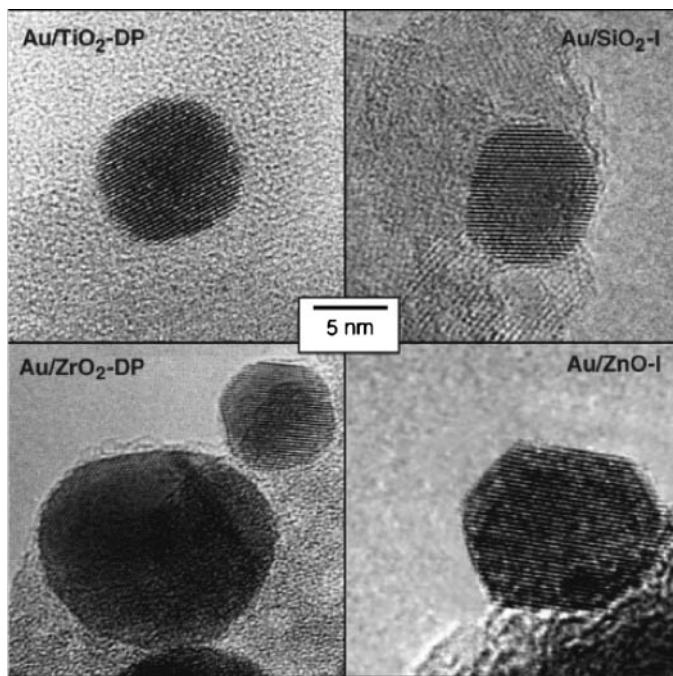
**Fig. 12.17** (a) CO oxidation turnover frequencies (TOFs) at 300 K as a function of the average size of the Au particles supported on a high surface area  $\text{TiO}_2$  support (Bamwenda *et al.* 1997). The  $\text{Au}/\text{TiO}_2$  catalysts were prepared by a deposition-precipitation method, and the average particle diameters were measured by transition electron microscopy (TEM). (b) Onset of high catalytic activity in CO oxidation in nanoparticles of gold (Lopez *et al.* 2004) (the three points marked by arrows are from the measurements of Lopez *et al.* (2004) and the remainder are from the work of several other investigators.) (c) Upper panel: The activity of CO oxidation at 350 K as a function of the Au particle size supported on  $\text{TiO}_2(110)$  assuming total dispersion of the Au. The CO:  $\text{O}_2$  mixture was 1:5 at a total pressure of 40 Torr. Activity is expressed as  $(\text{product molecules})/(\text{total Au atoms})^{-1} \text{ s}^{-1}$ ; Middle panel: Particle bandgap measured by scanning tunnelling spectroscopy (STS) as a function of the Au particle size; Lower panel: Relative population of the Au particles (two atom layers in height) that exhibited a bandgap of 0.2–0.6 V as measured by STS from  $\text{Au}/\text{TiO}_2(110)$ . (a) and (c) reproduced from Valden *et al.* (1998) with permission from AAAS. (b) reproduced from Lopez *et al.* (2004) with permission from Elsevier.



Chiorino 2000; Bond and Thompson 2000; Cosandey and Madey 2001; Haruta 2003; Kim *et al.* 2003; Schumacher *et al.* 2003; Meyer *et al.* 2004; Chen and Goodman 2006; Hashmi and Hutchings 2006). However, the structure–activity relation of Au nanocatalysts is still elusive and many issues are still being intensely debated. The most serious problem for the widespread industrial application of the Au nanocatalysts is their poor long-term stability towards sintering (Campbell *et al.* 2002; Schumacher *et al.* 2003; Yan *et al.* 2005; Ho and Yeung 2006).

It is today generally accepted that the high activity of the Au nanocatalysts is substantially correlated to the particle size (Bamwenda *et al.* 1997; Valden *et al.* 1998; Lee *et al.* 2004; Yoon *et al.* 2005), and a high catalytic activity is obtained only on nm-sized Au particles (Fig. 12.17(a)). This *size effect* was already reported in the late 1980s by Haruta and colleagues, who studied the oxidation of CO to  $\text{CO}_2$  on Au nanoclusters dispersed on a high surface area material such as powdered titanium dioxide ( $\text{TiO}_2$ ) (Bamwenda *et al.* 1997). Since then it has been shown that Au nanoparticles catalyze a surprising variety of chemical reactions (Hashmi and Hutchings 2006), and this finding has further spurred intense interest in this exciting area of nanocatalysis.

An outstanding example using surface-science model catalysts has been published by the Goodman group (Valden *et al.* 1998). A particle-size effect very similar to that originally reported by Haruta *et al.* was revealed (Fig. 12.17(c), upper panel). The high catalytic activity towards the CO oxidation reaction was suggested to be related to a transition from metal to non-metallic properties of the Au nanoparticles when the particle diameter was reduced to below  $\sim 3$  nm and the thickness dropped to two atomic layers (Fig. 12.17(c), middle and lower panel) (Valden *et al.* 1998). However, it is still



**Fig. 12.18** High-resolution transmission electron microscopes (HRTEM) images acquired using a JEM 4000 (400 kV) comparing representative gold particles on  $\text{TiO}_2$ ,  $\text{SiO}_2$ ,  $\text{ZrO}_2$  and  $\text{ZnO}$  supports “DP” means deposition–precipitation and “I” indicates impregnation method. Reproduced from Radnik *et al.* (2003) by permission of the PCCP Owner Societies.

debated to what extent the electronic properties of the Au particles influence the catalytic activity. Numerous research groups found similar size effects as those reported by the groups of Haruta and Goodman, thus confirming that the size of the particles is a crucial pre-requisite of the high catalytic activity (Fig. 12.17(b)).

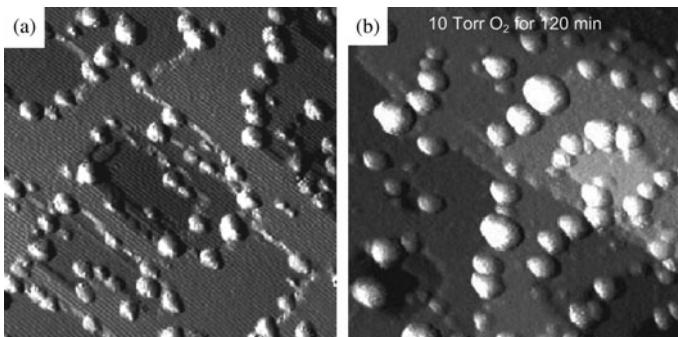
In addition to the characteristic size effect, there also exists a *support effect* in the sense that identical catalyst preparations using different supports lead to structurally different Au clusters, or that the oxide support material itself influences the catalysis of the Au nanoclusters. Support effects have been reported for the extensively studied CO oxidation reaction (Bond and Thompson 2000; Schubert *et al.* 2001; Meyer *et al.* 2004; Matthey *et al.* 2007), and in many cases reducible oxides are superior to non-reducible oxides (Schubert *et al.* 2001; Overbury *et al.* 2004). Similar support effects have also been reported for other reactions such as hydrogenation reactions. In hydrogenation reactions high selectivity is often an important issue (Hughes *et al.* 2005; Corma and Serna 2006). For the hydrogenation of acrolein Claus and coworkers compared the shape of the Au nanoparticles on different supports ( $\text{SiO}_2$ ,  $\text{TiO}_2$ ,  $\text{ZrO}_2$ ,  $\text{ZnO}$ ) after a treatment in hydrogen at 573 K (Fig. 12.18), and whereas on the  $\text{ZrO}_2$  and  $\text{ZnO}$  supports the Au nanoclusters are characterized by extended facets, Au particles supported on  $\text{SiO}_2$  and  $\text{TiO}_2$  were found to be nearly spherical (Radnik *et al.* 2003; Claus 2005). From these and additional studies where indium was used to decorate the outer faces of the gold particles, while the cluster edges remained uncovered, the higher turnover rate for the  $\text{Au}/\text{TiO}_2$  system compared to that of  $\text{Au}/\text{ZrO}_2$  was ascribed to more rounded gold

nanoparticles, which are characterized by a higher amount of low-coordinated surface sites such as edges and kinks (Mohr *et al.* 2003).

A tempting idea is that the catalytically active sites for the Au nanoparticles are the low-coordinated Au atoms either on the Au particle or on the interface between the support and the Au clusters. Similar to the work of Claus and coworkers Boccuzzi *et al.* suggested that the active sites for hydrogen ( $H_2$ ) dissociation and for  $H_2O$ -CO reactive interactions are located on the surface of the metallic gold particles (Boccuzzi *et al.* 1999). Furthermore, it has been proposed that the low-coordinated Au atoms on the Au-support interface are especially active for  $O_2$  activation (Daté and Haruta 2001; Haruta 2004; Molina and Hammer 2004; Molina *et al.* 2004), whereas for CO oxidation the CO molecules are preferentially adsorbed on Au atoms in corners and on edges of the Au particles (Lopez *et al.* 2004). In other studies a “gold-only” route has been proposed for the CO oxidation reaction, where the adsorbates only require low-coordinated Au atoms in order to make the reaction feasible (Chen and Goodman 2004, 2006; Remediakis *et al.* 2005; Janssens *et al.* 2007), and the oxide support only plays a minor role for the overall activity. Such a “gold-only” route could explain the apparent absence of oxide support effects on the catalytic activity of Au nanoclusters reported by some groups (Okumura *et al.* 1998; Lopez *et al.* 2004; Janssens *et al.* 2006).

A number of studies mainly performed on high-surface-area supported catalysts have reported that slightly oxidized gold ( $Au^{+\delta}$ ), in addition to metallic gold ( $Au^0$ ), is important to achieve high activity of dispersed Au catalysts (Bond and Thompson 2000; Costello *et al.* 2002; Fu *et al.* 2003, 2005; Guzman and Gates 2004; Meyer *et al.* 2004; Guzman *et al.* 2005; Hutchings *et al.* 2006; Miller *et al.* 2006; Wang and Hammer 2006). In parallel, several surface-science studies of model systems favored an interpretation where oxygen vacancies have been invoked as being responsible for the stabilization and activity of Au clusters on the oxide support (Wahlström *et al.* 2003; Chen and Goodman 2004; Remediakis *et al.* 2005; Wörz *et al.* 2005; Chretien and Metiu 2007). All these examples illustrate that the physical origin of the distinct catalytic activity of Au nanoparticles is indeed a rather controversial research area and many fundamental questions still have to be addressed before a more detailed and consistent understanding of the unique activity of Au nanocatalysts can be achieved. On the other hand, there is agreement in the literature that the adsorption and activation of molecular oxygen is an important, maybe even the most important, reaction step on supported Au nanocatalysts (Bondzie *et al.* 1999; Schubert *et al.* 2001; Häkkinen *et al.* 2003; Xu and Mavrikakis 2003; Meyer *et al.* 2004; Molina *et al.* 2004).

The high diversity within the literature regarding the structure–activity relationship may partly be explained by considering the different experimental conditions used in the various studies and the high sensitivity to the preparation and pre-treatment conditions of the Au nanocatalysts (Wang and Hammer 2007). Another reason for the lack of consensus on the origin of the enhanced activity of Au nanocatalysts is probably due to the inherent difficulties in studying supported Au catalysts because of its high susceptibility to poisoning and the low density of active sites (Kung *et al.* 2007). In order to resolve these issues it is advantageous to develop model systems that mimic the situation

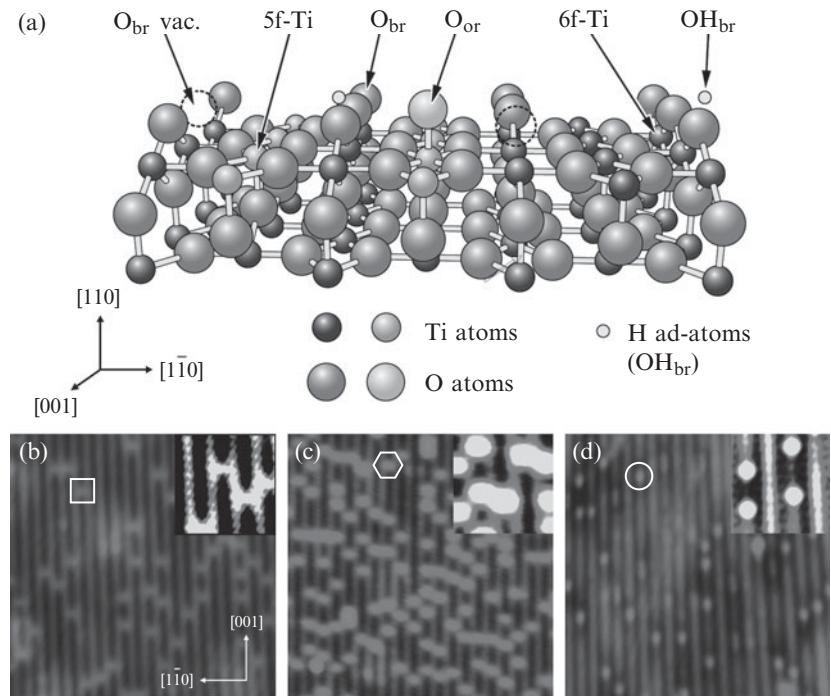


**Fig. 12.19** (a) An STM image of  $\text{Au}/\text{TiO}_2(110)$ -(1 × 1) after deposition of Au at 300 K and annealing at 850 K for 2 min. The Au coverage is 0.25 ML, and the image size is  $50 \times 50 \text{ nm}^2$ . (b) An STM images of  $\text{Au}/\text{TiO}_2(110)$ -(1 × 1) as prepared in (a) after 120 min of  $\text{O}_2$  exposure at 10 Torr. Image size is 50 nm by 50 nm. STM imaging after 120 min of  $\text{CO} : \text{O}_2$  (2 : 1) exposure at 10 Torr leads to similar results; Reproduced from Valden *et al.* (1998) with permission from AAAS.

of the real catalyst systems and can be more easily studied than real high-surface-area catalysts. Most efforts along this direction have focused on  $\text{TiO}_2$ -supported Au catalysts, and in the following, we will discuss a few such examples to illustrate the fundamental insight that has been achieved using this surface-science approach.

For model studies addressing  $\text{TiO}_2$ -supported Au catalysts, the most studied surface is the rutile  $\text{TiO}_2(110)$  – (1 × 1) surface (Diebold 2003; Wendt *et al.* 2005; Bowker 2006) that consists of alternating rows of fivefold-coordinated Ti (5f-Ti) atoms and protruding, twofold-coordinated bridging oxygen ( $\text{O}_{\text{br}}$ ) atoms (Fig. 12.20(a)). An example from the Goodman group is shown in Fig. 12.19. In this example the effect of  $\text{O}_2$  exposure on the Au nanocluster morphology was investigated in a sealed reactor chamber with a pressure of 10 Torr and at 300 K (Valden *et al.* 1998). After  $\text{O}_2$  exposure for 120 min (the time span after which the surface was deactivated in the course of the CO oxidation reaction), the Au cluster density was found to be greatly reduced as a result of sintering, leading to the formation of larger Au clusters. Identical changes of the Au cluster morphology were found after exposing the  $\text{Au}/\text{TiO}_2(110)$  surface to a  $\text{CO} : \text{O}_2$  mixture (2 : 1) at a total pressure of 10 Torr, whereas no changes were observed after exposing the  $\text{Au}/\text{TiO}_2(110)$  system to pure CO at 10 Torr and 300 K. Goodman and coworkers concluded that the  $\text{Au}/\text{TiO}_2(110)$  surface exhibits an exceptionally high reactivity towards  $\text{O}_2$  at 300 K that promotes the sintering of the Au nanocrystallites (Valden *et al.* 1998). Recently, Wahlström *et al.* and Matthey *et al.* used high-resolution STM to study the interesting  $\text{Au}/\text{TiO}_2(110)$  model catalyst in further detail (Wahlström *et al.* 2003; Matthey *et al.* 2007). The special focus in these studies was on the important question of how the Au particles are stabilized or destabilized on the oxide support material. Previously, it has often been suggested that point defects, e.g. oxygen vacancies and step edges, on the oxide support are important in this regard.

To shed light on this issue it is important to study the attachment of Au nanoclusters on  $\text{TiO}_2(110)$  surfaces in different oxidation states (Matthey *et al.* 2007). In the high-resolution STM images of three differently prepared  $\text{TiO}_2(110)$  surfaces (Fig. 12.20) the bright rows correspond to the Ti troughs, whereas geometrically protruding  $\text{O}_{\text{br}}$  atoms appear dark. It should be recalled that STM images reflect contours of constant local density of states (LDOS), and electronic effects are particularly important for clean

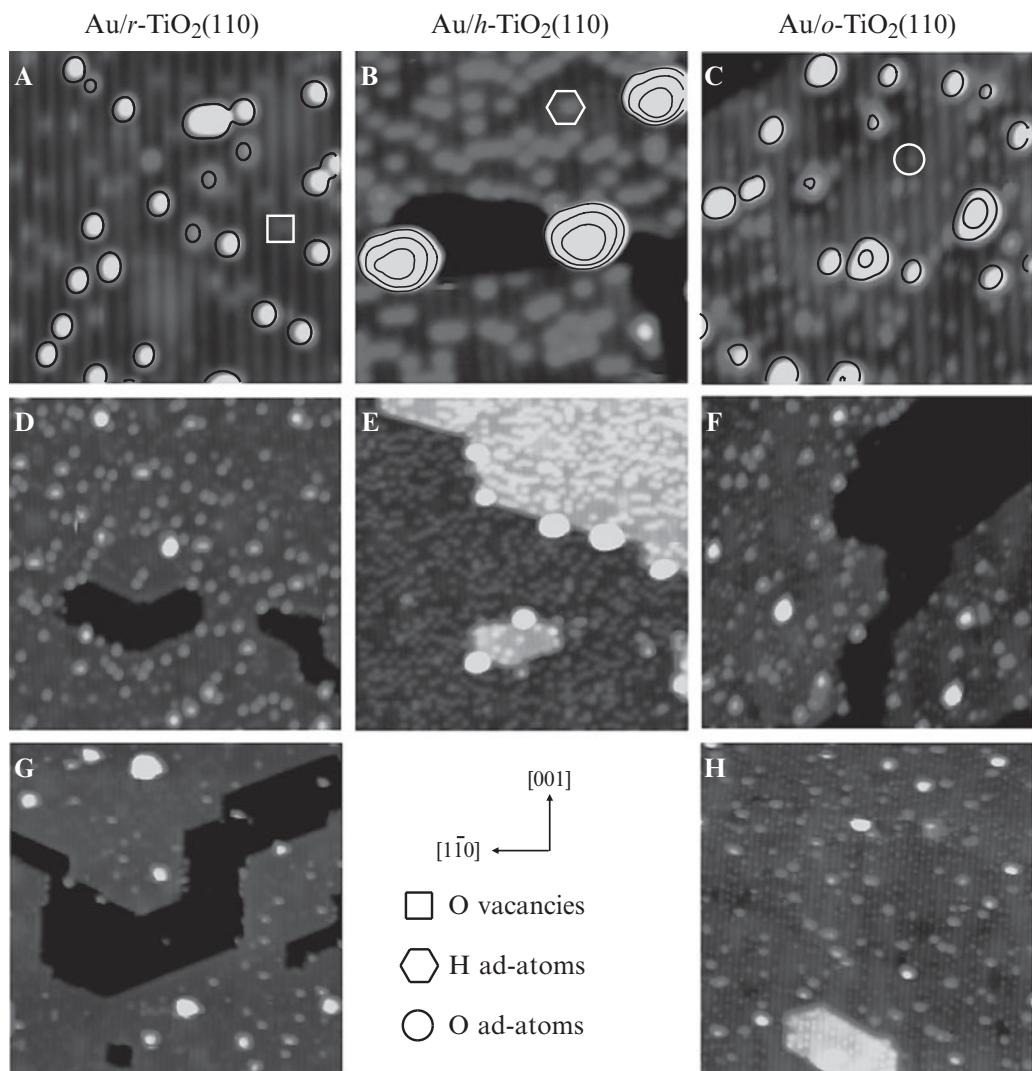


**Fig. 12.20** (a) Ball model of the  $\text{TiO}_2(110)$  surface. Large red balls represent O atoms, medium-sized black balls sixfold-coordinated Ti (6f-Ti), and medium-sized gray balls fivefold-coordinated surface Ti atoms (5f-Ti). Small light blue balls indicate H ad-atoms ( $\text{OH}_{\text{br}}$ ). The bridge-bonded O species ( $\text{O}_{\text{br}}$ ), single oxygen vacancies ( $\text{O}_{\text{br}}$  vac.), and the on-top bonded O species ( $\text{O}_{\text{ot}}$ ) are also indicated. STM images of  $r$ - (b),  $h$ - (c) and  $o$ - (d)  $\text{TiO}_2(110)$  surfaces before Au exposure. Size is  $13 \times 13 \text{ nm}^2$ . In the STM images, the symbols indicate  $\text{O}_{\text{br}}$  vacancies (square), H ad-atoms on  $\text{O}_{\text{br}}$  sites (hexagon) and  $\text{O}_{\text{ot}}$  adatoms in the Ti troughs (circle). Insets ( $3 \times 3 \text{ nm}^2$ ) show the point defects of interest enlarged. Reproduced from Matthey *et al.* (2007) with permission from AAAS.

and adsorbate-covered transition-metal-oxide surfaces. The inversion between geometry and STM appearance of the  $\text{TiO}_2(110)$  surface morphology is explained by a higher electronic LDOS along the Ti troughs compared with the density of states along the  $\text{O}_{\text{br}}$  rows (Diebold 2003). In the same manner, the faint protrusions observed by STM (Figs. 12.20(a) and (b)) in between the bright Ti troughs on clean reduced  $\text{TiO}_2(110)$  [ $r\text{-TiO}_2(110)$ ] surfaces are associated with  $\text{O}_{\text{br}}$  vacancies, i.e. missing O atoms in the surface layer (Wendt *et al.* 2005; Bikondoa *et al.* 2006). When water interacts with this reduced surface, it is found to dissociate in  $\text{O}_{\text{br}}$  vacancies (Wendt *et al.* 2005, 2006; Bikondoa *et al.* 2006; Zhang *et al.* 2006), resulting in the formation of bridging hydroxyl groups ( $\text{OH}_{\text{br}}$ ) that appear as protrusions between the Ti troughs, and these ( $\text{OH}_{\text{br}}$ ) appear brighter than the protrusions originating from  $\text{O}_{\text{br}}$  vacancies (Figs. 12.20(a) and (c)), and are thus easily distinguished from these. A hydroxylated  $\text{TiO}_2(110)$  surface with H ad-atoms, capping some of the  $\text{O}_{\text{br}}$  atoms [ $h\text{-TiO}_2(110)$ ], can thus be prepared. Additionally, an oxidized  $\text{TiO}_2(110)$  [ $o\text{-TiO}_2(110)$ ] surface can be prepared by letting  $\text{O}_2$  dissociate in  $\text{O}_{\text{br}}$  vacancies (Epling *et al.* 1998; Wendt *et al.* 2005), which results in an  $o\text{-TiO}_2(110)$  surface characterized by a number of O ad-atoms ( $\text{O}_{\text{ot}}$ ) in the Ti troughs and perfect  $\text{O}_{\text{br}}$  rows (Figs. 12.20(a) and (d)). The hereby created  $\text{O}_{\text{ot}}$  ad-atoms are located on top of 5f-Ti sites and show up in the STM images as protrusions (Wendt *et al.* 2005; Bikondoa *et al.* 2006).

The same amount of Au (3% ML) was subsequently deposited at room temperature (RT) on these three well-defined  $\text{TiO}_2(110)$  surfaces, the reduced with vacancies only, the hydrated with  $\text{OH}_{\text{br}}$  groups and the oxidized one

with a density of O<sub>ot</sub> ad-atoms identical to the one of initial vacancies, and quite different Au cluster morphologies appeared, as depicted by the STM images in Fig. 12.21. In the case of the *r*-TiO<sub>2</sub>(110) (Figs. 12.21(a) and (d)) and the *o*-TiO<sub>2</sub>(110) (Figs. 12.21(c) and (f)) surfaces, numerous small Au nanoclusters are distributed homogeneously on the terraces. In contrast, on the *h*-TiO<sub>2</sub>(110) surface, fairly large Au clusters (containing up to ca. 20 gold atoms) decorate preferentially the step edges of the TiO<sub>2</sub>(110) substrate (Figs. 12.21(b) and (e)). From these results it was concluded that the interaction

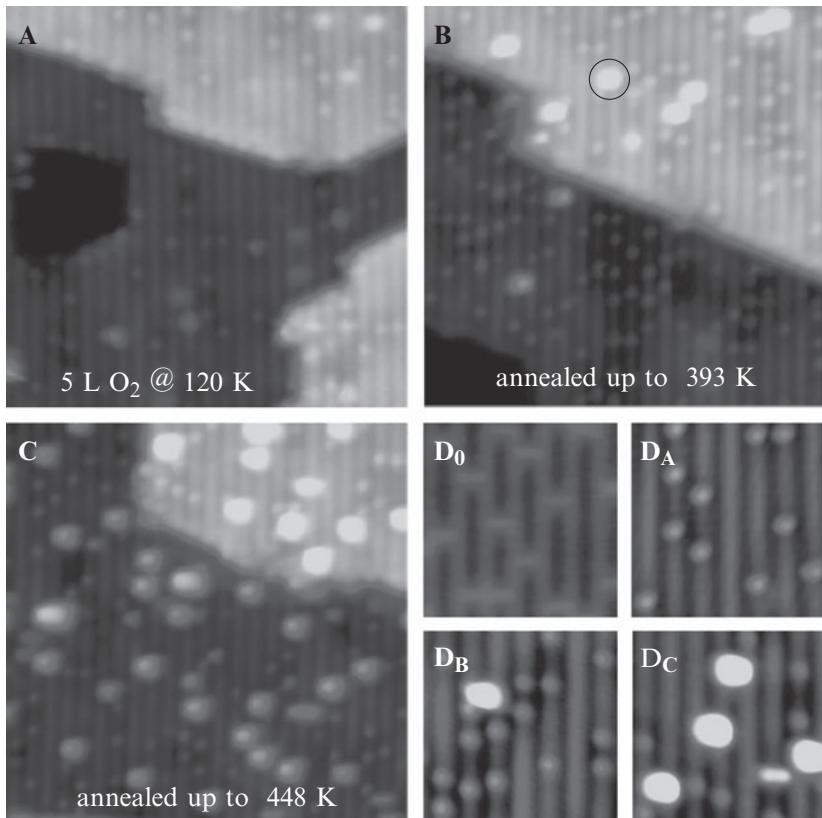


**Fig. 12.21** (a to f) STM images of the three different TiO<sub>2</sub>(110) surfaces after 3% ML Au exposure at RT, whereby 1 ML is defined as  $1.387 \times 10^{15}$  atoms per cm<sup>2</sup> corresponding to Au(111). Image sizes are  $13 \times 13$  nm<sup>2</sup> (a to c) and  $13 \times 13$  nm<sup>2</sup> (d to f), respectively. In (a) to (c) the heights of the Au clusters are indicated by contour lines at 0.12, 0.32 and 0.52 nm above the terrace. (g) STM image of the Au/r-TiO<sub>2</sub>(110) surface ( $35 \times 35$  nm<sup>2</sup>) prepared by 3% ML Au deposition at RT followed by heating up to 68 °C. (h) Like (g) but for the Au/o-TiO<sub>2</sub>(110) surface. Reproduced from Matthey *et al.* (2007) with permission from AAAS.

between Au clusters and the *h*-TiO<sub>2</sub>(110) surface is weaker than on the *r*- and the *o*-TiO<sub>2</sub>(110) surfaces, suggesting that indeed the point defects (O<sub>br</sub> vacancies and O<sub>ot</sub> ad-atoms, respectively) stabilize the Au clusters on the terraces. The obtained Au cluster distributions after Au exposure at RT are in line with DFT calculations, which found that on stoichiometric [*s*-TiO<sub>2</sub>(110)] and *h*-TiO<sub>2</sub>(110) Au monomers (Au<sub>1</sub>) diffuse readily in both the [001] and the [1̄10] directions, whereas Au<sub>1</sub> can be trapped at point defects on the terraces of the *r*- and *o*-TiO<sub>2</sub>(110) surfaces (Iddir *et al.* 2005; Wörz *et al.* 2005; Matthey *et al.* 2007).

To further compare the strength of the Au nanocluster attachment on *r*- and *o*-TiO<sub>2</sub>(110) surfaces, the Au/TiO<sub>2</sub>(110) surfaces were annealed to temperatures above RT (Matthey *et al.* 2007). Starting with *r*- and *o*-TiO<sub>2</sub>(110) surfaces as shown in Figs. 12.21(d) and (f), Au sintering was clearly evident for Au/*r*-TiO<sub>2</sub>(110) (Fig. 12.21(g)) when the sample was heated to 68°C, whereas no sintering was observed for the Au/*o*-TiO<sub>2</sub>(110) model system (Fig. 12.21(h)). These results show that the Au cluster–support bond strength is larger on the oxidized TiO<sub>2</sub>(110) surface with O<sub>ot</sub> ad-atoms than on the reduced TiO<sub>2</sub>(110) surface with O<sub>br</sub> vacancies. This trend was found to be valid irrespective of the precise Au cluster size and can be explained by invoking different bonding mechanisms between the Au clusters and the TiO<sub>2</sub>(110) support, depending on the oxidation state of the support (Matthey *et al.* 2007). If the TiO<sub>2</sub>(110) surface is oxidized [*o*-TiO<sub>2</sub>(110)], the Au nanoclusters are not only stabilized via covalent bonds, as in case of *r*-TiO<sub>2</sub>(110), but also through a stronger ionic bonding via Au–O bonds (Matthey *et al.* 2007). In light of these results, it can be argued that O<sub>br</sub> vacancies are not as important for the stabilization of dispersed Au clusters on TiO<sub>2</sub> surfaces under real reaction conditions, where high oxygen (and water) pressures are usually present, as previously anticipated. Rather, it appears that O-rich Au-support interfaces are important to stabilize the Au nanoclusters, which is probably of general relevance to produce stable, long-lived Au catalysts.

Given the importance of O-rich Au-support interfaces to stabilize the Au clusters on the support, it is interesting to study the adsorption and dissociation of O<sub>2</sub> on the clean *r*-TiO<sub>2</sub>(110) surfaces in more detail. As discussed above, O<sub>2</sub> dissociation occurs in the O<sub>br</sub> vacancies already at temperatures as low as 120 K (Fig. 12.22(a)), resulting in TiO<sub>2</sub>(110) surfaces [*o*-TiO<sub>2</sub>(110)] characterized by perfect O<sub>br</sub> rows and a number of isolated O<sub>ot</sub> ad-atoms in the Ti troughs with the same density as O<sub>br</sub> vacancies before the exposure. However, recently a novel non-vacancy related O<sub>2</sub> dissociation channel that occurs within the Ti troughs has been identified (Wendt *et al.* 2008). This additional dissociation channel is energetically more activated than the one associated with O<sub>br</sub> vacancies, since the density of O<sub>ot</sub> ad-atoms increased by a factor of ~2 only after annealing of the O<sub>2</sub> exposed crystals to 393 K (Fig. 12.22(b)). In addition, a number of small islands appeared on the terraces after annealing (one of which is indicated in Fig. 12.22(b) by a circle), and these islands are positioned in between two Ti troughs and show up in the STM images with a height of ~0.22 nm. Subsequent annealing to an even higher temperature (448 K) led to an increase in the density of the islands, and, in this case, islands with a height of ~0.32 nm were observed, in addition to



**Fig. 12.22** (a to c) STM images ( $15 \times 15 \text{ nm}^2$ ) acquired on a  $\text{TiO}_2(110)$  crystal characterized by an  $\text{O}_{\text{br}}$  vacancy density of  $\sim 4.5 \text{ \%ML}$ . (a) Exposed to  $5 \text{ L O}_2$  at  $120 \text{ K}$ , (b) subsequently annealed up to  $393 \text{ K}$ , and (c) further annealed up to  $448 \text{ K}$ . A circle in (b) indicates one of the newly formed  $\text{TiO}_x$  islands on the terraces with an STM height of  $\sim 0.22 \text{ nm}$ . (d) Zoom-in STM images ( $4.9 \times 4.9 \text{ nm}^2$ ) showing a clean  $r\text{-TiO}_2(110)$  surface with  $\text{O}_{\text{br}}$  vacancies ( $D_0$ ), and enlarged regions ( $D_A$  to  $D_C$ ) of oxidized and flash-annealed  $\text{TiO}_2(110)$  surfaces corresponding to the image shown in (a to c). Reproduced from Wendt *et al.* (2008) with permission from AAAS.

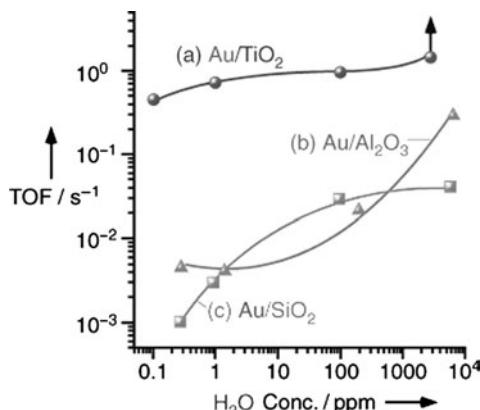
the islands of  $\sim 0.22 \text{ nm}$  height (Fig. 12.22(c)). The height of these islands of  $\sim 0.32 \text{ nm}$  closely resembles the step height on  $\text{TiO}_2(110)$  crystals, which is  $0.325 \text{ nm}$ . In accompanying XPS core level spectra (Wendt *et al.* 2008), exclusively Ti and O features were observed, which led to the conclusion that the new islands appearing on previously flat terraces (Figs. 12.22(b), (c), ( $D_B$ ) and ( $D_C$ )) are newly formed  $\text{TiO}_x$  structures, implying that Ti atoms have diffused from the near-surface region to the topmost surface layer where they reacted with  $\text{O}_{\text{ot}}$  ad-atoms and  $\text{O}_2$  molecules.

The results depicted in Fig. 12.22 show that the interaction of  $\text{O}_2$  with  $r\text{-TiO}_2(110)$  surfaces is much more complex than previously anticipated (Epling *et al.* 1998; Henderson *et al.* 1999; Rasmussen *et al.* 2004; Wendt *et al.* 2005; Kimmel and Petrik 2008). In addition to O vacancies surplus Ti ions on interstitial sites in the near-surface also appear to be important to describe reduced  $\text{TiO}_2$  surfaces in a more complete and coherent picture (Wendt *et al.* 2008). This conclusion is in accordance with several previous experimental findings such as the temperature-programmed static secondary ion mass spectrometry data reported by Henderson (1999), where the onset of Ti interstitial diffusion was found to occur at  $\sim 400 \text{ K}$ , in good agreement with the recent STM results discussed above. Furthermore, the results are in good agreement with previous STM results by Onishi *et al.* (Onishi and Iwasawa 1996), Bennett *et al.* (1999) and by Li *et al.* (1999), where  $\text{O}_2$ -induced growth

of new surface structures has been observed as well. That Ti interstitials are indeed an important feature of reduced  $\text{TiO}_2(110)$  samples is also supported by studies addressing the interaction of  $\text{O}_2$  with the clean  $r\text{-TiO}_2(110)$  surface at RT. In addition to the generally accepted vacancy associated dissociation channel, the new non-vacancy-related,  $\text{O}_2$  dissociation channel in the Ti troughs is evident from the high number of *paired*  $\text{O}_{\text{ot}}$  ad-atoms on next-nearest 5f-Ti sites appearing upon RT exposure (Wendt *et al.* 2008). Furthermore, the combined STM and PES results addressing the  $\text{O}_2$  dissociation on clean  $r\text{-TiO}_2(110)$  surfaces are further corroborated by a comprehensive set of DFT calculations (Wendt *et al.* 2008).

Within the Ti interstitial model a more complete picture emerges of how the stability of Au nanoclusters on the various  $\text{TiO}_2(110)$  surfaces depicted in Fig. 12.21 can be generalized. Since the  $\text{O}_{\text{ot}}$  adatoms serve as “glue” between the Au nanoclusters and the oxide support, reducible oxides are proposed to be better catalysts than those supported on non-reducible ones, because reducible oxides are more capable of forming O-rich Au/support interfaces. The latter comes about since only a reducible oxide can supply the electronic charge that is needed to stabilize and dissociate  $\text{O}_2$  molecules on the surface. In the case of rutile  $\text{TiO}_2(110)$  surfaces, the electronic charge stems from donor sites in the near-surface region and the bulk, that is from the Ti interstitials (Wendt *et al.* 2008). O-rich surface terminations have been identified also for several other reducible oxide supports including vanadium and iron oxide (Abu Hajja *et al.* 2006), which underlines the availability of electronic charge on these oxide materials and the above statement addressing the capability of forming O-rich Au/support interfaces. It should finally be mentioned that additional factors may also contribute to the higher catalytic activity of Au nanoclusters supported on reducible oxides compared to those supported on non-reducible oxides (Bocuzzo *et al.* 1996; Schubert *et al.* 2001; Overbury *et al.* 2004; Comotti *et al.* 2006). For example, Overbury *et al.* suggested that the support may influence the activity through the stabilization of subnanometer particles, the formation of active oxygen-containing intermediates, or the stabilization of confined, optimal Au structures (Overbury *et al.* 2004).

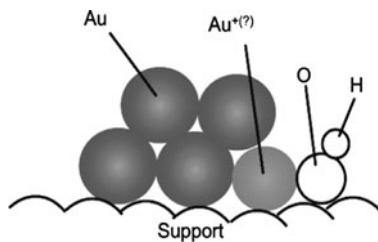
Another interesting example that further highlights the differences between reducible and non-reducible oxide supports is the vital role of moisture in the catalytic activity of supported Au nanoparticles (Daté and Haruta 2001; Kung *et al.* 2003; Daté *et al.* 2004). The effect of moisture in the reactant gas has been systematically investigated for the low-temperature oxidation of CO by  $\text{Au/TiO}_2$ ,  $\text{Au/Al}_2\text{O}_3$ , and  $\text{Au/SiO}_2$  over a wide range of  $\text{H}_2\text{O}$  concentrations (from about 0.1 to 6000 ppm  $\text{H}_2\text{O}$ ) by Haruta’s group (Daté *et al.* 2004). It was reported that moisture enhances the catalytic activity by at least two orders of magnitude, and that the effect of moisture is found to depend on the type of metal oxide (Fig. 12.23) (Daté *et al.* 2004). The effect of moisture becomes significant only above  $\sim 200$  ppm  $\text{H}_2\text{O}$  for  $\text{Au/Al}_2\text{O}_3$ , whereas the activity for  $\text{Au/SiO}_2$  diminishes considerably with a decrease in moisture to  $\sim 3$  ppm. The activity of  $\text{Au/TiO}_2$  at  $\sim 3000$  ppm  $\text{H}_2\text{O}$  is so high that it reaches 100% conversion of CO (Fig. 12.23(a)). The fact that moisture enhances the catalytic activity also supports the view that surface



**Fig. 12.23** Turnover frequencies per surface gold atom at 273 K for CO oxidation over (a)  $\text{Au}/\text{TiO}_2$ , (b)  $\text{Au}/\text{Al}_2\text{O}_3$  and (c)  $\text{Au}/\text{SiO}_2$  as a function of moisture concentration. The arrow means that the actual value of the reaction rate at about 3000 ppm  $\text{H}_2\text{O}$  is much higher than that shown by the solid circle (saturation of CO conversion). Reproduced from Daté *et al.* (2004). Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

$\text{O}$  vacancies can be ruled out as being of any relevance, because  $\text{H}_2\text{O}$  quickly reacts with surface  $\text{O}$  vacancies, thereby healing the  $\text{O}$  defect sites (Wendt *et al.* 2005). The role of moisture in oxidation reactions is probably to activate the oxygen and to decompose carbonate species (Daté and Haruta 2001; Kung *et al.* 2003; Daté *et al.* 2004). The differences between reducible and non-reducible oxide supports are most likely related to the different ways  $\text{H}_2\text{O}$  interacts with the various oxide supports. Daté *et al.* reported that the catalytic activity is determined by the amount of  $\text{H}_2\text{O}$  adsorbed on the catalyst surface (Daté *et al.* 2004), and proposed that the oxide support acts as a reservoir of water, and the actual catalytic reaction takes place at the Au-support interface.

A model of the active site of a supported Au catalyst under realistic operating conditions is depicted in Fig. 12.24 (Costello *et al.* 2002). This model highlights both the important role of metallic Au and the involvement of the Au/support interface, where cationic Au may exist even under reaction conditions. The model is similar to that proposed earlier by Bond and Thompson (2000), who additionally proposed the consumption of  $\text{O}$  atoms from the support, i.e. the creation of  $\text{O}$  vacancies at the Au/support interface, which, however, appears unlikely when considering more recent results (Molina and Hammer 2005). This active site model shown in Fig. 12.24 is in line with many experimental observations and also supported by recent DFT calculations (Wang and Hammer 2006). Numerous research groups have reported that metallic Au is required for the high catalytic activity, but whether or not cationic Au is also needed is still under intense debate. It is possible that the number of active sites on the Au nanoclusters is so small that the cationic Au in real high surface area catalysts is below the detection limit (Kung *et al.* 2007). A significant feature of the proposed active-site model is that “support effects” can be explained via differences in the formation and supply of hydroxyls on various support materials. Finally, the model is in line with the above STM results, wherein O-rich oxide supports were found to be advantageous, and the observed strong effect of moisture on the catalytic activity can be rationalized from the STM findings as well.



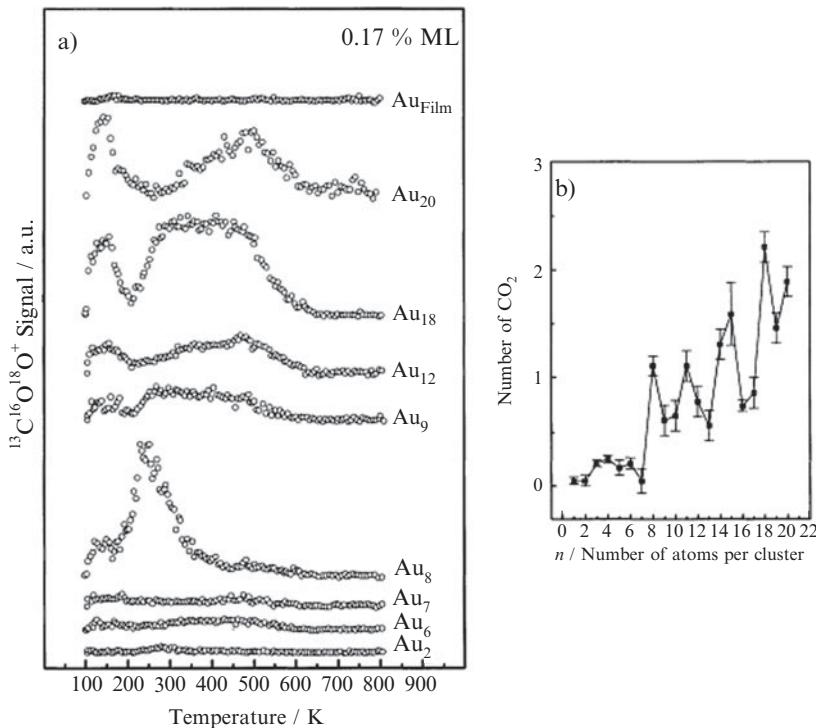
**Fig. 12.24** Proposed model for the active site of a supported gold catalyst incorporating cationic gold at the interface of the gold nanoparticle with the support and the involvement of a hydroxyl group. The current experimental evidence is to a large extent in agreement with this proposal. Reproduced with permission from Kung *et al.* (2007). Copyright 2007 Am Chem. Soc.

### 12.5.3 Model studies using size-selected gold nanoclusters on planar supports

Within the surface-science approach an even more specialized methodology is the use of size-selected metal clusters often deposited on oxide surfaces—an approach that has been developed in order to try to further disentangle the atomic-scale factors controlling the reactivity of nanocluster catalysts. Experimentally, the creation and deposition of size-selected metal clusters are rather challenging (Heiz *et al.* 1997; Heiz and Schneider 2001; Kemper *et al.* 2006; Buratto *et al.* 2007). In recent years an increasing number of articles dealing with monodispersed size-selected metal clusters, where the number of metal atoms per cluster is known, have been published (Lee *et al.* 2004, 2005; Tong *et al.* 2005; Yoon *et al.* 2005; Arenz *et al.* 2006; Lim *et al.* 2006, 2007; Buratto *et al.* 2007). It has thus become possible to study the catalytic reactions, e.g. CO oxidation, as a function of the particle size for clusters typically consisting of only a few atoms, often less than the 20 per cluster (Heiz *et al.* 1999; Lee *et al.* 2004, 2005; Yoon *et al.* 2005; Arenz *et al.* 2006).

An important concern in such studies is to maintain the identity of the produced monodispersed metal clusters upon deposition on a substrate. To achieve this goal, the size-selected metal clusters need to be deposited with low kinetic energy ( $<0.2$  eV/atom), to ensure a *soft landing*. Based on a variety of experimental techniques including STM (Brune *et al.* 1998), pulsed molecular beams (Judai *et al.* 2004) and absorption spectroscopy on small clusters embedded in a rare-gas matrix (Fedrigo *et al.* 1998) it is today established that a soft landing of small metal clusters such as  $\text{Au}_n$ ,  $\text{Pt}_n$  and  $\text{Pd}_n$ , with  $n$  being the number of atoms, is indeed possible.

A very detailed investigation of how the CO oxidation reaction depends on the size of  $\text{Au}_n$  clusters ( $1 \leq n \leq 20$ ) supported on defected thin MgO(100) films has been carried out by Heiz and coworkers (Fig. 12.25). Experimentally, temperature-programmed reaction (TPR) and Fourier transform infrared (FTIR) spectroscopy have been used, and the results have been compared with DFT calculations considering the stability and the charge state of the  $\text{Au}_n$  clusters (Sanchez *et al.* 1999; Häkkinen *et al.* 2003; Yoon *et al.* 2005; Arenz *et al.* 2006). To gather insight into the reaction mechanism Heiz and coworkers used isotopically labelled  $^{18}\text{O}_2$  and  $^{13}\text{CO}$  molecules. Since exclusively the  $^{13}\text{C}^{16}\text{O}^{18}\text{O}$  isotopomer was detected in the TPR experiments, the catalytic oxidation of CO involving oxygen atoms from the MgO substrate could be excluded. The data in Fig. 12.25 clearly show that the CO oxidation reaction, leading to the production of  $\text{CO}_2$ , depends strongly on the  $\text{Au}_n$  particle size.  $\text{Au}_n$  clusters up to  $\text{Au}_7$  on a defect-rich MgO film are inert, whereas  $\text{Au}_8$  clusters are catalytically active, and for larger  $\text{Au}_n$  clusters ( $n \leq 20$ ) a distinct size dependency can be seen (Fig. 12.25(a)). In Fig. 12.25(b) the total number of  $\text{CO}_2$  molecules formed per  $\text{Au}_n$  cluster is depicted against the number of atoms per cluster. Based on the observed non-linear reactivity as a function of size it was concluded that the number of reactive sites varies with the cluster size (Häkkinen *et al.* 2003; Yoon *et al.* 2005; Arenz *et al.* 2006). From a shift in the C–O stretch frequency of adsorbed CO it was further concluded that the reactivity of the  $\text{Au}_n$  clusters depends on cluster charging, a conclusion

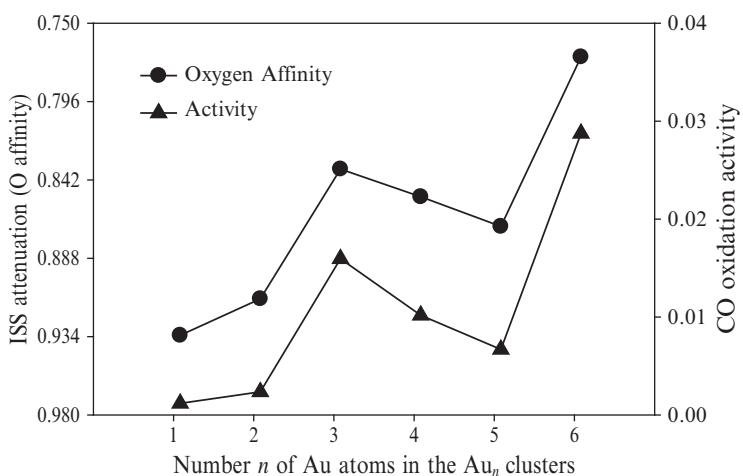


**Fig. 12.25** (a) Temperature-programmed reaction experiments for the CO oxidation on selected Au<sub>n</sub> clusters on defect-rich MgO(100) films. The model catalysts were saturated at 90 K with  $^{13}\text{CO}$  and  $^{18}\text{O}_2$ , and the isotopomer  $^{13}\text{C}^{18}\text{O}^{16}\text{O}$  was detected with a mass spectrometer, as a function of temperature; (b) Reactivities for Au<sub>n</sub> expressed as the number of formed  $\text{CO}_2$  per cluster. Reproduced from Arenz *et al.* (2006). Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

that is in line with the finding that Au<sub>8</sub> clusters supported on defect-poor MgO films were inert, whereas the same clusters on a defect-rich MgO films were active for CO oxidation (Yoon *et al.* 2005; Arenz *et al.* 2006). In accompanying DFT studies (Sanchez *et al.* 1999; Häkkinen *et al.* 2003; Yoon *et al.* 2005) the defect-rich MgO(100) films have been modelled via the introduction of surface layer O vacancies (F-centers) that serve as nucleation centers for the Au<sub>n</sub> clusters and as a source of the charge. It should be noted, however, that quite a variety of point-like and extended defects exist on MgO(100) surfaces and films (Barth and Henry 2003; Wendt *et al.* 2003; Sterrer *et al.* 2005; Freund and Goodman 2008), so that the F-centers considered by Heiz and colleagues (Sanchez *et al.* 1999; Abbet *et al.* 2001; Heiz and Schneider 2001; Häkkinen *et al.* 2003; Yoon *et al.* 2005; Arenz *et al.* 2006) are only one possible description among several others of the defect-rich MgO(100) film. In addition, the stability of the F-centers under reaction conditions is currently a disputed topic in the literature (Wang and Hammer 2007), and the charging of Au<sub>n</sub> clusters supported on MgO(100) films grown on a Mo(100) substrate can be strongly influenced by the Mo substrate (Ricci *et al.* 2006). Nevertheless, the work presented by Heiz and coworkers has become a textbook example, illustrating that sophisticated studies are today possible by using the surface-science approach and employing a combination of several experimental techniques and DFT calculations.

In addition to the utilization of thin MgO(100) films, rutile TiO<sub>2</sub>(110) crystals have also been used as model supports for the study of size-selected

$\text{Au}_n$  clusters. To study this popular model system, two research groups have used a variety of experimental techniques including temperature-programmed desorption (TPD), low-energy ion scattering spectroscopy (ISS) and STM (Lee *et al.* 2004, 2005; Tong *et al.* 2005; Buratto *et al.* 2007). Furthermore, Lee *et al.* used a pulse-dosing technique similar to that recently described by Judai *et al.* (2004) to gain further insight into the CO oxidation reaction on  $\text{Au}_n$  clusters with  $1 \leq n \leq 7$  (Lee *et al.* 2004; 2005). In these experiments, the  $\text{Au}_n/\text{TiO}_2(110)$  model catalysts were initially exposed to 600 L of  $^{18}\text{O}_2$  at RT and then exposed to  $\text{C}^{16}\text{O}$  through a pulsed valve/dosing tube with  $\sim 0.2\text{ L}$  (0.05 ML) pulses.  $\text{C}^{16}\text{O}$ ,  $\text{C}^{18}\text{O}$ , and  $\text{C}^{16}\text{O}^{18}\text{O}$  species have been detected to leave the surface by a mass spectrometer. Rather low  $\text{CO}_2$  production rates were found, but a strong size dependence was evident, with substantial activity for  $\text{Au}_n$  as small as three atoms, and a sharp increase in the activity for clusters larger than  $\text{Au}_5$  (Lee *et al.* 2004, 2005). In the work by Lee *et al.* an excellent anticorrelation exists between CO oxidation activity and the  $\text{O}_2$ -induced attenuation of Au-ISS signal (Fig. 12.26). As the Au-ISS intensity decreases upon 600 L  $\text{O}_2$  exposure (with respect to the signal before the  $\text{O}_2$  exposure) the CO-oxidation activity increases. As an example, for  $\text{Au}_6$  the attenuation of the Au-ISS intensity was found to be  $\sim 23\%$  and the CO oxidation activity was  $\sim 0.03\text{ C}^{16}\text{O}^{18}\text{O}$  molecules per Langmuir CO and deposited Au atom (Lee *et al.* 2005). As shown in Fig. 12.26, the anticorrelation is revealed on a cluster-by-cluster basis. Based on these results it can be concluded that the catalytic activity, that is the  $\text{CO}_2$  production, is correlated to the ability of the  $\text{Au}_n$  clusters to bind oxygen (Lee *et al.* 2005). From recent STM results (Wendt *et al.* 2008) it is known that  $\text{O}_2$  dissociates at RT both in the  $\text{O}_{\text{br}}$  vacancies as well as in the Ti troughs, as also discussed in the preceding section. Upon 600 L  $\text{O}_2$  exposure of the  $\text{Au}_n/\text{TiO}_2(110)$  model catalyst,  $\text{O}_2$  not only reacts with the  $\text{Au}_n$  clusters that are thought to be attached to  $\text{O}_{\text{br}}$  vacancies but also with uncovered regions on the  $\text{TiO}_2(110)$  surface, thus leading to a situation where  $\sim 10\%$  of the 5f-Ti atoms in the Ti troughs are occupied by

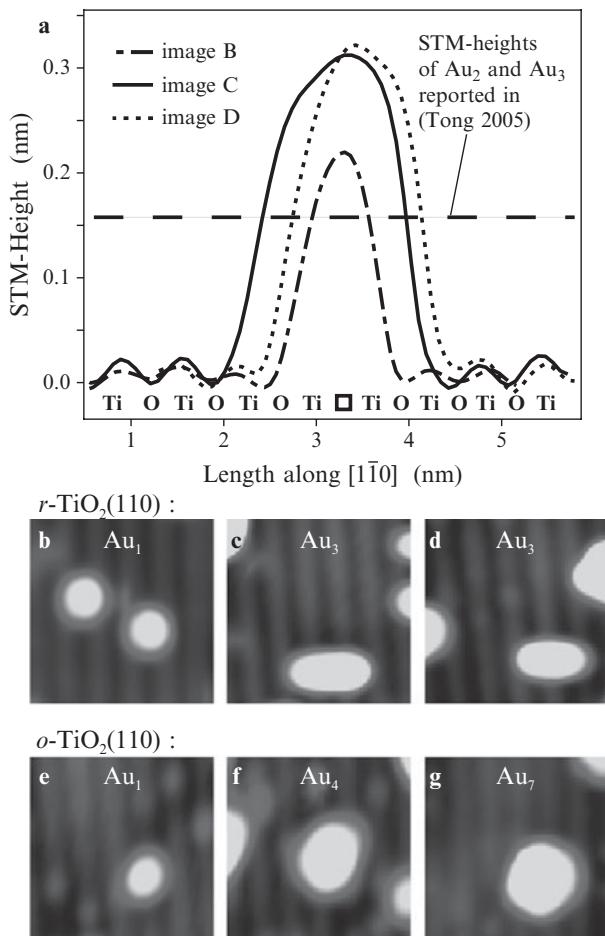


**Fig. 12.26** Correlation between CO oxidation activity (right scale) and the  $\text{O}_2$ -induced attenuation of the Au-ISS signal (left scale, inverted). Size dependence of CO oxidation activity on oxidized  $\text{TiO}_2(110)$  achieved by dosing of 600 L of  $^{18}\text{O}_2$  at room temperature. CO oxidation activity is given as  $\text{C}^{16}\text{O}^{18}\text{O}$  molecules produced per Langmuir CO per deposited Au atom. The uncertainty in comparing the samples was estimated to be  $\sim 20\%$ . Reproduced with permission from Lee *et al.* (2005), copyright 2005, American Institute of Physics.

$O_{ot}$  atoms (Wendt *et al.* 2008). Therefore, it is understandable that not only the Au-ISS signal attenuates upon 600 L  $O_2$  exposure but also the Ti-ISS signal (Lee *et al.* 2005). Lee *et al.* proposed that the attenuation of the Au-ISS intensity is caused by the stabilization of atomic O species on top of the  $Au_n$  clusters.

Buratto and coworkers used STM to study the  $Au_n/TiO_2(110)$  model system (Tong *et al.* 2005; Benz *et al.* 2006; Buratto *et al.* 2007). Size-selected  $Au_n$  clusters, with  $1 \leq n \leq 8$  were produced and deposited onto sputter-annealed  $TiO_2(110)$  surfaces at RT (Tong *et al.* 2005; Buratto *et al.* 2007). After deposition of 0.01 to 0.02 ML  $Au_1^+$  rather big Au clusters with STM heights of  $0.43 \pm 0.17$  nm and no small Au clusters were found on the  $TiO_2(110)$  substrate. However, for  $Au_n^+$  deposition (of about the same total coverage) with  $2 \leq n \leq 4$  Tong *et al.* reported the presence of much smaller Au clusters corresponding to STM heights of only  $0.16 \pm 0.05$  nm ( $Au_2$ ),  $0.15 \pm 0.02$  nm ( $Au_3$ ),  $0.15 \pm 0.02$  nm ( $Au_4$ ), respectively, which was explained by a low mobility and a flat-lying adsorption geometry of these clusters (Tong *et al.* 2005; Buratto *et al.* 2007). In contrast, for  $Au_n$  clusters corresponding to  $5 \leq n \leq 8$  these authors reported average heights of  $0.25 \pm 0.06$  nm,  $0.23 \pm 0.07$  nm,  $0.28 \pm 0.06$  nm, and  $0.27 \pm 0.09$  nm, respectively.

In a comparison of the STM results from Buratto's group (Tong *et al.* 2005; Buratto *et al.* 2007) with those published by the Besenbacher group, and discussed above (Wahlström *et al.* 2003; Matthey *et al.* 2007), it is interesting to note the STM heights of the  $Au_n$  nanoclusters. Striking similarities exist if the comparison is restricted to  $Au_n$  nanoclusters and hydrated  $TiO_2(110)$  surfaces, since in this case both groups obtained fairly large nanoclusters—most of them located at the step edges—after Au deposition with the substrate kept at RT. This comparison seems to imply that the  $TiO_2(110)$  surfaces exposed to  $Au_n$  with  $1 \leq n \leq 8$  in Buratto's group (Tong *et al.* 2005; Buratto *et al.* 2007) were indeed hydrated surfaces characterized by the presence of  $OH_{br}$  groups rather than by  $O_{br}$  vacancies. Quick hydration of sputter-annealed  $TiO_2(110)$  crystals is quite a common phenomenon when studying  $TiO_2(110)$  surfaces (Suzuki *et al.* 2000; Wendt *et al.* 2005; Bikonda *et al.* 2006; Wendt *et al.* 2006; Zhang *et al.* 2006; Park *et al.* 2008). Nevertheless, the conclusions by Tong *et al.* (2005) that  $Au_n$  nanoclusters with  $2 \leq n \leq 8$  landing on the surface have a low mobility should not be falsified by the hydration effects of the  $TiO_2(110)$  supports. However, what is puzzling, and presently not understood, is that the STM heights corresponding to  $Au_2$ ,  $Au_3$  and  $Au_4$  as reported by Buratto's group are even smaller than the STM heights obtained for  $Au_1$  attached to an  $O_{br}$  vacancy (Fig. 12.27) by the Besenbacher group (Wahlström *et al.* 2003; Matthey *et al.* 2007). Similarly, the  $Au_3$  nanoclusters identified by the Besenbacher group have a typical STM height of 0.3 nm (Figs. 12.27(a)–(d)), which is even higher than the  $0.26 \pm 0.08$  nm measured by Buratto's group for their  $Au_n$  clusters with  $5 \leq n \leq 8$ . Thus, the 2D to 3D transition when going from  $n = 4$  to  $n = 5$  as reported by Tong *et al.* (2005) can be debated, and clearly, further work is needed to resolve these issues.



**Fig. 12.27** (a) STM height profiles of typical  $\text{Au}_n$  clusters found on Au-exposed  $r\text{-TiO}_2(110)$  surfaces. The three clusters are ascribed to  $\text{Au}_n$  bound to  $\text{O}_{\text{br}}$  vacancies with  $n$  being 1 and 3, respectively. The difference between the profiles of the two elongated clusters is tentatively explained by symmetrical and unsymmetrical attachment of  $\text{Au}_3$  clusters to isolated  $\text{O}_{\text{br}}$  vacancies. A tunnelling voltage ( $V_T$ ) of  $+1.25$  V and a tunnelling current ( $I_T$ ) of  $\sim 0.1$  nA were used by Matthey *et al.* (2007), and Tong *et al.* (2005) used  $+1$  V  $\leqslant V_T \leqslant +2$  V and  $0.1$  nA  $\leqslant I_T \leqslant 0.2$  nA, respectively. (b to d) Zoom-in STM images ( $4 \times 4 \text{ nm}^2$ ) exemplifying  $\text{Au}_n$  clusters corresponding to the STM height profiles shown in (a). (e to g) Zoom-in STM images ( $4 \times 4 \text{ nm}^2$ ) showing small  $\text{Au}_n$  clusters attached in the vicinity of  $\text{O}_{\text{ot}}$  atoms that are tentatively assigned to  $\text{Au}_1$  (e),  $\text{Au}_4$  (f) and  $\text{Au}_7$  (g) on  $o\text{-TiO}_2(110)$ . Reproduced from Matthey *et al.* (2007) with permission from AAAS.

## 12.6 TEM studies of nanoclusters on high surface area supports

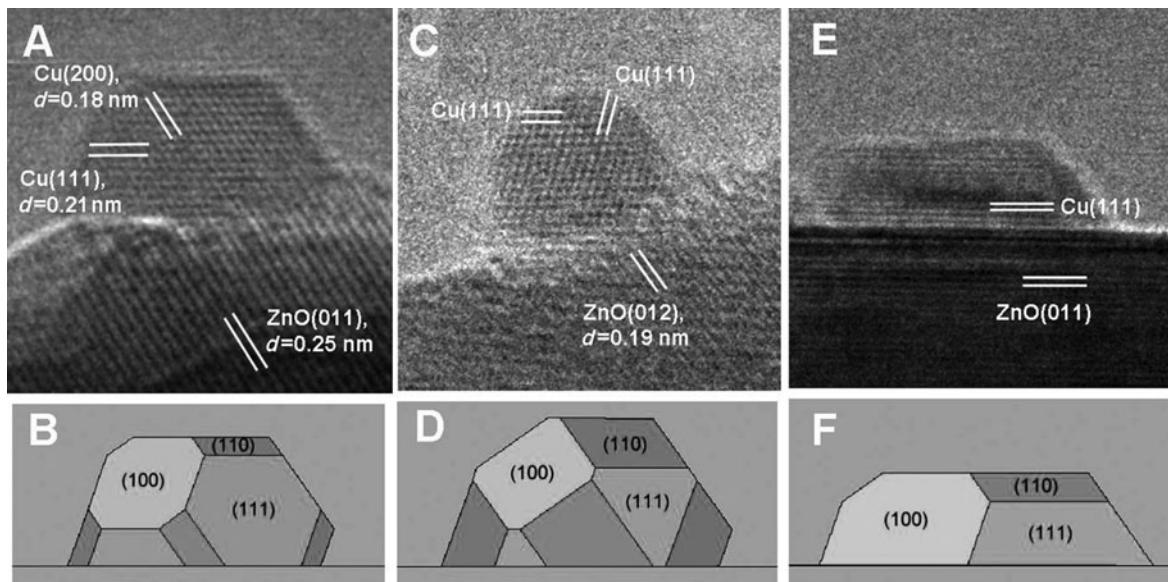
In the previous sections the focus has been on the study of model systems using the surface-science approach to heterogeneous catalysis. In the following, we will demonstrate that with the advance in the nanoscience characterization techniques it has become possible to study high surface area supported catalysts under reaction conditions at the atomic level. In particular, the development of atomic-resolution environmental transmission electron microscopes (ETEM) has represented a major breakthrough by enabling *in-situ* studies of high surface area supported catalysts in the mbar range and at high temperatures. Several outstanding examples have recently appeared. These examples show that it has become possible to directly probe the geometric, and in some cases also the compositional and electronic, structures of nanostructured catalysts with real space resolution down to the atomic level. This has opened up

new possibilities for obtaining structural information about the size, shape and detailed surface structure of nanomaterials, which is of utmost importance for the understanding of their catalytic properties. In the past, electron microscopy studies of catalysts have mainly been performed *ex situ*, where the catalysts are removed from the reaction environments and studied under the high-vacuum conditions in the electron microscope. This approach, however, has obvious limitations since catalysts in general respond dynamically to changes in the gas environment during the catalytic reaction. The new possibilities ensure that it has become possible to reveal the atomic-scale structural details of a catalyst in its working state at high pressures and high temperatures. The two examples discussed below very nicely illustrate the increased level of understanding obtained by performing microscopic characterization of the real catalysts as close to realistic reaction conditions as possible.

### 12.6.1 The Cu/ZnO methanol synthesis catalyst

Above, it was discussed how oxygen induces a surface reconstruction of the Cu(110) surface and that the surface of a catalyst in a reactive atmosphere cannot, in general, be considered a static checkerboard of adsorption sites, and distinct dynamic behavior is observed for real catalysts during reaction conditions. One such example is the industrial Cu/ZnO-based methanol synthesis catalyst that has become a prototype system for studying complex gas-dependent metal–support interactions. Previous *in-situ* studies using XAFS, XRD and FTIR in combination with online activity measurements have shown that the structure and activity of the supported Cu nanoparticles are strongly influenced by the gas environment (Clausen *et al.* 1994; Ovesen *et al.* 1997; Topsøe and Topsøe 1999; Grunwaldt *et al.* 2000).

With atomic-resolution ETEM it is possible to directly study the gas-induced structural dynamics of the Cu/ZnO-nanostructured catalyst in 3 dimensions, and this was recently done by Hansen *et al.* Figure 12.28 shows atomic-resolved ETEM images of Cu nanoparticles supported on ZnO. The lattice-resolved ETEM images and electron energy-loss spectroscopy data reveal the presence of metallic Cu nanoparticles with diameters in the range 3–6 nm (Fig. 12.28(a)). The response of the Cu nanoparticles to variations in the gas composition was studied by exposing the catalyst to either pure H<sub>2</sub> or more oxidizing/reducing conditions by adding H<sub>2</sub>O vapor or CO to the hydrogen gas, respectively. The ETEM images in Fig. 12.28 depict the equilibrium shapes obtained by the Cu nanoparticles in the three different gas phases, and directly reveal how the change in gas composition leads to drastic changes in the particle morphology. Such TEM studies have revealed the driving force behind the gas-induced structural transformations: addition of H<sub>2</sub>O results in Cu particles with a more spherical shape, i.e. the Cu nanoparticles are terminated by a higher fraction of (110) and (100) facets, relative to the more close-packed (111) facets as compared to the particles observed in pure hydrogen atmospheres. H<sub>2</sub>O is thus found to stabilize the more open (110) and (100) facets with respect to the close-packed (111) facet. At the interface the contact area between Cu nanoclusters and the ZnO support does not change



**Fig. 12.28** *In-situ* HRTEM images (a, c and e) of a Cu/ZnO catalyst in various gas environments together with the corresponding Wulff constructions (b, d and f) of the Cu nanocrystals. (a) The image was recorded at a pressure of 1.5 mbar of H<sub>2</sub> at 220 °C. (c) The image was obtained with the Cu nanocrystal exposed to a gas mixture of H<sub>2</sub> and H<sub>2</sub>O, H<sub>2</sub> : H<sub>2</sub>O = 3 : 1 at a total pressure of 1.5 mbar at 220 °C. (e) The image was obtained with the Cu nanocrystal exposed to a gas mixture of H<sub>2</sub> (95%) and CO (5%) at a total pressure of 5 mbar at 220 °C. Reproduced from Hansen *et al.* (2002) with permission from AAAS.

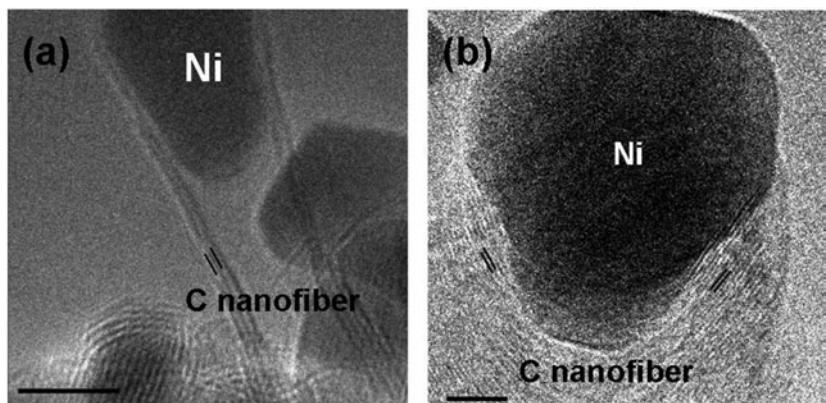
significantly, which reveals that H<sub>2</sub>O adsorption on the different exposed Cu facets is the main driving force for the gas-induced surface reconstructions and the resulting reshaping of the Cu nanoclusters. When the more reducing gas, CO, is added to the hydrogen gas distinct morphological changes occur, i.e. the Cu nanoclusters transform from spherical shapes into disc-like structures with an increased wetting of the ZnO support. On average, the interfacial area was observed to increase by about 50%, which indicates a large decrease in the interface energy. It has been suggested that CO mainly affects the ZnO surface energy, whereas the Cu surface energies are not perturbed (Clausen *et al.* 1994; Ovesen *et al.* 1997; Grunwaldt *et al.* 2000), which seems reasonable since the coverage of H and CO on the Cu surfaces is fairly low under the conditions used here (Clausen *et al.* 1994; Ovesen *et al.* 1997). It is thus found that the main effect of the CO is to change the oxygen content of the ZnO surface as a response to the changed oxidation potential of the gas phase (Clausen *et al.* 1994; Ovesen *et al.* 1997; Grunwaldt *et al.* 2000), and ETEM studies in combination with *in-situ* EELS measurements are found to support this interpretation (Wagner *et al.* 2003).

The *in-situ* electron microscopy results for the Cu/ZnO catalyst are in excellent agreement with previous *in-situ* XAFS data with respect to the adsorbate-induced dynamic morphology changes of the metallic Cu nanoclusters (Clausen *et al.* 1994; Grunwaldt *et al.* 2000). Furthermore, these ETEM results also provide additional and important new insight as the epitaxial relationship between Cu and ZnO can be identified from the lattice-resolved

ETEM images. It is revealed that the majority of the Cu nanoparticles are in contact with the porous ZnO support with their (111) facet, as also observed for Cu particles prepared by vapor deposition under ultrahigh-vacuum conditions onto various atomically clean ZnO surfaces (Campbell 1997). For the high surface area Cu/ZnO catalyst, the ZnO termination at the interface is observed to vary from particle to particle, which indicates a rather weak interaction between Cu and the porous ZnO support. Furthermore, from the lattice-resolved TEM images it is possible to make a Wulff construction for the Cu nanoparticles, as depicted in Fig. 12.28, and from this it is possible to determine quantitatively the relative metal (Cu) surface free energies and the interface free energy for the different catalytic relevant gas environments (Hansen *et al.* 2002). Previously, such quantitative information for supported nanoclusters has only been obtained from *ex-situ* microscopy observations (Henry 1998) or high-resolution STM studies under idealized UHV conditions (Hansen *et al.* 1999; Koplitz *et al.* 2003). The fact that these novel state-of-the-art ETEM studies of real catalysts allow us to determine such surface free energies quantitatively opens very important new perspectives, since the surface and interface energies permit a reconstruction of the full three-dimensional shape of the supported Cu nanoparticles, which in turn may provide important information on the concentration of the different types of surface sites (low-indexed planes, corners, steps, etc.) under the various reaction conditions (Fig. 12.28). This subsequently makes it possible to incorporate detailed information about the nanocluster morphology and the site-specific reactivity more directly into the microkinetic description of a catalytic reaction. Previous microkinetic descriptions of catalytic reactions and processes had to rely on assumptions with respect to the nature of the exposed surfaces and interfaces, as well as their dependence on the gas composition.

### 12.6.2 Ni-based catalysts for steam reforming: Carbon-whisker formation

One of the most important chemical reactions is the steam reforming process, where hydrocarbon molecules, mainly methane CH<sub>4</sub>, and water are converted into H<sub>2</sub> and CO (Rostrup-Nielsen *et al.* 2002). The usual catalyst for the reaction consists of Ni nanoclusters supported on MgAl<sub>2</sub>O<sub>4</sub>. As discussed in the section about the design of the Au-Ni steam reforming catalyst one of the main problems for this reaction is that Ni also catalyzes side reactions, leading to the formation of solid carbon, such as graphitic nanofibers, in the catalyst. To obtain atomic-scale insight into the carbon-filament formation and thereby find ways to prevent the process, ETEM has been applied in a dynamic mode to directly monitor the growth mechanism at the atomic scale (Baker *et al.* 1972; Helveg *et al.* 2004; Hofmann *et al.* 2007). The dynamic mode in which the gas-induced restructuring of the nanoclusters is directly visualized by recording time-lapsed image series and replaying these in the form of a movie has led to unprecedented new insight into the mechanistic details of gas-induced restructuring processes, such as reshaping, phase transformations, sintering and vapor deposition growth processes.

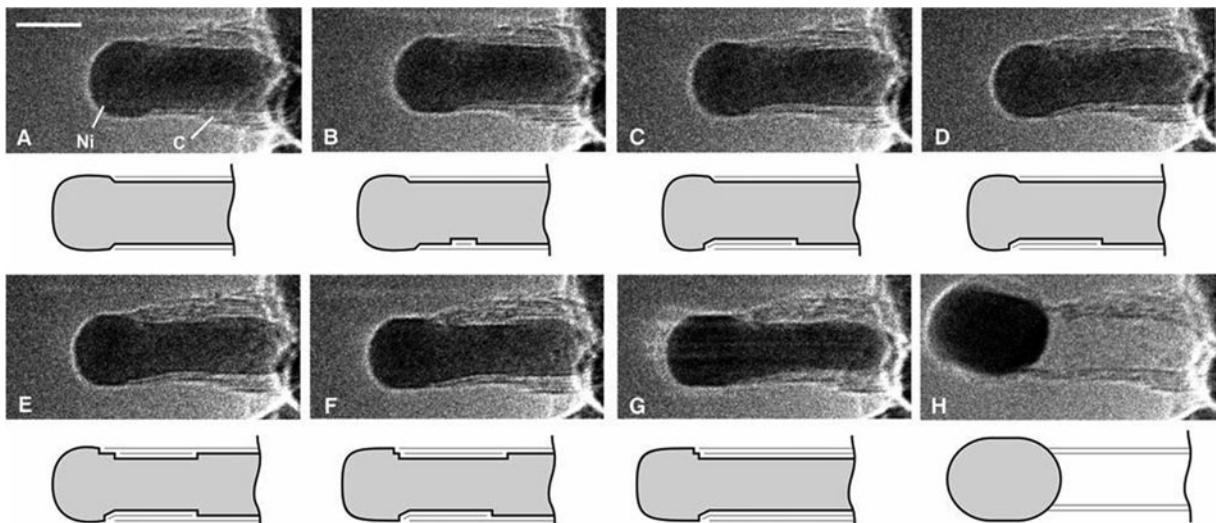


**Fig. 12.29** HRTEM images of (a) a multiwalled tubular carbon structure and (b) a whisker-type carbon nanofiber. The nanofibers are formed by exposing supported nickel nanocrystals to a mixture of  $\text{CH}_4 : \text{H}_2 = 1 : 1$  at a total pressure of about 2.0 mbar and a temperature of 500–540 °C. The lattice fringes in the nanofibers, indicated by the black lines, are separated by 0.34 nm, corresponding to the (002) lattice planes in graphite. Scale bar = 5 nm. Reproduced by permission from Macmillan Publishers Ltd: (Helveg *et al.* 2004), copyright 2004.

In the study by Helveg *et al.* (2004), the ETEM images reveal that the Ni nanoclusters, which have diameters of 5–20 nm, located at the front end of the nanofibers, undergo a reaction-induced reshaping and that the nanocluster shape acts as a template for the alignment of graphene layers into multiwalled carbon nanofiber structures (Fig. 12.29). The smaller Ni nanoclusters are found to obtain highly elongated shapes, and tubular nanofibers form with the graphene layers aligned along the fiber axis. However, the larger Ni clusters are found to become pear-shaped, and whisker-like graphitic nanofibers are observed to emerge with the graphene layers being inclined with respect to the fiber axis.

By means of ETEM movies the dynamics in the growth process was recorded, and Fig. 12.30 reveals a series of images from such an ETEM movie. It is observed that the initial compact equilibrium shape of the Ni nanocrystal ( $\sim 6$  nm in diameter) is transformed into a highly elongated shape, and this elongation appears to be correlated with the formation of more graphene sheets into a tubular structure at the graphene/Ni interface. From the dynamic TEM movies it is observed that this elongation of the Ni nanocrystals continues until an aspect ratio (length/width) of  $\sim 4$  is reached, at which stage the Ni nanoclusters abruptly, i.e. within less than  $\sim 0.5$  s, contracts into a spherical shape (Fig. 12.30(h)). This elongation/contraction scenario continues in a periodic manner as the nanofiber grows, and stops only if the graphene layers eventually encapsulate the Ni nanoparticles. The observed contraction is attributed to the fact that the increase in the Ni surface energy can no longer be compensated for by the energy gained when binding the graphitic fiber to the Ni surface.

A detailed analysis of the ETEM movies reveals that monoatomic step edges are present at the Ni surface, and that these are induced spontaneously in the course of the reaction as seen from Figs. 12.30(b)–(g). An additional graphene layer is observed in between the pair of such step-edges and the graphene layer grows as the Ni steps move concurrently towards the ends of the Ni cluster and vanish. From the ETEM movies one can thus conclude that for all sizes of Ni nanoparticles, an interfacial growth scenario where the step-edge-assisted graphene growth on the surface of the Ni nanoparticles is a general



**Fig. 12.30** *In-situ* HRTEM images sequence of a growing carbon nanofiber (movies can be found in Helveg *et al.* 2004). Images (a–h) illustrate one cycle in the elongation/contraction process. Drawings are included to guide the eye in locating the positions of monoatomic Ni step edges at the graphene Ni interface. The images are acquired *in situ* with  $\text{CH}_4 : \text{H}_2 = 1 : 1$  at a total pressure of 2.1 mbar with the sample heated to 536 °C. All images are obtained with a rate of 2 frames/s. Scale bar = 5 nm. Reproduced by permission from Macmillan Publishers Ltd: (Helveg *et al.* 2004), copyright 2004.

mechanism for the nickel-catalyzed growth of graphitic nanofibers under the steam reforming conditions is revealed. For the largest Ni nanoclusters with a diameter of up to 20 nm, the ETEM movies show that the Ni nanoparticles gradually become more stable and attain a pear shape.

These atomic-level observations extracted from the *in-situ* TEM movies were rationalized from an interplay with density-functional theory (DFT) calculations, which showed that methane dissociation is indeed facilitated at step edges and that carbon atoms bind stronger to step edges than to facet sites (Helveg *et al.* 2004; Abild-Pedersen *et al.* 2006). This finding suggests that the step-edge sites indeed act as the nucleation sites for graphene filaments, and the binding energy of carbon to a step site is found to be larger than the energy cost for the step formation, which explains why steps form spontaneously on the Ni surface (Bengaard *et al.* 2002; Helveg *et al.* 2004). The DFT calculations also show that carbon atoms are more stable in a graphene layer than at a step site, which indicates that there is an energetic driving force for graphene growth.

The atomic-resolution ETEM observations and DFT calculations demonstrate that step edges play an important role as nucleation sites for graphene growth on the surface of Ni nanoparticles. In this connection the findings support the general understanding of the effect of many different promoters in Ni-based steam reforming catalysts, e.g., potassium and sulfur promoter atoms are found to bind preferentially to Ni step edges, and these promoter elements thus suppress the graphite formation through a step-blocking mechanism (Bengaard *et al.* 2002; Rostrup-Nielsen *et al.* 2002). As already discussed, such a step-site blocking was demonstrated to provide control over the bond-breaking selectivity of higher hydrocarbons on nickel surfaces (Vang *et al.* 2005a, 2006).

## 12.7 Conclusions and outlook

In this chapter we have illustrated how it is possible today to obtain fundamental insight into the structure of catalysts and catalytic processes. In particular, with the development of new nanocharacterization tools it has become possible to provide atomic-scale insight into both high surface area catalysts under real reaction conditions and into model systems using the so-called surface-science approach. This possibility indeed represents major progress within catalysis research. In particular, the advancement within high-resolution microscopy techniques, i.e. transmission electron microscopy (TEM) for the study of real catalysts under reaction conditions, and scanning probe microscopy (STM and AFM), has made it possible to achieve some of the Holy Grails within catalysis research. With state-of-the-art *in-situ* TEM we can today “see catalysts” directly at the atomic scale in real time and under reaction conditions, and the picture of the catalyst as a dynamic medium has emerged, i.e. the morphology and structure of the catalyst change in response to the surrounding gaseous environment. Furthermore, from the study of numerous model systems using the so-called surface approach a fantastic knowledge base of the fundamental, sequential elementary processes typically involved in a catalytic process, i.e. the adsorption of reactants, the diffusion of these to active sites, docking to these sites followed by bond breaking and the formation of new bonds and a final desorption step of the reaction products has been developed. This new level of fundamental insight into catalytic processes has advanced our understanding of catalysis tremendously, and we are today approaching an era in which the design of new catalysts can be guided by *ab initio* fundamental insight rather than intuition and trial-and-error methods. This new approach has so far only led to the development of rather few new catalysts, but with the continuing development within nanocharacterization, nanosynthesis, and first-principles theoretical methods we are expecting many more new catalysts to be designed from first principles in the future.

The possibilities of catalyst design and synthesis of new nanoclusters and new nanostructured support materials go far beyond the use of standard metal-, metal-oxide- or metal-sulfide-based catalysts discussed in this chapter. Nanotechnology offers a variety of new methods for creating novel structures/materials that are impossible to synthesize with traditional techniques, and new paradigms within catalysis may be expected to follow the continuous development of nanostructured materials. As is the case for many other emerging science and technology areas, it is very difficult to predict in which direction nanoscience and technology will progress and harvest the largest successes, and hence it also difficult to state what impact we can expect nanoscience and technology to have on catalysis in the decades to come. However, one very promising, interesting and visionary route to follow is the method of *biomimetic catalysis* in which one tries to mimic the behavior of some of Nature’s own catalysts, e.g. enzymes. Enzymes have been developed through millions of years of evolution, and therefore it is not surprising that they are often superior to industrial catalysts, which are the outcome of roughly one century of development. A stunning example of the efficiency of nature’s catalysts is the natural enzyme nitrogenase,

which produces ammonia under ambient conditions in strong contrast to the industrial ammonia catalyst, which typically operates at  $\sim 400^\circ\text{C}$  and  $\sim 100$  atmospheres. The enzymes themselves are optimized for very specific biologic purposes and cannot directly be used for large-scale industrial production of chemicals. The idea is rather to mimic the active centers of the enzymes in nanostructured materials. Preliminary work has been done within this evolving research area (Hinnemann and Nørskov 2006), but many challenges are yet to be overcome before artificially made catalysts can compare to Nature's own. In this way, Nature reminds us that there is plenty of room for improvement of present state-of-the-art within catalysis, and nanoscience and technology will play a major part in realizing this great potential.

## References

- Abbet, S., Riedo, E., Brune, H., Heiz, U., Ferrari, A.M., Giordano, L., Pacchioni, G. *J. Am. Chem. Soc.* **123**, 6172 (2001).
- Abild-Pedersen, F., Nørskov, J.K., Rostrup-Nielsen, J.R., Sehested, J., Helveg, S. *Phys. Rev. B* **73**, 115419 (2006).
- Abu Haija, M., Guimond, S., Romanishyn, Y., Uhl, A., Kuhlenbeck, H., Todorova, T.K., Ganduglia-Pirovano, M.V., Dobler, J., Sauer, J., Freund, H.J. *Surf. Sci.* **600**, 1497 (2006).
- Ackermann, M., Robach, O., Walker, Quiros, C., Isern, H., Ferrer, S. *Surf. Sci.* **557**, 21 (2004).
- Arenz, M., Landman, U., Heiz, U. *Chem. Phys. Chem.* **7**, 1871 (2006).
- Baker, R.T.K., Barber, M.A., Harris, P.S., Feates, F.S., Waite, R.J. *J. Catal.* **26**, 51 (1972).
- Bamwenda, G.R., Tsubota, S., Nakamura, T., Haruta, M. *Catal. Lett.* **44**, 83 (1997).
- Barth, C., Henry, C.R. *Phys. Rev. Lett.* **91**, 196102 (2003).
- Barth, J.V., Brune, H., Ertl, G., Behm, R. *Phys. Rev. B* **42**, 9307 (1990).
- Bäumer, M., Libuda, J., Neyman, K.M., Rosch, N., Rupprechter, G., Freund, H.J. *Phys. Chem. Chem. Phys.* **9**, 3541 (2007).
- Bengaard, H.S., Nørskov, J.K., Sehested, J., Clausen, B.S., Nielsen, L.P., Molenbroek, A.M., Rostrup-Nielsen, J.R. *J. Catal.* **209**, 365 (2002).
- Bennett, R.A., Stone, P., Price, N.J., Bowker, M. *Phys. Rev. Lett.* **82**, 3831 (1999).
- Benz, L., Tong, X., Kemper, P., Metiu, H., Bowers, M.T., Buratto, S.K. *J. Phys. Chem. B* **110**, 663 (2006).
- Besenbacher, F., Chorkendorff, I., Clausen, B.S., Hammer, B., Molenbroek, A.M., Nørskov, J.K., Stensgaard, I. *Science* **279**, 1913 (1998).
- Besenbacher, F., Sprunger, P.T., Ruan, L., Olesen, L., Stensgaard, I., Lægsgaard, E. *Top. Catal.* **1**, 325 (1994).
- Bikondoa, O., Pang, C.L., Ithnin, R., Muryn, C.A., Onishi, H., Thornton, G. *Nature Mater.* **5**, 189 (2006).
- Bocciuzzi, F., Chiorino, A. *J. Phys. Chem. B* **104**, 5414 (2000).

- Boccuzzi, F., Chiorino, A., Manzoli, M., Andreeva, D., Tabakova, T. *J. Catal.* **188**, 176 (1999).
- Boccuzzi, F., Chiorino, A., Tsubota, S., Haruta, M. *J. Phys. Chem.* **100**, 3625 (1996).
- Bollinger, M.V., Lauritsen, J.V., Jacobsen, K.W., Nørskov, J.K., Helveg, S., Besenbacher, F. *Phys. Rev. Lett.* **87**, 196803 (2001).
- Bond, G.C., Thompson, D.T. *Gold Bull.* **33**, 41 (2000).
- Bondzie, V.A., Parker, S.C., Campbell, C.T. *Catal. Lett.* **63**, 143 (1999).
- Bouwens, S.M.A.M., Van Veen, J.A.R., Koningsberger, D.C., De Beer, V.H.J., Prins, R. *J. Phys. Chem.* **95**, 123 (1991).
- Bowker, M. *Curr. Opin. Solid State Mater Sci.* **10**, 153 (2006).
- Boyes, E.D., Gai, P.L. *Ultramicroscopy* **67**, 219 (1997).
- Brune, H., Giovannini, M., Bromann, K., Kern, K. *Nature* **394**, 451 (1998).
- Buatier De Mongeot, F., Scherer, M., Gleich, B., Kopatzki, E., Behm, R.J. *Surf. Sci.* **411**, 249 (1998).
- Buratto, S.K., Bowers, M.T., Metiu, H., Manard, X., Tong, X., Benz, L., Kemper, P., Chrétien, S. in *Atomic Clusters: From Gas Phase to Deposited Structures*, (ed.) Woodruff, D.P. (Elsevier, 2007) p. 151.
- Byskov, L.S., Nørskov, J.K., Clausen, B.S., Topsøe, H. *J. Catal.* **187**, 109 (1999).
- Campbell, C.T. *Surf. Sci. Rep.* **27**, 1 (1997).
- Campbell, C.T., Ertl, G., Kuipers, H., Segner, J. *J. Chem. Phys.* **73**, 5862 (1980).
- Campbell, C.T., Parker, S.C., Starr, D.E. *Science* **298**, 811 (2002).
- Chen, D.A., Friend, C.M. *Surf. Sci.* **371**, 131 (1997).
- Chen, M.S., Goodman, D.W. *Science* **306**, 252 (2004).
- Chen, M.S., Goodman, D.W. *Catal. Today* **111**, 22 (2006).
- Chorkendorff, I., Niemantsverdriet, J.W. *Concepts of Modern Catalysis and Kinetics* (Wiley-VCH, Weinheim, 2003).
- Chretien, S., Metiu, H. *J. Chem. Phys.* **126**, (2007).
- Christoffersen, E., Liu, P., Ruban, A., Skriver, H.L., Nørskov, J.K. *J. Catal.* **199**, 123 (2001).
- Ciobica, I.M., Santen, R.A.V. *J. Phys. Chem. B* **107**, 3808 (2003).
- Claus, P. *Appl. Catal. A* **291**, 222 (2005).
- Clausen, B.S., Lengeler, B., Candia, R., Als-Nielsen, J., Topsøe, H. *Bull. Soc. Chim. Belg.* **90**, 1249 (1981).
- Clausen, B.S., Schiøtz, J., Gråbæk, L., Ovesen, C.V., Jacobsen, K.W., Nørskov, J.K., Topsøe, H. *Top. Catal.* **1**, 367 (1994).
- Clausen, B.S., Topsøe, H., Frahm, R. *Adv. Catal.* **42**, 315 (1998).
- Comotti, M., Li, W.C., Spliethoff, B., Schuth, F. *J. Am. Chem. Soc.* **128**, 917 (2006).
- Corma, A., Serna, P. *Science* **313**, 332 (2006).
- Cosandey, F., Madey, T.E. *Surf. Rev. Lett.* **8**, 73 (2001).
- Costello, C.K., Kung, M.C., Oh, H.S., Wang, Y., Kung, H.H. *Appl. Catal. A* **232**, 159 (2002).
- Coulman, D.J., Wintterlin, J., Behm, R.J., Ertl, G. *Phys. Rev. Lett.* **64**, 1761 (1990).
- Creemer, J.F., Helveg, S., Hoveling, G.H., Ullmann, S., Molenbroek, A.M., Sarro, P.M., Zandbergen, H.W. *Ultramicroscopy* **108**, 993 (2008).

- Cristol, S., Paul, J.F., Payen, E., Bougeard, D., Clémendot, S., Hutschka, F. *J. Phys. Chem. B* **104**, 11220 (2000).
- Dahl, S., Logadottir, A., Egeberg, R.C., Larsen, J.H., Chorkendorff, I., Törnquist, E., Nørskov, J.K. *Phys. Rev. Lett.* **83**, 1814 (1999).
- Daté, M., Haruta, M. *J. Catal.* **201**, 221 (2001).
- Daté, M., Okumura, M., Tsubota, S., Haruta, M. *Angew. Chem. Int. Ed.* **43**, 2129 (2004).
- Datye, A.K., Hansen, P.L., Helveg, S. in *Handbook of Heterogeneous Catalysis*, (eds) Ertl, G., Knözinger, H., Schüth, F. and Weitkamp, J. (Wiley-VCH, New York, 2008) p. 803.
- Davies, J.C., Nielsen, R.M., Thomsen, L.B., Chorkendorff, I., Logadottir, A., Lodzianna, Z., Nørskov, J.K., Li, W.X., Hammer, B., Longwitz, S.R., Schnadt, J., Vestergaard, E.K., Vang, R.T., Besenbacher, F. *Fuel Cells* **4**, 309 (2004).
- Diebold, U. *Surf. Sci. Rep.* **48**, 53 (2003).
- Epling, W.S., Peden, C.H.F., Henderson, M.A., Diebold, U. *Surf. Sci.* **413**, 333 (1998).
- Ertl, G. *Surf. Sci.* **6**, 208 (1967).
- Ertl, G. *Angew. Chem. Int. Ed.* **47**, 3524 (2008).
- Evans, J., Puig-Molina, A., Tromp, M. *MRS Bull.* **32**, 1038 (2007).
- Fedrigo, S., Harbich, W., Buttet, J. *Phys. Rev. B* **58**, 7428 (1998).
- Ferrer, S., Ackermann, M.D., Lundgren, E. *MRS Bull.* **32**, 1010 (2007).
- Freund, H.J., Goodman, D.W. in *Handbook of Heterogeneous Catalysis*, (eds) Ertl, G., Knözinger, H., Schüth, F. and Weitkamp, J. (Wiley-VCH, New York, 2008) p. 1309.
- Fu, L., Wu, N.Q., Yang, J.H., Qu, F., Johnson, D.L., Kung, M.C., Kung, H.H., David, V.P. *J. Phys. Chem. B* **109**, 3704 (2005).
- Fu, Q., Saltsburg, H., Flytzani-Stephanopoulos, M. *Science* **301**, 935 (2003).
- Gai, P.L., Boyes, E.D., Helveg, S., Hansen, P.L., Giorgio, S., Henry, C.R. *MRS Bull.* **32**, 1044 (2007).
- Gambardella, P., Sljivancanin, Z., Hammer, B., Blanc, M., Kuhnke, K., Kern, K. *Phys. Rev. Lett.* **87**, 056103 (2001).
- Gasteiger, H.A., Markovic, N., Ross, P.N., Cairns, E.J. *J. Phys. Chem.* **98**, 617 (1994).
- Giorgio, S., Joao, S.S., Nitsche, S., Chaudanson, D., Sitja, G., Henry, C.R. *Ultramicroscopy* **106**, 503 (2006).
- Gosselink, J.W. *CATTECH* **4**, 127 (1998).
- Greeley, J., Mavrikakis, M. *Nature Mater.* **3**, 810 (2004).
- Greeley, J., Nørskov, J.K., Mavrikakis, M. *Ann. Rev. Phys. Chem.* **53**, 319 (2002).
- Grunwaldt, J.-D., Molenbroek, A.M., Topsøe, N.-Y., Topsøe, H., Clausen, B.S. *J. Catal.* **194**, 452 (2000).
- Grunwaldt, J.D., Clausen, B.S. *Top. Catal.* **18**, 37 (2002).
- Gustafson, J., Mikkelsen, A., Borg, M., Lundgren, E., Kohler, L., Kresse, G., Schmid, M., Varga, P., Yuhara, J., Torrelles, X., Quiros, C., Andersen, J.N. *Phys. Rev. Lett.* **92**, 126102 (2004).
- Guzman, J., Carrettin, S., Corma, A. *J. Am. Chem. Soc.* **127**, 3286 (2005).
- Guzman, J., Gates, B.C. *J. Am. Chem. Soc.* **126**, 2672 (2004).
- Hammer, B. *Phys. Rev. Lett.* **83**, 3681 (1999).
- Hammer, B., Nørskov, J.K. *Nature* **376**, 238 (1995).

- Hammer, B., Nørskov, J.K. *Adv. Catal.* **45**, 71 (2000).
- Hansen, K.H., Worren, T., Stempel, S., Lægsgaard, E., Bäumer, M., Freund, H.-J., Besenbacher, F., Stensgaard, I. *Phys. Rev. Lett.* **83**, 4120 (1999).
- Hansen, P.L., Wagner, J.B., Helveg, S., Rostrup-Nielsen, J.R., Clausen, B.S., Topsøe, H. *Science* **295**, 2053 (2002).
- Haruta, M. in *Science and Technology in Catalysis 2002* (Kodansha Ltd, Tokyo, 2003) p. 31.
- Haruta, M. *Gold Bull.* **37**, 27 (2004).
- Haruta, M., Kobayashi, T., Sano, H., Yamada, N. *Chem. Lett.* **16** 405 (1987).
- Haruta, M., Yamada, N., Kobayashi, T., Iijima, S. *J. Catal.* **115** 301 (1989).
- Hashmi, A.S.K., Hutchings, G.J. *Angew. Chem. Int. Ed.* **45**, 7896 (2006).
- Heiz, U., Sanchez, A., Abbet, S., Schneider, W.D. *J. Am. Chem. Soc.* **121**, 3214 (1999).
- Heiz, U., Schneider, W.D. *Crit. Rev. Solid State Mater. Sci.* **26**, 251 (2001).
- Heiz, U., Vanollie, F., Trento, L., Schneider, W.D. *Rev. Sci. Instrum.* **68**, 1986 (1997).
- Helveg, S., Lauritsen, J.V., Lægsgaard, E., Stensgaard, I., Nørskov, J.K., Clausen, B.S., Topsøe, H., Besenbacher, F. *Phys. Rev. Lett.* **84**, 951 (2000).
- Helveg, S., Lopez-Cartes, C., Sehested, J., Hansen, P.L., Clausen, B.S., Rostrup-Nielsen, J.R., Abild-Pedersen, F., Nørskov, J.K. *Nature* **427**, 426 (2004).
- Henderson, M.A. *Surf. Sci.* **419**, 174 (1999).
- Henderson, M.A., Epling, W.S., Perkins, C.L., Peden, C.H.F., Diebold, U. *J. Phys. Chem. B* **103**, 5328 (1999).
- Hendriksen, B.L.M., Frenken, J.W.M. *Phys. Rev. Lett.* **89**, 046101 (2002).
- Henry, C.R. *Surf. Sci. Rep.* **31**, 231 (1998).
- Hinnemann, B., Nørskov, J.K. *Top. Catal.* **37**, 55 (2006).
- Hinnemann, B., Nørskov, J.K., Topsøe, H. *J. Phys. Chem. B* **109**, 2245 (2005).
- Ho, K.Y., Yeung, K.L. *J. Catal.* **242**, 131 (2006).
- Hofmann, S., Sharma, R., Ducati, C., Du, G., Mattevi, C., Cepek, C., Cantoro, M., Pisana, S., Parvez, A., Cervantes-Sodi, F., Ferrari, A.C., Dunin-Borkowski, R., Lizzit, S., Petaccia, L., Goldoni, A., Robertson, J. *Nano Lett.* **7**, 602 (2007).
- Honkala, K., Hellman, A., Remediakis, I.N., Logadottir, A., Carlsson, A., Dahl, S., Christensen, C.H., Nørskov, J.K. *Science* **307**, 555 (2005).
- Hughes, M.D., Xu, Y.J., Jenkins, P., McMorn, P., Landon, P., Enache, D.I., Carley, A.F., Attard, G.A., Hutchings, G.J., King, F., Stitt, E.H., Johnston, P., Griffin, K., Kiely, C.J. *Nature* **437**, 1132 (2005).
- Hutchings, G.J., Hall, M.S., Carley, A.F., Landon, P., Solsona, B.E., Kiely, C.J., Herzing, A., Makkee, M., Moulijn, J.A., Overweg, A., Fierro-Gonzalez, J.C., Guzman, J., Gates, B.C. *J. Catal.* **242**, 71 (2006).
- Häkkinen, H., Abbet, W., Sanchez, A., Heiz, U., Landman, U. *Angew. Chem. Int. Ed.* **42**, 1297 (2003).
- Iddir, H., O gut, S., Browning, N.D., Disko, M.M. *Phys. Rev. B* **72**, 081407 (2005).
- Igarashi, H., Fujino, T., Zhu, Y., Uchida, H., Watanabe, M. *Phys. Chem. Chem. Phys.* **3**, 306 (2001).

- Jacobsen, C.J.H., Dahl, S., Clausen, B.S., Bahn, S., Logadottir, A., Nørskov, J.K. *J. Am. Chem. Soc.* **123**, 8404 (2001).
- Janssens, T.V.W., Carlsson, A., Puig-Molina, A., Clausen, B.S. *J. Catal.* **240**, 108 (2006).
- Janssens, T.V.W., Clausen, B.S., Hvolbæk, B., Falsig, H., Christensen, C.H., Bligaard, T., Nørskov, J.K. *Top. Catal.* **44**, 15 (2007).
- Jensen, F., Stensgaard, I., Besenbacher, F., Klink, C. *Vacuum* **41**, 337 (1990).
- Jensen, J.A., Rider, K.B., Chen, Y., Salmeron, M., Somorjai, G.A. *J. Vac. Sci. Technol. B* **17**, 1080 (1999).
- Judai, K., Abbet, S., Worz, A.S., Heiz, U., Henry, C.R. *J. Am. Chem. Soc.* **126**, 2732 (2004).
- Kabe, T., Ishihara, A., Qian, W. *Hydrodesulfurization and Hydrogenation – Chemistry and Engineering* (Wiley-VCH, Kodansha, 1999).
- Kemper, P., Kolmakov, A., Tong, X., Lilach, Y., Benz, L., Manard, M., Metiu, H., Buratto, S.K., Bowers, M.T. *Int. J. Mass Spectrom.* **254**, 202 (2006).
- Ketteler, G., Ogletree, D.F., Bluhm, H., Liu, H.J., Hebenstreit, E.L.D., Salmeron, M. *J. Am. Chem. Soc.* **127**, 18269 (2005).
- Kim, T.S., Stiehl, J.D., Reeves, C.T., Meyer, R.J., Mullins, C.B. *J. Am. Chem. Soc.* **125**, 2018 (2003).
- Kimmel, G.A., Petrik, N.G. *Phys. Rev. Lett.* **100**, 196102 (2008).
- Koplitz, L.V., Dulub, O., Diebold, U. *J. Phys. Chem. B* **107**, 10583 (2003).
- Kuk, Y., Chua, F.M., Silverman, P.J., Meyer, J.A. *Phys. Rev. Rev. B* **41** 12393 (1990).
- Kung, H.H., Kung, M.C., Costello, C.K. *J. Catal.* **216**, 425 (2003).
- Kung, M.C., Davis, R.J., Kung, H.H. *J. Phys. Chem. C* **111**, 11767 (2007).
- Lauritsen, J.V., Besenbacher, F. *Adv. Catal.* **50**, 97 (2006).
- Lauritsen, J.V., Bollinger, M.V., Lægsgaard, E., Jacobsen, K.W., Nørskov, J.K., Clausen, B.S., Topsøe, H., Besenbacher, F. *J. Catal.* **221**, 510 (2004a).
- Lauritsen, J.V., Helveg, S., Lægsgaard, E., Stensgaard, I., Clausen, B.S., Topsøe, H., Besenbacher, F. *J. Catal.* **197**, 1 (2001).
- Lauritsen, J.V., Nyberg, M., Nørskov, J.K., Clausen, B.S., Topsøe, H., Lægsgaard, E., Besenbacher, F. *J. Catal.* **224**, 94 (2004b).
- Lauritsen, J.V., Nyberg, M., Vang, R.T., Bollinger, M.V., Clausen, B.S., Topsøe, H., Jacobsen, K.W., Lægsgaard, E., Nørskov, J.K., Besenbacher, F. *Nanotech.* **14**, 385 (2003).
- Lauritsen, J.V., Vang, R.T., Besenbacher, F. *Catal. Today* **111**, 34 (2006).
- Lauritsen, J.V., Vang, R.T., Besenbacher, F. in *Applied Scanning Probe Methods VII*, (eds) Bushan, B. and Fuchs, H. (Springer-Verlag, Berlin Heidelberg, 2007) p. 197.
- Lee, S.S., Fan, C.Y., Wu, T.P., Anderson, S.L. *J. Am. Chem. Soc.* **126**, 5682 (2004).
- Lee, S.S., Fan, C.Y., Wu, T.P., Anderson, S.L. *J. Chem. Phys.* **123**, 124710 (2005).
- Li, M., Hebenstreit, W., Gross, L., Diebold, U., Henderson, M.A., Jennison, D.R., Schultz, P.A., Sears, M.P. *Surf. Sci.* **437**, 173 (1999).
- Lim, D.C., Dietsche, R., Bubek, M., Ketterer, T., Gantefor, G., Kim, Y.D. *Chem. Phys. Lett.* **439**, 364 (2007).

- Lim, D.C., Lopez-Salido, I., Dietsche, R., Bubek, M., Kim, Y.D. *Angew. Chem. Int. Ed.* **45**, 2413 (2006).
- Liu, Z.P., Hu, P. *J. Am. Chem. Soc.* **125**, 1958 (2003).
- Logadottir, A., Moses, P.G., Hinnemann, B., Topsøe, N.Y., Knudsen, K.G., Topsøe, H., Nørskov, J.K. *Catal. Today.* **111**, 44 (2006).
- Logadottir, A., Rod, T.H., Nørskov, J.K., Hammer, B., Dahl, S., Jacobsen, C.J.H. *J. Catal.* **197**, 229 (2001).
- Longwitz, S.R., Schnadt, J., Vestergaard, E.K., Vang, R.T., Lægsgaard, E., Stensgaard, I., Brune, H., Besenbacher, F. *J. Phys. Chem. B* **108**, 14497 (2004).
- Lopez, N., Janssens, T.V.W., Clausen, B.S., Xu, Y., Mavrikakis, M., Bligaard, T., Nørskov, J.K. *J. Catal.* **223**, 232 (2004).
- Louwers, S.P.A., Prins, R. *J. Catal.* **133**, 94 (1992).
- Lundgren, E., Gustafson, J., Mikkelsen, A., Andersen, J.N., Stierle, A., Dosch, H., Todorova, M., Rogal, J., Reuter, K., Scheffler, M. *Phys. Rev. Lett.* **92**, 046101 (2004).
- Matthey, D., Wang, J.G., Wendt, S., Matthiesen, J., Schaub, R., Lægsgaard, E., Hammer, B., Besenbacher, F. *Science* **315**, 1692 (2007).
- Mcewen, J.S., Payne, S.H., Kreuzer, H.J., Kinne, M., Denecke, R., Steinrück, H.P. *Surf. Sci.* **545**, 47 (2003).
- McIntyre, B.J., Salmeron, M., Somorjai, G.A. *J. Vac. Sci. Technol. A* **11**, 1964 (1993).
- Meyer, R., Lemire, C., Shaikhutdinov, S.K., Freund, H.J. *Gold Bull.* **37**, 72 (2004).
- Miller, J.T., Kropf, A.J., Zha, Y., Regalbuto, J.R., Delannoy, L., Louis, C., Bus, E., Van Bokhoven, J.A. *J. Catal.* **240**, 222 (2006).
- Mohr, C., Hofmeister, H., Radnik, J., Claus, P. *J. Am. Chem. Soc.* **125**, 1905 (2003).
- Molina, L.M., Hammer, B. *Phys. Rev. B* **69**, 155424 (2004).
- Molina, L.M., Hammer, B. *Appl. Catal. A* **291**, 21 (2005).
- Molina, L.M., Rasmussen, M.D., Hammer, B. *J. Chem. Phys.* **120**, 7673 (2004).
- Moses, P.G., Hinnemann, B., Topsøe, H., Nørskov, J.K. *J. Catal.* **248**, 188 (2007).
- Nielsen, L.P., Besenbacher, F., Stensgaard, I., Lægsgaard, E., Engdahl, C., Stoltze, P., Jacobsen, K.W., Nørskov, J.K. *Phys. Rev. Lett.* **71**, 754 (1993).
- Niemann, W., Clausen, B.S., Topsøe, H. *Catal. Lett.* **4**, 355 (1990).
- Oetjen, H.F., Schmidt, V.M., Stimming, U., Trila, F. *J. Electrochem. Soc.* **143**, 3838 (1996).
- Ogletree, D.F., Bluhm, H., Lebedev, G., Fadley, C.S., Hussain, Z., Salmeron, M. *Rev. Sci. Instrum.* **73**, 3872 (2002).
- Okumura, M., Nakamura, S., Tsubota, S., Nakamura, T., Azuma, M., Haruta, M. *Catal. Lett.* **51**, 53 (1998).
- Onishi, H., Iwasawa, Y. *Phys. Rev. Lett.* **76**, 791 (1996).
- Österlund, L., Rasmussen, P.B., Thostrup, P., Lægsgaard, E., Stensgaard, I., Besenbacher, F. *Phys. Rev. Lett.* **86**, 460 (2001).
- Overbury, S.H., Ortiz-Soto, L., Zhu, H.G., Lee, B., Amiridis, M.D., Dai, S. *Catal. Lett.* **95**, 99 (2004).
- Ovesen, C.V., Clausen, B.S., Schiøtz, J., Stoltze, P., Topsøe, H., Nørskov, J.K. *J. Catal.* **168**, 133 (1997).

- Ozensoy, E., Goodman, D.W. *Phys. Chem. Chem. Phys.* **6**, 3765 (2004).
- Ozensoy, E., Hess, C., Goodman, D.W. *Top. Catal.* **V28**, 13 (2004).
- Park, J.B., Conner, S.F., Chen, D.A. *J. Phys. Chem. C* **112**, 5490 (2008).
- Peters, K.F., Walker, C.J., Steadman, P., Robach, O., Isern, H., Ferrer, S. *Phys. Rev. Lett.* **86**, 5325 (2001).
- Prins, R. in *Handbook of Heterogeneous Catalysis*, (eds) Ertl, G., Knözinger, H. and Weitkamp, J. (VHC, Weinheim, 1997) p. 1908.
- Prins, R., De Beer, V.H.J., Somorjai, G.A. *Catal. Rev. Sci. Eng.* **31**, 1 (1989).
- Quiros, C., Robach, O., Isern, H., Ordejon, P., Ferrer, S. *Surf. Sci.* **522**, 161 (2003).
- Radnik, J., Mohr, C., Claus, P. *Phys. Chem. Chem. Phys.* **5**, 172 (2003).
- Rasmussen, M.D., Molina, L.M., Hammer, B. *J. Chem. Phys.* **120**, 988 (2004).
- Rasmussen, P.B., Hendriksen, B.L.M., Zeijlemaker, H., Ficke, H.G., Frenken, J.W.M. *Rev. Sci. Instrum.* **69**, 3879 (1998).
- Raybaud, P., Hafner, J., Kresse, G., Kasztelan, S., Toulhoat, H. *J. Catal.* **190**, 128 (2000a).
- Raybaud, P., Hafner, J., Kresse, G., Kasztelan, S., Toulhoat, H. *J. Catal.* **189**, 129 (2000b).
- Remediakis, I.N., Lopez, N., Nørskov, J.K. *Angew. Chem. Int. Ed.* **44**, 1824 (2005).
- Reuter, K., Frenkel, D., Scheffler, M. *Phys. Rev. Lett.* **93**, 116105 (2004).
- Reuter, K., Scheffler, M. *Phys. Rev. B* **65**, 035406 (2002).
- Ricci, D., Bongiorno, A., Pacchioni, G., Landman, U. *Phys. Rev. Lett.* **97**, 036106 (2006).
- Rostrup-Nielsen, J.R., Rostrup-Nielsen, T. *CATTECH* **6**, 150 (2002).
- Rostrup-Nielsen, J.R., Sehested, J., Nørskov, J.K. *Adv. Catal.* **47**, 65 (2002).
- Rupprechter, G., Bruce, C.G.A.H.K. in *Advanced Catalysis* (Academic Press, 2007), 133.
- Rupprechter, G., Dellwig, T., Unterhalt, H., Freund, H.J. *J. Phys. Chem. B* **105**, 3797 (2001).
- Rupprechter, G., Weilach, C. *J. Phys. Cond. Mat.* **20**, (2008).
- Rössler, M., Geng, P., Wintterlin, J. *Rev. Sci. Instrum.* **76**, 023705 (2005).
- Salmeron, M., Schlögl, R. *Surf. Sci. Rep.* **63**, 169 (2008).
- Sanchez, A., Abbet, S., Heiz, U., Schneider, W.D., Häkkinen, H., Barnett, R.N., Landman, U. *J. Phys. Chem. A* **103**, 9573 (1999).
- Schmid, M., Reicho, A., Stierle, A., Costina, I., Klikovits, J., Kostelnik, P., Dubay, O., Kresse, G., Gustafson, J., Lundgren, E., Andersen, J.N., Dosch, H., Varga, P. *Phys. Rev. Lett.* **96**, 146102 (2006).
- Schubert, M.M., Hackenberg, S., Van Veen, A.C., Muhler, M., Plzak, V., Behm, R.J. *J. Catal.* **197**, 113 (2001).
- Schumacher, B., Plzak, V., Kinne, M., Behm, R.J. *Catal. Lett.* **89**, 109 (2003).
- Shuxian, Z., Hall, W.K., Ertl, G., Knözinger, H. *J. Catal.* **100**, 167 (1986).
- Somorjai, G.A. *Surface Chemistry and Catalysis* (Wiley, New York, 1994).
- Sterrer, M., Fischbach, E., Risse, T., Freund, H.J. *Phys. Rev. Lett.* **94**, 186101 (2005).
- Stoltze, P. *Prog. Surf. Sci.* **65**, 65 (2000).

- Stoltze, P., Nørskov, J.K. *Phys. Rev. Lett.* **55**, 2502 (1985).
- Studt, F., Abild-Pedersen, F., Bligaard, T., Sørensen, R.Z., Christensen, C.H., Nørskov, J.K. *Science* **320**, 1320 (2008).
- Su, X.C., Cremer, P.S., Shen, Y.R., Somorjai, G.A. *Phys. Rev. Lett.* **77**, 3858 (1996).
- Suzuki, S., Fukui, K., Onishi, H., Iwasawa, Y. *Phys. Rev. Lett.* **84**, 2156 (2000).
- Taylor, H.S. *Proc. R. Soc. London, Ser. A* **108**, 105 (1925).
- Tersoff, J., Hamann, D.R. *Phys. Rev. Lett.* **50**, 1998 (1983).
- Tersoff, J., Hamann, D.R. *Phys. Rev. B* **31**, 805 (1985).
- Thostrup, P., Vestergaard, E.K., An, T., Lægsgaard, E., Besenbacher, F. *J. Chem. Phys.* **118**, 3724 (2003).
- Tong, X., Benz, L., Kemper, P., Metiu, H., Bowers, M.T., Buratto, S.K. *J. Am. Chem. Soc.* **127**, 13516 (2005).
- Topsøe, H., Clausen, B.S., Massoth, F.E. *Hydrotreating Catalysis* (Springer Verlag, Berlin-Heidelberg, 1996).
- Topsøe, H., Hinemann, B., Nørskov, J.K., Lauritsen, J.V., Besenbacher, F., Hansen, P.L., Hytoft, G., Egeberg, R.G., Knudsen, K.G. *Catal. Today* **107–08**, 12 (2005).
- Topsøe, N.-Y., Topsøe, H. *J. Electron. Spectrox. Relat. Phenom.* **39**, 11 (1986).
- Topsøe, N.-Y., Topsøe, H. *J. Mol. Catal. A: Chem.* **141**, 95 (1999).
- Valden, M., Lai, X., Goodman, D.W. *Science* **281**, 1647 (1998).
- Vang, R.T., Honkala, K., Dahl, S., Vestergaard, E.K., Schnadt, J., Lægsgaard, E., Clausen, B.S., Nørskov, J.K., Besenbacher, F. *Nature Mater.* **4**, 160 (2005a).
- Vang, R.T., Honkala, K., Dahl, S., Vestergaard, E.K., Schnadt, J., Lægsgaard, E., Clausen, B.S., Nørskov, J.K., Besenbacher, F. *Surf. Sci.* **600**, 66 (2006).
- Vang, R.T., Lægsgaard, E., Besenbacher, F. *Phys. Chem. Chem. Phys.* **9**, 3460 (2007).
- Vang, R.T., Wang, J.G., Knudsen, J., Schnadt, J., Lægsgaard, E., Stensgaard, I., Besenbacher, F. *J. Phys. Chem. B* **109**, 14262 (2005b).
- Vestergaard, E.K., Thostrup, P., An, T., Lægsgaard, E., Stensgaard, I., Hammer, B., Besenbacher, F. *Phys. Rev. Lett.* **88**, 259601 (2002).
- Wagner, J.B., Hansen, P.L., Molenbroek, A.M., Topsøe, H., Clausen, B.S., Helveg, S. *J. Phys. Chem. B* **107**, 7753 (2003).
- Wahlström, E., Lopez, N., Schaub, R., Thostrup, P., Rønnau, A., Africh, C., Lægsgaard, E., Nørskov, J.K., Besenbacher, F. *Phys. Rev. Lett.* **90**, 026101 (2003).
- Wang, J.G., Hammer, B. *Phys. Rev. Lett.* **97**, 136107 (2006).
- Wang, J.G., Hammer, B. *Top. Catal.* **44**, 49 (2007).
- Wendt, S., Kim, Y.D., Goodman, D.W. *Prog. Surf. Sci.* **74**, 141 (2003).
- Wendt, S., Matthiesen, J., Schaub, R., Vestergaard, E.K., Lægsgaard, E., Besenbacher, F., Hammer, B. *Phys. Rev. Lett.* **96**, 066107 (2006).
- Wendt, S., Schaub, R., Matthiesen, J., Vestergaard, E.K., Wahlström, E., Rasmussen, M.D., Thostrup, P., Molina, L.M., Lægsgaard, E., Stensgaard, I., Hammer, B., Besenbacher, F. *Surf. Sci.* **598**, 226 (2005).
- Wendt, S., Sprunger, P.T., Lira, E., Madsen, G.K.H., Li, Z.S., Hansen, J.Ø., Matthiesen, J., Blekinge-Rasmussen, A., Lægsgaard, E., Hammer, B., Besenbacher, F. *Science* **320**, 1755 (2008).
- Whitehurst, D.D., Isoda, T., Mochida, I. *Adv. Catal.* **42**, 345 (1998).

- Wiesendanger, R. *Scanning Probe Microscopy and Spectroscopy* (Cambridge University Press, Cambridge, 1994).
- Wintterlin, J., Trost, J., Renisch, S., Schuster, R., Zambelli, T., Ertl, G. *Surf. Sci.* **394**, 159 (1997a).
- Wintterlin, J., Völkening, S., Janssens, T.V.W., Zambelli, T., Ertl, G. *Science* **278**, 1931 (1997b).
- Wörz, A.S., Heiz, U., Cinquini, F., Pacchioni, G. *J. Phys. Chem. B* **109**, 18418 (2005).
- Xu, Y., Mavrikakis, M. *J. Phys. Chem. B* **107**, 9298 (2003).
- Yan, W.F., Mahurin, S.M., Pan, Z.W., Overbury, S.H., Dai, S. *J. Am. Chem. Soc.* **127**, 10480 (2005).
- Yoon, B., Häkkinen, H., Landman, U., Wörz, A.S., Antonietti, J.M., Abbet, S., Judai, K., Heiz, U. *Science* **307**, 403 (2005).
- Zambelli, T., Trost, J., Wintterlin, J., Ertl, G. *Phys. Rev. Lett.* **76**, 795 (1996a).
- Zambelli, T., Wintterlin, J., Trost, J., Ertl, G. *Science* **273**, 1688 (1996b).
- Zangwill, A. *Physics at Surfaces* (Cambridge University Press, Cambridge, 1988).
- Zeuthen, P., Skyum, L. *Hydrocarbon Eng.* **11**, 65 (2004).
- Zhang, Z., Bondarchuk, O., Kay, B.D., White, J.M., Dohnalek, Z. *J. Phys. Chem. B* **110**, 21840 (2006).

# 13

# Bifunctional nanomaterials for the imaging and treatment of cancer

13.1 Introduction	474
13.2 Thermal ablative therapy in cancer	475
13.3 Nanomaterial applications	481
13.4 Gold nanoshells and nanorods	492
13.5 Iron-oxide nanoparticles	496
13.6 Conclusions and future directions	496
Acknowledgments	497
References	497

*A. Burke, D. Carroll, F.M. Torti, and S.V. Torti*

## 13.1 Introduction

Cancer represents a significant burden with regards to both health and economy the world over. The World Health Organization projects global cancer incidence to increase 50% to 15 million new cases by 2020 (Stewart and Kleihues 2003). In the United States alone, the National Cancer Institute documented 1 444 920 new cases resulting in 559 650 disease-related deaths in 2007 (Ries *et al.* 2007). While incremental improvements in conventional therapeutic approaches have increased overall survival, for many cancers the benefits of treatment can be modest. For example, NCI data indicates that modern treatments fail nearly 40% of cancer patients each year (Ries *et al.* 2007). Thus, there is ample rationale for the search for novel and more efficacious cancer therapies.

The application of nanoscale materials for the diagnosis and treatment of disease (termed “nanomedicine”) represents one such novel approach to the treatment of cancer (Ferrari 2005; Jain 2008). Following the discovery of carbon nanotubes in 1991 by Iijima, the field of nanotechnology has expanded rapidly, leading to the synthesis and characterization of a variety materials whose principal dimensions are in the nanometer range (Iijima 1991; Dai 2002). Dissimilar to the macroscopic formulation of a given material, the nanomaterial may exhibit beneficial emergent properties that can significantly enhance its therapeutic utility.

Several nanomaterials possess properties desirable for a cancer therapy, and exploration of these materials as anticancer agents has begun. Those that have received the most attention include encapsulated iron oxides, single- and multiwalled carbon nanotubes, gold nanorods and gold nanoshells (Hirsch *et al.* 2003; Kam *et al.* 2005; Huang *et al.* 2006; Torti *et al.* 2007). This chapter reviews these nanomaterials as they seek to offer clinically relevant,

minimally invasive therapeutic options for the treatment of local, non-resectable cancers. Each approach utilizes a unique photosensitizing nanomaterial capable of generating heat in response to exposure to electromagnetic radiation, resulting in the thermal ablation of surrounding tissue. Tumor specificity is achieved through the use of conjugated antibody targeting or passive accumulation via the enhanced permeability and retention effect. In addition, each material can serve as a contrast agent when used in conjunction with existing optical or magnetic resonance imaging modalities, permitting the *in vivo*-visualization and location of the target lesion. Such nanoscale materials specifically fabricated for the treatment of cancer have a unique potential to provide integrated imaging and treatment modalities.

## 13.2 Thermal ablative therapy in cancer

Ablation of cancerous lesions and consequent clinical remission of the disease is the express goal of any cancer treatment undertaken with curative intent. Surgery, chemotherapy, and radiation are the tools most commonly used to attain this goal. For most malignancies, surgery, if appropriate, remains the standard of care, as it is the pre-dominant modality through which cure can be achieved (Lenhard *et al.* 2001). However, sizable subpopulations of cancer patients are ineligible for surgery, due to factors such as location and extent of the primary lesion, poor performance status or other preclusive comorbidities (Lenhard *et al.* 2001). In recent decades, several approaches have been developed that exploit the differing sensitivities of normal and malignant tissue to a thermal insult to treat such patients (Overgaard and Suit 1979; Dickson and Calderwood 1980). Perhaps the most developmentally advanced applications of this approach (all are clinically available) are interstitial laser coagulation, high-intensity focused ultrasound, radio-frequency and microwave ablation (Izzo 2003; Nikfarjam and Christophi 2003; Kelekis *et al.* 2006; McDannold *et al.* 2006). These techniques are only moderately invasive and have generally tolerable toxicities, making them more broadly applicable to the cancer-patient population. Each therapy utilizes a distinct carrier wave to transduce energy through the body to the target lesion, whereupon the energy is absorbed and converted to heat. The heating of cells to supraphysiologic temperatures induces distinct, temperature-dependent alterations in morphology, metabolism and viability that will be addressed in a subsequent section of this chapter.

Thermal treatment of cancer can be divided into two basic types: hyperthermic and thermal ablative. The principal distinctions between the two lie in the range of temperatures achieved, the duration of treatment and their clinical indications for use. Hyperthermic cancer therapy has been an effective adjuvant treatment (used in conjunction with chemo- and radiotherapy) for many cancers and ongoing clinical trials continue to validate the technique and generate new indications for its use (Datta *et al.* 1990; Dahl and Overgaard 1995; Overgaard *et al.* 1996; Snead *et al.* 1998). Hyperthermic therapy seeks to elevate the temperature of either the whole body or isolated regions of the body to a range between 40–46 °C from 1 to several hours.

This has the direct cytotoxic effects described elsewhere in this chapter in addition to enhancing the efficacy of subsequent regimens of chemo- and radiotherapy (Dewhirst *et al.* 1997; Dahl *et al.* 1999). Alternatively, thermal ablative therapy (several of the techniques of this therapy are described above) utilizes differing energy transducers to rapidly heat a small area of tissue (<8 cm in diameter) to much higher temperatures (55–90 °C) for a short period of time, generally <10 min per treatment (Nikfarjam *et al.* 2005). To date, uses of nanomaterials in cancer therapy have focused on thermal tumor ablation.

Experimental thermal therapies employing nanomaterials must be designed with consideration for the biological environment with which they must contend. This chapter will begin with an introduction to the processes in the tumor that these therapies are designed to disrupt and a discussion of countermeasures a tumor can employ to attenuate these effects. These factors bear directly on the success or failure of the therapy and must be accounted for in the design and capabilities of all nano-based thermal therapies.

### **13.2.1 Mechanisms of thermotoxicity**

Broadly speaking, a therapy (especially when considering a cancer therapy) operates on two primary principles: (1) that there exists a biological target in the organism where the treatment exerts its intended effect, and (2) the sensitivity of the true target differs (ideally appreciably) from that of non-targets such that it gives rise to a useful therapeutic index. Thermal therapies satisfy these criteria, since transformed cells exhibit enhanced susceptibility to supraphysiologic temperatures (Overgaard and Suit 1979; Dickson and Calderwood 1980; Fajardo *et al.* 1980; Henle and Dethlefsen 1980; Everts 2007). Analogous to therapeutic radiation exposure, heating a cell induces a multitude of alterations in cellular processes as diverse as cytoskeletal dynamics, enzyme structure and function and overall metabolism (Dewey 1989; Harmon *et al.* 1990; Heisterkamp *et al.* 1999; Hildebrandt *et al.* 2002; Nikfarjam *et al.* 2005; Roti 2008). Thus, despite decades of investigation, the definitive proximate cause of heat-induced cell death remains elusive. Several lines of evidence do, however, implicate selective events that correlate well with cell death, and these are discussed in the following sections.

#### *13.2.1.1 The targets of heat*

Cellular constituents most often associated with hyperthermic cell death include alterations in the structure and function of cellular plasma membranes, cytoskeleton, DNA and proteins (Dewey 1989; Heisterkamp *et al.* 1999; Hildebrandt *et al.* 2002; Nikfarjam *et al.* 2005; Coffey *et al.* 2006; Roti 2008).

#### *13.2.1.2 Heat-induced alterations of the plasma membranes*

The plasma membrane of a cell and its various membrane-bound organelles serve the vital purpose of defining the extent of the cytosol and, through its general impermeability, permitting the establishment of essential electrochemical

gradients through which work may be done. Thus, the integrity of the plasma membrane is essential to the survival of any cell.

Elevating the temperature of a cell serves to alter the fluidity of its membranes and destabilizes several of these essential electrochemical gradients (Konings and Ruifrok 1985). Indeed, several reports indicate that exposure to hyperthermic and thermal ablative temperatures decreases the functionality of many membrane-bound ion pumps resulting in marked increases in intracellular sodium and calcium concentrations (Konings and Ruifrok 1985; Vidair and Dewey 1986; Majda *et al.* 1994; Hildebrandt *et al.* 2002). Of these changes, perhaps the influx of calcium is the most significant, resulting in a marked decline in mitochondrial membrane potential ( $\Delta\Psi_m$ ) and stalling of ATP synthase (Klaassen 2008). As ion concentrations equilibrate with the extracellular space, sensitive proteins begin to denature and aggregate. Such an event in the mitochondrial membrane is thought to drive the mitochondrial permeability transition, leading to the release of various pro-death factors (such as cytochrome *c* and the protein Smac/Diablo) (Kroemer *et al.* 1998; Lemasters *et al.* 2002; Kim *et al.* 2003a,b). Higher cellular temperatures, such as those seen during thermal ablative treatment, can expedite this process through lysis of the plasma membrane and direct exposure of the cytosol to the extracellular milieu. The end result of this cascade is either apoptotic (hyperthermia) or necrotic (thermal ablative) cell death (Harmon *et al.* 1990) (discussed further below).

#### 13.2.1.3 Heat-induced alterations to DNA

During the thermal treatment of cancer, DNA damage is induced both directly and indirectly as determined by the severity of the heat shock (Roti 2008). In the thermal ablative temperature range (55–90 °C) the effect on DNA is direct; resulting in the denaturation of the antiparallel strands and possible strand breakage due to the unstabilized abnormal conformation (Dewey *et al.* 1971; Defer *et al.* 1977; Lewis 1977). In the hyperthermic ranges (40–46 °C) the role of DNA damage is less well defined. It is well known that the hyperthermic heat shock of a cell traversing S-phase results in chromosomal aberrations (Dewey *et al.* 1990; Wong *et al.* 1993). Although some reports have documented the formation of  $\gamma$ H2AX complexes (a phosphate-modified histone that is a surrogate marker for DNA double-strand breaks) following temperatures in the hyperthermic range (Olive 2004; Takahashi *et al.* 2004), others have failed to observe DNA fragmentation through pulsed gel electrophoresis in heat-shocked cells, despite the detection of  $\gamma$ H2AX complexes in those same cells (Hunt *et al.* 2007). Thus, the role of DNA damage in hyperthermia is uncertain, although it is of clear importance in the response to heat exposure in the thermoablative range.

#### 13.2.1.4 Heat-induced alterations to proteins

Of all the cellular macromolecules that could be critical targets of heat, proteins are perhaps the most likely, due to their abundance and the variety of vital functions they perform, as well as to their dependence on weak molecular interactions for structural and functional integrity (Nelson and Cox 2005).

The nuclear matrix (Berezney and Coffey 1974) is a protein network that provides the structural support and spatial regulation necessary to manage a cell's  $5 \times 10^9$  base pairs of DNA as well as an additional mass of associated proteins that is nearly 6 times that of the DNA alone (Roti 2008). This matrix facilitates most vital DNA-related events such as transcription, replication and repair. Differential scanning calorimetry of isolated cell nuclei has demonstrated that the nuclear matrix is one of the most thermally sensitive cellular structures, exhibiting marked changes in response to temperatures as low as 40 °C (Lepock 2003, 2004, 2005). Additionally, as other nuclear proteins unfold and aggregate in response to heat shock, these aggregates bind the nuclear matrix and prevent normal DNA–matrix interactions. These events are especially critical during the S-phase of the cell cycle when DNA replication occurs. At this time, the nuclear matrix serves to scaffold an open DNA conformation and facilitate the assembly of the replicative machinery. Heat-induced alterations to this process are thought to be one of the causative mechanisms of replication fork stalls and DNA strand breaks observed in heat-shocked cells. This likely underlies the observation that S-phase cells are most sensitive to hyperthermic cell death (Wong *et al.* 1993). This has special significance in regard to the treatment of cancer, where many tumors have high mitotic indices and thus elevated S-phase cell fractions.

Another critical protein target of heat is the cytoskeleton. This structure is composed of three principal elements: microtubules, intermediate filaments and actin filaments, and is responsible for cell morphology, movement and various signalling processes. All three cytoskeletal constituents are susceptible to damage by heat through a function that varies by thermal dose and cell type. Studies have shown that at moderate hyperthermic heat doses (43 °C) microtubules depolymerize, leading to the disruption of microtubule organization centers and mitotic spindles (Coss *et al.* 1982; Coss and Linnemanns 1996). Such an event would have significant consequences in an M-phase cell, and may represent a contributing factor in the death of mitotic cells. Loss of pro-survival signalling mediated by heat-induced shedding of integrin adhesion molecules and depolymerization of actin fibers may also contribute to hyperthermic cell death (Shih-Horng Huang 1999).

### 13.2.1.5 Terminal events: Apoptosis or necrosis

Pathways of cell death have been broadly grouped into apoptotic or necrotic (Buja *et al.* 1993; Lenhard *et al.* 2001). Unlike necrosis, apoptosis is a highly ordered, energy-dependent process that does not induce a host inflammatory response (Bursch *et al.* 1992; Klaassen 2008). Necrosis generally results from overwhelming cellular trauma, which is believed to trigger cell death at least in part by disrupting mitochondrial function, triggering bioenergetic catastrophe and ultimately cell lysis (Lemasters *et al.* 2002; Kim *et al.* 2003a,b). The resulting loss of integrity and dissemination of various bioactive, formerly cytoplasmic macromolecules throughout the interstitium leads to the recruitment of inflammatory immune cells (neutrophils and macrophages).

Available data suggest that both death processes are induced by heat treatments. At hyperthermic temperatures (below 46 °C) apoptosis pre-dominates (Harmon *et al.* 1990). This has the benefit of not inducing the recruitment of inflammatory cells that, along with removing necrotic cellular debris, also release various pro-growth, pro-angiogenic cytokines that can enhance the expansion and recovery of malignant cells not destroyed in the initial treatment (Bursch *et al.* 1992; Cain 2003). However, therapies that work predominantly through apoptotic mechanisms often fail, as tumors invariably contain cell populations insensitive to apoptotic stimuli (Boehrer *et al.* 2006; Eberle *et al.* 2007). For thermal therapies operating in the ablative temperature range, cellular resistance is not a concern (see below for supracellular mechanisms of resistance) as research has demonstrated they work almost entirely through coagulative necrosis (Heisterkamp *et al.* 1999; Nikfarjam *et al.* 2005a,b). As the necrotic process does not require cellular “cooperation” to induce death, the entire tumor is generally susceptible to the pro-death thermal insult (Lemasters *et al.* 2002; Kim *et al.* 2003). However, if the ablated region fails to encompass the entire tumor, the residual disease may be stimulated to recur through the eventual influx of inflammatory cells. These will likely remain important considerations for any heat-based cancer therapy.

### 13.2.2 Mechanisms of thermoresistance

When faced with thermal-therapy-induced cytotoxic insults, cancerous cells do not passively transition to apoptosis or necrosis. On the contrary, they trigger evolutionarily conserved cytoprotective mechanisms designed to attenuate cytotoxic effects and maintain viability (Williams *et al.* 2007). Two of these figure prominently in the failure of heat-based therapies: the heat-sink effect and heat-shock proteins (Nikfarjam *et al.* 2005).

#### 13.2.2.1 Heat-sink effect

The efficacy of thermal-ablation therapies is dependent on their ability to uniformly reach and maintain a given temperature in a target tissue volume for a defined length of time (Heisterkamp *et al.* 1999). This can be hindered when the target treatment area abuts a large, well-perfused vessel where the transfer of heat into the blood facilitates its dissipation throughout the body (Patterson *et al.* 1998; Lu *et al.* 2002, 2003). This is termed the heat-sink effect. This event is widely encountered in highly perfused tissues such as liver and kidney. Current strategies to mitigate this effect include the surgical isolation and temporary occlusion of the offending vessel, either by clamping or catheter-mediated embolization. The resultant decrease in local blood flow enhances tissue heating and helps to decrease the likelihood of undertreated regions and local recurrence.

#### 13.2.2.2 Heat-shock proteins

A family of stress-inducible proteins known collectively as heat-shock proteins (HSPs) plays a central role in a cell’s response to thermal insult. The protein constituents of this diverse family range in size from 10 to >100 kDa and

exist at basal levels throughout the various cellular compartments where they serve a variety of housekeeping functions (Rylander *et al.* 2005). Induction of HSPs can occur in response to a diverse array of stimuli from environmental stressors like heat shock, a pathologic febrile state or to normal developmental conditions such as cytokine exposure (Schlesinger 1990). Induction of these proteins is believed to occur through a mechanism involving release from cytoplasmic binding proteins into the nucleus, where they induce transcription of cytoprotective proteins, including the heat-shock proteins themselves (Kiang and Tsokos 1998; Morimoto and Santoro 1998; Wang *et al.* 2003; Rylander *et al.* 2005). For the purposes of thermotolerance, two HSPs in particular have been implicated: HSP 27 and HSP 70.

The thermoprotective effect of HSP expression is believed to result from their ability to serve as chaperones to misfolded proteins (Alberts *et al.* 1994; Rylander *et al.* 2005). This function prevents the lethal accumulation of denatured protein aggregates in the cell. To that same end, HSP may also be involved in the trafficking of misfolded/denatured proteins to other organelles where they are fated for either refolding or degradation (Rylander *et al.* 2005).

HSPs represent a means by which cancerous cells escape thermally induced death. This protective state is termed thermotolerance and describes the condition in which a cell previously exposed to elevated but sublethal temperatures can resist the effects of further heating for an extended period of time (Subjeck *et al.* 1982a,b; Li *et al.* 1995). The duration of the resistive period varies as a direct function of the intensity of the initial thermal insult, with higher initial temperature exposures leading to more protracted periods of thermotolerance that may last as long as 5 days (Wang *et al.* 2003). During this time, cells are not only desensitized to additional temperature increases, but also to other pro-apoptotic stimuli (Fortin *et al.* 2000; Latchman 2001; Rylander *et al.* 2005). Indeed, the elevated expression level of HSPs during the thermotolerant period has been implicated in the resistance of several cancers to both chemo- and radiotherapy, likely through their ability to attenuate the pro-apoptotic effects of p53 (Levine 1989; Levine *et al.* 1991; Tomei and Cope 1991; Ciocca *et al.* 1993; Samali and Cotter 1996; Rylander *et al.* 2005).

It is important to note that the varied activities of the HSPs cannot protect a cell against an acute, high-temperature (55–90 °C) exposure, as would be produced during an optimal course of thermoablative therapy (Nikfarjam *et al.* 2005). However, current thermoablative techniques and temperature-monitoring strategies conspire to make it difficult to ensure that such temperatures are uniformly reached throughout the entire tumor volume, especially at the marginal extremes of the target neoplasm, and it is in this cell population that the protective effects of HSPs are most worrisome. These undertreated cells on the tumor periphery that are now both thermotolerant and chemo/radioresistant are primed to become sites of treatment failure and, ultimately, disease recurrence.

Interventions preventing or greatly attenuating the ability of a transformed cell to mount an effective heat-shock response could be of great utility in boosting the efficacy of heat-based cancer therapies (Morimoto and Santoro

1998; Calderwood and Asea 2002), and are currently being explored (Rossi *et al.* 2006; Sanguino *et al.* 2008).

## 13.3 Nanomaterial applications

### 13.3.1 Nanomaterials in thermal therapy

As summarized above, thermal therapy is a commonly accepted technique in the treatment of cancer. Heat is used to both enhance the efficacy of other therapies (hyperthermia) as well as in direct tumor ablation. Limitations of current approaches include inadvertent treatment of non-malignant adjacent tissue, incomplete treatment of tumor margins, and inadequate heat delivery, with consequent activation of resistance mechanisms.

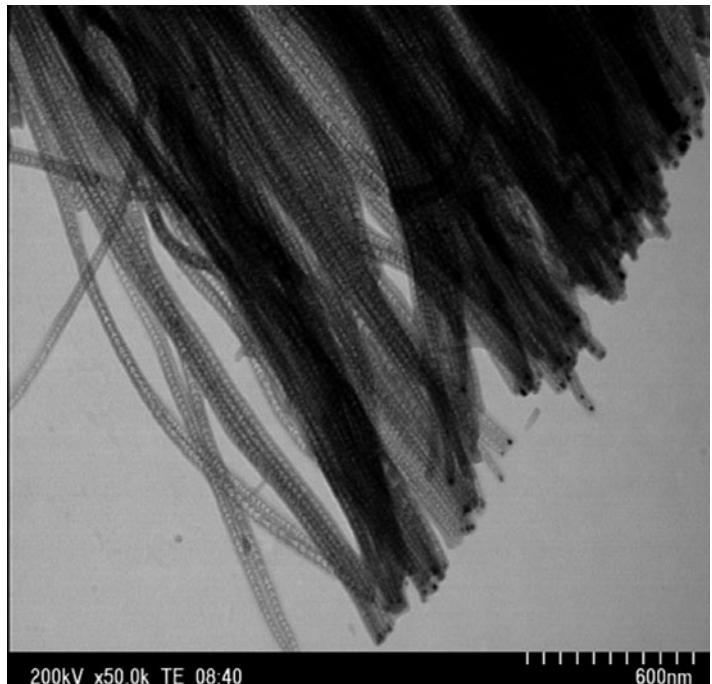
Due to their capacity to deliver substantial heat in a non-invasive manner, their potential to deliver multiple rounds of heat therapy, and the ability to utilize their imaging potential for more precise localization of heat delivery, nanomaterials are being actively investigated for their potential to improve the thermal treatment of cancer. Although this is a relatively new area of investigation, several different types of nanomaterials have been explored in this regard. Each of these is described in greater detail below (Table 13.1).

#### 13.3.1.1 Physical properties of single- and multiwalled carbon nanotubes

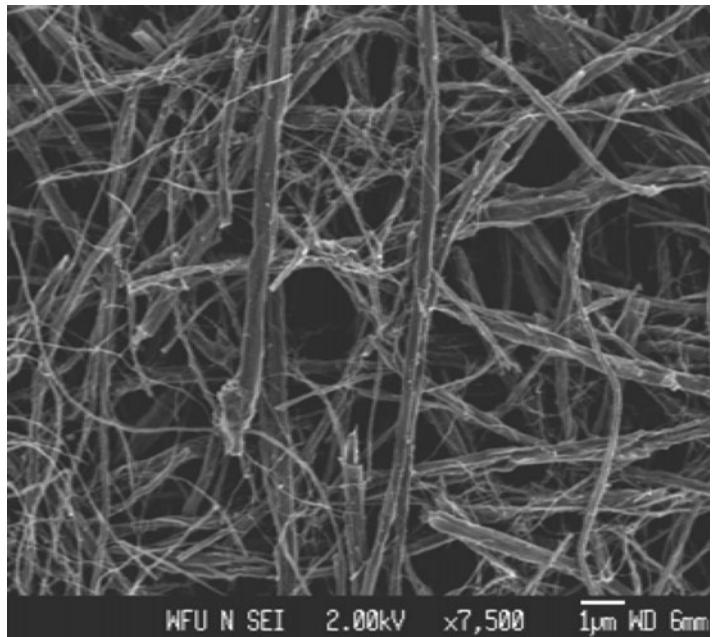
Carbon nanotubes (CNTs) are cylindrical, carbon molecules that range from 1.4 nm in diameter for a single tube (SWNT) (Fig. 13.1), to 30 nm to 50 nm for concentrically arranged tubes (MWNT) (Fig. 13.2), with lengths in both cases of micrometers or more (Iijima 1991). These structures exhibit extraordinary strength; CNTs are the strongest known fibers derived from their carbon–carbon graphitic bonds. They also exhibit unique electrical properties including ballistic transport, and are highly efficient conductors of heat.

**Table 13.1** Physical properties of nanoparticles designed for thermal ablative cancer therapies.

Particle	Size	Composition	References
Nanoshells	50–500 nm	Silica core surrounded by metal (usually gold) shell	Hirsch <i>et al.</i> (2006)
Gold Nanorods	50–100 nm	Reduced gold	Huang <i>et al.</i> (2007)
Super-paramagnetic Iron Oxide	5–10 nm	Magnetite ( $\text{Fe}_3\text{O}_4$ ) and Maghemite ( $\gamma\text{-Fe}_2\text{O}_3$ )	Gneveckow <i>et al.</i> (2004)
Single-Wall Carbon Nanotubes	1.4 nm diameter; lengths to micrometers	Carbon	Iijima (1991) Kam <i>et al.</i> (2005)
Multiwalled Carbon Nanotubes	30–50 nm diameters; Lengths to micrometers	Primarily carbon; may be doped with nitrogen, boron, etc.	Torti <i>et al.</i> (2007) Liu <i>et al.</i> (2005)



**Fig. 13.1** As-grown SWNTs imaged by TEM. D. Carroll, unpublished.



**Fig. 13.2** Multiwalled nanotubes. D. Carroll, unpublished.

The specific heat and thermal conductivity numbers of CNTs are among the highest known in materials (Brigger *et al.* 2002).

The unique electronic properties of carbon nanotubes come from the quantum confinement of electrons around the circumference of the nanotube. Radially, electrons are confined to reside on the monolayer thickness of the graphene sheets making up the tube. Because there are various ways in which the lattice of the single-walled nanotube can be twisted down its axis, the overall unit cell of the carbon lattice can be composed of very large or very small numbers of graphene hexagons. This means that nanotubes can range from being very good metals to being semiconductors with rather large bandgaps; extraordinary for a single-element molecule (Brigger *et al.* 2002). In the case of multiwalled nanotubes, the concentric walls of the tubes usually have one or more that are metals, so all multiwalled nanotubes are metals or small-bandgap semiconductors.

The interaction of light with the nanotubes structure is of particular importance in a number of applications, including biomedical. The photophysical effects related to low dimensionality and quantum confinement of electrons in single-walled nanotubes have been studied through Raman scattering, fluorescence, non-linear optical properties, etc., yet this is still quite an unexplored area of research (Hui 2005). What we do know is that SWNTs exhibit a wide variety of different vibrational modes, which result from the large number of carbon atoms in the unit cell of the nanotube. In general, more phonon modes appear as the nanotube diameter and the size of the unit cell increases, meaning that the ability of heat transport and heat capacity is unique to the nanotube structure.

Likewise in MWNTs a large number of phonon modes mean that these structures are exceptional heat conductors. However, in contrast to the SWNT, the larger numbers of electrons available in the MWNT for transport, together with the small bandgap, or metallic behavior suggests that these materials act much more like a classical antenna; coupling highly efficiently to radiation when the length of the nanotubes are comparable to half that of the wavelength of the incident radiation (Hanson 2005). This means the electrons in the nanotube can more easily become excited and begin releasing excess energy in the form of heat. Further, radiation captured in this way can be scattered (or transferred) efficiently to the surrounding media creating enhanced dielectric loss in the material, serving to enhance the heating of any host material in which the nanotube is embedded.

The substitutional doping of carbon nanotubes is a desirable method to control the electronic properties of the tubes by the chemistry rather than the specific geometry of the tubes. This allows for the engineering electronic band structure to increase the overall conductivity (hence the antenna behavior) of the nanotube. Carbon nanotubes can be doped in different ways, including intercalation of electron donors like alkali metals or acceptors like halogens, substitutional doping, encapsulating in the interior space, coating on the surface, molecular absorption, covalent sidewall functionalization, etc. Substitution of the carbon atoms with electron donors like nitrogen (N) or acceptors like boron (B) is the method of doping we employed. It has been found that the impurity level induced by N is located 0.27 eV below the bottom

of the conduction bands and the B-induced level is 0.16 eV above the top of the valence bands (Schonenberger and Forro 2000).

For pure carbon nanotubes, the valence and conduction bands appear to be symmetric about the Fermi level. The metallic behavior of B-doped nanotubes characterized by the local density of states (LDOS) can be investigated by scanning tunnelling microscopy and spectroscopy. It has been found that undoped carbon nanotubes show a small bandgap (semiconducting or semi-metallic behavior), whereas for the B-doped MWNTs the bandgap is filled from the valence-band side with a prominent acceptor-like peak near the Fermi level—this indicates that the outermost tube cylinders are structurally perfect with few defects. For the undoped tubes, the DOS is symmetric around the bandgap for the 1st and 2nd van Hove singularities—asymmetries arise due to the mixing of p and s orbitals (Dai 2002). Doping of the tubes leads to a lowering of the Fermi level into the valence band of the undoped tube. The stronger the doping, the stronger is also the shift of the Fermi level. In relation to transport properties it has also been found that this lowering of the Fermi level by substituted B dopants also increases the number of conduction channels without introducing strong carrier scattering. B-doped nanotubes show metallic behavior with weak electron–phonon coupling and the resistance increases at lower temperatures. Substitutional doping also significantly enhances the value of third-order optical non-linear coefficients. For doped nanotubes this enhancement can be by an order of magnitude with respect to C<sub>60</sub>. This is particularly useful for photonic applications that require large second hyperpolarizabilities.

Finally, carbon nanotubes are also good candidates for optical-limiting applications as they exhibit non-linear scattering. Enhanced optical-limiting behaviors of B- and N-doped nanotubes are observed by varying the incident energy and measuring the transmitted energy. In comparison with the non-linear transmittance versus incident fluence of pure and B- or N-doped carbon nanotubes at 532 nm and 1064 nm, doped nanotubes are found to have better optical-limiting properties (lower threshold values) than pure nanotubes (Schonenberger and Forro 2000).

### 13.3.1.2 *Applications of single- and multiwalled carbon nanotubes*

Given the multitude of exploitable properties detailed above, it is perhaps not surprising that carbon nanotubes of every variety have been studied as potential means to enhance the treatment of human malignancies (Kim 2007) (Table 13.2). A review of the literature suggests that two forms, the single- and multiwalled variants, have progressed the furthest in development, and this section will focus on these exclusively.

### 13.3.1.3 *Anticancer activity of carbon nanotubes in vitro*

The efficacy of both single- and multiwalled carbon nanotubes in the thermal treatment of cancer was initially demonstrated in cell culture. In 2005, the laboratory of Dr. Hongjie Dai published an initial study on the use of single-walled carbon nanotubes (SWNT) for the thermal treatment of cancer. This publication demonstrated that single-walled nanotubes efficiently absorb radiation in the near-infra-red range ( $\lambda = 700\text{--}1100\text{ nm}$ ) and convert it to heat, which is dissipated in the surrounding medium. Absorbance was robust enough

**Table 13.2** Use of nanoparticles in anticancer thermal therapy.

Particles	Applications	Imaging modality	Targeting strategies	Advantages	Limitations	Developmental stage
Nanoshells <sup>1</sup>	-Imaging contrast agent -Photothermal ablation	Optical Coherence Tomography	-EPR effect -Antibody conjugation -Direct injection	-Tunable EM absorption and reflection spectra -Ease of gold conjugation chemistry -Uniform particle sizes	Not proven visible by standard clinical imaging modalities	Phase I Clinical Trial: Head and neck cancer
Nanorods <sup>2</sup>	-Imaging contrast agent -Photothermal ablation	Optical Coherence Tomography	-Antibody conjugation	-Tunable EM absorption and reflection spectrums -Ease of gold conjugation chemistry -Uniform particle sizes	Not proven visible by standard clinical imaging modalities	Pre-Clinical: Cell-culture systems
Iron-Oxide Nanocrystals <sup>3</sup>	-Imaging contrast agent -Thermal ablation	MRI, CT	-Direct injection	-No depth of penetration issues with alternating magnetic fields	Requires direct injection	Phase I/II Clinical Trials: Brain and prostate cancers
Single-Wall Carbon Nanotubes <sup>4</sup>	-Imaging agent -Photothermal ablation -Drug carrier	Raman Spectroscopy	-EPR effect -Antibody conjugation -Targeting Peptides	-Potential for tuneable absorption spectrum. -Low toxicity	Particle sizes not uniform	Pre-Clinical: Animal models
Multiwalled Carbon Nanotubes <sup>5</sup>	-Imaging contrast agent -Photothermal Ablation -Drug carrier	MRI	-Direct injection -EPR effect -Antibody conjugation	-Efficient heat generation -Doping adds new functionalities -Low toxicity	Particle sizes not uniform	Pre-Clinical: Animal models

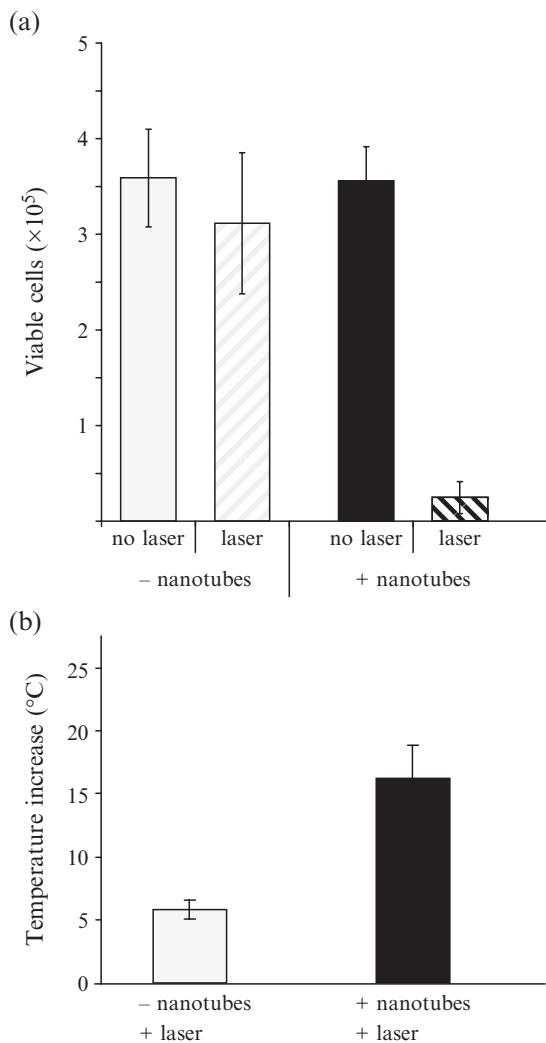
Key References: 1) Hirsch *et al.* (2006); 2) Huang *et al.* (2005); 3) Gnevcekow *et al.* (2004); 4) Kam *et al.* (2005); and 5) Torti *et al.* (2007).

to raise the temperature of the SWNT-containing solution to over 55 °C (i.e. into the established thermoablative range) during 60 s of exposure to 808-nm laser radiation (Kam *et al.* 2005). The significance of strong absorbance in this wavelength range is considerable, since biological bodies themselves generally have weak absorbance in the near-infra-red segment of the EM spectrum (Konig 2000; Weissleder 2001). This property is fundamental to the application of this and other nanomaterials (indeed, most studies discussed in this chapter utilize NIR to stimulate the release of heat from their respective nanomaterials) that seek to treat non-superficial cancerous lesions *in vivo*.

This initial study also demonstrated that the heat generated following laser radiation of SWNT was sufficient to kill HeLa cells (a well-characterized cervical carcinoma cell line) that had internalized SWNT during an incubation period (Kam *et al.* 2005). Additionally, SWNTs could be selectively targeted to HeLa cells that overexpressed the folate receptor ( $\text{FR}^+$ ) by conjugating a folate moiety to polyethylene glycol-functionalized (PEGylated) SWNT. Folate-functionalized SWNTs were preferentially taken up by the  $\text{FR}^+$  cells, and these cells were killed with considerable selectivity upon illumination with 808-nm laser radiation (Kam *et al.* 2005). In a later study, it was also demonstrated that functionalized SWNTs can be conjugated with platinum-based therapeutics that can enhance cell kill in when used in conjunction with NIR-stimulated heating (Feazell *et al.* 2007). Given the structural similarities between SWNT and MWNT, this approach would likely be applicable to multiwalled carbon nanotube (MWNT)-based cancer therapeutics as well.

Recently, it has been demonstrated that SWNTs can be heated through the use of a non-invasive radio-frequency field, and that this capability can be used to successfully kill cancer cells *in vitro* and VX liver tumors *in vivo* (Gannon *et al.* 2007). This demonstration suggests SWNTs may be capable of non-invasively treating tumors in any part of the body, a capability currently not shared by NIR laser-based treatments.

Similar to SWNT, MWNT also have demonstrated utility in the realm of experimental cancer thermotherapy. The coupling of MWNT to NIR and its application to killing of cancer cells was recently demonstrated (Torti *et al.* 2007). In keeping with the predictions of classical antenna theory, this paper confirmed that MWNT with a length of 330 nm failed to generate appreciable heat in response to illumination with a 1064-nm laser, while MWNT with average lengths of 700 and 1100 nm, respectively, readily raised the temperature of the solution to over 50 °C (Torti *et al.* 2007). Cell-culture experiments using CRL 1932 cells (an established human clear cell carcinoma line) demonstrated that cells treated with either 700 or 1100-nm MWNT and NIR were readily killed, whereas treatment with 330-nm MWNT and NIR failed to elicit a cell-death response. There was no toxic effect of either MWNT or NIR when used individually (Fig. 13.3). MWNTs exhibit two potential advantages: first, due to the unique structure of MWNTs they absorb NIR more efficiently than SWNTs, thus requiring less incident radiation to generate an equal increase in temperature. In addition, they do not require internalization by the target cell to induce its death (Torti *et al.* 2007). An additional potential advantage of



**Fig. 13.3** The combination of MWNT and laser irradiation induces a temperature increase and cell death of human kidney cancer cells. S. Torti, D. Carroll, unpublished.

MWNT is their ability to be modified by dopants such as nitrogen or boron, which may augment their thermal properties, although to date this has not been explicitly investigated (Liu *et al.* 2005).

#### 13.3.1.4 Pharmacokinetics and biodistribution of carbon nanotubes

Pharmacological properties are critical in the assessment of any therapeutic material, including nanomaterials. Several groups have investigated the disposition of both single- and multiwalled nanotubes in rodent models following intravenous delivery.

Early studies focused on the distribution of nanotubes in non-disease-state model systems. Singh *et al.* described an approach by which ammonium-functionalized single- and multiwalled carbon nanotubes were covalently modified with diethylenetriaminepentaacetic (DTPA) dianhydride chelated  $^{111}\text{In}$ .

Dilute solutions were injected intravenously into cohorts of mice and tissues were harvested at 0.5, 3 and 24 h time points. These tissues were then read by a gamma counter (Singh *et al.* 2006). Surprisingly, both nanomaterials demonstrated little uptake by the reticuloendothelial system (RES), with renal excretion being the dominant form of systemic clearance. Moderate uptake was also seen in blood, muscle, skin and bone 30 min postinjection, with minimal activity observed at later time points. From this data the authors calculated a blood circulation half-life of 3–3.5 h (Singh *et al.* 2006).

A more recent publication utilizing Raman spectroscopy to measure tissue distribution of PEG-functionalized SWNT came to rather different conclusions (Liu *et al.* 2007). Contrary to the previous report, Liu *et al.* noted extended blood circulation times in excess of 20 h, with eventual nanotube clearance performed almost *exclusively* by organs of the RES (specifically liver and spleen) (Liu *et al.* 2008). In addition, significant quantities of nanotubes were detected in these organs up to three months postinjection. Encouragingly, the authors noted no resultant pathologies in the liver or spleen as determined by histological examination at the 100-day conclusion of the study. Together, these reports suggest that the method of nanomaterial functionalization strongly impacts its final disposition in the body. Thus, each new approach to functionalization will likely require new pharmacokinetic assessments.

Both of the previous studies were performed in tumor-free mice. Since the eventual goal is the targeting of nanomaterials to tumors, studies have also been initiated to assess the distribution of nanotubes in tumor-bearing models (Liu *et al.* 2007). Using PEGylated and radiolabelled SWNTs, mice bearing subcutaneous tumor xenographs were injected intravenously and the tissue distribution was tracked by positron emission tomography. Concordant with results described above (Liu 2008), blood-circulation times were extensive, and excretion was primarily hepatobiliary in nature. They also demonstrated that about 5% of the injected dose of nanotubes passively accumulated in the tumors, and that this could be increased to over 15% if the nanotubes were conjugated to RGD, a small peptide that targets integrins (Liu *et al.* 2007). These reports suggest that systemic delivery of nanotubes is a viable option. Section 13.3.1.6 will discuss ongoing efforts to enhance the uptake of these materials by their target lesions.

### 13.3.1.5 Toxicology of nanotubes

The inherent toxicity of any agent is a key attribute of its overall profile as a therapeutic. The debate over the intrinsic biocompatibility of carbon nanotubes has intensified following recent reports comparing long nanotubes (15–20  $\mu\text{m}$ ) to asbestos particles (Poland *et al.* 2008). At present, the toxicologic profiles of SWNT and MWNT remain poorly understood, in part due to the diversity of nanotube preparations, the difficulty in obtaining pure, homogeneous samples, the uncertain contribution of coating and functionalization to toxicity and biodistribution, and the unclear impact of contaminants such as residual catalyst or solvent to observed toxic effects.

Nanotubes are not new materials: MWNT have been observed in ice samples dating from the Neolithic Stone Age, and are present as combustion products in both indoor and outdoor environments (Lam *et al.* 2006). Nevertheless, their deliberate manufacture has raised concerns over hazards posed by occupational exposure. Toxicologic studies have focused primarily on the lung, since inhalation is the most likely route of occupational exposure, and epidemiological studies of nanoscale exhaust particles in the air have already demonstrated an association between materials of this scale and cardiovascular and respiratory morbidity and mortality (Samet *et al.* 2000). Early *in-vitro* studies using rat macrophage and human lung cell lines demonstrated that exposing cells to doses of SWNT and MWNT ranging from 5–100 µg/mL was not acutely toxic and did not result in marked decreases in cell viability, despite appreciable uptake of the material into the cytosol (Pulskamp *et al.* 2007). The authors did note that unpurified nanotubes increased intracellular levels of reactive oxygen species and attenuated mitochondrial membrane potential; however, these effects were not observed in cells treated with purified nanotubes and the authors attributed those toxicities to residual metal catalyst on the nanotube exterior and not the tubes themselves.

More recently, studies using a human mesothelioma cell line reached markedly different conclusions (Wick *et al.* 2007). In this study, doses of SWNT between 7.5–30 µg/mL substantially reduced cell viability in a dose-dependent manner, to below 25% of control for the highest dose. An important result was that the degree of nanotube agglomeration affected the outcome, with large aggregates decreasing viability to a greater extent than small bundles of nanotubes. This observation highlights the need for extensive characterization of the carbon nanotubes under investigation, since small changes in material state can significantly impact the study outcome.

Recent *in-vivo* studies have generally supported the conclusion that non-functionalized carbon nanotubes are detrimental to their host organism when directly instilled in the lung. For example, mice tracheally instilled with a 500 µg/kg bolus of SWNT developed multifocal granulomas with extensive macrophage infiltration and lung inflammation over the 14-day course of the study (Chou *et al.* 2008). Studies by others with doses ranging up to 10 mg/kg in both mice and rats showed similar trends, with animals developing extensive granulomas and fibrosis (Muller *et al.* 2005; Shvedova *et al.* 2005). Carrero-Sanchez *et al.* were able to replicate these findings with high doses (1–5 mg/kg) of multiwalled carbon nanotubes instilled tracheally. Interestingly, nitrogen-doped MWNTs failed to induce granuloma formation and only provoked moderate inflammation in mice challenged with the highest dose (Carrero-Sanchez *et al.* 2006).

In contrast to these studies, MWNT administered by inhalation did not lead to significant lung inflammation or tissue damage (Mitchell *et al.* 2007). Unlike previous studies that relied on single bolus tracheal instillations of carbon nanotubes, Mitchell and colleagues exposed mice by an inhalation chamber (a more directly relevant exposure technique) to doses of MWNTs up to the comparable allowed occupational limit of human exposure to nuisance dust (5 mg/m<sup>3</sup>: equivalent to a calculated dose of 2.7 mg/kg for the study animals). After 14 days of exposure, mice were sacrificed and extensively examined.

Surprisingly, there was no difference in the number of inflammatory cells between the lungs of control and MWNT-exposed mice or any overt lung damage as determined by histological analysis, although there was an overall decline in immune cell function. The authors suggested that the lack of toxic effect in the lung was likely due to the lower doses used. However, as their highest dose was already 100 times greater than that detected in a recent industrial hygiene report (Maynard *et al.* 2004) they questioned the necessity of using higher doses due to the enhanced risk of false-positive results.

In a recent provocative study, 50 µg of long (15–20 µm) MWNTs injected intraperitoneally into mice induced an inflammatory reaction in the mesothelium similar to that seen in asbestos-injected mice, suggesting, by extension, that long MWNTs might be similarly carcinogenic to that tissue (Poland *et al.* 2008). The relevance of this finding to currently proposed biomedical applications (including those detailed in this chapter) of carbon nanotubes is uncertain, since no studies have suggested using nanotubes longer than 1 µm; a length that failed to provoke any detrimental response in this study. Nevertheless, this result underscores the need for more extensive longitudinal studies that may uncover such long-term toxicities.

Overall, current data regarding the toxicity and biocompatibility of carbon nanotubes are sparse and conflicting. There does appear to be a pre-ponderance of evidence that high doses (>0.5 mg/kg) directly instilled in a target tissue such as the lung can provoke a toxic response. The data are less clear at more therapeutically relevant doses. Additionally, agents used to coat and functionalize nanotubes are likely to have a large impact on their toxicity, and the contribution of these agents is only beginning to be appreciated (Zhou *et al.* 2008). There is also a disconnect between nanotube preparations used for toxicity studies and those used for therapeutic intent: for example, the toxicology studies detailed above used either uncoated nanotubes or nanotubes solubilized with atypical surfactants (PS80); in contrast, nanotube solutions described elsewhere in this chapter that are being investigated as cancer therapies rely on PEG or Pluronic surfactants to coat and functionalize the nanotubes. At a minimum, coating agents will affect bundling, which has been shown to affect toxicity (Wick *et al.* 2007); likely, coatings will also affect other key parameters that influence toxicity, such as clearance and biodistribution. Thus, in addition to the general need for more extensive toxicologic studies, there is a need for studies that directly focus on the precise materials that are envisioned as therapeutic agents. These will be essential for the future translation of investigational carbon-nanotube-based cancer therapies into clinical application.

#### 13.3.1.6 *In-vivo targeting strategies employing nanotubes*

To date, two methods have pre-dominated in the effort to selectively target intravenously delivered nanomaterials to tumors; passive targeting and active targeting through the use of conjugated targeting peptides.

The first strategy is a passive approach that takes advantage of what is termed the enhanced permeability and retention effect (EPR) to deposit nanomaterials in the tumor bed. The EPR effect derives from structural

abnormalities in tumor vasculature: tumor vessels are characterized by discontinuous endothelial tight junctions, and a lack of uniform mural cell reinforcement, which together lead to vessels of poor integrity that are highly permeable to large molecules. This action is complemented by deficiencies in tumor lymphatic structure, such that molecules that have extravasated in the tumor are not readily removed through lymphatic drainage (Maeda *et al.* 2000; Maeda 2001). The dissimilar competencies of normal and tumor-associated vessels can be exploited to selectively target materials of a particular size range to a tumor, while sparing normal tissues. Indeed, it has been shown that particles with apparent molecular masses >45 kDa or sizes in the range of hundreds of nanometers (these sizes exceed the renal clearance threshold, preventing them from being rapidly cleared from circulation by glomerular filtration) can achieve concentrations tenfold greater than both plasma and normal tissue within 24 h of intravenous administration (Matsumura and Maeda 1986; Maeda and Matsumura 1989; Maeda *et al.* 2003). With specific reference to its application in the targeting of systemically delivered carbon nanotubes, Liu *et al.* demonstrated that 5% of an intravenous dose of PEG-functionalized SWNTs passively accumulated in an experimentally induced rodent tumor within 6 h of induction (Liu *et al.* 2007).

The second strategy is an active targeting strategy, in which a tumor-targeting ligand is attached to the nanotube in order to direct the nanotube to the tumor. For example, by covalently modifying PEGylated SWNTs with RGD peptides (these bind tightly to the angiogenesis and metastasis-related integrin  $\alpha_v\beta_3$ ) tumor accumulation of SWNTs amounting to 15% of the total injected dose was achieved (Liu *et al.* 2007). While this technique has great potential for the active targeting of drugs to a given tissue, it relies on either the tumor-specific expression or tumor overexpression of a known target.

#### 13.3.1.7 Use of carbon nanotubes in tumor imaging

The refinement of existing medical imaging modalities (primarily MRI, CT and PET) has greatly enhanced their utility in detecting indwelling tumors and tracking their response to therapy. In relation to the proposed use of carbon nanotubes (particularly MWNTs) for the thermal treatment of cancer, MRI has capabilities that best complement the envisioned therapeutic strategies. Two in particular (its amenability to contrast enhancement and use in non-invasive thermometry) will be briefly considered.

The spatial resolution of proton magnetic resonance is the physical basis of the medical imaging modality known as MRI. Following the application of the pulsed magnetic field and the consequent excitation of the proton to a high-energy state two relaxation times, termed  $T_1$  and  $T_2$ , can be determined through specific pulse sequences as the proton falls back to low-energy equilibrium. These relaxation times are a function of the proton's immediate environment, and thus change from tissue to tissue and between normal and malignant. Contrast agents can be used to further highlight areas of the body that preferentially uptake or exclude the material. This technique has proven especially useful in detecting indwelling tumors, as their dysfunctional vasculatures facilitate the accumulation of contrast agent in the diseased tissue.

Gadolinium and iron are  $T_1$  and  $T_2$  contrast agents, respectively, that have both been successfully incorporated into single- and multiwalled carbon nanotubes (Sitharaman *et al.* 2005; Bai *et al.* 2008) (Fig. 13.3). Incorporation was achieved either during the growth phase or in postgrowth processing, with the result being the creation of MR contrast agents that are several times more efficient than those currently in clinical use (Sitharaman *et al.* 2005; Son *et al.* 2006; Bai *et al.* 2008). If one considers this functionality in conjunction with the tumor-targeting strategies detailed above, it is not difficult to imagine the use of these materials to both enhance the detection, treatment and follow-up imaging of a given tumor.

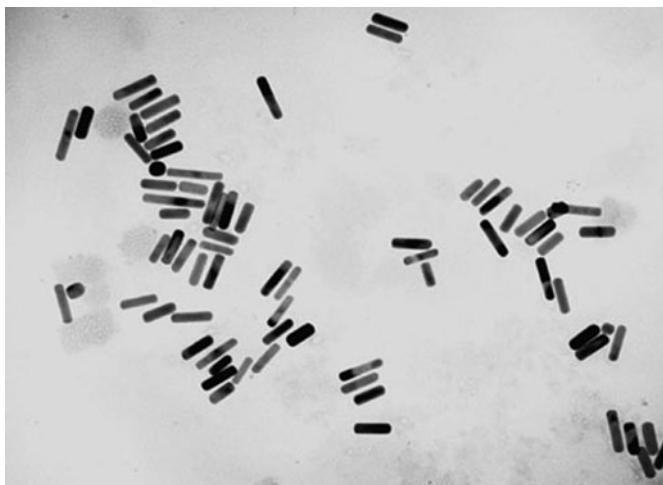
One additional synergy can be envisioned in the combined use of magnetic resonance imaging with nanotube-mediated thermal-ablative therapy. Recent advancements in the design of new MR pulse sequences have made non-invasive thermometry a reality (Samset 2006; de Senneville *et al.* 2007). These techniques are readily applicable to 1.5 and 3 Tesla clinical MR equipment. Due to minimal postacquisition processing, use of MR thermometry during therapy with nanomaterials has the potential to allow near-real-time temperature mapping, minimizing the risk of undertreating regions of tumor, such as those abutting well-perfused vessels. MR-based thermometry will help by providing the clinician a means by which the anticipated effectiveness of the intervention can be readily assessed, giving them the option to adjust the therapy midtreatment as the need arises. With this capability it becomes reasonable to envision the use of nanotubes as a minimally invasive, multimodal therapeutic approach to the treatment of non-resectable solid cancers.

## 13.4 Gold nanoshells and nanorods

### 13.4.1 Physical properties of gold nanoshells and nanorods

Nanoscale formulations of gold and other noble metals exhibit unique properties that make them useful for various biomedical applications (Fig. 13.4). Two in particular, gold nanoshells and gold nanorods, have emerged in recent years as viable contenders in the application of nanomaterial-mediated photothermal therapy.

Gold nanoshells consist of dielectric silica cores onto which gold colloids of variable thickness are adsorbed (Hirsch *et al.* 2006). These structures range in size from 50–500 nm with alterations in growth conditions determining the size range of the final product. Nanoshells exhibit a particularly robust surface plasmon resonance (the oscillation of conducting electrons at the gold surface interface) in response to incident electromagnetic radiation, resulting in the generation and transduction of heat into their immediate surroundings. By varying the principal dimensions of the silica core and outer gold shell, this effect can be tuned to respond maximally at a discrete portion of the EM spectrum. This property is of particular use when considering the application of nanoshells as mediators of photothermal therapy *in vivo*. Untuned gold nanoparticles exhibit strong absorbance in the visible segment of the EM spectrum, limiting their utility *in vivo* due to the natural abundance of biological



**Fig. 13.4** Gold nanorods imaged by TEM. Used with permission of Dr. Mostafa, A. El-Sayed group, Laser Dynamics Laboratory, Georgia Institute of Technology.

chromophores that absorb in this range (Huang *et al.* 2007). By increasing the diameter of the silica core and varying the thickness of the gold shell, the absorption spectrum of these particles can be redshifted into the NIR, where tissue absorption is minimal. Additionally, nanoshells can be further tuned to scatter rather than absorb NIR, making them applicable as contrast agents for optical tomography imaging techniques.

Perhaps not surprisingly, gold nanorods retain many of the same physical properties as nanoshells due to the commonality of their external composition. Like nanoshells, nanorods exhibit strong surface plasmon resonance in response to incident light and the absorption spectrum can be controlled by altering the aspect ratio (length/width) of the material, with a larger ratio redshifting peak absorption into the near-infra-red range (Huang *et al.* 2007).

Reduced gold is a stable and generally inert material highly suitable for medical applications. It has long been used in dentistry for the manufacture of biocompatible prosthetics that have little demonstrated local toxicities. There is some concern over systemic dosing of gold nanoparticles, as colloidal gold is a well established nephrotoxin and inducer of peripheral neuropathies (Klaassen 2008). These are concerns that will have to be addressed in future toxicology studies as both materials progress to *in-vivo* applications.

Finally, conjugation chemistries involving gold have been described extensively, making the process of attaching targeting moieties fairly straightforward, assuming the proper functional groups are accessible. Thiol-containing proteins have a demonstrated affinity for gold particles, to which they form very tight, non-covalent bonds (Nuzzo *et al.* 1987). This property is useful for the purposeful conjugation of desired polypeptides for targeting or to enhance biocompatibility. However, it may also enhance the rate of opsonization and subsequent RES clearance when these particles are used in living systems. PEGylation or other means of “stealthing” may constitute one way to address this concern.

#### 13.4.1.1 *Anticancer activity of gold nanoparticles in vitro*

The use of gold nanoparticles for the photothermal treatment of cancer was initially described by Hirsch *et al.* in 2003. In this study 130-nm PEG-passivated gold nanoshells were incubated with SK-BR-3 breast cancer cells and exposed to 820 nm at 35 W/cm<sup>2</sup> for seven minutes. Calcein fluorescence and dextran exclusion assays demonstrated complete cell kill within a region corresponding to the area impacted by the laser beam (Hirsch *et al.* 2003). Huang *et al.* generated similar findings through the use of gold nanorods. Normal human keratinocytes and the oral squamous cell carcinoma cell line HSC 3 were incubated gold nanorods conjugated to an anti-EGFR antibody (an antibody directed against the epidermal growth factor receptor). Prior to 800-nm laser irradiation (up to 20 W/cm<sup>2</sup> for four minutes), cells were washed with buffered saline to remove unbound nanorods. Following laser exposure, cell viability was assessed by trypan blue exclusion. HSC 3 cells, which overexpress EGFR, were readily killed at lower laser powers, while normal keratinocytes were relatively unharmed, due to the low expression of that receptor (Huang *et al.* 2006). This group went on to demonstrate that their cell lines required temperatures in the range of 70–80 °C to effect complete cell death, and that these temperatures could be reached in nanorod-loaded cells following four minutes of near-infra-red laser irradiation (Huang *et al.* 2006).

A novel targeting strategy for gold nanoshells is through the use of monocytes/macrophages as a “Trojan horse” (Choi *et al.* 2007). Using a tumor spheroid model system of breast cancer, Choi and colleagues demonstrated that macrophages “pre-loaded” (through incubation) with gold nanoshells will migrate into a simulated tumor mass, and that the spheroid can then be ablated by irradiating it with a near-infrared laser (Choi *et al.* 2007). Additional work will be needed to confirm the utility of this strategy *in vivo*; however, it represents a potential method through which the limitations of previously discussed targeting schemes may be overcome.

#### 13.4.1.2 *Anticancer activity of gold nanoparticles in vivo*

Initial work described by Hirsch *et al.* demonstrated that in a mouse model, solutions of gold nanoshells injected directly into a subcutaneous tumor followed by near-infra-red laser illumination for 4–6 min resulted in a temperature increase within the tumor of over 35 °C. Control tumors that did not receive an injection of nanoshells saw average temperature increases of only 7 °C (Hirsch *et al.* 2003). Histological analyses revealed that the zone of thermal injury in the nanoshell-treated tumors extended about 4 mm below the dermal surface. The authors noted that this could likely be increased through better distribution of the nanoshells throughout the tumor volume and longer treatment times (Hirsch *et al.* 2003).

Following this initial report, additional studies explored the impact of nanoshell-mediated thermal therapy on tumor regression and overall survival. O’Neal *et al.* demonstrated that PEGylated nanoshells could be injected intravenously in tumor-bearing mice and that within 6 h they would passively accumulate in the tumors in therapeutically relevant concentrations (a function of

the enhanced permeability and retention effect discussed in a previous section). Subsequent exposure of these same tumors (diameters at time of treatment 3–5 mm) to an 810-nm laser at 4 W/cm<sup>2</sup> for 4 min induced complete remission of the animals for the entire length of follow-up (3 months) (O’Neal *et al.* 2004). Similar results were observed by a separate group using a heterotopic murine model of prostate cancer, and by yet another group using a colon cancer xenograft system (Gobin *et al.* 2007; Stern *et al.* 2008).

#### 13.4.1.3 Use of gold nanoshells and nanorods in cancer imaging

Dissimilar to the carbon-nanotube-based nanotherapeutics discussed in previous sections, gold nanomaterials are invisible to most clinically relevant imaging modalities (MRI and CT). However, their usefulness in optics-based imaging techniques has been shown and several papers have documented applications of these particles for *in-vitro* analyses.

The early detection of cancer is often cited as one of the best predictors of successful treatment. To accomplish this, screening assays need to be developed that have the sensitivity to distinguish a small population of transformed cells from a background of normal tissue constituents. Current techniques that involve the use of oncoprotein-directed antibodies conjugated to conventional chromophores lack the signal intensity necessary to generate the desired sensitivity. Two recent publications suggest that antibody conjugated gold nanorods may offer a promising solution. Huang *et al.* discovered that anti-EGFR conjugated nanorods could be used to successfully identify malignant oral squamous carcinoma cells from their normal counterparts. It was shown that following the incubation of the targeted nanorods with the two cell types and subsequent analysis by Raman spectroscopy, the cancerous cells exhibited a distinct polarized Raman spectra that was absent in the normal samples (El-Sayed *et al.* 2005; Huang *et al.* 2007; Oyelere *et al.* 2007). A similar study involving anti-HER2 (a clinically relevant breast cancer marker) targeted nanoshells demonstrated their ability to selectively identify HER-2<sup>+</sup> breast cancer cells from normal cells (Loo *et al.* 2005). These initial studies suggest such strategies may become useful in the screening of patient tissue biopsies and other clinical samples.

Optical coherence tomography (OCT) is a relatively new clinical imaging technique that analyzes the reflections of low coherence light off of underlying tissue components to provide cross-sectional images of the subsurface down to micrometer resolution (Huang *et al.* 1991; Fujimoto *et al.* 1995). The utility of this system can be further enhanced through the use of contrast agents that serve to highlight specific biological structures. Gobin and colleagues demonstrated that gold nanoshells have the capacity to serve as OCT contrast agents due to their ability to reflect and scatter light in a tunable manner (Gobin *et al.* 2007). While OCT has been used to image superficial cancers such as gastrointestinal and oral mucosal, its use in the detection of deeper-seated tumors is constrained by a maximum imaging depth of 2 mm (Li *et al.* 2000). Beyond this threshold, the reflected light is too greatly attenuated by the intervening tissue to be of use.

### 13.5 Iron-oxide nanoparticles

Iron-oxide nanoparticles have received considerable attention as a vehicle for the generation of tumor-specific heat. Because these particles have been extensively reviewed recently (Kumar 2006), they are not discussed in detail here. Briefly, this technique involves the application of alternating magnetic fields to nanoscale crystals of iron oxide ( $\text{Fe}_3\text{O}_4$  or  $\gamma\text{-Fe}_2\text{O}_3$ ) to generate heat through the process of hysteresis and relaxation losses (Jordan *et al.* 1993). In this approach, ultrafine (5–20 nm) iron-oxide crystals are coated with a functionalizing material (either dextran or aminosilane) to prevent aggregation and suspended in saline to create a magnetic fluid (MF). This fluid is directly injected into a target tissue and is later heated when placed in a magnetic-field applicator (Gnevackow *et al.* 2004). Additionally, the MF functions as a contrast agent in both CT and MR imaging protocols. The primary advantage of this technique over other nanomaterial-based thermoablative methods described earlier in this chapter is its ability to theoretically treat any region of the body. Unlike NIR-based therapies, this method does not suffer from depth of penetration issues as the body is functionally transparent to magnetic fields. This approach has received considerable support in Europe, where it has rapidly transitioned into the clinic. Phase I trials have already been completed in glioblastoma multiforme, prostate, esophageal, pancreatic and various sites of recurrent cancers and Phase II trials are underway for both glioblastoma and prostate cancers (MagForce 2008). However, no trials are open in the United States.

### 13.6 Conclusions and future directions

The nanoparticles described in this review have each demonstrated potential as multifunctional agents for the treatment of cancer. The promised package of tumor specificity, therapeutic efficacy and imaging enhancement makes a compelling argument for their continued development. Indeed, in recognizing the significant impact nanomaterials will have in the diagnosis and treatment of cancer, the National Cancer Institute has pledged support of research into nanomedicine in cancer through its Alliance for Nanotechnology in Cancer. However, it is clear that several areas of concern must be addressed before these technologies can make the desired “bench to bedside” transition.

Fundamental to the success of these therapies will be their ability to demonstrate an acceptable toxicity profile. It is clear that both oncologists and their patients favor efficacy over safety, and it is encouraging that the limited data that is currently available from studies in mice has not pointed to dramatic toxicity of any of the nanomaterials highlighted in this review (SWNT, MWNT, nanoshells, nanorods and encapsulated iron oxides). However, data is still sparse, and precludes the drawing of firm conclusions on toxicity. An additional limitation is the current inability to reproducibly manufacture identical batches of nanoparticles (more a concern for carbon-nanotube-based therapies) for assessment. Advancements in manufacturing and purification techniques will likely redress this in the future; however, it remains a concern in the near term.

Future work will also likely include expanded research on the use of these particles, particularly carbon nanotubes, as targetable transporters of chemotherapeutics and for use as chemo/radiosensitizers through their ability to generate localized heat. Published studies have already demonstrated the ability of carbon nanotubes to be filled with drugs such as carboplatin, and indicated that these formulations effectively inhibit the growth of cancers (Hampel *et al.* 2008). Confirmation of the effectiveness of this strategy *in vivo* will be necessary; however, this is likely to be a very promising application.

Additionally, work will be required to integrate the imaging and treatment elements inherent in these new strategies. MRI appears to be the most promising imaging modality, as it provides high-resolution anatomical images, offers a thermometric capability that synergizes well with proposed thermal treatments, and does not suffer from a limited depth of penetration that constrains techniques like optical coherence tomography. Of course, the strong magnetic field required for MR imaging will necessitate the use of only non-ferromagnetic materials in laser delivery.

Although the use of nanomaterials in cancer therapy is still in the early stages, it is encouraging that Nanospectra Biosciences has recently received FDA approval to begin a Phase I clinical trial in recurrent or refractory head and neck cancer patients with a nanoshell-based photothermal therapy (Nanospectra 2008). The lessons learned from this and future trials will undoubtedly foster the creation of a new generation of purpose-built nanomaterials with enhanced capabilities for the imaging and effective treatment of cancer.

## Acknowledgments

This work was supported in part by National Institutes of Health grants R01 CA12842 (SVT) and T32 CA079448 (AB). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

## References

- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. (eds), *Molecular Biology of the Cell* (Garland Publishing, New York, 1994).
- Bai, X., Son, S.J., Zhang, S., Liu, W., Jordan, E.K., Frank, J.A., Venkatesan, T., Bok Lee, S. *Nanomedicine* **3**, 163 (2008).
- Berezney, R., Coffey, D.S. *Biochem. Biophys. Res. Commun.* **60**, 1410 (1974).
- Boehrer, S., Nowak, D., Hoelzer, D., Mitrou, P.S., Kai Uwe, C. *Curr. Pharmaceut. Des.* **12**, 111 (2006).
- Brigger, I., Dubernet, C., Couvreur, P. *Adv. Drug. Deliv. Rev.* **54**, 631 (2002).
- Buja, L.M., Eigenbrodt, M.L., Eigenbrodt, E.H. *Arch. Pathol. Lab. Med.* **117**, 1208 (1993).

- Bursch, W., Oberhammer, F., Schulte-Hermann, R. *Trends. Pharmacol. Sci.* **13**, 245 (1992).
- Cain, K. *Drug Metab. Rev.* **35**, 337 (2003).
- Calderwood, S.K., Asea, A. *Int. J. Hyperthermia* **18**, 597 (2002).
- Carrero-Sanchez, J.C., Elias, A.L., Mancilla, R., Arrellin, G., Terrones, H., Laclette, J.P., Terrones, M. *Nano Lett.* **6**, 1609 (2006).
- Choi, M.R., Stanton-Maxey, K.J., Stanley, J.K., Levin, C.S., Bardhan, R., Akin, D., Badve, S., Sturgis, J., Robinson, J.P., Bashir, R., Halas, N.J., Clare, S.E. *Nano Lett.* **7**, 3759 (2007).
- Chou, C.C., Hsiao, H.Y., Hong, Q.S., Chen, C.H., Peng, Y.W., Chen, H.W., Yang, P.C. *Nano Lett.* **8**, 437 (2008).
- Ciocca, D.R., Clark, G.M., Tandon, A.K., Fuqua, S.A., Welch, W.J., McGuire, W.L. *J. Natl. Cancer Inst.* **85**, 570 (1993).
- Coffey, D.S., Getzenberg, R.H., DeWeese, T.L. *J. Am. Med. Assoc.* **296**, 445 (2006).
- Coss, R.A., Dewey, W.C., Bamburg, J.R. *Cancer Res.* **42**, 1059 (1982).
- Coss, R.A., Linnemans, W.A. *Int. J. Hyperthermia* **12**, 173 (1996).
- Dahl, O., Overgaard, J. *Acta Oncol.* **34**, 1075 (1995).
- Dahl, O., Dalene, R., Schem, B.C. *Mella, O. Acta Oncol.* **38**, 863 (1999).
- Dai, H. *Acc. Chem. Res.* **35**, 1035 (2002).
- Datta, N.R., Bose, A.K., Kapoor, H.K., Gupta, S. *Int. J. Hyperthermia* **6**, 479 (1990).
- de Senneville, B.D., Mougenot, C., Quesson, B., Dragonu, I., Grenier, N., Moonen, C.T. *Eur. Radiol.* **17**, 2401 (2007).
- Defer, N., Kitzis, A., Kruh, J., Brahms, S., Brahms, J. *Nucleic Acids Res.* **4**, 2293 (1977).
- Dewey, W.C., Westra, A., Miller, H.H., Nagasawa, H. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **20**, 505 (1971).
- Dewey, W.C. *Radiat. Res.* **120**, 191 (1989).
- Dewey, W.C., Li, X.L., Wong, R.S. *Radiat. Res.* **122**, 268 (1990).
- Dewhirst, M.W., Prosnitz, L., Thrall, D., Prescott, D., Clegg, S., Charles, C., MacFall, J., Rosner, G., Samulski, T., Gillette, E., LaRue, S. *Semin. Oncol.* **24**, 616 (1997).
- Dickson, J.A., Calderwood, S.K. *Ann. NY Acad. Sci.* **335**, 180 (1980).
- Eberle, J., Kurbanov, B.M., Hossini, A.M., Trefzer, U., Fecker, L.F. *Drug Resist. Updates* **10**, 218 (2007).
- El-Sayed, I.H., Huang, X., El-Sayed, M.A. *Nano Lett.* **5**, 829 (2005).
- Everts, M. *Expert Rev. Med. Dev.* **4**, 131 (2007).
- Fajardo, L.F., Egbert, B., Marmor, J., Hahn, G.M. *Cancer* **45**, 613 (1980).
- Feazell, R.P., Nakayama-Ratchford, N., Dai, H., Lippard, S.J. *J. Am. Chem. Soc.* **129**, 8438 (2007).
- Ferrari, M. *Nat. Rev. Cancer* **5**, 161 (2005).
- Fortin, A., Raybaud-Diogene, H., Tetu, B., Huot, J., Landry, J., Deschenes R. *Int. J. Radiat. Oncol. Biol. Phys.* **46**, 1259 (2000).
- Fujimoto, J.G., Brezinski, M.E., Tearney, G.J., Boppart, S.A., Bouma, B., Hee, M.R., Southern, J.F., Swanson, E.A. *Nature Med.* **1**, 970 (1995).

- Gannon, C.J., Cherukuri, P., Yakobsen, B., Cognet, L., Kanzius, J., Kittrell, C., Weisman, R.B., Pasquali, M., Schmidt, H.K., Smalley, R.E., Curley, S.A. *Cancer* **110**, 2654 (2007).
- Gneveckow, U., Jordan, A., Scholz, R., Brub, V., Waldfner, N., Ricke, J., Feussner, A., Hildebrandt, B., Rau, B., Wust, P. *Med. Phys.* **31**, 1444 (2004).
- Gobin, A.M., Lee, M.H., Halas, N.J., James, W.D., Drezek, R.A., West, J.L. *Nano Lett.* **7**, 1929 (2007).
- Hampel, S., Kunze, D., Haase, D., Kramer, K., Rauschenbach, M., Ritschel, M., Leonhardt, A., Thomas, J., Oswald, S., Hoffmann, V., Buchner, B. *Nanomedicine* **3**, 175 (2008).
- Hanson, G.W. *IEEE Trans. Antenn. Propag.* **53**, 3426 (2005).
- Harmon, B.V., Corder, A.M., Collins, R.J., Gobe, G.C., Allen, J., Allan, D.J., Kerr, J.F. *Int. J. Radiat. Biol.* **58**, 845 (1990).
- Heisterkamp, J., van Hillegersberg, R.I.N.I., I.J. *Lasers Surg. Med.* **25**, 257 (1999).
- Henle, K.J., Dethlefsen, L.A. *Ann. NY Acad. Sci.* **335**, 234 (1980).
- Hildebrandt, B., Wust, P., Ahlers, O., Dieing, A., Sreenivasa, G., Kerner, T., Felix, R., Riess, H. *Crit. Rev. Oncol. Hematol.* **43**, 33 (2002).
- Hirsch, L.R., Stafford, R.J., Bankson, J.A., Sershen, S.R., Rivera, B., Price, R.E., Hazle, J.D., Halas, N.J., West, J.L. *Proc. Natl. Acad. Sci. USA* **100**, 13549 (2003).
- Hirsch, L.R., Gobin, A.M., Lowery, A.R., Tam, F., Drezek, R.A., Halas, N.J., West, J.L. *Ann. Biomed. Eng.* **34**, 15 (2006).
- Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A. *et al. Science* **254**, 1178 (1991).
- Huang, X., El-Sayed, I.H., Qian, W., El-Sayed, M.A. *J. Am. Chem. Soc.* **128**, 2115 (2006a).
- Huang, X., Jain, P.K., El-Sayed, I.H., El-Sayed, M.A. *Photochem. Photobiol.* **82**, 412 (2006b).
- Huang, X., El-Sayed, I.H., Qian, W., El-Sayed, M.A. *Nano Lett.* **7**, 1591 (2007a).
- Huang, X., Jain, P.K., El-Sayed, I.H., El-Sayed, M.A. *Lasers Med. Sci.* (2007b).
- Hui, N. University of Wisconsin-Madison (2005). [www.tahan.com/charlie/nanosociety/course201/nanos/NH.pdf](http://www.tahan.com/charlie/nanosociety/course201/nanos/NH.pdf). Accessed 06-27-09.
- Hunt, C.R., Pandita, R.K., Laszlo, A., Higashikubo, R., Agarwal, M., Kitamura, T., Gupta, A., Rief, N., Horikoshi, N., Baskaran, R., Lee, J.H., Lobrich, M., Paull, T.T., Roti Roti, J.L., Pandita, T.K. *Cancer Res.* **67**, 3010 (2007).
- Iijima, S. *Nature* **354**, 56 (1991).
- Izzo, F. *Ann. Surg. Oncol.* **10**, 491 (2003).
- Jain, K. *Medi. Princ. Pract.* **17**, 89 (2008).
- Jordan, A., Wust, P., Fahling, H., John, W., Hinz, A., Felix, R. *Int. J. Hypertherm.* **9**, 51 (1993).
- Kam, N.W., O'Connell, M., Wisdom, J.A., Dai, H. *Proc. Natl. Acad. Sci. USA* **102**, 11600 (2005).
- Kelekis, A.D., Thanos, L., Mylona, S., Ptohis, N., Malagari, K., Nikita, A., Christodoulidou, J., Kelekis, N. *Eur. Radiol.* **16**, 2471 (2006).
- Kiang, J.G., Tsokos, G.C. *Pharmacol. Ther.* **80**, 183 (1998).
- Kim, J.S., He, L., Lemasters, J.J. *Biochem. Biophys. Res. Commun.* **304**, 463 (2003a).
- Kim, J.S., Qian, T., Lemasters, J.J. *Gastroenterology* **124**, 494 (2003b).

- Kim, K.Y. *Nanomedicine* **3**, 103 (2007).
- Klaassen, C. (ed.) *Casarett and Doull's Toxicology: The Basic Science of Poisons* (McGraw-Hill Medical, New York, 2008).
- Konig, K. *J. Microsc.* **200**, 83 (2000).
- Konings, A.W., Ruifrok, A.C. *Radiat. Res.* **102**, 86 (1985).
- Kroemer, G., Dallaporta, B., Resche-Rigon, M. *Annu. Rev. Physiol.* **60**, 619 (1998).
- Kumar, C. (ed.) *Nanomaterials for Cancer Therapy* (Wiley-VCH, Weinheim, 2006).
- Lam, C.W., James, J.T., McCluskey, R., Arepalli, S., Hunter, R.L. *Crit. Rev. Toxicol.* **36**, 189 (2006).
- Latchman, D. *Cardiovasc. Res.* **51**, 637 (2001).
- Lemasters, J.J., Qian, T., He, L., Kim, J.S., Elmore, S.P., Cascio, W.E., Brenner, D.A. *Antioxid. Redox. Signal* **4**, 769 (2002).
- Lenhard, R., Osteen, R., Gansler, T. (eds), *The American Cancer Society's Clinical Oncology* (Blackwell, Williston, 2001).
- Lepock, J.R. *Int. J. Hyperthermia* **19**, 252 (2003).
- Lepock, J.R. *Int. J. Hyperthermia* **20**, 115 (2004).
- Lepock, J.R. *Methods* **35**, 117 (2005).
- Levine, A.J. *Princess Takamatsu Symp.* **20**, 221 (1989).
- Levine, A.J., Momand, J., Finlay, C.A. *Nature* **351**, 453 (1991).
- Lewis, P.N. *Can. J. Biochem.* **55**, 736 (1977).
- Li, G.C., Mivechi, N.F., Weitzel, G. *Int. J. Hyperthermia* **11**, 459 (1995).
- Li, X.D., Boppart, S.A., Van Dam, J., Mashimo, H., Mutinga, M., Drexler, W., Klein, M., Pitris, C., Krinsky, M.L., Brezinski, M.E., Fujimoto, J.G. *Endoscopy* **32**, 921 (2000).
- Liu, J., Webster, S., Carroll, D.L. *J. Phys. Chem. B* **109**, 15769 (2005).
- Liu, Z., Cai, W., He, L., Nakayama, N., Chen, K., Sun, X., Chen, X., Dai, H. *Nature Nanotechnol.* **2**, 6 (2007).
- Liu, Z., Davis, C., Cai, W., He, L., Chen, X., Dai, H. *Proc. Natl. Acad. Sci. USA* **105**, 1410 (2008).
- Loo, C., Hirsch, L., Lee, M.H., Chang, E., West, J., Halas, N., Drezek, R. *Opt. Lett.* **30**, 1012 (2005).
- Lu, D.S., Raman, S.S., Vodopich, D.J., Wang, M., Sayre, J., Lassman, C. *AJR Am. J. Roentgenol.* **178**, 47 (2002).
- Lu, D.S., Raman, S.S., Limanond, P., Aziz, D., Economou, J., Busuttil, R., Sayre, J. *J. Vasc. Interv. Radiol.* **14**, 1267 (2003).
- Maeda, H. Matsumura, Y. *Crit. Rev. Ther. Drug Carrier Syst.* **6**, 193 (1989).
- Maeda, H., Wu, J., Sawa, T., Matsumura, Y., Hori, K. *J. Control Release* **65**, 271 (2000).
- Maeda, H. *Adv. Enzyme Regul.* **41**, 189 (2001).
- Maeda, H., Fang, J., Inutsuka, T., Kitamoto, Y. *Int. Immunopharmac.* **3**, 319 (2003).
- MagForce [www.magforce.de/english](http://www.magforce.de/english). Accessed: 06-27-08.
- Majda, J.A., Gerner, E.W., Vanlandingham, B., Gehlsen, K.R., Cress, A.E. *Exp. Cell Res.* **210**, 46 (1994).
- Matsumura, Y., Maeda, H. *Cancer Res.* **46**, 6387 (1986).

- Maynard, A.D., Baron, P.A., Foley, M., Shvedova, A.A., Kisin, E.R., Castranova, V. *J. Toxicol. Environ. Health A* **67**, 87 (2004).
- McDannold, N., Tempany, C.M., Fennelly, F.M., So, M.J., Rybicki, F.J., Stewart, E.A., Jolesz, F.A., Hynynen, K. *Radiology* **240**, 263 (2006).
- Mitchell, L.A., Gao, J., Wal, R.V., Gigliotti, A., Burchiel, S.W., McDonald, J.D. *Toxicol. Sci.* **100**, 203 (2007).
- Morimoto, R.I., Santoro, M.G. *Nat. Biotechnol.* **16**, 833 (1998).
- Muller, J., Huaux, F., Moreau, N., Misson, P., Heilier, J.F., Delos, M., Arras, M., Fonseca, A., Nagy, J.B., Lison, D. *Toxicol. Appl. Pharmacol.* **207**, 221 (2005).
- Nanospectra Biosciences www.nanospectra.com. Accessed: 06-27-08.
- Nelson, D.L., Cox, M. (eds), *Lehninger Principles of Biochemistry* (W.H. Freeman, New York, 2005).
- Nikfarjam, M., Christophi, C. *Br. J. Surg.* **90**, 1033 (2003).
- Nikfarjam, M., Malcontenti-Wilson, C., Christophi, C. *J. Gastroint. Surg.* **9**, 410 (2005a).
- Nikfarjam, M., Muralidharan, V., Christophi, C., *J. Surg. Res.* **127**, 208 (2005b).
- Nuzzo, R.G., Fusco, F.A., Allara, D.L. *J. Am. Chem. Soc.* **109**, 2358 (1987).
- O'Neal, D.P., Hirsch, L.R., Halas, N.J., Payne, J.D., West, J.L. *Cancer Lett.* **209**, 171 (2004).
- Olive, P. *Methods Cell Biol.* **75**, 355 (2004).
- Overgaard, J., Suit, H.D. *Cancer Res.* **39**, 3248 (1979).
- Overgaard, J., Gonzalez Gonzalez, D., Hulshof, M.C., Arcangeli, G., Dahl, O., Mella, O., Bentzen, S.M. *Int. J. Hyperthermia* **12**, 3 (1996).
- Oyelere, A.K., Chen, P.C., Huang, X., El-Sayed, I.H., El-Sayed, M.A. *Bioconjug. Chem.* **18**, 1490 (2007).
- Patterson, E.J., Scudamore, C.H., Owen, D.A., Nagy, A.G., Buczkowski, A.K. *Ann. Surg.* **227**, 559 (1998).
- Poland, C.A., Duffin, R., Kinloch, I., Maynard, A., Wallace, W.A.H., Seaton, A., Stone, V., Brown, S., MacNee, W., Donaldson, K. *Nature Nano., Advanced Online Publication* (2008).
- Pulskamp, K., Diabate, S., Krug, H.F. *Toxicol. Lett.* **168**, 58 (2007).
- Ries, L., Harkins, D., Krapcho, M., Mariotto, A., Miller, B., Feuer, E., Clegg, L., Eisner, M., Horner, M., Howlader, N., Hayat, M., Hankey, B., Edwards, B. (National Cancer Institute, 2007).
- Rossi, A., Ciafre, S., Balsamo, M., Pierimarchi, P., Santoro, M.G. *Cancer Res.* **66**, 7678 (2006).
- Roti, J.L.R. *Int. J. Hyperthermia* **24**, 3 (2008).
- Rylander, M.N., Feng, Y., Bass, J., Diller, K.R. *Ann. NY Acad. Sci.* **1066**, 222 (2005).
- Samali, A., Cotter, T.G. *Exp. Cell Res.* **223**, 163 (1996).
- Samet, J.M., Dominici, F., Curriero, F.C., Coursac, I., Zeger, S.L. *N. Engl. J. Med.* **343**, 1742 (2000).
- Samset, E. *Minim. Invasive Ther. Allied Technol.* **15**, 36 (2006).
- Sanguino, A., Lopez-Berestein, G., Sood, A. *Mini. Rev. Med. Chem.* **8**, 248 (2008).

- Schlesinger, M. (ed.), *Stress Proteins: Induction and Function* (Springer-Verlag, Berlin, 1990).
- Schonenberger, C., Forro, L. *Phys. World* **6**, (2000).
- Shih-Horng Huang, K.-J.Y. J.-C.W. K.-J.C. S.-M.W. *J. Cell. Biochem.* **75**, 327 (1999).
- Shvedova, A.A., Kisin, E.R., Mercer, R., Murray, A.R., Johnson, V.J., Potapovich, A.I., Tyurina, Y.Y., Gorelik, O., Arepalli, S., Schwegler-Berry, D., Hubbs, A.F., Antonini, J., Evans, D.E., Ku, B.K., Ramsey, D., Maynard, A., Kagan, V.E., Castranova, V., Baron, P. *Am. J. Physiol. Lung Cell Mol. Physiol.* **289**, L698 (2005).
- Singh, R., Pantarotto, D., Lacerda, L., Pastorin, G., Klumpp, C., Prato, M., Bianco, A., Kostarelos, K. *Proc. Natl. Acad. Sci. USA* **103**, 3357 (2006).
- Sitharaman, B., Kissell, K.R., Hartman, K.B., Tran, L.A., Baikarov, A., Rusakova, I., Sun, Y., Khant, H.A., Ludtke, S.J., Chiu, W., Laus, S., Toth, E., Helm, L., Merbach, A.E., Wilson, L.J. *Chem. Commun. (Camb)*, 3915 (2005).
- Sneed, P.K., Stauffer, P.R., McDermott, M.W., Diederich, C.J., Lamborn, K.R., Prados, M.D., Chang, S., Weaver, K.A., Spry, L., Malec, M.K., Lamb, S.A., Voss, B., Davis, R.L., Wara, W.M., Larson, D.A., Phillips, T.L., Gutin, P.H. *Int. J. Radiat. Oncol. Biol. Phys.* **40**, 287 (1998).
- Son, S.J., Bai, X., Nan, A., Ghandehari, H., Lee, S.B. *J. Control. Rel.* **114**, 143 (2006).
- Stern, J.M., Stanfield, J., Kabbani, W., Hsieh, J.T., Cadeddu, J.A. *J. Urol.* **179**, 748 (2008).
- Stewart, B. Kleihues, P. (eds), *World Cancer Report* (IARC, 2003).
- Subjeck, J.R., Sciandra, J.J., Chao, C.F., Johnson, R.J. *Br. J. Cancer Suppl.* **5**, 127 (1982a).
- Subjeck, J.R., Sciandra, J.J., Johnson, R.J. *Br. J. Radiol.* **55**, 579 (1982b).
- Takahashi, A., Matsumoto, H., Nagayama, K., Kitano, M., Hirose, S., Tanaka, H., Mori, E., Yamakawa, N., Yasumoto, J., Yuki, K., Ohnishi, K., Ohnishi, T. *Cancer Res.* **64**, 8839 (2004).
- Tomei, L.D., Cope, F.O. (ed.), *Apoptosis: The Molecular Basis of Cell Death* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1991).
- Torti, S.V., Byrne, F., Whelan, O., Levi, N., Ucer, B., Schmid, M., Torti, F.M., Akman, S., Liu, J., Ajayan, P.M., Nalamasu, O., Carroll, D.L. *Int. J. Nanomed.* **2**, 707 (2007).
- Vidair, C.A., Dewey, W.C. *Radiat. Res.* **105**, 187 (1986).
- Wang, S., Diller, K.R., Aggarwal, S.J. *J. Biomech. Eng.* **125**, 794 (2003).
- Weissleder, R. *Nature Biotechnol.* **19**, 316 (2001).
- Wick, P., Manser, P., Limbach, L.K., Dettlaff-Weglikowska, U., Krumeich, F., Roth, S., Stark, W.J., Bruinink, A. *Toxicol. Lett.* **168**, 121 (2007).
- Williams, D.L., Pittman, T.L., Deshotel, M., Oby-Robinson, S., Smith, I., Husson, R. *J. Bacteriol.* **189**, 8818 (2007).
- Wong, R.S., Kapp, L.N., Krishnaswamy, G., Dewey, W.C. *Radiat. Res.* **133**, 52 (1993).
- Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., Gao, S., Zhang, Q., Qu, G., Chen, Y., Liu, G., Zhang, B., Yan, B. *Nano Lett.* **8**, 859 (2008).

# Nanoparticles in medicine

D. Maysinger, P. Kujawa, and J. Lovrić

14

## 14.1 Introduction

Nanomedicine is a fast-growing new scientific discipline in which various applications of nanotechnological products are expected to bring about breakthroughs in healthcare. It offers new solutions for therapeutic interventions and very likely tools that will achieve earlier diagnoses of pathologies than is possible using currently available means. There is considerable enthusiasm for the introduction into medicine of advanced and well-characterized nanotechnological products that will be especially effective in fighting cancer (Ferrari 2005; Duncan 2006), diabetes (Khafagy El *et al.* 2007), neurodegenerative disorders (Silva 2006) and cardiovascular diseases (Kong and Goldschmidt-Clermont 2005). For example, innovative nanoparticulate drug-delivery systems could markedly improve cancer therapy by guiding drugs to tumor cells, coincidentally reducing untoward effects on healthy tissues. Indeed, some of these nanoparticle approaches are currently in clinical trials, and early results seem promising. Certainly the survival rate of cancer patients would be likely improved by the early detection of abnormal cells, and the monitoring of disease progression during the course of therapeutic intervention. These outcomes could be achieved by employing highly sensitive and reliable imaging tools based on nanoprobes.

Nanoprobes and nanosensors are being developed for an early diagnosis of other diseases aside from cancer, e.g. diabetes. Currently, most patients are asymptomatic in the early stages of this disease, the norm being relatively late diagnosis and expensive therapy thereafter. Early diagnosis and consequent lifestyle changes could significantly slow down the progression of diabetes. Significant research efforts are also in progress to develop nanoparticles that contain antidiabetic drugs and/or insulin capable of a feedback-modulated release to control glucose homeostasis. An early detection of neurodegenerative changes would also allow the early diagnosis of the common neurodegenerative disorders, leading to earlier therapeutic intervention to slow down or halt their progression. Currently, by the time that such neurodegenerative disorders are diagnosed there is already an extensive and irreversible neuronal damage. Design of non-invasive imaging agents and drug-delivery systems

14.1 Introduction	503
14.2 Current problems with use of nanoparticles in medicine	513
14.3 Nanoparticle–cell interactions	519
14.4 Nanoparticles as imaging tools in animals and humans	525
14.5 Conclusions	530
Acknowledgments	530
References	531

capable of crossing the blood/brain barrier is thus a research focus of many laboratories, both in academia and in industry, the objective being diagnosis and intervention in the earliest possible stages of central nervous system disorders.

New nanobiotechnological tools could also improve the survival of patients with cardiovascular problems by enabling an early detection of vascular plaques at high resolution that would distinguish unstable plaques before rupture and clogging of blood vessels. Similarly, advances in nanomedicine and regenerative medicine will likely play a crucial role in the detection and management of cardiac infarcts and strokes. “Intelligent nanomaterials” applied in these pathologies would not only serve as stimulus-responsive drug containers but would also “mark” the sites of injury in such a way that local adult stem cells would be attracted to them, as would administered cultured cells.

The promise of nanomedicine is to make available different nanosystems whose novel, usually size-dependent, physical, chemical and/or biological properties are exploited to combat the disease of interest.

One kind of particulate systems represents a vast array of either metallic, semiconductor, polymeric, protein or lipid nanoparticles that can be exploited for diagnosis and treatment of various diseases. These nanoparticles do not respond to the biological stimuli but mostly serve as drug containers, with or without targeting moieties, or they fluoresce. Fluorescent nanoparticles have been mainly explored for single-cell imaging and imaging of “diseased” (e.g. cancer) tissue. The second kind of systems includes biologically active polymeric drugs, dendrimers, as well as polymer–drug and polymer–protein water-soluble conjugates.

In this chapter we will focus on the first category of nanosystems, namely nanoparticles. First, we describe general issues related to physicochemical and biological properties of different nanoparticles. In the second part, we will discuss the current problems associated with the use of nanoparticles in medicine and suggest some solutions. In the third section, interaction of nanoparticles and cells and factors determining these interactions are discussed. Some examples of nanoparticles discussed in this chapter are shown in Fig. 14.1. The final section provides some examples of new approaches for real-time imaging of experimental animals that could be useful, complementary methods for evaluations of effectiveness (or toxicity!) of novel nanomaterials and nanomedicines.

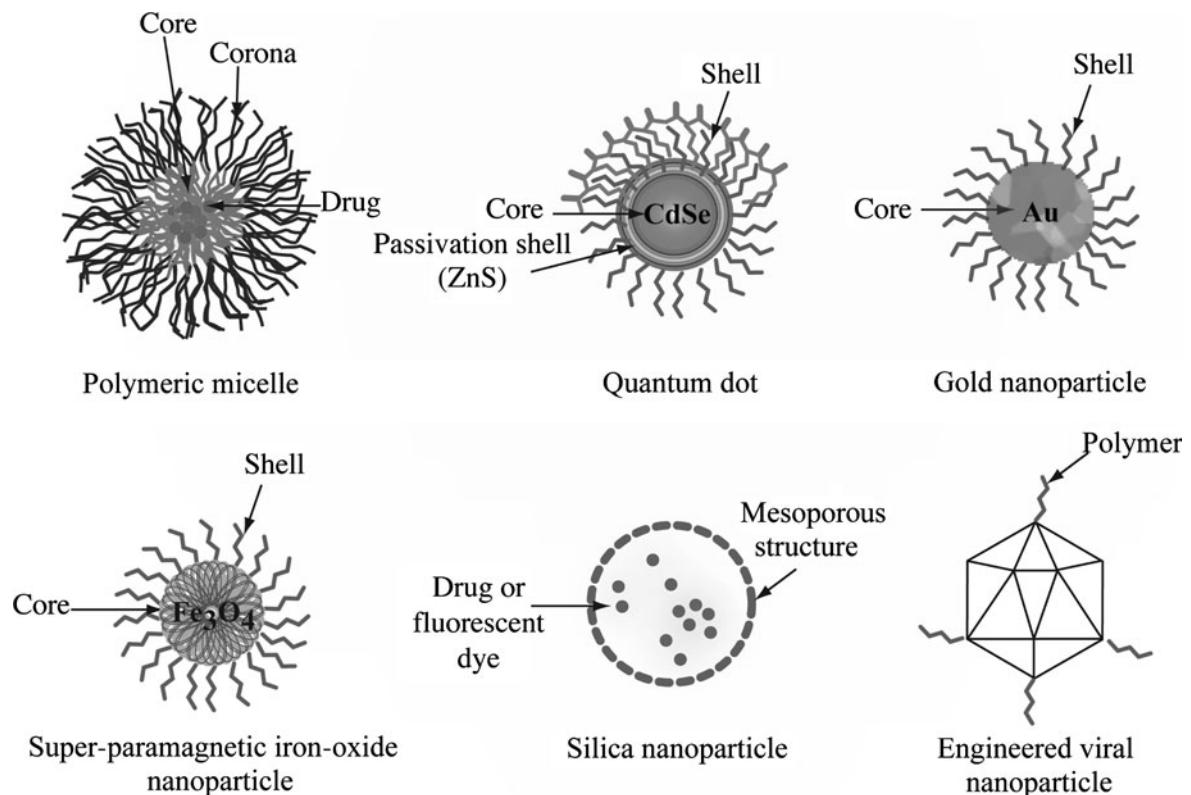
### **14.1.1 Nanoparticles in research and clinical investigations**

Drug-delivery systems currently dominate nanomedicine (Wagner *et al.* 2006; Zhang *et al.* 2008), but significant progress has been also made in development of *in-vivo* imaging agents (ESF 2005). Examples of some of the nanoparticles already on the market, or in the clinical or pre-clinical investigations are shown in Tables 14.1–14.3.

**Table 14.1** Examples of nanoparticles on the market

Product	Nanoparticle/Drug	Indication	Company
Drug delivery			
Abraxane	Albumin-bound paclitaxel particles	Breast cancer	Abraxis BioScience, AstraZeneca
AmBisome	Liposome/Amphotericin B	Fungal infections	Gilead Sciences
DaunoXome	Liposome/Daunorubicin	Kaposi sarcoma	Gilead Sciences
Depocyt	Liposome/Cytarabine	Cancer	SkyePharma
DepotDur	Liposome/Morphine	Postsurgical analgesia	SkyePharma
Doxil	PEGylated liposome/Doxorubicin	Metastatic ovarian cancer, Kaposi sarcoma	OrthoBiotech, Schering-Plough
Estrasorb	Micelles/Estradiol	Menopausal therapy	Novavax
Myocet	Liposome/Doxorubicin	Breast cancer	Zeneus Pharma
Visudyne	Liposome/Verteporfin	Macular degeneration	QLT, Novartis
<i>In-vivo</i> imaging			
Feridex/Endorem	Iron-oxide nanoparticles	Liver tumors	Advanced magnetics, Guerbet
Gastromark/Lumirem	Iron-oxide nanoparticles	Abdominal structures	Advanced magnetics, Guerbet
Resovist	Iron-oxide nanoparticles	Liver tumors	Schering

PEG, polyethyleneglycol.

**Fig. 14.1** Some representative classes of nanoparticles as potential nanodiagnostic and nanotherapeutic means in medicine

**Table 14.2** Examples of nanoparticles/nanomedicines in clinical trials

Product	Nanoparticle/Drug	Indication	Company
Drug delivery			
AeroLEF	Liposome/fentanyl	Postoperative analgesic	Delex Therapeutics
AI-850	Paclitaxel nanoparticles in porous, hydrophilic matrix	Solid tumors	Acusphere
L-Annamycin	Liposome/annamycin	Acute lymphocytic leukemia, acute myeloid leukemia	Callisto
Basulin	Polyglutamate polymer dotted with insulin	Type I diabetes	Flamel Technologies
Biovant	Nanosized calcium phosphate	Vaccine component	BioSante Pharmaceuticals
Cyclosert-camptothecin	Cyclodextrin nanoparticles	Metastatic solid tumors	InsertTherapeutics
CYT-6091 (Aurimmune)	TNF $\alpha$ -bound PEGylated colloidal gold particles	Solid tumors	CytImmune Sciences
ProLindac	HMPA copolymer-DACH/platinate	Ovarian cancer	Acess Pharmaceuticals
INGN-401	Liposome	Metastatic lung cancer	Introgen
Transdrug	PIHCA polymer nanoparticles/doxorubicin	Liver cancer	BioAlliance Pharma
NB-001	Nanoemulsion-based therapy	Herpes labialis	NanoBio
NeoLipid LEP-ETU	Liposome/Paclitaxel	Cancer	NeoPharm
Onco TCS	Liposome/vincristine	Non-Hodgkin's lymphoma	Inex, Enzon
OSI-211	Liposome/lurtotecan	Ovarian cancer	OSI Pharmaceuticals
Sarcodoxome	Liposome/doxorubicin	Soft-tissue sarcoma	GP-Pharm
SLIT Cisplatin	Liposome/cisplatin	Progressive osteogenic sarcoma	Transave
SP1049C	Pluronic block copolymer/doxorubicin	Esophageal carcinoma	Supratek Pharma
SPL 7013 (VivaGel)	Dendrimer/Microbicidal drug	STD prevention	Starpharma
NK911	PEGylated polymeric micelles/Doxorubicin	Cancer	NanoCarrier Co.
<i>In-vivo imaging</i>			
Combidex (Ferumoxtran-10)	Iron-oxide nanoparticles	Cancer	Advanced magnetics

DACH, diaminocyclohexane; HMPA, hydroxypropylmethacrylamide; PIHCA, polyisohexyl-cyanoacrylate; STD, sexually transmitted diseases; TNF $\alpha$ , tumor necrosis factor-alfa.

**Table 14.3** Examples of nanoparticles in pre-clinical investigations

Product	Nanoparticle/Drug	Indication	Company
Drug delivery			
AuroLase	Gold nanoshell	Head and neck cancer	Nanospectra Biosciences
CYT-21001 (Auritol)	Taxol and TNF $\alpha$ -bound PEGylated colloidal gold nanoparticles	Solid tumors	CytImmune Sciences
PNDS-SN38	Polymeric micelle/SN38	Colon cancer	Labopharm
PNDS-propofol	Polymeric micelle/Propofol	Anesthesia	Labopharm
MRX-951	Branching block copolymer self-assembling nanoparticles	Cancer	ImaRx Therapeutics
TNT-Anti-Ep-CAM	Polymer-coated iron oxide	Solid tumors	Triton Biosystems
DF1	Dendritic fullerene	Chemoprotection	Carbon Nanotechnology
C3	Fullerene derivative	Chemoprotection	Carbon Nanotechnology
<i>In-vivo imaging</i>			
Dendrimer-Magnevist complex	PAMAM dendrimer	MRI imaging agent	Dendritic Nanotechnologies
Biocjugated nanoparticles	Quantum dots encapsulated with ABC triblock copolymer	Cancer	Emory-Georgia Tech Nanotechnology Center

PAMAM, polyamidoamine.

### 14.1.2 Polymeric nanoparticles

Polymeric nanoparticles are defined as polymeric matrix-like colloidal particles, characterized by the ability to provide targeted and predictable controlled release of the drug (Moghimi *et al.* 2001; Panyam and Labhasetwar 2003). The drug of interest is either encapsulated within polymer matrix or adsorbed or conjugated onto the surface of nanoparticle. A number of different polymers, both natural (gelatin, alginate, chitosan, etc.) and synthetic (polylactide-polyglycolide copolymers, polyacrylates, polycaprolactones, etc.) have been utilized for the formulation of polymeric nanoparticles. By changing the formulation composition or method of formulation, the nanoparticle size, charge, hydrophobicity/hydrophilicity and drug release can be easily controlled. A diverse array of therapeutic agents, such as lipophilic or hydrophilic drugs, oligonucleotides, DNA, peptides, proteins and antigens, can be encapsulated in polymeric nanoparticles (Bertin *et al.* 2006). The drug from the polymer matrix is released through the diffusion of the drug in the matrix and by the degradation of the polymer matrix. Polymers used for the preparation of nanoparticles are usually biodegradable and their biocompatibility has been extensively studied both in animals and humans. Cell-permeable and biocompatible polymeric nanoparticles are promising nanotools for imaging of cellular processes in single cells (Kim *et al.* 2006).

### 14.1.3 Polymeric micelles

Polymeric micelles are formed via self-assembly of amphiphilic or oppositely charged copolymers in aqueous medium. They usually have spherical core/shell structure where the hydrophobic core acts as a reservoir for the entrapped drugs, proteins, or DNA. Hydrophilic corona faces the biological medium and enhances biocompatibility of the system. Physicochemical versatility of micellar structures allows one to develop tailor-made nanocarriers that are designed taking into account chemical and pharmacological properties of the incorporated drug, route of administration and nature of the disease. By optimizing the chemical structure of the core-forming part of the polymer the drug-entrapment efficiency and stability of the micellar aggregates can be improved, as well as control of drug-release kinetics from the carrier can be achieved. The chemical structure of the polymer can also be altered in order to change the biological destination of the carrier or to incorporate stimuli-responsive units that allows one to optimize the release of encapsulated compounds and to enhance the targeting efficiency of the drug.

Research on applications of polymeric micelles as drug-delivery vehicles was pioneered by Ringsdorf *et al.* in the 1980s, followed by Kataoka's group through the development of doxorubicin (DOX)-conjugated block copolymer (Kataoka *et al.* 2000; Nakanishi *et al.* 2001), and Kabanov's group (Kabanov *et al.* 2002), who developed polymeric micelles with non-covalently incorporated drug. Originally, polymeric micelles were investigated as an intravenous delivery system but nowadays their applications in oral drug delivery are also considered (Croy and Kwon 2006; Nishiyama and Kataoka 2006). Several polymer-based nanomedicines are presently on the market and many clinical

investigations are in progress to show their usefulness in the treatment of various diseases and especially in anticancer therapy (Le Garrec *et al.* 2004b; Park *et al.* 2008). This is because polymeric micelles are able to accumulate efficiently in tumor sites due to the enhanced permeability and retention (EPR) effect (Maeda 2001). The EPR effect results from the fact that blood vessels in pathological tissues have a higher permeability compared to the healthy tissue. Moreover, lymphatic drainage is not fully developed in tumors. These features lead to an efficient accumulation of polymeric micelles in tumor and other inflamed tissues, provided that they circulate long enough in the bloodstream. An increased circulation time of nanocarriers can be achieved by reducing their non-specific recognition and uptake by the reticuloendothelial system. It has been shown that this effect is greatly enhanced by the presence of hydrophilic chains (e.g. poly(ethylene glycol)) on the surface of micelles or other nanoparticles (Torchilin 1998).

#### 14.1.4 Quantum dots

Quantum dots (QDs) are nanoparticles made of a hundred to a thousand atoms. They have unique optical and electrical properties strongly dependent on their size (Alivisatos 1996). QDs are unique because their emission maxima shift to higher wavelengths when their size increases. No matter what their size, QDs can be excited within a relatively large range of wavelengths. The emission spectra are rather narrow, allowing for the simultaneous detection of multicolored QDs upon illumination with a single-wavelength light source. The size and morphology of QDs is easily tunable. QDs have exceptionally narrow size distributions. Their resistance to photobleaching, which permits prolonged visualization of the target, is also a valuable addition to their properties (Dahan *et al.* 2003). In addition, due to the electron dense QDs cores, they are suitable for analyses by electron microscopy. All these QD properties makes them particularly suitable for studying their intracellular fate by dual imaging, i.e. confocal and electron microscope. These two microscopic techniques are complementary and provide both the information about the QD distribution in different cellular organelles (EM) and a dynamics of their movements in living cells (Larson *et al.* 2003).

Due to their outstanding properties QDs are expected to “revolutionize” live-cell imaging, *in-vivo* imaging of tissues in animals and diagnostic analyses (Medintz *et al.* 2005; Michalet *et al.* 2005). For these purposes QDs have to be coated with hydrophilic ligands that render them “water soluble”. High-quality QDs are mostly synthesized using high-temperature solution reactions and are capped with hydrophobic organic ligands. As such, they are “insoluble” in water and the first hurdle to overcome is to preserve their fluorescence by making them “water soluble”. Various methods were reported to achieve this goal, e.g. exchange of native ligands with hydrophilic ones, encapsulation within amphiphilic polymers or phospholipid micelles (reviewed in Medintz *et al.* 2005). There is also a need for the development of new capping ligands that can improve the stability of QDs in biological environments and ease the coupling with targeting moieties. Modular ligands based on poly(ethylene

glycol) that provide stable and highly water soluble QDs with an easy access to various biological entities were recently described (Susumu *et al.* 2007). Due to the large surface-to-volume ratio, a QD's surface can be decorated with targeting molecules and also therapeutic agents, creating multifunctional nanoparticles for simultaneous imaging and therapy (Chen *et al.* 2005; Derfus *et al.* 2007; Tan *et al.* 2007). Recently, QD-aptamer-doxorubicin conjugates were reported as targeted cancer imaging, therapy and sensing system (Bagalkot *et al.* 2007). Fluorescence resonance energy transfer can be exploited for bio-sensing and delivery of anticancer drugs specifically to tumor cells (Bakalova *et al.* 2004). Recently, modified QDs with attached ligands excitable in the near-infrared spectral region were prepared and tested as photosensitizers for cancer therapy. Similar, more sophisticated QD-sensitizers are particularly attractive because of their near-infrared excitation, allowing for deep penetration into tissues and sensitization at these sites (Dayal and Burda 2008).

### 14.1.5 Gold nanoparticles

Gold nanoparticles (AuNP) have been used for centuries in the arts because of their vivid colors. Their appeal is in strictly defined shapes and sizes as well as their relatively high biocompatibility. However, only limited data has shown how systematic alteration of NP shape, size, charge or drug ligand can determine the fate of AuNPs in specific cell types. Recently, Chan's group reported on herceptin-modified gold nanoparticles significantly taken up by SK-BR-3 cells, expressing high number of Erb receptors (Jiang *et al.* 2008). The study demonstrates that engineered gold nanoparticles with well-defined sizes can selectively induce membrane receptor internalization and downregulate their expression level. As a consequence, downstream signalling pathways are also affected. Hence, the studies support the view that nanostructures such as gold nanoparticles, are not simply passive bystanders in the cell or around the cell but rather that they can alter cell functions.

Synthetic procedures yielding precise shapes, sizes and surfaces of AuNP have only recently been achieved (reviewed in Jennings and Strouse 2007). Popular analytical determination methods of gold in the biosciences are mostly based on plasmon resonance (SPR) and detection of shifts in SPR bands when AuNPs interact with proteins (Taton *et al.* 2000). Significant contributions to AuNP applications have been made by several groups (Taton *et al.* 2000; Hirsch *et al.* 2003; Yun *et al.* 2005; Chithrani *et al.* 2006; Huang *et al.* 2006).

In summary, gold nanoparticles are of great interest for biomedical applications due to their (i) relatively easy preparation with controlled size, morphology and surface properties; (ii) facile conjugation to a variety of biomolecular ligands; (iii) strong absorbing and scattering properties; and (iv) biocompatibility. Spherical AuNPs have the advantage of small size, but their use *in vivo* has been limited by their absorption in the visible region. Visible light cannot penetrate deep into body tissue and can cause photochemical decomposition. In contrast, near-infrared light (NIR) penetrates deep within tissues, and since

NIR absorption is away from the region of biomolecular excitation transitions, NIR does not cause photochemical damage. For instance, gold nanoshells have an optical resonance in the near-infrared region of light and the resonance wavelength can be increased by increasing the total nanoshell size or the core-to-shell radius. Gold nanorods also possess a tuneable optical resonance in NIR region, but at much smaller effective sizes than nanoshells. Gold nanorods seem to have the most superior NIR absorption and scattering (tuneable by changing the aspect ratio—length divided by width) at much smaller particle sizes and their absorption and scattering coefficients are an order of magnitude higher than those for nanoshells and nanospheres (Chen *et al.* 2006; Huang *et al.* 2006; Oyelere *et al.* 2007).

#### 14.1.6 Iron-oxide nanoparticles

Super-paramagnetic iron-oxide nanoparticles (spion) are small  $\gamma\text{-Fe}_2\text{O}_3$  or  $\text{Fe}_3\text{O}_4$  particles with the size of a core around 6–15 nm. Upon appropriate surface coating with organic or inorganic materials, these nanoparticles can be dispersed into suitable solvents forming homogeneous suspensions, called ferrofluids. Such a suspension can interact with external magnetic field and be positioned to a specific region, enabling, e.g. magnetic resonance imaging or magnetic-field-assisted physical drug targeting in cancer therapy. For these applications, the particles must show combined properties of supermagnetism, biocompatibility, and functional surface. The number of possible spion applications in medicine has significantly increased over the last couple of years (Gupta and Gupta 2005; Neuberger *et al.* 2005). In the field of clinical medicine, spions have been used *in vivo* for gene transport, magnetic resonance imaging, hyperthermia, and radiotherapy. Furthermore, iron-oxide nanoparticles are used for immunomagnetic separation of cells, proteins, nucleic acids, bacteria, and viruses.

Recently, a lot of research efforts are focused on multifunctional nanomedicines, which combine the properties of a few classes of nanoparticles. For instance, a multifunctional platform for targeted drug delivery has recently been described (Nasongkla *et al.* 2006). It consists of polymeric micelles that can actively target tumors by molecular recognition of cancer-specific markers,  $\alpha\text{v}\beta 3$ -integrins, by cRGD ligands attached to the micellar corona. The micelles are loaded with doxorubicin, an anticancer drug, and with spions, which are applied as efficient magnetic resonance imaging contrast agents. This integrated nanoparticulate system opens many opportunities for targeted delivery as well as the use of magnetic resonance imaging to monitor tumor-targeting efficiency.

When spions are exposed to an alternating magnetic field, their magnetic moment oscillates and the energy of these oscillations is released in the form of heat. This property has been used for hyperthermic treatments of cancers, since tumor cells are highly susceptible to high temperatures (42–46 °C). Currently, clinical trials in patients with brain and prostate tumors are being conducted using local magnetic hyperthermia in combination with radiotherapy (Johannsen *et al.* 2005, 2007; Maier-Hauff *et al.* 2007).

### 14.1.7 Silica and organically modified silica nanoparticles

The unique features of silica and organically modified silica (ORMOSIL) nanoparticles are their chemical and thermal stability, mesoporous structure with tuneable pore sizes and well-defined surface properties. Mesoporous silica nanoparticles were found to be efficient in intracellular delivery of membrane-impermeable proteins (Slowing *et al.* 2007). ORMOSIL nanoparticles surface functionalized by amino groups for binding DNA, however, demonstrated their potential in gene delivery (Bharali *et al.* 2005a; Roy *et al.* 2005). FloDots, silica nanoparticles doped with fluorescent dyes represent an important probe for *in-vivo* diagnostics (Yao *et al.* 2006). Dye molecules can be either luminescent organic or inorganic molecules dispersed within the silica matrix. These novel nanoparticles are extremely luminescent and are highly photostable (Wang *et al.* 2005). Silica provides dispersion of nanoparticles in water, and its surface can be modified to contain functional groups. Examples of thousand-fold increases in fluorescence intensities of FloDots over that of pure dyes have been reported (Yao *et al.* 2006). Similar to QDs, FloDots can be analyzed both by fluorescent and electron microscopy. One of the most promising applications of FloDots is for detection of different microorganisms in biological fluids and possibly *in vivo*.

### 14.1.8 Nanotubes

Carbon nanotubes are fullerene molecules, highly ordered graphene sheets, which are rolled-up and can be single-walled or multiwalled (Martin and Kohli 2003; Bianco *et al.* 2005). Their diameter ranges from one to tens of nanometers, while their length can range from a few micrometers to hundred of micrometers or more. Due to their structural, mechanical, electrical and optical properties, nanotubes are explored as biosensors (Chen *et al.* 2003), drug-delivery systems (Kam and Dai 2005; Kam *et al.* 2005a) and even new therapeutic agents (Kam *et al.* 2005b). For biomedical purposes carbon nanotubes are functionalized, which renders them hydrophilic, and as such, nanotubes can easily cross cell membranes and deliver their cargo to intracellular sites (Kostarelos *et al.* 2007). Nanotubes functionalized with phospholipids bearing polyethylene-glycol are surprisingly stable *in vivo*, they exhibit long blood circulation times and low uptake by RES (Liu *et al.* 2007b). An efficient targeting of integrin-positive tumors was achieved with PEG-modified nanotubes on their surface with arginine-glycine-aspartic acid peptide as a targeting moiety (Liu *et al.* 2007b).

### 14.1.9 Dendrimers

Dendrimers are polymeric molecules composed of multiple perfectly branched monomers that emanate radially from a central core (Lee *et al.* 2005). The number of branch points determines the generation of dendrimer. Low-generation dendrimers have a floppy, disc-like structure, while higher generations (>4th generation) have a globular or even spherical conformation. The uniqueness of these tree-like macromolecules in comparison with other

polymers is their monodispersity in size. A monodispersed product is desirable for synthetic, experimental and therapeutic reproducibility. Additionally, of great importance is the fact that their size, shape, branching length or density and surface functionality is easily controlled (Svenson and Tomalia 2005). The first and the most studied class of dendrimers are polyamidoamine (PAMAM) dendrimers, while other types are based on polyamides (polypeptides), poly(aryl ethers), polyesters, carbohydrates and DNA. Dendrimers have a high potential for drug (Malik *et al.* 1999; Kukowska-Latallo *et al.* 2005) and gene delivery (Haensler and Szoka 1993; Tang *et al.* 1996), as well as for *in-vivo* imaging (Kobayashi and Brechbiel 2003). Cargo can be either encapsulated into the interior of dendrimers or chemically attached or physically adsorbed onto the dendrimer surface (Patri *et al.* 2005). Dendrimers can also be exploited as photonic oxygen sensors (Dunphy *et al.* 2002). The concentration of oxygen in tumor tissue is an important parameter for determining appropriate therapy. Dendrimers seem to be able to increase the efficacy of anticancer therapy by altering the oxygen concentration in neoplasm. Dendrimers even without any conjugated compounds were found to be biologically active. This dendrimer property was explored in the context of infectious diseases, e.g. to stimulate the removal of prion proteins in the infected cells (Supattapone *et al.* 1999), and also, to block adsorption and entrance of viruses into cells (Bourne *et al.* 2000).

#### 14.1.10 Liposomes

Liposome-based formulations are successfully used in medicine and pharmacy mainly as drug-delivery vehicles (Torchilin 2005). The main problem encountered during the early stages of development in this area has been related to the fast clearance of the liposomes from the bloodstream and the leaky structure of lipid bilayer, which made controlled drug release to the surrounding medium difficult. A suitable release rate and residence time in the bloodstream of liposomal formulations have been achieved by the modification of lipids with poly(ethylene glycol) and addition of specific excipients (Harashima and Kiwada 1996). Although several liposome-based formulations are currently available on the market (e.g. Doxil®, Caelyx®, or DaunoXome®) (Gill *et al.* 1996; Gabizon *et al.* 2003), recent studies have indicated that their repeated dosing can result in accumulation in the liver and acute hypersensitivity (Judge *et al.* 2006). This makes one of the long-standing paradigms of nanomedicine, that poly(ethylene glycol) modification of nanoparticles does not result in immunogenic and toxic response, questionable and needed to be verified.

#### 14.1.11 Engineered viral nanoparticles and artificial viruses

Viruses are natural nanoparticles characterized by robust protein cages with well-defined geometry and remarkable uniformity. A variety of viruses have been widely investigated as delivery vectors in gene therapy, and still remain unprecedented due to significantly higher transfection efficiencies

than any other artificial nanoparticulate delivery system. Moreover, only viruses can grant site-specific integration in a host genome. Nevertheless, their application in gene therapy is limited because of the possibility of host infection and immunological reactions. Many strategies are in development to overcome these problems, including multivalent polymer coating of viral particles (Green *et al.* 2004) and generation of mutant viral capsids with new properties (Maheshri *et al.* 2006). Another limitation is a lack of selectivity in transducing only cells of interest, but substantial progress in modifying the surface of viral vectors using diverse techniques now allows targeting to many cells (Waehler *et al.* 2007). Recently, the idea of making nanoparticles artificial viruses was born, and it is believed that these nanosystems will greatly impact gene therapy. The challenge is to design an artificial carrier that equals or exceeds its viral counterpart in terms of transfection efficiency, yet that is safe to use, target-cell specific, non-immunogenic and relatively inexpensive (Mastrobattista *et al.* 2006).

Viral particles are also investigated as platforms for diagnostic imaging (Manchester and Singh 2006). Furthermore, multifunctional viral nanoparticles for the delivery and imaging can be developed by designing hybrid eukaryotic/prokaryotic viral particles (Hajitou *et al.* 2006) or by encapsidating nanoparticles such as quantum dots and gold nanoparticles (Dixit *et al.* 2006; Loo *et al.* 2006).

## 14.2 Current problems with use of nanoparticles in medicine

The main problems related to the application of nanoparticles today are: (i) nanoparticle stability (shelf-life stability or stability in a biological system after administration); (ii) surface chemistry (iii) penetration through biological barriers including intracellular membranes; (iv) lack of selectivity for particular cell type; (v) unexplored clearance from human body and (vi) toxicity.

### 14.2.1 Stability

Stability of nanoparticles is a very important issue as unstable nanoparticles for example can release their cargo before reaching their target, whereas too stable or non-biodegradable nanoparticles could accumulate in the body with possible negative consequences on patient health. Nanoparticle stability can be differentiated at four levels: (i) physical stability, (ii) stability in circulation, (iii) intracellular stability and (iv) stability after performing their function in the body. Stability is highly dependent on nanoparticle type and closely related to nanoparticle surface properties.

For biomedical applications nanoparticles have to be dispersible in aqueous media. Problems that can occur in such media are either agglomeration of nanoparticulate system or its disintegration. Certain nanoparticles, such as micelles, self-assemble in aqueous media, whereas QDs, often synthesized in an organic solvent, have to be modified to be dispersed in water. The composition of water media is a crucial factor determining nanoparticle

stability. For example, polymeric nanoparticles, whose colloidal stability in an aqueous solution was ensured for weeks, coagulated in the cell-culture medium (Lorenz *et al.* 2006). Different types of cell-culture media, as well as absence or presence of serum, were found to affect differently stability of supermagnetic iron-oxide nanoparticles (Petri-Fink *et al.* 2008). The fluorogenic-based approach was used to address the integrity of micelles under complex biological conditions (Savic *et al.* 2006). One way to enhance the micelle stability is by using shell cross-linking and formation of shell cross-linked knedel-like (SCKs) nanoparticles (Rossin *et al.* 2005).

A crucial role in determining the stability of QDs and other metallic nanoparticles in an aqueous environment belongs to the organic molecules attached to their surface. Loss of the surface ligands ultimately leads to the precipitation of QDs and loss of their luminescence. Different processes can affect the stability of surface ligands, and it was shown that the desorption of hydrophilic thiols, common surface ligands used, can be caused by their protonation (Aldana *et al.* 2005) or by their photooxidation (Aldana *et al.* 2001). Deprotonated thiols, thiolates, are bound to cadmium chalcogenide QDs, so if the pH at the QD/ligand interface decreases to a certain value, the ligands get protonated and detach from the QD surface. The precipitation of QDs coated with hydrophilic thiols occurs in a relatively low pH range, between 2 and 7, depending on the size and chemical composition of QDs. Moreover, prolonged UV illumination of QDs can compromise their stability by inducing photooxidation of surface ligands. Increasing the thickness and packing density of ligand shell can delay the initiation process of photooxidation. Preparation of QDs with ligands relatively resistant to photooxidation will enable QDs to fulfill their capacity for long-term cellular or *in-vivo* imaging. Due to its inorganic nature, the QD core itself is relatively resistant to photodegradation, however, prolonged exposure to UV light also affects the core integrity (Derfus *et al.* 2004).

Agglomeration and degradation of nanoparticles can occur in circulation leading to nanoparticle sequestration in the reticulo-endothelial system (RES). RES is composed of monocytes and macrophages that are responsible for phagocytosis and removal of cellular debris, pathogens, foreign macromolecules and particles from the bloodstream. One of the solutions to this problem is the PEGylation of the nanoparticle surface that then reduces the uptake by RES and prolongs the plasma half-life of nanosystems.

Depending on the pathway of internalization, nanoparticles can be exposed to different cellular microenvironments, which can affect their stability in different ways. After completing their function, nanoparticles have to be cleared from the body. Multiple factors can determine their elimination, the most important being their size, charge, shape and biodegradability (Soo Choi *et al.* 2007).

### 14.2.2 Surface properties

The surface of nanoparticles plays a crucial role in nanoparticle behavior. Firstly, properties of the groups present on the surface or adsorbed molecules

determine the stability of nanoparticulate system. Molecules on the surface are also responsible for the nanoparticle's interaction with cells, and they can determine the fate of the nanoparticle itself in the circulation and destine the nanoparticle to a certain cell population or cell organelle.

Important factors in view of surface groups are their size, charge, packing density and targeting capability. The hydrodynamic size of the nanoparticle is logically dependent on the size of surface molecules themselves, their structure and flexibility. Then again, hydrodynamic diameter can vary due to the change in the environment. For example, the hydrodynamic diameter of polyethyleneimine (PEI) coated QDs reduces from 15.3 nm to 13 nm by slight alkalization of the medium (Duan and Nie 2007).

The charge of nanoparticle surface influences its interaction with the cell surface and consequently internalization. If nanoparticles are positively charged, in contact with cells their interaction is due to the electrostatic interactions with negatively charged glycocalyx on cell membranes. Charge can also be exploited for carrying the cargo inside the cell. Positively charged nanoparticles can carry negatively charged DNA molecules (De La Fuente *et al.* 2008) while negatively charged nanoparticles can bind positively charged proteins at their surface (Mattossi *et al.* 2000).

The packing density of molecules at the surface of nanoparticles is an important factor for the integrity of nanoparticles. For example, in the case of quantum dots whose core is sensitive to oxidation, a higher density of surface molecules will slow down the diffusion of oxygen molecules and prevent core degradation (Aldana *et al.* 2001). Furthermore, the attachment and the density of biomolecules at the surface of nanoparticles also affects nanoparticle stability, which is, for example, in the case of QDs evident as change in the fluorescence intensity, surface charge, colloidal stability, and biological interactions (Clarke *et al.* 2008).

### 14.2.3 Crossing biological barriers

From the point of administration to their intended target, nanoparticles have to overcome certain biological/physiological barriers. Depending on the route of administration, the endothelium or epithelium represent major barriers that limit the accessibility of nanoparticles to their intended target. However, there are other obstacles that have to be overcome before nanoparticles reach their final destination. For instance, after intravenous administration, nanoparticles in the blood usually interact with blood proteins, which can lead to the formation of protein corona. This protein corona can then largely define the biological destiny of the nanoparticle (Cedervall *et al.* 2007). First, the adsorption of the proteins from the class of opsonins at the nanoparticle surface, known as the process of opsonization, can take place. This process leads to the sequestration of nanoparticles by RES. Opsonins are capable of interacting with specific plasma membrane receptors on monocytes and tissue macrophages, hence promoting nanoparticle recognition by these cells (Moghimi *et al.* 2001). Depending on the nanoparticle surface properties such as charge, nanoparticles attract different types of opsonins as well as

other plasma proteins. Adsorption of plasma proteins can further affect the biodistribution of nanoparticles. If not cleared from the blood, nanoparticles have to then cross the vascular endothelium. Extravasation of nanoparticles can take paracellular and transcellular routes (trancytosis). The paracellular route is limited due to the tight junctions ( $<2$  nm) between endothelial cells and the underlying basement membrane (Garnett and Kallinteri 2006). The transcellular route includes endocytosis of nanoparticles at the apical membrane and exocytosis at basal membrane. Continuous endothelium contains distinct flask-shaped invaginations in plasma membrane called caveolae and it is thought that they may provide a trafficking pathway into and possibly across cells (Schnitzer 2001; McIntosh *et al.* 2002). However in certain disease states, such as cancer and inflammation, integrity of the endothelium is compromised and it can become permeable for nanoparticles. When they reach the target tissue, nanoparticles have to travel through the extracellular matrix to reach the target cells. The extracellular matrix represents an additional barrier on the journey of nanoparticles to their destination. A transport of nanoparticles in different cells and biological fluids probably occurs both by diffusion and convection, and could be slowed down by non-specific interaction with extracellular matrix components and high interstitial fluid pressures.

#### 14.2.4 Selectivity

The attachment of specific biomolecules to the surface provides nanoparticles with the ability to home in to specific sites in the body. Specific targeting of nanoparticles has been accomplished by the attachment of antibodies (Gao *et al.* 2004), small molecules (e.g., transferrin, folic acid) (Bartlett *et al.* 2007), aptamers (Gu *et al.* 2008), or various peptide ligands (Simberg *et al.* 2007). These biomolecules usually target receptors on the cell surface, which are specific for certain cell types or overexpressed in disease states. For example, antibodies against tumor antigens, such as prostate-specific membrane antigen (Gao *et al.* 2004), or overexpressed HER2 receptor (Park *et al.* 2002; Wu *et al.* 2003) were exploited to achieve active targeting of quantum dots and liposomes to cancer cells *in vitro* and *in vivo*. New cell-specific targeting ligands can be discovered through phage display peptide libraries (Arap *et al.* 2002; Laakkonen *et al.* 2002). This technology is used to obtain defined peptide sequences that interact with a particular molecule. Moreover, the injection of a peptide display phage library into the bloodstream of an animal enables isolation and identification of peptides that have the ability to home in to particular organ or tissue (Pasqualini and Ruoslahti 1996). Using a tumor-homing peptide identified by *in-vivo* screening of a phage display peptide library, Simberg *et al.* constructed biomimetic nanoparticles that not only home in to tumors but also amplify their own homing. Iron-oxide nanoparticles and liposomes coated with this tumor-homing peptide accumulated in tumor vessels, where they induced additional local clotting, thereby producing new binding sites for more nanoparticles (Simberg *et al.* 2007).

Many drugs, and moreover proteins, DNA and siRNA require cellular internalization for efficacy, so cell-specific ligands that provide not only targeting

but the internalization of the nanoparticulate carrier will be essential for positive therapeutic outcome. For example, investigation of biodistribution and functional efficacy of targeted and non-targeted siRNA nanoparticles revealed that the attachment of transferrin to the nanoparticle surface has only negligible effect on the biodistribution, but significantly enhances the internalization of nanoparticles (Bartlett *et al.* 2007). This study suggests that the primary advantage of targeted nanoparticles is associated with cellular uptake in tumor cells rather than overall tumor localization.

Enhanced cellular internalization of nanoparticles can be achieved by the attachment of cell-penetrating peptides (CPPs) (Lewin *et al.* 2000; Torchilin *et al.* 2001; Zhao *et al.* 2002). The CPPs are defined by their ability to translocate across the plasma membrane and reach cytoplasmic and/or nuclear compartments in live cells (Joliot and Prochiantz 2004; Stewart *et al.* 2008). This was first shown for HIV Tat protein and has since expanded to non-natural peptides that share this property. Organelle-specific CPPs have also been generated, providing a means to target nanoparticles to specific subcellular destinations. Targeting to organelles can be achieved by signal peptides, such as nuclear localization sequence, that are used by cell machinery to target new translated peptides to their correct destination in the cell. A class of mitochondria-penetrating peptides (MPPs) was recently designed and engineered, to match conventional CPPs in cellular uptake but display strong mitochondrial localization (Horton *et al.* 2008). The MPPs represent an important tool to deliver nanoparticles to these organelles whose function is altered in the variety of disease states.

An attractive feature of nanoparticles is their ability to accommodate a large number of targeting ligands at their surface enabling multivalent binding of nanoparticle to their target (Gao *et al.* 2004). It can facilitate binding and enhance internalization, in particular when specific receptors are present on the cell surface at very low concentration.

### 14.2.5 Clearance of nanoparticles from the human body

Unless biodegradable into biocompatible components, nanoparticles have to be cleared from the human body within a reasonable time. The major routes of excretion for intact nanoparticles are through liver, into bile and feces, and through the kidneys into urine. Excretion of nanoparticles into bile is an extremely slow and inefficient process. The critical factor for renal excretion is most likely hydrodynamic diameter. Using intravenously administered QDs in rodents as model system, Soo Choi and colleagues have precisely defined the requirements for renal filtration and urinary excretion of spherical inorganic, metal-containing nanoparticles (Soo Choi *et al.* 2007). The hydrodynamic diameter can increase by more than 15 nm in the circulation and this change occurs due to the serum protein adsorption onto the charged nanoparticle surface. Only a final hydrodynamic diameter less than 5.5 nm resulted in rapid and efficient urinary excretion of QDs. Thus, nanoparticles less than 5.5 nm but charged are not efficiently eliminated from the body, which can lead to their accumulation and consequential toxicity. Several studies were recently completed showing

different destinies of nanoparticles in living animals (Akerman *et al.* 2002; Fischer *et al.* 2006). Results from these studies suggest that metal nanoparticles can be retained in the body and could therefore interfere with medical tests such as X-ray imaging. Therefore, similar kinetic studies must be undertaken in primates and eventually in humans prior to their use in clinics.

#### 14.2.6 Toxicity

Recently, a significant number of investigations have focused on the potential toxic effects of nanoparticles (Hoet *et al.* 2004; Oberdorster *et al.* 2005; Nel *et al.* 2006). The novel and unique properties of nanoparticles, which have been enthusiastically explored for their advancement in therapeutics and diagnostics, could also be the source of undesirable effects on biological systems (Stone and Donaldson 2006). Each type of nanoparticle most likely specifically interacts with cellular systems and the molecular bases of these interactions remain to be elucidated. It is of great importance to identify the key factors that can be used to predict nanoparticle toxicity and we should use our knowledge of the toxicity of environmental pollutants in this size range to complement our understanding of engineered nanoparticles.

The majority of polymeric materials used to form polymeric micelles and nanoparticles is synthesized under relatively harsh conditions, at elevated temperatures and in the presence of organic solvents and additives. Also, in order to slow down the degradation process various additives, e.g. stabilizers or antioxidants, are frequently used in polymer production. One must therefore consider if these additives, products of polymer degradation or depolymerization can induce immunological and pharmacological response upon administration (Dobrovolskaia and McNeil 2007).

Unfortunately, because of the lack of systematic studies as well as often unavailable information on the additives and contaminants present in polymer samples there is relatively little information on the cytotoxic effect of polymeric materials. *In-vitro* cytotoxicity for micelles of various polymers was assessed by MTT tests (Le Garrec *et al.* 2004a, 2005). The experiments performed against murine C26, EMT-6 and human OVCAR-3 cell lines show that the inhibitory concentration of polymer leading to 50% cell-growth inhibition or death ( $IC_{50}$ ) is an order of magnitude higher ( $IC_{50} = 1.0\text{--}4.4\text{ g/L}$ , depending on cell line) than that of polyethoxylated castor oil Cremophor<sup>®</sup> EL ( $IC_{50} = 0.7\text{--}0.9\text{ g/L}$ ), often used in various pharmaceutical formulations, favoring the use of polymeric excipients. Acute toxicity of polymeric micelles in mice after intravenous injection has been minimal. Even at such a high dose as 16 g/kg, polymer is well tolerated and exhibits no apparent side-effects (Le Garrec *et al.* 2005).

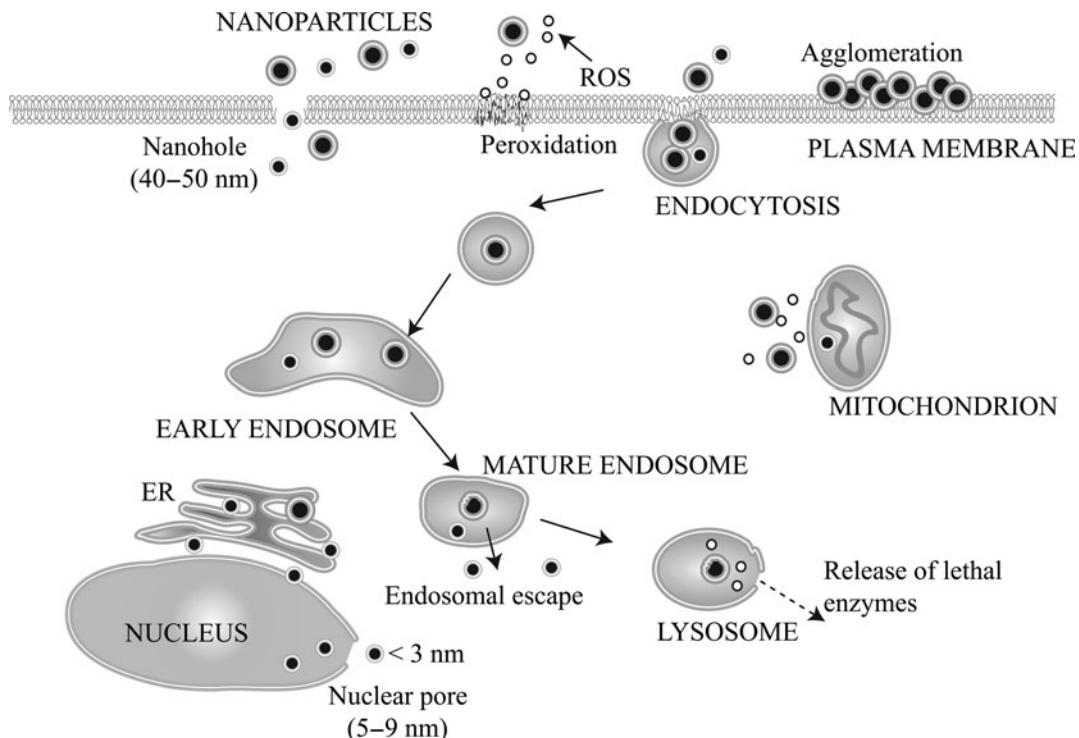
While completely benign in certain paradigms (Jaiswal *et al.* 2003; Gao *et al.* 2004), nanoparticles including quantum dots, fullerenes and nanotubes have induced negative effects in others (Oberdorster 2004). The factors that play important roles include nanoparticle size, chemical composition, surface structure, shape, solubility and aggregation. Cell death by nanoparticles is a relatively unexplored area of research. Further studies are required to

understand the mechanisms involved in nanoparticle cell death, and to discover the ways to prevent it (Hild *et al.* 2008; Lewinski *et al.* 2008). We should also keep in mind that nanoparticles could have an impact on cellular function even if they are non-toxic (Choi *et al.* 2008). Studying all these effects will be important to the safe preparation, handling and utilization of nanoparticles in science and beyond.

### 14.3 Nanoparticle–cell interactions

A crucial factor determining the usefulness of nanoparticles in medicine is an indepth understanding of their interactions with cells and their constitutive components (Fig. 14.2).

The first contact between nanoparticles and cells occurs at the plasma membrane. Depending on their size, composition, surface charge, and hydrophilicity or hydrophobicity nanoparticles were shown to induce certain alterations in



**Fig. 14.2** Nanoparticle–cell interactions. Interactions of nanoparticles with plasma membrane include: (i) the invagination and budding of the pieces of membrane leading to the formation of endocytic vesicle and nanoparticle internalization; a process known as endocytosis; (ii) formation of transient nanoholes; (iii) ROS-induced membrane peroxidation; (iv) agglomeration of nanoparticles at plasma membrane. Once in endocytic vesicles, nanoparticles can either stay stable for a certain period of time, or escape from endocytic vesicles, or be degraded in lysosomes. The escape is dependent on the nanoparticle surface properties such as charge, and it happens due to intravesicular change of pH; “proton sponge effect”. Nanoparticle-induced ROS can damage the membrane of vesicles leading to the escape of nanoparticles but also to the release of lethal enzymes into the cytosol. Nanoparticles can be localized in the vicinity of mitochondria or even enter mitochondria. Nanoparticle-induced ROS formation can lead to further damage of various membranes in particular those of mitochondria. Nanoparticles were found to colocalize with endoplasmatic reticulum (ER). If smaller than 3 nm, some nanoparticles are able to enter nuclei by the passive diffusion through the nuclear pore.

the integrity of the plasma membrane (Unfried *et al.* 2007). Experimental data suggests that nanoparticles with positively charged groups at their surface, such as cationic polymer polyethylenimine (PEI) or polyamidoamine (PAMAM) dendrimers, can induce physical disruption of the plasma membrane. This disruption is characterized by formation of transient nanoscale holes and results in enhanced membrane permeability (Hong *et al.* 2004, 2006). The proposed hole-formation mechanisms were suggested based on the data obtained by atomic force microscopy on model membranes (Mecke *et al.* 2004) and by diffusion of cytosolic enzymes outside the cell upon the interaction with positively charged nanoparticles. The ability to form membrane holes was more pronounced for the polymers or nanoparticles with higher density of charged groups. However, if the damage to the cell membrane is too severe to "patch up", it will probably lead to cell lysis and cell death (Goodman *et al.* 2004). Nanoparticles that were able to induce formation of nanoholes were also able to enter the cell, leading to the conclusion that they could internalize through nanoholes. In addition to the hole formation, nanoparticles were found to precipitate at the cell surface. Agglomeration of nanoparticles at the cell surface can result in the impairment of cell function by hindering normal exchange and communication of the cell with its environment (Kirchner *et al.* 2005).

Nanoparticles are candidates for cellular uptake by endocytosis (Jones *et al.* 2003). Endocytosis is a process in which macromolecules or particles are carried into a cell in membrane-bound vesicles derived by invagination and pinching off of the pieces of the plasma membrane (Conner and Schmid 2003). This process occurs by multiple mechanisms and is usually categorized in two broad categories; phagocytosis and pinocytosis.

Phagocytosis is the process by which large particles are taken up and is conducted primarily by specialized cells, such as macrophages, monocytes and neutrophils. Pinocytosis occurs in all the cells by at least four different mechanisms: macropinocytosis, clathrin-dependent endocytosis, caveolin-dependent endocytosis and clathrin- and caveolin-independent endocytosis (Conner and Schmid 2003; Mayor and Pagano 2007). Each pathway varies in the cargo it transports, the regulation of the entry of the cargo, the mechanism of vesicle formation, the protein machinery that facilitates the endocytic process and intracellular destination of delivered cargo. Postendocytic itineraries of cargo are very complex; the initial site of delivery is sorting endosomes, from which cargo can be transported to an endocytic recycling compartment and recycled out. Sorting endosomes mature in late endosomes, from which cargo can be delivered to Golgi apparatus and later also recycled to plasma membrane. Late endosomes mature to lysosomes, which represent terminal degradative compartments of the endocytic pathway. Moreover, on this multifaceted endocytic pathway cargo can escape from endosomes that then results in cytoplasmic delivery. In polarized epithelial cells additional sorting mechanisms operate that lead to transcytosis across the cell (Maxfield and McGraw 2004).

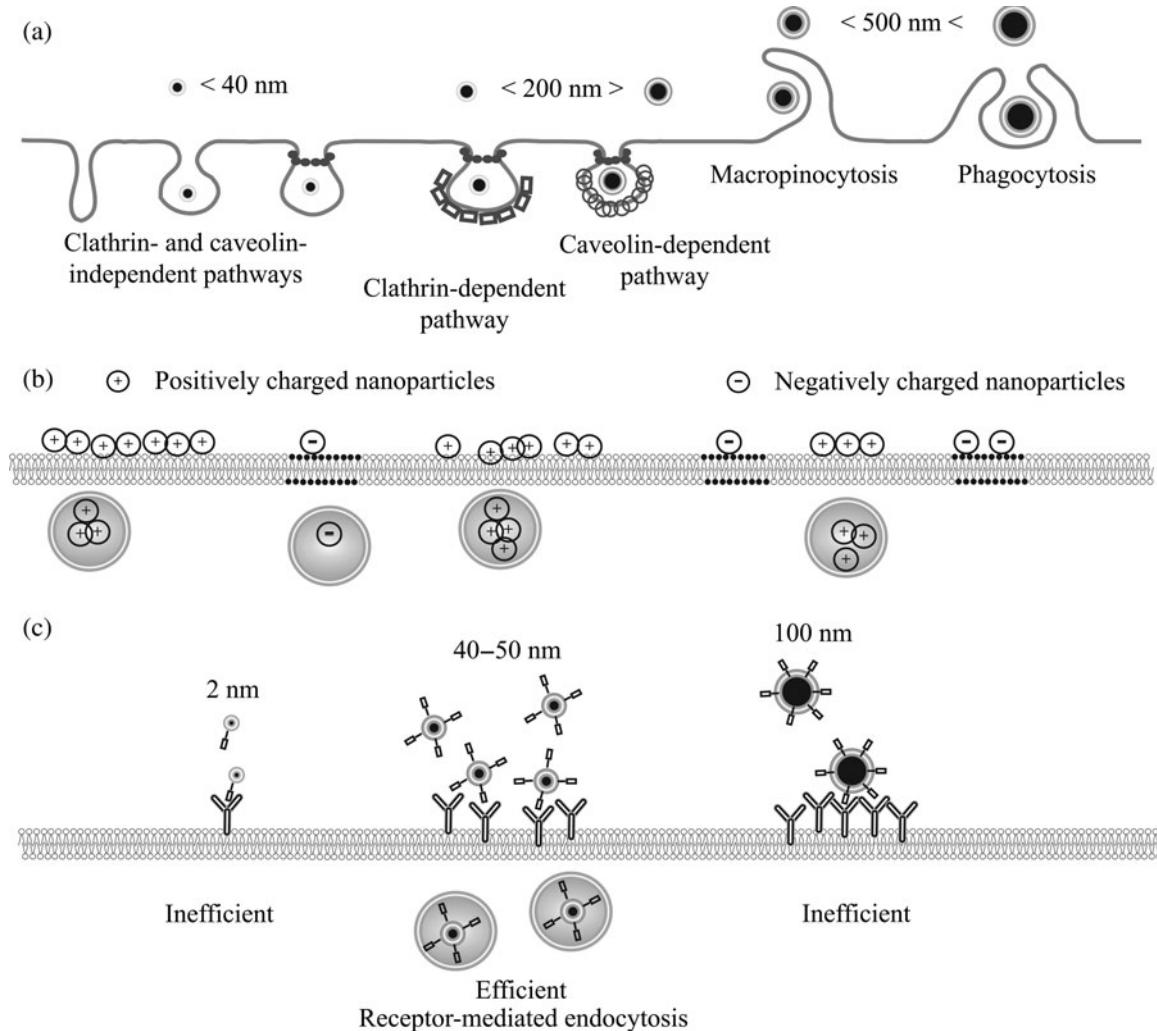
The nature of the nanoparticle undoubtedly determines the endocytic pathway through which nanoparticles gain entry into the cell. The nanoparticle size and surface charge as well as the presence of specific ligands attached at the nanoparticle surface are crucial players that dictate nanoparticle internalization

(Fig. 14.3). Nanoparticle size was found to strongly affect the efficiency of cellular uptake, the mechanism of endocytosis and the subsequent sorting along the endocytic pathway (Rejman *et al.* 2004). In this study smaller nanoparticles were found to be taken up more efficiently and rapidly but mostly ended up in lysosomes (50, 100 and 200 nm). The main operating mechanism of endocytosis up to the size of 200 nm was clathrin dependent, and the increase in size caused the shift to a caveolin-dependent one (Fig. 14.3(a)). However, nanoparticles of 500 nm never reached the lysosomes. Lai *et al.* found significant differences in cellular internalization and the fate of polymeric nanoparticles of 24 and 43 nm. Smaller nanoparticles were internalized in clathrin- and caveolin-independent manner and were found to access a privileged non-degradative intracellular pathway that leads to perinuclear accumulation (Lai *et al.* 2008). This prolonged retention of nanoparticles near the nucleus may be exploited for improved intracellular delivery of therapeutics to a nucleus.

Electrostatic interactions between nanoparticles and plasma membrane or the membranes of endocytic compartments can play a decisive role for the nanoparticle's cellular fate (Fig. 14.3(b)). The interaction of charged nanoparticles with the membranes is thought to be non-specific and it was described as two-step internalization (Wilhelm *et al.* 2002). The first step is the adsorption of nanoparticles onto the cell surface followed by a subsequent step of internalization. This process is termed the adsorptive endocytosis (Wilhelm *et al.* 2003). Though considered non-specific interaction, charged nanoparticles probably interact with certain membrane nanodomains that can behave in a manner similar to receptors. Plasma membranes possess large negatively charged domains but also limited cationic sites allowing interaction of both positively and negatively charged nanoparticles. Accordingly, the surface charge was found to affect the mechanism and final destination in the cell (Harush-Frenkel *et al.* 2007, 2008). Significant differences were also found in the internalization efficiency of oppositely charged nanoparticles with usually better results for positively charged ones (Miller *et al.* 1998). Moreover, positively charged nanoparticles were found to avoid the degradative lysosomal pathway.

Specific interactions of nanoparticles with the plasma membrane are mediated by binding of ligand molecules on the nanoparticle surface to their receptor molecules on the plasma membrane (Fig. 14.3(c)). Specific ligands then carry nanoparticles in the cell by receptor-mediated endocytosis, and, moreover, can grant the internalization to a specific population of cells such as cancer cells. Such a ligand is for example folic acid, a high-affinity ligand to folate receptor that is efficiently internalized into the cell through the receptor-mediated endocytosis even when conjugated with a wide variety of molecules (Lu *et al.* 2004). Folic acid attached to the surface of quantum dots enables internalization of these nanoparticles in folate receptor overexpressing tumor cells, while their uptake is significantly lower in folate receptor-negative cells (Bharali *et al.* 2005b).

However, receptor-mediated endocytosis is also strongly size, charge and shape dependent. Gao *et al.* proposed a mathematical model for the receptor-mediated endocytosis without assessing the charge effects, and they deduced a



**Fig. 14.3** Internalization of nanoparticles is dependent on the size, charge and the presence of specific ligands at nanoparticle surface. (a) The mechanism of endocytosis depends in part on nanoparticle size. Nanoparticles larger than 500 nm can mainly be phagocytosed by specialized cells. This schematic represents internalization of fluorescent beads and polymeric nanoparticles (Rejman *et al.* 2004; Lai *et al.* 2008). Internalization of nanoparticles by endocytosis also depends on other properties such as core composition, charge and ligand at the nanoparticle surface. (b) Charged nanoparticles are considered to be internalized by adsorptive endocytosis. Positive nanoparticles are more efficiently internalized than negatively charged nanoparticles due to the pre-dominant existence of negatively charged domains of plasma membrane (gray) in comparison to scarce positively charged domains (black). (c) Receptor-mediated endocytosis is highly size dependent (Jiang *et al.* 2008).

threshold radius below which endocytosis would not occur (Gao *et al.* 2005). For cylindrical and spherical nanoparticles these values were 15 and 30 nm, respectively. Aoyama and coworkers experimentally came up with a similar finding that an optimal nanoparticle size for receptor-mediated endocytosis is 25 nm (Aoyama *et al.* 2003; Nakai *et al.* 2003; Osaki *et al.* 2004). Chan's group has demonstrated that engineered nanoparticles of well-defined sizes can selectively induce membrane receptor internalization to downregulate their expression (Jiang *et al.* 2008). In the size range between 2 to 100 nm,

antibody-coated gold and silver nanoparticles of 40 and 50 nm showed the greatest internalization and, moreover, the ability to alter the signalling process essential for basic cell function. The explanation for this size effect lies in the antibody density on the nanoparticle surface that increases proportionally with the increase of the nanoparticle radius. More ligands at the surface allow cross-linking of the receptors, which can make a process of endocytosis more efficient. Smaller nanoparticles lack this cross-linking ability, and also dissociate from the receptor before being engulfed by the membrane due to the low binding activity (Jiang *et al.* 2008). Even though larger nanoparticles possess significantly higher amount of ligands at the surface, the inability to compensate for the depletion of the receptors within the binding area could slow down the process of internalization and make it inefficient (Gao *et al.* 2005).

Affirmation of the significance of charge comes from another mathematical model that showed that attractive/repulsive forces play an important role in binding to the receptor (Decuzzi and Ferrari 2007). Repulsive interactions make nanoparticle endocytosis more difficult, while attractive, non-specific interactions favor endocytosis.

A cell membrane is very susceptible to peroxidation that compromises membrane integrity and can eventually lead to cell death. Recent data suggest that an important aspect of the mechanism of cytotoxicity induced by nanoparticles could be the production of reactive oxygen species (ROS) (Stone and Donaldson 2006; Xia *et al.* 2006; Unfried *et al.* 2007). The exact mechanisms for the production of these reactive species are not well understood; usually high levels of free metal ions (Green and Howman 2005) lead to the formation of ROS and oxidative damage (Choi *et al.* 2007). Some types of nanoparticles, such as QDs, due to their electronic configuration have the ability to spontaneously induce ROS. At the same time, it seems that cells themselves can be the source of ROS in contact with nanoparticles lacking this intrinsic property (Xia *et al.* 2006).

After nanoparticles enter the cell they are usually trapped in vesicles, endosomes and lysosomes. In order to reach other cellular targets the strategy is that they escape from these organelles. One way to achieve the escape is to disrupt the organelle membrane. Polymer PEI mentioned earlier has the ability of endosomal membrane disruption, which was in this case ascribed to a “proton sponge effect” (Neu *et al.* 2005). This effect is based on the presence of a large number of amine groups on PEI polymer leading to proton absorption in acid organelles, which results in an osmotic pressure buildup and membrane disruption. Polymeric or metallic nanoparticles with PEI or its derivatives at their surface were shown to be released in cytoplasm after endocytosis (Duan and Nie 2007). However, the organelle disruption can result in cytotoxicity, especially if lysosomal membrane is disrupted and degradative enzymes released in cytoplasm. By fine tuning of the surface coating chemistry the toxic effect can potentially be evaded and the desirable intracellular fate of nanoparticles accomplished. Lysosomes are also very sensitive to oxidative stress, and this property is exploited in photodynamic therapy (Reiners *et al.* 2002; Guicciardi *et al.* 2004). Lysosomal membranes can then be potentially damaged by nanoparticle-produced ROS internalized by endocytosis.

This can also lead to the release of lethal enzymes and consequent cell death.

Nanoparticles were found to interact with mitochondria. For example, confocal microscopy showed colocalization of fluorescently labelled non-functionalized block copolymer micelles with mitochondria (Savic *et al.* 2003; Stewart *et al.* 2008) while electron microscopy revealed few gold-labelled micelles inside mitochondria (Maysinger *et al.* 2007b). Attachment of mitochondrial localization signal peptide to the surface of quantum dots or specific oligonucleotide probes to TiO<sub>2</sub> nanoparticles has directed these nanoparticles specifically to mitochondria after the internalization (Hoshino *et al.* 2004; Paunesku *et al.* 2007). Mitochondria play a central role in energy metabolism; they are responsive to even small stresses in multiple ways and moreover, play a critical role in cell death. Negative interactions of nanoparticles and mitochondria can lead to the release of proteins from the intermembrane space of mitochondria that are lethal to cells once released into the cytosol. When cells are exposed to nanoparticles (e.g. quantum dots), which can induce generation of ROS, mitochondria are among the first and most sensitive organelles affected. If nanoparticles are not capable of producing ROS themselves, dysfunctional mitochondria in stressed cells could be the major source of ROS produced in cells (Li *et al.* 2003). In quantum-dot injured cells, reduction of mitochondrial membrane potential and swelling of mitochondria can be detected (Lovric *et al.* 2005b; Choi *et al.* 2007).

Interactions of nanoparticles with endoplasmic reticulum (ER) and Golgi apparatus are less explored. Savic and colleagues suggested colocalization of polymeric micelles with these organelles too (Savic *et al.* 2003) and the nature of these interactions is being explored. Several studies suggest that polymeric nanoparticles with neutral surfaces and in low micromolar concentrations do not cause irreversible damage to ER and Golgi, whereas metallic nanoparticles that leak Cd<sup>2+</sup> ions, can cause marked disorganization of these organelles.

Non-functionalized nanocrystals can exploit a cell's active transport machinery and be delivered into the cells and even the nucleus (Nabiev *et al.* 2007). Quantum dots linked to nuclear localization signal peptide (Hoshino *et al.* 2004) but intriguingly also unfunctionalized quantum dots were found in cell nuclei (Lovric *et al.* 2005a; Ryman-Rasmussen *et al.* 2007a). The nuclear access for unfunctionalized quantum dots was size dependent; only quantum dots of size around 2 nm were allowed to enter. Nabiev *et al.* made a step forward and discovered that quantum dots are actively transported to the perinuclear area and enter the nucleus where they display confined areas of compact concentration highlighting nucleoli (Nabiev *et al.* 2007). Again, this was a characteristic of quantum dots of 2 nm, while bigger ones did not show nuclear localization. It was suggested that these nanoparticles enter a nucleus through the nuclear pore, and indeed, the inhibition of passive diffusion through a nuclear pore by thapsigargin has blocked the nuclear internalization of quantum dots. The most remarkable finding is that the smallest, negatively charged QDs specifically target histones in cell nuclei by a multistep process. The nuclear entry seems to be mediated via nuclear-pore complexes. Moreover, quantum dots used in this study demonstrated a strong affinity for purified histone protein. More recent studies (Choi *et al.* 2008) provide further

evidence that minute amounts of small (<5 nm diameter) QDs can interfere with histones and induce epigenetic changes. These results from other small QDs and nanomaterials that could unintentionally reach the nucleus could have far-reaching undesirable consequences related to cell functions. Similar studies are warranted to uncover possible histone modifications with or without subsequent gene activations. This is an exciting and emerging new branch in nanotoxicology that we propose to call “nanoepigenetics”.

An important question is how nanoparticles move across the cytoplasm. The movement could be just passive diffusion or active transport along microtubules or actin filaments. The movement will most likely depend on their nature, including size and charge, as well as nanoparticle fate, i.e. their sequestration in vesicles or free association with cytoplasmic components. Quantum dots were shown to be actively transported in vesicles along the microtubules from the cell surface to the perinuclear area (Nabiev *et al.* 2007).

## 14.4 Nanoparticles as imaging tools in animals and humans

Many non-functionalized and some functionalized fluorescent quantum dots were thoroughly characterized in cell lines, but very few studies were done in live animals. Table 14.4 summarizes most of the recent studies illustrating the first attempts to use quantum dots for imaging whole animals and living cells.

**Table 14.4** Quantum dots for *in-vivo* imaging

Particle core	Surface modification	Concentration and size	Findings	Reference
CdSe/ZnS	Peptides, PEG	100–200 µg in 0.1–0.2 mL	QDs can be targeted to specific tissues. PEG coating can prevent non-selective accumulation of QDs in reticuloendothelial tissues.	(Akerman <i>et al.</i> 2002)
CdSe/ZnS	Micelle	20 nM – 1 µM 18–28 nm	QDs were visualized through the skin of live mice and in capillaries hundreds of micrometers beneath the skin.	(Larson <i>et al.</i> 2003)
CdTe/CdSeTOPO	Oligomeric—polydentate phosphines	10–400 pmol 15–20 nm	QDs can be stabilized in serum. Near-infrared emitting QDs can be used for sentinel lymph-node mapping allowing for real time complete image guidance during surgery of lymph node 1 cm deep.	(Kim <i>et al.</i> 2004)
CdSe/ZnS	Amphiphilic poly(acrylic acid) polymer (m = methoxy, c = carboxy) PEG: mPEG-750 QDs, mPEG-5000 QDs, cPEG-3400 QDs	50–500 pmol in 50–200 µL	QDs accumulate primarily in the endosomes of the liver, but also in the spleen and bone. Fluorescence can be detected for up to 4 months. Fluorescence after 1 month is the same as after 24 h. mPEG 5000 QDs remain in circulation longer and as such accumulate less.	(Ballou <i>et al.</i> 2004)

(cont.)

**Table 14.4 (Continued)**

Particle core	Surface modification	Concentration and size	Findings	Reference
CdSe/ZnS TOPO	COOH, PEG, PEG + Prostate-specific membrane antigen (PSMA)	6 nmol 20–30 nm	QDs accumulate in tumors both due to enhanced permeability and retention of tumors, as well as due to antibody–antigen specific binding. These findings hold true upon subcutaneous QD injection as well as systemic administration.	(Gao <i>et al.</i> 2004)
CdSe/ZnS TOPO	DHLA	10 pmol	Tumor cells containing QDs were injected into mice and tumor cell death was expected but not observed. QDs within cells did not impede tumor formation.	(Voura <i>et al.</i> 2004)
CdS:Mn/ZnS	TAT peptide	0.75 mL of 10 mg/ml 3.1 nm	Multifunctional QDs that possess fluorescence, radio opacity and paramagnetic properties can be synthesized. These QDs can be used to label brain tissue.	(Santra <i>et al.</i> 2005)
CdMgTe/Hg	BSA	100 µL of 1:1.5 QD, heparin lock flush 100 U/mL 5 nm	QDs can be used as a contrast agent with NIR fluorescence. QD-photobleaching and degradation was not observed even after continuous excitation for 1 h.	(Morgan <i>et al.</i> 2005)
Cds/ZnS TOPO	PEG-phospholipid micelles, PEG-PE-TAT	4 µM	QDs can be utilized with multiphoton microscopy techniques.	(Stroh <i>et al.</i> 2005)
CdSe/ZnS	Streptavidin	100 nM, 1.7 nL or 1 µM	QDs were used for lineage tracing of embryonic zebra fish. Non-stationary clusters of QDs in the circulation can outline newly formed blood vessels by fluorescing in the endothelial wall.	(Rieger <i>et al.</i> 2005)
InAs/ZnSe	DHLA, DHLA-PEG	5.3 nm	Particularly small QDs were synthesized and found to circulate <i>in vivo</i> for several minutes prior to penetrating into the interstitial fluid.	(Zimmer <i>et al.</i> 2006)
CdSe/ZnS	mPEG	2 µM in 50 mM borate buffer ~35 nm	QDs can diffuse within the rat neocortical extracellular space. Therefore, QDs were used to determine the upper size limit of particles that pass the BBB.	(Thorne and Nicholson 2006)
CdSe/ZnS	Arginine-Glycine-Aspartic acid	200 pmol 15–20 nm	QDs can be specifically targeted to a particular tumor marker <i>in vitro</i> , <i>ex vivo</i> and <i>in vivo</i> .	(Cai <i>et al.</i> 2006)
CdTe/CdSe	Oligomeric Phosphines and HSA800, or ICG:HSA	100 pmol – 1 nmol (stock of 1 µM in PBS) 15 nM	Type II NIR QDs can be used as part of organic/inorganic tracers for NIR fluorescence-guided mapping of sentinel lymph-node mapping.	(Tanaka <i>et al.</i> 2006)
CdTe/ZnS	Polymer with carboxy groups	8–16 pmol 6–12 nm	Two different QDs are used to simultaneously image two distinct lymphatic drainage and their trafficking to a lymph node.	(Hama <i>et al.</i> 2007)
CdSe/ZnS	Captopril	7 mg/kg	QDs can be conjugated to an ACE inhibitor (antihypertensive drug), to study and analyze the pharmacokinetic and pharmacodynamic actions of the drug.	(Manabe <i>et al.</i> 2006)
CdSe/ZnS	Anti-HER 2 antibody	2 µM total 100 µL	QDs conjugated to an antibody were monitored to better understand the trafficking of a tumor antibody conjugated to a nanoparticle. Six modes of antibody delivery were identified.	(Tada <i>et al.</i> 2007)

**Table 14.4 (Continued)**

Particle core	Surface modification	Concentration and size	Findings	Reference
CdSe/ZnS	AminoPEG	3–17 nmol	QDs when delivered in nanomolar doses are phagocytized by macrophages and microglia that infiltrate gliomas.	(Jackson <i>et al.</i> 2007)
CdSe/CdS	PEG	37 nm	Intradermally injected QDs are found to move through the lymphatic ducts to lymph nodes, eventually depositing in the liver, kidney, spleen and other lymph nodes. Cadmium and selenium were also found in these regions.	(Gopee <i>et al.</i> 2007)
CdSe & CdTe	carboxyPEG	10 $\mu$ L of 12–24 pmol 15–19 nm	Five different QDs were used to simultaneously image five different lymphatic flows and their trafficking to different lymph nodes.	(Kobayashi <i>et al.</i> 2007)
CdSe/ZnS	carboxyPEG	8.2 nM	Human mesenchymal stem cells were loaded with QDs permitting the long-term imaging of cells.	(Rosen <i>et al.</i> 2007)
CdSe/ZnS	Alpha fetoprotein antibody	0.4 nmol	QDs conjugated to anti-AFP enable active tumor targeting and spectroscopic hepatoma imaging.	(Yu <i>et al.</i> 2007)
CdSe/ZnS	Phospholipid coat Carboxy conjugated	8 fL, 0.005 pg	QD-labelled neural stem and progenitor cells (NSPCs) continue to develop normally. QDs can be used to effectively study <i>in-vivo</i> fate mapping, cellular migration and NSPC differentiation.	(Slotkin <i>et al.</i> 2007)
CdSe/ZnS & CdTe	PEGylated and non-PEGylated	30 nmol 2.67–8 $\mu$ l	Assessment of QD-induced activation of astrocytes in the brain of transgenic mice with the luciferase gene driven by the murine astrocyte promoter.	(Maysinger <i>et al.</i> 2007a)
CdSe/ZnS	In solid lipid nanoparticles (SLN)	4 nm in SLN – 90 nm	QDs encapsulated within SLN remain stable and fluorescent.	(Liu <i>et al.</i> 2007a)
CdSe/ZnS	AminoPEG-mAb (anti-PECAM, anti-VECAM, anti-ICAM, QD-IgG <sub>1</sub> )	500 nM (0.2 mL)	mAb-QDs targeted to the cell adhesion molecules (CAMs) PECAM-1, ICAM-1, and VCAM-1 on the retinal endothelium were imaged <i>in vivo</i> using fluorescence angiography in a rat model of diabetes.	(Jayagopal <i>et al.</i> 2007)
CdSe/ZnS	PEG	10 nM	Successful <i>in-vivo</i> multiplex imaging and trafficking of mouse embryonic stem cells labelled with QDs.	(Lin <i>et al.</i> 2007)
CdSe/ZnS	Amino-PEG	5 pmol/g	<i>In-vivo</i> fluorescence imaging with bioluminescent tumor monitoring, provided valuable information for lesion localization.	(Inoue <i>et al.</i> 2007)
CdSe/ZnS	DOTA–QD–RGD with <sup>125</sup> I-echistatin as the integrin $\alpha_v\beta_3$ -specific radioligand	20 pmol 20 nm	Quantitative evaluation of the tumor-targeting efficacy of a dual-function QD-based probe with PET and NIRF imaging. Non-invasive PET with radiolabelled QD-conjugates permit reliable measures of <i>in-vivo</i> probe distribution.	(Cai <i>et al.</i> 2007)
CdSe/(ZnCdS)	Cysteine	1 mL or 3 $\mu$ L 3.6 nm	The small size of QD-Cys nanoparticles permits renal clearance.	(Liu <i>et al.</i> 2007a)
CdHgTe	Thiol capped	2–17.5 $\mu$ g/g 7 nm	Low doses of QDs appeared not to be toxic even after 3 months, whereas the high doses were lethal. QDs were found to accumulate in the liver. QDs were excreted via intestine and not via kidney. QDs can be a good angiographic contrast agent in the NIR range.	(Chen <i>et al.</i> 2008)

(cont.)

**Table 14.4** (Continued)

Particle core	Surface modification	Concentration and size	Findings	Reference
CdSeTe/ZnS	EGF	10 pmol	Pharmacokinetics of EGF-QD studied show the three phases of interactions of QDs with the tumor. A measurable contrast enhancement of tumors is visible 4 h postsystemic administration of EGF-QD probes.	(Diagarad Jane <i>et al.</i> 2008)

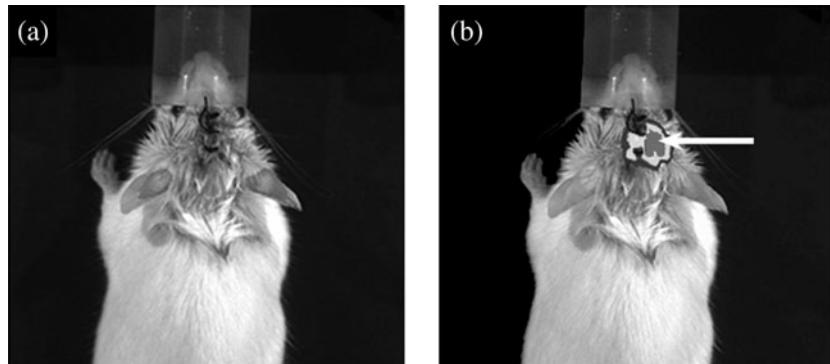
Imaging of living cells was rather successful and provided a wealth of knowledge providing seminal information of selected quantum dots. Such studies partly answered some of the key biological questions related to quantum-dot uptake, subcellular distribution and functional consequences. The studies attempted to answer the following questions: Which type of quantum dot can be taken up? Is it the size, charge or some other surface properties that determine the extent of the uptake? Are quantum dots taken up, retained or eliminated from the cells or remain there and get degraded? Despite the convincing data these results are limited to particular cell type and particular type of quantum dots and many more studies are required to gather information for other types of cells and nanoparticles.

Studies related to quantum dots in whole animals are relatively sparse. One of the reasons for this is that imaging of the whole animal using fluorescent nanoparticles in general, remains difficult, because of strong autofluorescence of tissues due to endogenous chromophores (e.g. collagens, porphyrins and flavins) and the stability of the core and corona ligands. In most studies using CdSe/ZnS the issue of core stability within the relatively short time periods (hours) did not seem to pose a problem. However, non-covalently attached surface molecules might have been partly lost *en route* towards to desired destination sites. It is also unknown to what degree quantum dots interact with extracellular and cellular proteins and get quenched. A newer class of quantum dots with superior photophysical properties, which can at least in part overcome these limitations by emitting in near-infrared regions, were prepared and tested in rodents (le Masne De Chermont *et al.* 2007). Using NIR-emitting nanocrystals one can avoid the interference noise from the animal autofluorescence. An example of such quantum dots is QD705 (i.e. a quantum dot that is characterized by a maximum of luminescence intensity at around 705 nm), which were intravenously injected in mice. Results from these studies show that they can be found after 24 h in liver, spleen and to a lesser degree in other organs. Analyses of urine and feces did not reveal significant quantities in these specimens, suggesting that most of the QD705 stayed in the body. Concordant with such a notion are results obtained by Fischer *et al.* (2006). It seems that surface-modified quantum dots with poly(ethylene glycol) are biocompatible both with living cells in cultures (Ryman-Rasmussen *et al.* 2007a) and whole animals. Reports on distribution and their pharmacokinetic properties show that cadmium selenide quantum dots are sequestered in several organs after

i.v. administration (Fischer *et al.* 2006), they accumulate in lymph nodes (Frangioni *et al.* 2007) and solid tumors (Ballou *et al.* 2007). A recent study showed how size and charge of QDs affect their renal clearance (Soo Choi *et al.* 2007). Selected types of QDs were injected intravenously in rodents and the key results showed that zwitterionic or neutral organic coatings prevented adsorption of serum proteins and did not markedly impair renal excretion. Moreover, these studies also show that final hydrodynamic diameter <5.5 nm provides the appropriate size for urinary excretion and elimination of QDs from the body.

Studies by Inoue *et al.* showed that quantum dots allowed for long-term and repeated observation of the reticulendothelial system in mice, despite a gradual decline of signal intensity (Inoue *et al.* 2007). Combined imaging of quantum dots by fluorescence and injected bioluminescent luciferase expressing tumor cells (luminescence) provided a way of concomitant detection of tumor cells and quantum dots nanoparticles. The concept of combining bioluminescence and fluorescence in living cells was first illustrated by Rao's group, whereas self-illuminating quantum dots were tested in living cells and whole animals (So *et al.* 2006a,b). The results from these studies provided evidence for the usefulness of quantum dots covalently bound with luciferase. When the conjugates were exposed to luciferase substrate coelentrazine, the energy released by substrate catabolism was transferred to nanoparticles by bioluminescence energy transfer, resulting in significant quantum-dot luminescence. The protocol for making these self-illuminating quantum dots provided by So *et al.* is widely applicable to other conjugation products that can be utilized for different biosensing purposes. Another example of combining bioluminescence and fluorescence came from studies in transgenic mice expressing luciferase under control of astrocyte promoter GFAP (glial fibrillary acidic protein) (Maysinger *et al.* 2007a). These studies were sparked by the common observation that reactive oxygen species (ROS) are formed in cells exposed to certain types of quantum dots, e.g. cadmium telluride (Lovric *et al.* 2005b; Choi *et al.* 2007). The consequences of ROS formation are different depending on cell types, concentration and duration of exposure to the quantum dots, as well as several physical and chemical properties of nanoparticles (Seleverstov *et al.* 2006; Xia *et al.* 2006; Zanello *et al.* 2006; Cambi *et al.* 2007; Ryman-Rasmussen *et al.* 2007b).

Biocompatibility of quantum dots in live animals, especially after repeated imaging sessions, has been shown only in a few examples. Studies by Maysinger *et al.* focus on astrocyte response to different types of quantum dots. Astrocytes are the principle macroglial cell type in the brain and their activation is one of the key components of the cellular responses to stress and brain injuries. The passage from the quiescent to reactive astrocytes is associated with strong upregulation of the intermediate filament, glial fibrillary acidic protein (GFAP) (Ridet *et al.* 1997; Pekny and Nilsson 2005). Current methods of astrocyte and microglia detection are mainly based on immunocytochemistry. However, in recent years, imaging strategies employing reporter molecules have been developed to study biological processes as they occur in living animals. For instance, a transgenic mouse that carries the luciferase gene under the transcriptional control of murine GFAP promoter and upregulation



**Fig. 14.4** *In-vivo* fluorescence and bioluminescence imaging of the brain. Representative pseudocolor fluorescence images of QD 705 stereotactically injected in the right parietal cortex of GFAP-luc reporter mouse (the coordinates for injection were: 2 mm posterior to bregma, 2 mm lateral to the midline and 1.5 mm below the surface of the skull). The images were acquired using IVIS system. Note: (a)—control without injected nanoparticles. (b) Astrocyte activation (arrow) and fluorescence of QD705 are detectable within the same brain region.

of GFAP or luciferase expression in response to quantum dots was analyzed non-invasively in living animals using biophotonic imaging. When injected subcutaneously QD705 do not reach the brain but if injected in the parenchyma they can be seen for at least one week even when as little as  $2 \mu\text{L}$  of  $8 \mu\text{M}$  solution is injected (Fig. 14.4). In contrast, bioluminescence is transient, indicating that glial cells are only transiently activated as the signal intensity gradually declines after 24 h.

In summary, these studies suggest that transgenic animals expressing luciferase protein under different cell specific and tissue specific promoters could be useful model systems for dynamic, *in-vivo* assessment of nanotoxicity.

## 14.5 Conclusions

The most rewarding applications of nanomaterials, including fluorescent, luminescent, super-paramagnetic and other types of nanoparticles (both functionalized and non-functionalized) will depend on our judicious selection of these nanomaterials. The cumulative effects of small doses of nanoparticles in the human body cannot be ignored due to their possible epigenetic and genetic effects. Application of quantum dots for diagnostic and analytical purposes of biological fluids and easily accessible tissues seems to be particularly promising for applications of several types of functionalized nanoparticles with biosensing abilities. Several examples of the current uses, advantages and limitations of nanomaterials discussed in this chapter, suggest that nanoparticles are promising nanotools in medicine and that our increasing understanding of their possible adverse effects in the biological systems will help us to overcome hurdles and use them safely.

## Acknowledgments

This work was supported by NSERC, CIHR, ADUSA, JDRF, and the Ministry of Science, Education and Sports, Republic of Croatia (006-0061117-1244).

## References

- Akerman, M.E., Chan, W.C., Laakkonen, P., Bhatia, S.N., Ruoslahti, E. *Proc. Natl. Acad. Sci. USA* **99**, 12617 (2002).
- Aldana, J., Lavelle, N., Wang, Y., Peng, X. *J. Am. Chem. Soc.* **127**, 2496 (2005).
- Aldana, J., Wang, Y.A., Peng, X. *J. Am. Chem. Soc.* **123**, 8844 (2001).
- Alivisatos, A.P. *Science* **271**, 933 (1996).
- Aoyama, Y., Kanamori, T., Nakai, T., Sasaki, T., Horiuchi, S., Sando, S., Niidome, T. *J. Am. Chem. Soc.* **125**, 3455 (2003).
- Arap, W., Kolonin, M.G., Trepel, M., Lahdenranta, J., Cardo-Vila, M., Giordano, R.J., Mintz, P.J., Ardel, P.U., Yao, V.J., Vidal, C.I., Chen, L., Flamm, A., Valtanen, H., Weavind, L.M., Hicks, M.E., Pollock, R.E., Botz, G.H., Bucana, C.D., Koivunen, E., Cahill, D., Troncoso, P., Baggerly, K.A., Pentz, R.D., Do, K.A., Logothetis, C.J., Pasqualini, R. *Nature Med.* **8**, 121 (2002).
- Bagalkot, V., Zhang, L., Levy-Nissenbaum, E., Jon, S., Kantoff, P.W., Langer, R., Farokhzad, O.C. *Nano Lett.* **7**, 3065 (2007).
- Bakalova, R., Ohba, H., Zhelev, Z., Ishikawa, M., Baba, Y. *Nature Biotechnol.* **22**, 1360 (2004).
- Ballou, B., Ernst, L.A., Andreko, S., Harper, T., Fitzpatrick, J.A., Waggoner, A.S., Bruchez, M.P. *Bioconjug. Chem.* **18**, 389 (2007).
- Ballou, B., Lagerholm, B.C., Ernst, L.A., Bruchez, M.P., Waggoner, A.S. *Bioconjug. Chem.* **15**, 79 (2004).
- Bartlett, D.W., Su, H., Hildebrandt, I.J., Weber, W.A., Davis, M.E. *Proc. Natl. Acad. Sci. USA* **104**, 15549 (2007).
- Bertin, P.A., Gibbs, J.M., Shen, C.K., Thaxton, C.S., Russin, W.A., Mirkin, C.A., Nguyen, S.T. *J. Am. Chem. Soc.* **128**, 4168 (2006).
- Bharali, D.J., Klejbor, I., Stachowiak, E.K., Dutta, P., Roy, I., Kaur, N., Bergey, E.J., Prasad, P.N., Stachowiak, M.K. *Proc. Natl. Acad. Sci. USA* **102**, 11539 (2005a).
- Bharali, D.J., Lucey, D.W., Jayakumar, H., Pudavar, H.E., Prasad, P.N. *J. Am. Chem. Soc.* **127**, 11364 (2005b).
- Bianco, A., Kostarelos, K., Prato, M. *Curr. Opin. Chem. Biol.* **9**, 674 (2005).
- Bourne, N., Stanberry, L.R., Kern, E.R., Holan, G., Matthews, B., Bernstein, D.I. *Antimicrob. Agents. Chemother.* **44**, 2471 (2000).
- Cai, W., Chen, K., Li, Z.B., Gambhir, S.S., Chen, X. *J. Nucl. Med.* **48**, 1862 (2007).
- Cai, W., Shin, D.W., Chen, K., Gheysens, O., Cao, Q., Wang, S.X., Gambhir, S.S., Chen, X. *Nano Lett.* **6**, 669 (2006).
- Cambi, A., Lidke, D.S., Arndt-Jovin, D.J., Figgdr, C.G., Jovin, T.M. *Nano Lett.* **7**, 970 (2007).
- Cedervall, T., Lynch, I., Lindman, S., Berggard, T., Thulin, E., Nilsson, H., Dawson, K.A., Linse, S. *Proc. Natl. Acad. Sci. USA* **104**, 2050 (2007).
- Chen, A.A., Derfus, A.M., Khetani, S.R., Bhatia, S.N. *Nucleic. Acids. Res.* **33**, e190 (2005).
- Chen, C.C., Lin, Y.P., Wang, C.W., Tzeng, H.C., Wu, C.H., Chen, Y.C., Chen, C.P., Chen, L.C., Wu, Y.C. *J. Am. Chem. Soc.* **128**, 3709 (2006).
- Chen, H., Wang, Y., Xu, J., Ji, J., Zhang, J., Hu, Y., Gu, Y. *J. Fluoresc.* **18**, 801 (2008).

- Chen, R.J., Bangsaruntip, S., Drouvalakis, K.A., Kam, N.W., Shim, M., Li, Y., Kim, W., Utz, P.J., Dai, H. *Proc. Natl. Acad. Sci. USA* **100**, 4984 (2003).
- Chithrani, B.D., Ghazani, A.A., Chan, W.C. *Nano Lett.* **6**, 662 (2006).
- Choi, A.O., Brown, S.E., Szyf, M., Maysinger, D. *J. Mol. Med.* **86**, 291 (2008).
- Choi, A.O., Cho, S.J., Desbarats, J., Lovric, J., Maysinger, D. *J. Nanobiotechnol.* **5**, 1 (2007).
- Clarke, S.J., Hollmann, C.A., Aldaye, F.A., Nadeau, J.L. *Bioconjug. Chem.* **19**, 562 (2008).
- Conner, S.D., Schmid, S.L. *Nature* **422**, 37 (2003).
- Croy, S.R., Kwon, G.S. *Curr. Pharm. Des.* **12**, 4669 (2006).
- Dahan, M., Levi, S., Luccardini, C., Rostaing, P., Riveau, B., Triller, A. *Science* **302**, 442 (2003).
- Dayal, S., Burda, C. *J. Am. Chem. Soc.* **130**, 2890 (2008).
- De La Fuente, M., Seijo, B., Alonso, M.J. *Gene Ther.* **15**, 668 (2008).
- Decuzzi, P., Ferrari, M. *Biomaterials* **28**, 2915 (2007).
- Derfus, A.M., Chen, A.A., Min, D.H., Ruoslahti, E., Bhatia, S.N. *Bioconjug. Chem.* **18**, 1391 (2007).
- Derfus, A.M., Chen, W.C.W., Bhatia, S.N. *Nano Lett.* **4**, 11 (2004).
- Diagaradjane, P., Orenstein-Cardona, J.M., N., E.C.-C., Deorukhkar, A., Shentu, S., Kuno, N., Schwartz, D.L., Gelovani, J.G., Krishnan, S. *Clin. Cancer Res.* **14**, 731 (2008).
- Dixit, S.K., Goicochea, N.L., Daniel, M.C., Murali, A., Bronstein, L., De, M., Stein, B., Rotello, V.M., Kao, C.C., Dragnea, B. *Nano Lett.* **6**, 1993 (2006).
- Dobrovolskaia, M.A., McNeil, S.E. *Nature Nanotechnol.* **2**, 469 (2007).
- Duan, H., Nie, S. *J. Am. Chem. Soc.* **129**, 3333 (2007).
- Duncan, R. *Nat. Rev. Cancer* **6**, 688 (2006).
- Dunphy, I., Vinogradov, S.A., Wilson, D.F. *Anal. Biochem.* **310**, 191 (2002).
- ESF *Nanomedicine*. An ESF—European Medical Research Councils (EMRC) Forward Look Report. Strasbourg, European Science Foundation (2005).
- Ferrari, M. *Nat. Rev. Cancer* **5**, 161 (2005).
- Fischer, H.C., Liu, L.C., Pang, K.S., Chan, W.C.W. *Adv. Func. Mater.* **16**, 1299 (2006).
- Frangioni, J.V., Kim, S.W., Ohnishi, S., Kim, S., Bawendi, M.G. *Methods Mol. Biol.* **374**, 147 (2007).
- Gabizon, A., Shmeeda, H., Barenholz, Y. *Clin. Pharmacokinet.* **42**, 419 (2003).
- Gao, H., Shi, W., Freund, L.B. *Proc. Natl. Acad. Sci. USA* **102**, 9469 (2005).
- Gao, X., Cui, Y., Levenson, R.M., Chung, L.W., Nie, S. *Nature Biotechnol.* **22**, 969 (2004).
- Garnett, M.C., Kallinteri, P. *Occup. Med. (Lond)* **56**, 307 (2006).
- Gill, P.S., Wernz, J., Scadden, D.T., Cohen, P., Mukwaya, G.M., Von Roenn, J.H., Jacobs, M., Kempin, S., Silverberg, I., Gonzales, G., Rarick, M.U., Myers, A.M., Shepherd, F., Sawka, C., Pike, M.C., Ross, M.E. *J. Clin. Oncol.* **14**, 2353 (1996).
- Goodman, C.M., McCusker, C.D., Yilmaz, T., Rotello, V.M. *Bioconjug. Chem.* **15**, 897 (2004).

- Gopee, N.V., Roberts, D.W., Webb, P., Cozart, C.R., Siiiton, P.H., Warbritton, A.R., Yu, W.W., Colvin, V.L., Walker, N.J., Howard, P.C. *Toxicol. Sci.* **98**, 249 (2007).
- Green, M., Howman, E. *Chem. Commun. (Camb.)* **121**, (2005).
- Green, N.K., Herbert, C.W., Hale, S.J., Hale, A.B., Mautner, V., Harkins, R., Hermiston, T., Ulbrich, K., Fisher, K.D., Seymour, L.W. *Gene Ther.* **11**, 1256 (2004).
- Gu, F., Zhang, L., Teply, B.A., Mann, N., Wang, A., Radovic-Moreno, A.F., Langer, R., Farokhzad, O.C. *Proc. Natl. Acad. Sci. USA* **105**, 2586 (2008).
- Guicciardi, M.E., Leist, M., Gores, G.J. *Oncogene* **23**, 2881 (2004).
- Gupta, A.K., Gupta, M. *Biomaterials* **26**, 3995 (2005).
- Haensler, J., Szoka, Jr. F.C., *Bioconjug. Chem.* **4**, 372 (1993).
- Hajitou, A., Trepel, M., Lilley, C.E., Soghomonyan, S., Alauddin, M. M., Marini, F.C., 3rd, Restel, B.H., Ozawa, M.G., Moya, C.A., Rangel, R., Sun, Y., Zaoui, K., Schmidt, M., Von Kalle, C., Weitzman, M.D., Gelovani, J.G., Pasqualini, R., Arap, W. *Cell* **125**, 385 (2006).
- Hama, Y., Koyama, Y., Urano, Y., Choyke, P.L., Kobayashi, H. *J. Invest. Dermatol.* **127**, 2351 (2007).
- Harashima, H., Kiwada, H. *Adv. Drug Deliv. Rev.* **19**, 425 (1996).
- Harush-Frenkel, O., Debottin, N., Benita, S., Altschuler, Y. *Biochem. Biophys. Res. Commun.* **353**, 26 (2007).
- Harush-Frenkel, O., Rozentur, E., Benita, S., Altschuler, Y. *Biomacromolecules* **9**, 435 (2008).
- Hild, W.A., Breunig, M., Goepferich, A. *Eur. J. Pharm. Biopharm.* **68**, 153 (2008).
- Hirsch, L.R., Stafford, R.J., Bankson, J.A., Sershen, S.R., Rivera, B., Price, R.E., Hazle, J.D., Halas, N.J., West, J.L. *Proc. Natl. Acad. Sci. USA* **100**, 13549 (2003).
- Hoet, P.H., Bruske-Hohlfeld, I., Salata, O.V. *J. Nanobiotechnol.* **2**, 12 (2004).
- Hong, S., Bielinska, A.U., Mecke, A., Keszler, B., Beals, J.L., Shi, X., Balogh, L., Orr, B.G., Baker, Jr. J.R., Banaszak Holl, M.M. *Bioconjug. Chem.* **15**, 774 (2004).
- Hong, S., Leroueil, P.R., Janus, E.K., Peters, J.L., Kober, M.M., Islam, M.T., Orr, B.G., Baker, Jr. J.R., Banaszak Holl, M.M. *Bioconjug. Chem.* **17**, 728 (2006).
- Horton, K.L., Stewart, K.M., Fonseca, S.B., Guo, Q., Kelley, S.O. *Chem. Biol.* **15**, 375 (2008).
- Hoshino, A., Fujioka, K., Oku, T., Nakamura, S., Suga, M., Yamaguchi, Y., Suzuki, K., Yasuhara, M., Yamamoto, K. *Microbiol. Immunol.* **48**, 985 (2004).
- Huang, X., El-Sayed, I.H., Qian, W., El-Sayed, M.A. *J. Am. Chem. Soc.* **128**, 2115 (2006).
- Inoue, Y., Izawa, K., Yoshikawa, K., Yamada, H., Tojo, A., Ohtomo, K. *Eur. J. Nucl. Med. Mol. Imaging.* **34**, 2048 (2007).
- Jackson, H., Muhammad, O., Daneshvar, H., Nelms, J., Popescu, A., Vogelbaum, M.A., Bruchez, M., Toms, S.A. *Neurosurgery* **60**, 524 (2007).
- Jaiswal, J.K., Mattoucci, H., Mauro, J.M., Simon, S.M. *Nature Biotechnol.* **21**, 47 (2003).
- Jayagopal, A., Russ, P.K., Haselton, F.R. *Bioconjug. Chem.* **18**, 1424 (2007).
- Jennings, T., Strouse, G. *Adv. Exp. Med. Biol.* **620**, 34 (2007).
- Jiang, W., Kim, B.Y.S., Rutka, J.T., Chan, W.C. *Nature Nanotechnol.* **3**, 145 (2008).

- Johannsen, M., Gneveckow, U., Eckelt, L., Feussner, A., Waldofner, N., Scholz, R., Deger, S., Wust, P., Loening, S.A., Jordan, A. *Int. J. Hyperthermia* **21**, 637 (2005).
- Johannsen, M., Gneveckow, U., Thiesen, B., Taymoorian, K., Cho, C. H., Waldofner, N., Scholz, R., Jordan, A., Loening, S.A., Wust, P. *Eur. Urol.* **52**, 1653 (2007).
- Joliot, A., Prochiantz, A. *Nat. Cell. Biol.* **6**, 189 (2004).
- Jones, A.T., Gumbleton, M., Duncan, R. *Adv. Drug Deliv. Rev.* **55**, 1353 (2003).
- Judge, A., Mcclintock, K., Phelps, J.R., Maclachlan, I. *Mol. Ther.* **13**, 328 (2006).
- Kabanov, A.V., Batrakova, E.V., Alakhov, V.Y. *J. Control Release* **82**, 189 (2002).
- Kam, N.W., Dai, H. *J. Am. Chem. Soc.* **127**, 6021 (2005).
- Kam, N.W., Liu, Z., Dai, H. *J. Am. Chem. Soc.* **127**, 12492 (2005a).
- Kam, N.W., O'connell, M., Wisdom, J.A., Dai, H. *Proc. Natl. Acad. Sci. USA* **102**, 11600 (2005b).
- Kataoka, K., Matsumoto, T., Yokoyama, M., Okano, T., Sakurai, Y., Fukushima, S., Okamoto, K., Kwon, G.S. *J. Control Release* **64**, 143 (2000).
- Khafagy El, S., Morishita, M., Onuki, Y., Takayama, K. *Adv. Drug Deliv. Rev.* **59**, 1521 (2007).
- Kim, K., Lee, M., Park, H., Kim, J.H., Kim, S., Chung, H., Choi, K., Kim, I.S., Seong, B.L., Kwon, I.C. *J. Am. Chem. Soc.* **128**, 3490 (2006).
- Kim, S., Lim, Y.T., Soltesz, E.G., De Grand, A.M., Lee, J., Nakayama, A., Parker, J.A., Mihaljevic, T., Laurence, R.G., Dor, D. M., Cohn, L.H., Bawendi, M.G., Frangioni, J.V. *Nature Biotechnol.* **22**, 93 (2004).
- Kirchner, C., Liedl, T., Kudera, S., Pellegrino, T., Munoz Javier, A., Gaub, H.E., Stolzle, S., Fertig, N., Parak, W.J. *Nano Lett.* **5**, 331 (2005).
- Kobayashi, H., Brechbiel, M.W. *Mol. Imaging.* **2**, 1 (2003).
- Kobayashi, H., Hama, Y., Koyama, Y., Barrett, T., Regino, C.A., Urano, Y., Choyke, P.L. *Nano Lett.* **7**, 1711 (2007).
- Kong, D.F., Goldschmidt-Clermont, P.J. *Trends. Cardiovasc. Med.* **15**, 207 (2005).
- Kostarelos, K., Lacerda, L., Pastorin, G., Wu, W., Wieckowski, S., Luangsivilay, J., Godefroy, S., Pantarotto, D., Briand, J.P., Muller, S., Prato, M., Bianco, A. *Nature Nanotechnol.* **2**, 108 (2007).
- Kukowska-Latallo, J.F., Candido, K.A., Cao, Z., Nigavekar, S.S., Majoros, I.J., Thomas, T.P., Balogh, L.P., Khan, M.K., Baker, Jr. J.R., *Cancer Res.* **65**, 5317 (2005).
- Laakkonen, P., Porkka, K., Hoffman, J.A., Ruoslahti, E. *Nature Med.* **8**, 751 (2002).
- Lai, S.K., Hida, K., Chen, C., Hanes, J. *J. Control Release* **125**, 107 (2008).
- Larson, D.R., Zipfel, W.R., Williams, R.M., Clark, S.W., Bruchez, M.P., Wise, F.W., Webb, W.W. *Science* **300**, 1434 (2003).
- Le Garrec, D., Gori, S., Karkan, D., Luo, L., Lessard, D.G., Smith, D., Ranger, M., Yessine, M.A., Leroux, J.C. *J. Drug Deliv. Sci. Technol.* **15**, 115 (2005).
- Le Garrec, D., Gori, S., Luo, L., Lessard, D., Smith, D.C., Yessine, M.A., Ranger, M., Leroux, J.C. *J. Control Release* **99**, 83 (2004a).
- Le Garrec, D., Ranger, M., Leroux, J.C. *Am. J. Drug Deliv.* **1**, 15 (2004b).
- Le Masne De Chermont, Q., Chaneac, C., Seguin, J., Pelle, F., Maitrejean, S., Jolivet, J.P., Gourier, D., Bessodes, M., Scherman, D. *Proc. Natl. Acad. Sci. USA* **104**, 9266 (2007).
- Lee, C.C., Mackay, J.A., Frechet, J.M., Szoka, F.C. *Nature Biotechnol.* **23**, 1517 (2005).

- Lewin, M., Carlesso, N., Tung, C.H., Tang, X.W., Cory, D., Scadden, D.T., Weissleder, R. *Nature Biotechnol.* **18**, 410 (2000).
- Lewinski, N., Colvin, V., Drezek, R. *Small* **4**, 26 (2008).
- Li, N., Sioutas, C., Cho, A., Schmitz, D., Misra, C., Sempf, J., Wang, M., Oberley, T., Froines, J., Nel, A. *Environ Health Perspect.* **111**, 455 (2003).
- Lin, S., Xie, X., Patel, M.R., Yang, Y.H., Li, Z., Cao, F., Gheysens, O., Zhang, Y., Gambhir, S.S., Rao, J.H., Wu, J.C. *BMC Biotechnol.* **7**, 67 (2007).
- Liu, W., Choi, H.S., Zimmer, J.P., Tanaka, E., Frangioni, J.V., Bawendi, M. *J. Am. Chem. Soc.* **129**, 14530 (2007a).
- Liu, Z., Cai, W.B., He, L.N., Nakayama, N., Chen, K., Sun, X.M., Chen, X.Y., Dai, H.J. *Nature Nanotechnol.* **2**, 47 (2007b).
- Loo, L., Guenther, R.H., Basnayake, V.R., Lommel, S.A., Franzen, S. *J. Am. Chem. Soc.* **128**, 4502 (2006).
- Lorenz, M.R., Holzapfel, V., Musyanovych, A., Nothelfer, K., Walther, P., Frank, H., Landfester, K., Schrezenmeier, H., Mailander, V. *Biomaterials* **27**, 2820 (2006).
- Lovric, J., Bazzi, H.S., Cuie, Y., Fortin, G.R., Winnik, F.M., Maysinger, D. *J. Mol. Med.* **83**, 377 (2005a).
- Lovric, J., Cho, S.J., Winnik, F.M., Maysinger, D. *Chem. Biol.* **12**, 1227 (2005b).
- Lu, Y., Segal, E., Leamon, C.P., Low, P.S. *Adv. Drug Deliv. Rev.* **56**, 1161 (2004).
- Maeda, H. *Adv. Enzyme Regul.* **41**, 189 (2001).
- Maheshri, N., Koerber, J.T., Kaspar, B.K., Schaffer, D.V. *Nature Biotechnol.* **24**, 198 (2006).
- Maier-Hauff, K., Rothe, R., Scholz, R., Gneveckow, U., Wust, P., Thiesen, B., Feussner, A., Von Deimling, A., Waldoefner, N., Felix, R., Jordan, A. *J. Neurooncol.* **81**, 53 (2007).
- Malik, N., Evagorou, E.G., Duncan, R. *Anticancer Drugs* **10**, 767 (1999).
- Manabe, N., Hoshino, A., Liang, Y.Q., Goto, T., Kato, N., Yamamoto, K. *IEEE Trans. Nanobiosci.* **5**, 263 (2006).
- Manchester, M., Singh, P. *Adv. Drug Deliv. Rev.* **58**, 1505 (2006).
- Martin, C.R., Kohli, P. *Nat. Rev. Drug Discov.* **2**, 29 (2003).
- Mastrobattista, E., Van Der Aa, M.A., Hennink, W.E., Crommelin, D.J. *Nat. Rev. Drug Discov.* **5**, 115 (2006).
- MattoSSI, H., Mauro, J.M., Goldman, E.R., Anderson, G.P., Sundar, V.C., Mikulec, F.V., Bawendi, M.G. *J. Am. Chem. Soc.* 12142 (2000).
- Maxfield, F.R., McGraw, T.E. *Nat. Rev. Mol. Cell Biol.* **5**, 121 (2004).
- Mayor, S., Pagano, R.E. *Nat. Rev. Mol. Cell Biol.* **8**, 603 (2007).
- Maysinger, D., Behrendt, M., Lalancette-Hebert, M., Kriz, J. *Nano Lett.* **7**, 2513 (2007a).
- Maysinger, D., Lovric, J., Eisenberg, A., Savic, R. *Eur. J. Pharm. Biopharm.* **65**, 270 (2007b).
- Mcintosh, D.P., Tan, X.Y., Oh, P., Schnitzer, J.E. *Proc. Natl. Acad. Sci. USA* **99**, 1996 (2002).
- Mecke, A., Uppuluri, S., Sassanella, T.M., Lee, D.K., Ramamoorthy, A., Baker, Jr. J.R., Orr, B.G., Banaszak Holl, M.M. *Chem. Phys. Lipids* **132**, 3 (2004).
- Medintz, I.L., Uyeda, H.T., Goldman, E.R., MattoSSI, H. *Nature Mater.* **4**, 435 (2005).

- Michalet, X., Pinaud, F.F., Bentolila, L.A., Tsay, J.M., Doose, S., Li, J.J., Sundaresan, G., Wu, A.M., Gambhir, S.S., Weiss, S. *Science* **307**, 538 (2005).
- Miller, C.R., Bondurant, B., Mclean, S.D., McGovern, K.A., O'Brien, D.F. *Biochemistry* **37**, 12875 (1998).
- Moghim, S.M., Hunter, A.C., Murray, J.C. *Pharmacol. Rev.* **53**, 283 (2001).
- Morgan, N.Y., English, S., Chen, W., Chernomordik, V., Russo, A., Smith, P.D., Gandjbakhche, A. *Acad. Radiol.* **12**, 313 (2005).
- Nabiev, I., Mitchell, S., Davies, A., Williams, Y., Kelleher, D., Moore, R., Gun'ko, Y.K., Byrne, S., Rakovich, Y.P., Donegan, J.F., Sukhanova, A., Conroy, J., Cottell, D., Gaponik, N., Rogach, A., Volkov, Y. *Nano Lett.* **7**, 3452 (2007).
- Nakai, T., Kanamori, T., Sando, S., Aoyama, Y. *J. Am. Chem. Soc.* **125**, 8465 (2003).
- Nakanishi, T., Fukushima, S., Okamoto, K., Suzuki, M., Matsumura, Y., Yokoyama, M., Okano, T., Sakurai, Y., Kataoka, K. *J. Control Release* **74**, 295 (2001).
- Nasongkla, N., Bey, E., Ren, J., Ai, H., Khemtong, C., Guthi, J.S., Chin, S.F., Sherry, A.D., Boothman, D.A., Gao, J. *Nano Lett.* **6**, 2427 (2006).
- Nel, A., Xia, T., Madler, L., Li, N. *Science* **311**, 622 (2006).
- Neu, M., Fischer, D., Kissel, T. *J. Gene. Med.* **7**, 992 (2005).
- Neuberger, T., Schopf, B., Hofmann, H., Hofmann, M., Von Rechenberg, B. *J. Magn. Magn. Mater.* **293**, 483 (2005).
- Nishiyama, N., Kataoka, K. *Pharmacol. Ther.* **112**, 630 (2006).
- Oberdorster, E. *Environ. Health Perspect.* **112**, 1058 (2004).
- Oberdorster, G., Oberdorster, E., Oberdorster, J. *Environ. Health Perspect.* **113**, 823 (2005).
- Osaki, F., Kanamori, T., Sando, S., Sera, T., Aoyama, Y. *J. Am. Chem. Soc.* **126**, 6520 (2004).
- Oyelere, A.K., Chen, P.C., Huang, X., El-Sayed, I.H., El-Sayed, M.A. *Bioconjug. Chem.* **18**, 1490 (2007).
- Panyam, J., Labhasetwar, V. *Adv. Drug Deliv. Rev.* **55**, 329 (2003).
- Park, J.H., Lee, S., Kim, J.H., Park, K., Kim, K., Kwon, I.C. *Prog. Polym. Sci.* **33**, 113 (2008).
- Park, J.W., Hong, K., Kirpotin, D.B., Colbern, G., Shalaby, R., Baselga, J., Shao, Y., Nielsen, U.B., Marks, J.D., Moore, D., Papahadjopoulos, D., Benz, C.C. *Clin. Cancer Res.* **8**, 1172 (2002).
- Pasqualini, R., Ruoslahti, E. *Nature* **380**, 364 (1996).
- Patri, A.K., Kukowska-Latallo, J.F., Baker, Jr. J.R. *Adv. Drug Deliv. Rev.* **57**, 2203 (2005).
- Paunesku, T., Vogt, S., Lai, B., Maser, J., Stojicevic, N., Thurn, K.T., Osipo, C., Liu, H., Legnini, D., Wang, Z., Lee, C., Woloschak, G.E. *Nano Lett.* **7**, 596 (2007).
- Pekny, M., Nilsson, M. *Glia* **50**, 427 (2005).
- Petri-Fink, A., Steitz, B., Finka, A., Salaklang, J., Hofmann, H. *Eur. J. Pharm. Biopharm.* **68**, 129 (2008).
- Reiners, Jr. J.J., Caruso, J.A., Mathieu, P., Chelladurai, B., Yin, X.M., Kessel, D. *Cell Death Differ.* **9**, 934 (2002).
- Rejman, J., Oberle, V., Zuhorn, I.S., Hoekstra, D. *Biochem. J.* **377**, 159 (2004).
- Ridet, J.L., Malhotra, S.K., Privat, A., Gage, F.H. *Trends Neurosci.* **20**, 570 (1997).

- Rieger, S., Kulkarni, R.P., Darcy, D., Fraser, S.E., Koster, R.W. *Dev. Dyn.* **234**, 670 (2005).
- Rosen, A.B., Kelly, D.J., Schuldt, A.J., Lu, J., Potapova, I.A., Doronin, S.V., Robichaud, K.J., Robinson, R.B., Rosen, M.R., Brink, P.R., Gaudette, G.R., Cohen, I.S. *Stem Cells* **25**, 2128 (2007).
- Rossin, R., Pan, D., Qi, K., Turner, J.L., Sun, X., Wooley, K.L., Welch, M.J. *J. Nucl. Med.* **46**, 1210 (2005).
- Roy, I., Ohulchanskyy, T.Y., Bharali, D.J., Pudavar, H.E., Mistretta, R.A., Kaur, N., Prasad, PN. *Proc. Natl. Acad. Sci. USA* **102**, 279 (2005).
- Ryman-Rasmussen, J.P., Riviere, J.E., Monteiro-Riviere, N.A. *J. Invest. Dermatol.* **127**, 143 (2007a).
- Ryman-Rasmussen, J.P., Riviere, J.E., Monteiro-Riviere, N.A. *Nano Lett.* **7**, 1344 (2007b).
- Santra, S., Yang, H., Holloway, P.H., Stanley, J.T., Mericle, R.A. *J. Am. Chem. Soc.* **127**, 1656 (2005).
- Savic, R., Eisenberg, A., Maysinger, D. *J. Drug Target* **14**, 343 (2006).
- Savic, R., Luo, L., Eisenberg, A., Maysinger, D. *Science* **300**, 615 (2003).
- Schnitzer, J.E. *Adv. Drug Deliv. Rev.* **49**, 265 (2001).
- Seleverstov, O., Zabirnyk, O., Zscharnack, M., Bulavina, L., Nowicki, M., Heinrich, J.M., Yezhelyev, M., Emmrich, F., O'regan, R., Bader, A. *Nano Lett.* **6**, 2826 (2006).
- Silva, G.A. *Nat. Rev. Neurosci.* **7**, 65 (2006).
- Simberg, D., Duza, T., Park, J.H., Essler, M., Pilch, J., Zhang, L., Derfus, A.M., Yang, M., Hoffman, R.M., Bhatia, S., Sailor, M.J., Ruoslahti, E. *Proc. Natl. Acad. Sci. USA* **104**, 932 (2007).
- Slotkin, J.R., Chakrabarti, L., Dai, H.N., Carney, R.S., Hirata, T., Bregman, B.S., Gallicano, G.I., Corbin, J.G., Haydar, T.F. *Dev. Dyn.* **236**, 3393 (2007).
- Slowing, Ii, Trewyn, B.G., Lin, V.S. *J. Am. Chem. Soc.* **129**, 8845 (2007).
- So, M.K., Loening, A.M., Gambhir, S.S., Rao, J. *Nature Protoc.* **1**, 1160 (2006a).
- So, M.K., Xu, C., Loening, A.M., Gambhir, S.S., Rao, J. *Nature Biotechnol.* **24**, 339 (2006b).
- Soo Choi, H., Liu, W., Misra, P., Tanaka, E., Zimmer, J.P., Itty Ipe, B., Bawendi, M.G., Frangioni, J.V. *Nature Biotechnol.* **25**, 1165 (2007).
- Stewart, K.M., Horton, K.L., Kelley, S.O. *Org. Biomol. Chem.* **6**, 2242 (2008).
- Stone, V., Donaldson, K. *Nature Nanotechnol.* **1**, 23 (2006).
- Stroh, M., Zimmer, J.P., Duda, D.G., Levchenko, T.S., Cohen, K.S., Brown, E.B., Scadden, D.T., Torchilin, V.P., Bawendi, M.G., Fukumura, D., Jain, R.K. *Nature Med.* **11**, 678 (2005).
- Supattapone, S., Nguyen, H.O., Cohen, F.E., Prusiner, S.B., Scott, M.R. *Proc. Natl. Acad. Sci. USA* **96**, 14529 (1999).
- Susumu, K., Uyeda, H.T., Medintz, I.L., Pons, T., Delehanty, J.B., Mattossi, H. *J. Am. Chem. Soc.* **129**, 13987 (2007).
- Svenson, S., Tomalia, D.A. *Adv. Drug Deliv. Rev.* **57**, 2106 (2005).
- Tada, H., Higuchi, H., Wanatabe, T.M., Ohuchi, N. *Cancer Res.* **67**, 1138 (2007).
- Tan, W.B., Jiang, S., Zhang, Y. *Biomaterials* **28**, 1565 (2007).

- Tanaka, E., Choi, H.S., Fujii, H., Bawendi, M.G., Frangioni, J.V. *Ann. Surg. Oncol.* **13**, 1671 (2006).
- Tang, M.X., Redemann, C.T., Szoka, Jr. F.C. *Bioconjug. Chem.* **7**, 703 (1996).
- Taton, T.A., Mirkin, C.A., Letsinger, R.L. *Science* **289**, 1757 (2000).
- Thorne, R.G., Nicholson, C. *Proc. Natl. Acad. Sci. USA* **103**, 5567 (2006).
- Torchilin, V.P. *J. Microencapsul.* **15**, 1 (1998).
- Torchilin, V.P. *Nat. Rev. Drug Discov.* **4**, 145 (2005).
- Torchilin, V.P., Rammohan, R., Weissig, V., Levchenko, T.S. *Proc. Natl. Acad. Sci. USA* **98**, 8786 (2001).
- Unfried, K., Albrecht, C., Klotz, L.O., Von Mikecz, A., Grether-Beck, S., Schins, R.P.F. *Nanotoxicology* **1**, 52 (2007).
- Voura, E.B., Jaiswal, J.K., Matoussi, H., Simon, S.M. *Nature Med.* **10**, 993 (2004).
- Waehler, R., Russell, S.J., Curiel, D.T. *Nature Rev. Genet.* **8**, 573 (2007).
- Wagner, V., Dullaart, A., Bock, A.K., Zweck, A. *Nature Biotechnol.* **24**, 1211 (2006).
- Wang, L., Yang, C., Tan, W. *Nano Lett.* **5**, 37 (2005).
- Wilhelm, C., Billotey, C., Roger, J., Pons, J.N., Bacri, J.C., Gazeau, F. *Biomaterials* **24**, 1001 (2003).
- Wilhelm, C., Gazeau, F., Roger, J., Pons, J.N., Bacri, J.C. *Langmuir* **18**, 8148 (2002).
- Wu, X., Liu, H., Liu, J., Haley, K.N., Treadway, J.A., Larson, J. P., Ge, N., Peale, F., Bruchez, M.P. *Nature Biotechnol.* **21**, 41 (2003).
- Xia, T., Kovochich, M., Brant, J., Hotze, M., Sempf, J., Oberley, T., Sioutas, C., Yeh, J.L., Weisner, M.R., Nel, A.E. *Nano Lett.* **6**, 1794 (2006).
- Yao, G., Wang, L., Wu, Y., Smith, J., Xu, J., Zhao, W., Lee, E., Tan, W. *Anal. Bioanal. Chem.* **385**, 518 (2006).
- Yu, X., Chen, L., Li, K., Li, Y., Xiao, S., Luo, X., Liu, J., Zhou, L., Deng, Y., Pang, D., Wang, Q. *J. Biomed. Opt.* **12**, 014008 (2007).
- Yun, C.S., Javier, A., Jennings, T., Fisher, M., Hira, S., Peterson, S., Hopkins, B., Reich, N.O., Strouse, G.F. *J. Am. Chem. Soc.* **127**, 3115 (2005).
- Zanello, L.P., Zhao, B., Hu, H., Haddon, R.C. *Nano Lett.* **6**, 562 (2006).
- Zhang, L., Gu, F.X., Chan, J.M., Wang, A.Z., Langer, R.S., Farokhzad, O.C. *Clin. Pharmacol. Ther.* **83**, 761 (2008).
- Zhao, M., Kircher, M.F., Josephson, L., Weissleder, R. *Bioconjug. Chem.* **13**, 840 (2002).
- Zimmer, J.P., Kim, S.W., Ohnishi, S., Tanaka, E., Frangioni, J.V., Bawendi, M.G. *J. Am. Chem. Soc.* **128**, 2526 (2006).

# **Nanostructured probes to enhance optical and vibrational spectroscopic imaging for biomedical applications**

# **15**

<b>15.1 Introduction</b>	<b>539</b>
<b>15.2 Background</b>	<b>542</b>
<b>15.3 Theoretical modelling: NanoLAMPs</b>	<b>548</b>
<b>15.4 Design</b>	<b>564</b>
<b>15.5 Conclusion</b>	<b>566</b>
<b>References</b>	<b>567</b>

*Anil K. Kodali and Rohit Bhargava*

## **15.1 Introduction**

The probability of successful treatment of diseases like carcinoma is highly dependent on their early detection. It is now being increasingly recognized that diagnosis by novel disease classifications, effective treatment by personalized medicine and functional pathology by case-specific understanding are all based on the ability to obtain molecular expression profiles of malignancies. In this milieu, ultrasensitive and high-throughput profiling of multiple biomolecular species becomes critical. Optical spectroscopic imaging techniques are especially attractive for high-throughput profiling due to the richness of biochemical and structural information they provide and the relatively accessible instrumentation. For ultrasensitive molecular detection, however, the sensitivity of optical responses is often not high enough. This requires either improving the instrumentation for highly sensitive and specific recording or amplifying the probe signal. A combination of the approaches and enhancements in computation or modelling may also be used for maximal effect. Contemporary optical modalities and associated probe-assisted variants are, unfortunately, limited in the number of different species that could be simultaneously targeted in a biofluid or in a tissue. Poor signal intensity, broad spectral responses, photobleaching, cellular toxicity and complicated fabrication are some of the other barriers that limit their applicability.

Vibrational spectroscopy (both infrared absorption and Raman scattering) provides structural and chemical information of molecules based on

their vibrational transitions. Its high molecular specificity makes it exceptionally suitable as a transduction mechanism for multiplexed biosensing. Further, imaging techniques based on vibrational spectroscopy offer stable signals and may be suitable for both *in-vivo* and *ex-vivo* use. In Raman spectroscopy, inelastic scattering of incident radiation due to its interaction with the electron cloud of a functional group is detected. The inelasticity is a result of nuclei moving during the interaction of incident photons with the electron cloud. Such scattering events are significantly rarer compared to elastic (Rayleigh) scattering, resulting in extremely low cross-sections (about  $10^{-29}$  to  $10^{-31}$  cm $^2$ /molecule). The Raman response hence needs to be enhanced by about 14 to 15 orders of magnitude to achieve molecular sensitivity. On the other hand, infrared (IR) spectroscopy detects the loss or absorption of incident IR frequencies that match those of vibrational transitions due to these interactions. Although the absorption cross-sections are about 9 orders of magnitude higher than the Raman scattering cross-sections, the IR signal is still not strong enough for monomolecular characterization, due to instrument limitations and simple thermal fluctuations at room temperature.

Both techniques have made considerable progress towards exceptionally high sensitivity in recent years, however, with the availability of better components and computers. In particular, the advances in imaging instrumentation now allow for thousands of pixels of data to be acquired in much the same time as single spectra a few years ago. These emerging capabilities have driven instrumentation towards scanning large areas and/or more sophisticated analyses from smaller volumes, thereby increasing the need for more sensitive analyses. Advanced instrumentation, such as multiple excitation wavelength sources in the case of Raman spectroscopy, can be used to enhance signals. While increasing the expense and complexity, such techniques are often still limited in the highest sensitivity level attained. Alternatively, engineered probes and surfaces can be used to enhance signals for ultrasensitive detection. For example, nanostructured substrates (nanofeatured surfaces and nanoparticle probes) can be used as enhancing agents, based on the optical enhancement effects in their vicinity.

Enhanced vibrational spectroscopy of molecules adsorbed on or in the vicinity of metal (inorganic) particles and/or surfaces is termed: surface-enhanced infrared absorption (SEIRA) spectroscopy and surface-enhanced Raman spectroscopy (SERS) for IR and Raman modalities, respectively. Together, SERS and SEIRA can be broadly categorized under a common modality termed surface-enhanced vibrational spectroscopy (SEVS). The experimental signal enhancement observed in SEVS, is a compounded contribution of several factors. The primary and dominant effects in determining the spectral enhancement characteristics of the analyte molecule are the surface plasmon resonance of the substrate and its variations due to surface roughness. In addition, resonance effects of the analyte molecule, its charge-transfer coupling with the substrate and other chemical interactions play a contributing role. Such complex interactions have often resulted in case-specific results of enhancement and defied general design principles for experimentally reliable SERS and SEIRA since their discovery about three decades ago. Efforts to understand

the enhancement effects and harness them using rationally designed substrates, where undertaken, have proven to be very effective. Reports published about 10 years ago indicating the possibility of single-molecule detection using SERS and monomolecular film characterization using SEIRA also renewed the interest in improving the reliability of SEVS substrates. Current SEVS substrates can be divided into two major categories based on methodology for fabrication: substrates based on surfaces with nanoengineered features and substrates based on colloidal solutions of nanoparticles.

Nanostructured surfaces, when carefully designed, offer controllability and reliability in enhancement. These applications represent the true promise of SEVS in that the enhancements can reasonably be predicted, the desired substrates fabricated and the results consistently validated. Patterned substrates, unfortunately, are limited in the achievable sensitivity levels and are especially not suitable for *in-vivo* applications. Metal nanoparticle aggregate (colloidal) probes, in contrast, are ideal for *in-vivo* applications. Further, the giant enhancements reported for molecules adsorbed on particle surfaces and in interstices provide an excellent opportunity for single-molecule sensitivity. The detection reliability of target molecules is, however, affected by target-adsorption variability on the nanoparticle probe surface and the nature of the interstitial space in which it is localized. Hence, one avenue that has been suggested is to avoid direct measurement of the desired analyte and improve data reproducibility by detecting a SERS probe associated with the target molecule(s).

Concentric-layered probes with symmetry in interspersing reporting molecules within metal layers are one such implementation of the concept of a Raman spectral probe based on SERS. The spherical symmetry offers reliability of the optical response and negates polarization effects. Recently, some two-layered metal–dielectric probes have been proposed and their applicability for a variety of applications is demonstrated. The reported expertise for two-layered probes, however, is lacking in a rational theoretical design and in the required level of sensitivity. We propose multilayered metal nanoparticles as a paradigm for probes with reliable and reproducible as well as optimal SEVS response. Hereafter, these probes will be referred to as nano-LAMPs (nano-layered metal probes). These nanoparticles have Raman/IR-active reporting layers interspersed within metal layers. The reporting layers are made of organic molecules of structural similarity but can be widely changed from probe to probe for different vibrational patterns, enabling multiplexed detection. Nano-LAMPs also respond to multiple excitations and offer higher optical tunability and enhancement per probe than the two-layered or three-layered probes. In this chapter, we propose the use of the nano-LAMP configuration as a paradigm for multiplexed SERS, propose a theoretical design framework based on classical electromagnetic (EM) scattering calculations and provide example predictions for various configurations.

As background to the developed models, we describe various breakthrough findings reported in SERS and SEIRA in Section 15.2. Different types of substrates and contrast agents used in realizing the enhancement and theories proposed to explain these findings are also reviewed. In Section 15.3, the configurations of nano-LAMPs are discussed. The EM calculations to evaluate

the potential for SEVS within these structures are provided. Example results are presented demonstrating their optical resonances and tunability. The theoretical design principles and optimization are briefly discussed in Section 15.4. Last, a few techniques to fabricate multilayered nanoparticles and some issues with respect to fabrication are presented.

## 15.2 Background

Several studies have been reported in the fields of SEVS and nanoparticles and their biomolecular applications. A comprehensive review of the history of this enormous body of literature cannot possibly be included in this chapter. We therefore choose to highlight only some of the reports that are relevant to understanding the background and current status of these fields with respect to the proposed concept of nano-LAMPs.

### 15.2.1 SEVS

#### 15.2.1.1 SERS

The SERS effect was first demonstrated for pyridine molecules adsorbed on roughened silver electrodes (Fleischman *et al.* 1974; Albrecht *et al.* 1977; Jeanmarie *et al.* 1977). The high Raman signal observed in 1974 was thought to be due to the increase in the number of pyridine molecules adsorbed on the increased surface area of roughened silver electrodes. Later experiments, in 1977, however, demonstrated that the surface enrichment by itself is not sufficient to explain the  $\sim 10^5$  enhancement and it was proposed that the Raman scattering efficiency was enhanced. Jeanmarie *et al.* (1977) suggested that the enhanced intensity was a result of amplified electric fields in the vicinity of metal surfaces. In this line of thinking, Efrima *et al.* (1978) considered the absorbate molecule as a classical dipole at a smooth planar metal surface. The dependence of enhancement on excitation wavelength was explained using multiple scattering calculations of higher-order harmonics between the dipole and metal surface. The effect of roughness was taken into account by a study conducted later (Moskovits 1978) in which roughened metal substrates were modelled as a 2D colloid of metal spheres on top of a smooth metal surface. The collective resonances due to excitations of conduction-electron resonances were evaluated to determine the effect of the degree of roughness on enhancement. Very soon, intense Raman scattering of pyridine molecules adsorbed on gold and colloidal sols was reported (Creighton *et al.* 1979). Apart from paving the way for metal colloidal sols to be used as SERS substrates, this study corroborated the explanations for enhancement based on EM calculations.

These first EM theory predictions of Raman enhancement for molecules near metal nanoparticles were based on simple models, such as isolated spheres and spheroids (Gersten *et al.* 1980; Kerker *et al.* 1980). Gersten *et al.* (1980) pointed out the curvature effects on the enhancement termed as “lightning-rod effects”. On a parallel track to EM theory, Albrecht *et al.* (1977) proposed that enhancement is due to resonance Raman scattering, which is a

result of electronic interaction of adsorbate molecules with the metal surface. Smardzewski *et al.* (1979) claimed that the giant enhancement is restricted to the spectra of adsorbed species in the immediate vicinity (“first-layer”) of the metal surface. In this study, the Raman spectrum of a thin film of deuterium-enriched pyridine on silver was examined. It was observed that approximately 1000 times as much pyridine had to be deposited on the top of this thin film to yield Raman intensities similar to it. Based on these kinds of studies it was proposed that the adsorbate had to be bound on an atomic-scale roughness for strong enhancements to be possible (Otto *et al.* 1980). This view, termed chemical theory, proposed that the SERS effect was a resonance Raman phenomenon due to bonding of the adsorbate with the surface.

Zwemer *et al.* in 1980, considered a setup similar to that of Smardzewski *et al.* (1979) with a monolayer of pyridine on an iodine-roughened silver surface followed by deposition of deuterated pyridine on the top. They reported that both species contribute to the enhanced spectra and SERS certainly extends beyond the first layer, as predicted by EM theories. This opinion was further popularized by other experimental findings in 1980 that large enhancements persist for molecule–metal separations as high as 100 Å (Murray *et al.* 1980; Rowe *et al.* 1980). Such findings led to a further controversy and debate pertaining to the contribution of electromagnetic and chemical effects.

The contradicting theories and its immense potential made SERS an academically fascinating field during these early years. At the same time, a flurry of experiments were conducted and the SERS effect was reported on several molecules and molecular groups like: cyanide complexes, 2,6-lutidine, methylpyridines, isonicotinic acid, hydrocarbons, cytochrome C, ruthenium red, pyrazine, nucleic acids, polystyrene, adenine mononucleotides, azides, benzene, acetonitrile, bile pigments, isoquinoline, cyclohexane etc. The first SERS substrates used in most of these experiments were metal (mostly silver) electrode surfaces subjected to alternating oxidation reduction cycles. Though the initial experiments were done using silver electrodes, by the early 1980s SERS was reported using substrates made of various coinage and alkali metals, including gold, copper, lithium, sodium, potassium, indium aluminum, platinum, cadmium, palladium and rhodium. In addition to different metals, SERS substrates were also fabricated using different approaches. An excellent review by Moskovits (1985) lists fabrication procedures of SERS-active systems proposed by various groups. Table 15.1 summarizes some of these surface-based and nanoparticle-based SERS systems in no particular order.

During 1980–85, several reports were published further validating the EM contribution to SERS enhancement. These early works of EM theory have been thoroughly reviewed (Metiu *et al.* 1984; Schatz 1984; Moskovits 1985). Liao *et al.* (1981) measured Raman spectra of cyanide deposited on substrates with lithographically obtained silver ellipsoids. They validated the dependence of SERS enhancement on aspect ratio and dielectric constant of the ambience as predicted by EM theory. The agreement of EM theory predictions was demonstrated on silver-island films by Weitz *et al.* (1982). Furthermore, some studies showed that SERS excitation spectra obtained for different molecules using rough films were similar. This led to the further belief that resonance strongly

**Table 15.1** Initial SERS substrates

Substrate	Fabrication procedure(s)
Roughened metal surfaces	1) Oxidation-reduction cycles of electrode surfaces 2) Ion bombardment <i>in vacuo</i> 3) Mechanical polishing 4) Acid etching 5) Metal-halide films exposed to UV-radiation 6) Metal deposition on rough CaF <sub>2</sub> films
Island films	Vapor deposition of silver on warm glass/quartz substrate
Cold-deposited films	Vapor deposition of silver on a cold substrate below 30 K
Spheroid assembly gratings	Glancing-angle evaporation of silver onto SiO <sub>2</sub> posts produced by microlithography
Metal gratings	Surface modulation of silver using lithography
Silver colloidal sols	1) Reduction of dissolved metal salt in aqueous or non-aqueous medium 2) Quenching the metal vapor in an argon atmosphere and condensing on a low-temperature surface
Silver powders	AgNO <sub>3</sub> reduced with glucose in aqueous KOH
Colloidal platinum	K <sub>2</sub> PtCl <sub>6</sub> reduced by sodium citrate
Rhodium particles	Rhodium salt reduction

depends on the variation of localized plasmon on the surface geometry of metal substrate. Some of these EM models were, however, based on calculations for isolated metallic spheroids; hence, they failed to account for important features of SERS spectra due to coupling effects in systems like cold-deposited films and aggregated colloids. The electrostatic EM calculations based on coupled particles (Aravind *et al.* 1981; Albano *et al.* 1983) also suggested that the interstices between metal features/particles are high-field locations or “hot-spots”. In the case of spheres the fields in interstices were shown to exceed those near single spheres by an order of magnitude (Liver *et al.* 1984).

The EM models assume that the total enhancement is a product of enhancement in the incident and emitted fields at their respective frequencies. Thus, an order of magnitude increase in the field translates into four orders of magnitude enhancement in the Raman scattering signal. Some of the earlier EM models were based on electrostatic calculations and hence did not include retardation effects due to higher-order multipoles. The calculations made by Kerker *et al.* (1980) on spherical particles were, however, electrodynamic in nature and accounted for these. Barber *et al.* (1983) evaluated fields on surfaces of prolate spheroids using electrodynamic calculations. Despite the fact that EM theory enjoyed success in explaining many effects, it couldn’t evolve into a comprehensive model to explain the SERS effect. Moskovits (1985) reviewed a list of experimental SERS observations EM theory failed to explain. Some of these are: differences in SERS intensities of some molecules of equally strong ordinary Raman spectra, differences between relative intensities of many lines in a SERS spectrum of a molecule, lack of correlation between SERS excitation spectra and absorption spectra in colloidal silver, absence of signal of

electrolytes. Avouris *et al.* (1981) reported a weak absorption in the electron energy-loss spectrum of several molecules adsorbed on silver consistent with their relative SERS enhancement. This absorption was interpreted as a charge-transfer band leading to a charge-transfer mechanism. Based on other similar observations, a theory for the chemical contribution to the enhancement was proposed. This theory proposed that the weakness of the SERS effect in some adsorbates compared to other molecules was due to lack of a charge-transfer band. The differences in the relative intensities of features of SERS spectra were attributed to the fact that the observed spectra are resonance spectra of a metal–molecule complex instead of the adsorbate itself. In 1984, Murray *et al.* suggested that the chemical effect can be included in EM theory models by using the polarizability of the metal–molecule complex rather than that of the molecule in the pertinent equations. Zeman *et al.* (1987) proposed corrections for electrodynamic and surface-scattering effects in EM models based on electrostatics. Over the next few years the field of SERS continued to attract several studies on the theoretical front (in improving and extending on EM and chemical theories) and on the experimental front (in improving the reliability and extending the applications of SERS).

Surface-enhanced resonance Raman spectroscopy (SERRS) of rhodamine 6G on colloidal silver particles was reported in 1984 (Hildebrandt *et al.* 1984). The colloidal particles are anion activated to form closely spaced interstices with huge enhancements. Theoretically, colloidal aggregates were modelled as fractal clusters and the role of scaling properties of the fractal on Raman scattering was examined (Stockman *et al.* 1992). In a series of papers, Xu *et al.* (1993, 1994) developed and applied analytical expressions for calculating the primary Raman–Stokes moments using the amplitudes of scattered multipolar fields. They further made calculations for CO and benzene physisorbed on pairs and linear chains of particles made of silver, platinum, germanium and SiO. By this time, numerical methods like finite-element methods, discrete dipole approximation, finite-difference time-domain method and T-matrix method (Barber *et al.* 1990; Kunz *et al.* 1993; Yang *et al.* 1995) along with exact methods based on Mie scattering theory were all used to account for the enhancement based on EM fields. Kneipp *et al.* (1995) proposed the possibility of single-molecule SERRS using colloids. Very shortly afterwards, Kneipp *et al.* (1997) and Nie *et al.* (1997) reported the first findings of single-molecule surface-enhanced Raman scattering (SM-SERS). The former group demonstrated large effective cross-sections ( $10^{-17}$  to  $10^{-16}\text{ cm}^2/\text{molecule}$ ) of crystal violet molecule in aqueous colloidal silver solution, while the latter group demonstrated a similar order of enhancement for rhodamine 6G molecules adsorbed on selected nanoparticles. These observations were further corroborated on Hemoglobin molecules using isolated silver nanoparticles and on rhodamine 6G molecules using silver nanocrystals (Michaels *et al.* 1999; Xu *et al.* 1999). These reports accompanied by the potential widespread applicability of a SM-SERS modality fuelled the interest of several research groups in the SERS theory of colloidal nanoparticle substrates. Based on the prospect of huge enhancements of SERS using colloidal nanoparticles, self-assembled nanoparticle aggregates and composite nanoparticles were of interest with a hope to control the reliability of enhancement. Oldenburg

*et al.* (1999) fabricated core-shell nanoparticles made of dielectric core and metal shell nanoparticles (called nanoshells) using colloidal growth chemistry. Nanoshells confirmed the optical tunability of particles with a combination of dielectric and metal layers, an idea initially proposed by Neeves *et al.* (1989). Around the same time, programmed assemblies of gold nanoparticles using DNA linkers were proposed (Alivisatos *et al.* 1996; Mirkin *et al.* 1996). The gold nanoparticle assemblies were later adapted as probes for genomic applications (Cao *et al.* 2002).

While these were seemingly bottom-up applications in which the enhancement was achieved using sequential building blocks, effect, several advances were made in top-down nanofabrication procedures, which involved large-scale patterning of surfaces. One significant approach was the use of lithography to produce nanoscale features. One of the most common ones is based on natural microscale lithography (Deckman *et al.* 1982). In this technique, random and/or ordered submicrometer columnar structures were produced using colloidal particles as etching/or deposition masks. Based on extending this approach, nanosphere lithography to obtain nanometric periodic particle arrays was proposed (Hulteen *et al.* 1995), which was later improved to create arrays of structures of different geometries. During recent years, significant progress has been made in creating different types of SERS substrates. Some of the lithography-based substrates reported of late are by Baumberg *et al.* (2005), Perney *et al.* (2006) and Qin *et al.* (2006). Some of the recent and popular particle self-assembly-based substrates are nanoshells (Oldenburg *et al.* 1999; Jackson *et al.* 2004), dye-embedded nanoparticles (Doering *et al.* 2003; Ansari *et al.* 2005), protein-functionalized nanoparticles (Thaxton *et al.* 2006) and coupled nanowires (Tao *et al.* 2005). From the perspective of the discussion here, a very interesting idea was to symmetrically structure materials on the nanoscale. Xu (2005) proposed the idea of multilayered core-shell metal nanospheres to induce large optical fields at the core that could be used for very high Raman enhancements. No theoretical basis of designing optimal structures, however, based on this principle has been reported so far. In SM-SERS, understanding the quantitative specificity of Raman enhancement is incomplete unless the charge complexes and chemical mechanism are considered. In designing the optimal structures, however, one can hope that with a given set of materials and metal substrates the chemical enhancement factor remains constant and instead optimize the nanoscale geometry of the substrates.

### 15.2.1.2 SEIRA

<sup>1</sup>Though the term attenuate total reflectance finds common usage, we contend that the correct etiology should dictate that the ATR configuration be termed attenuated total reflection for the geometry and AT reflectance for the measurement.

Soon after the first experiments of SERS, SEIRA was also observed in attenuated total reflection (ATR) geometry<sup>1</sup> in thin films of p-nitrobenzoic acid deposited on silicon substrates with a thin metal (either gold or silver) layer evaporated on the top (Hartstein *et al.* 1980). The enhanced bands reported were incorrectly assigned to C-H stretching modes of aromatic molecules in this report, while much later they were found to be due to contaminants (Jensen *et al.* 1999, 2000). Also, the limited transparency of the ATR prism used in the experiment allowed for only spectral features above 2800 cm<sup>-1</sup> to be recorded. This led Hatta *et al.* (1982) to measure features below 2000 cm<sup>-1</sup> for nitrobenzoate deposited on a silver layer, demonstrating the existence of

enhancement. A multilayer Fresnel formula was used to calculate the fields at the metal air interface in such a configuration (Ueba *et al.* 1982) suggesting that enhanced IR absorption is due to the surface plasmon polariton of the metal surface. This was not adequate in explaining certain polarization-dependent observations (Osawa *et al.* 1986). Nakao *et al.* (1986) reported enhanced IR spectra for polymer samples using silver, nickel, palladium and platinum films. The maximum enhancement in this experiment was observed at an angle different from the plasmon angle and the authors suggested that the effects of multiple reflections in the metal film, a decrease in penetration depth and chemical interactions at the sample metal interface might contribute to absorption enhancement. Hatta *et al.* (1984, 1985) reported that the short-ranged enhancement and polarization effects could be explained by considering the collective electron resonance of small metal islands in the films based on EM theory. Later, it was shown that the inplane component of the electric field in the metal film can enhance IR absorption of monolayer of stearic acid consistent with the island nature and EM theory (Kamata *et al.* 1987).

Wadayama *et al.* (1988) showed that enhanced IR absorption of thiocynate ( $\text{SCN}^-$ ) ions on Au film is restricted to the immediate vicinity and reaches a maximum at negative electrode potential. Based on this observation, a charge-transfer mechanism was suggested for IR enhancement. Hatta *et al.* (1989) further tried to distinguish EM and chemical enhancement mechanisms and noticed that for films of polycyanoacrylate deposited on silver films, enhancement extends up to a thickness of 5 nm. Using Langmuir–Blodgett (LB) monolayers of a rhodamine derivative on silver-island films, Muraki *et al.* (1989) demonstrated the long-range EM enhancement characteristic of SEIRA. EM computational approaches proposed for SERS by Zeman *et al.* (1987) were general for the surface-enhanced mechanism and were also applicable to SEIRA.

In the normal configuration, SEIRA was demonstrated in transmission mode specifically for p-nitrobenzoic acid on top of silver-island films evaporated on a  $\text{CaF}_2$  substrate (Osawa *et al.* 1991). This study suggested and examined three contributions to SEIRA: long-ranged EM enhancement due to the metal-island films, enhanced vibrational polarizability due to chemical interactions of metal surface and molecules (charge-transfer mechanism) and orientation effects of vibrational dipoles of molecules. SEIRA in the reflection configuration was reported for thin organic films on low reflective substrates by evaporating thin silver films either on the sample surface or on the substrates (Nishikawa *et al.* 1993). Enhanced IR spectra of LB monolayers on metal-island films were reported in both reflection and transmission configurations (Johnson *et al.* 1995). These authors showed that a complete surface-enhanced vibrational characterization (SERS and SEIRA) can be achieved using a single sample substrate. Also, in the reflection configuration using effective medium theory, in which the island and LB film are modelled as a composite film, they showed the effect of distance of the thin film from the metal and the angle of incidence on enhancement.

Most of the thin metal films mentioned above were prepared by using a high-vacuum evaporator. During the last ten years, different SEIRA-active systems have been reported. Kang *et al.* (1998) have demonstrated SEIRA

**Table 15.2** Reviews on biomolecular applications of SERS

Review	Application
Dou <i>et al.</i> (1999)	Metal colloids for quantitative analysis of biomolecules
Kneipp <i>et al.</i> (2002)	Biomedical spectroscopy and biophysics of molecules and processes
Vo-Dinh (2003)	Near-field SERS probes and SERS-based bioassays
Pinzaru <i>et al.</i> (2004)	Quantitative analysis in pharmaceuticals
Grow <i>et al.</i> (2003)	Biochip technology for pathogen detection
Smith <i>et al.</i> (2001)	SERS applications in forensic sciences, drugs, explosives, etc.
Schlucker (2009)	Biomedical applications of SERS nanoparticle probes
Hering <i>et al.</i> (2008)	SERS in biochemical analyses
Porter <i>et al.</i> (2008)	Bioassay platforms utilizing SERS

using silver colloidal particles spread on a transparent substrate. Further metal particle layers were obtained on transparent substrates by techniques like silane-induced immobilization of laser-ablated silver colloids, sputtering of metals like gold, silver and platinum. Nanoparticle films of Ag, Au, Cu and Pt were also successfully prepared using chemical deposition. Apart from coinage metals, SEIRA was reported on island films and other substrates of metals like Sn, Pb, Fe, Pt, Ni, Pd, Rh, Ir, etc. The enhancement in transition metals was characterized by using phonon resonances. Periodic reviews on SEIRA (Aroca *et al.* 2004; Osawa *et al.* 1997) provide an excellent listing of various theories used to model SEIRA and various SEIRA substrates fabricated. The progress made in recent years has been well reviewed in Aroca (2006).

### 15.2.2 Biomolecular applications of SEVS

Table 15.2 gives information about some recent reviews of biological fields of applications of SERS. This is by no means a complete list of such reviews but only provides a few pointer reviews to important directions in the field. The applications of SEIRA for biomedical analyses are relatively rarer, due in part to the significantly lower enhancements observed. For current SEIRA applications, readers are referred to Aroca (2006).

## 15.3 Theoretical modelling: NanoLAMPs

“... although much remains to be discovered both theoretically and experimentally, we understand a great deal about the origin of SERS, certainly for the knowledgeable experimentalist to design workable SERS systems and successful SERS experiments and otherwise make useful predictions. The notion that the basic physics behind SERS is either absent or wrong, is therefore a bewildering shibboleth that often creates doubt in the ability to design SERS-based applications, rationally.” Kneipp *et al.* (2006)

“There is a consensus on the electromagnetic origin and fundamental properties of the signal enhancement of SERS, as assisted by surface plasmon excitation on certain nanostructures. Thus, the presence of this component in the observed enhanced intensity will define the observed spectrum as SERS spectrum. Defining SERS in terms of one of the components of the observed enhanced intensity may, at first, seem limited and narrow. However, this definition provides the basis for a full discussion

of the observations and also a guide for the experimentalist to tune experimental conditions according to the ultimate goal of their research project. The definition does not necessarily imply that the plasmon-assisted contribution ought to be the largest; other resonances may contribute and, in some cases, produce dominant contributions. However, it is the presence of the plasmon resonance that will define the observed spectral intensities as a SERS spectrum".  
Aroca (2006)

The two excerpts quoted above, brilliantly sum up our motivation and philosophy in devising nano-LAMPs for SEVS. Nano-LAMPs are nanospheres with a metal core surrounded by a number of metal and dielectric shells. The dielectric shells can be embedded with organic molecules that act as reporters. For Raman applications, dyes like rhodamine 6G can also be used as reporters to exploit the resonance effect. The outermost metal or dielectric layer is covered with a protective glass layer to shield from the aqueous ambience in biological samples. Previously, this encapsulation technique has been used for Raman probes based on single gold and silver nanoparticles tagged with reporting molecules (Mulvaney *et al.* 2003). Each layer of a nano-LAMP has its own reporting characteristics and can be uniquely identified. The outer glass coat is conjugated with antibodies that bind to the specific receptor sites through complimentary antigens. Hence, the framework of this design provides for flexibility in design and options for increasing enhancements to specified levels. Figure 15.1 shows an application for tissue imaging based on nano-LAMPs; the inset depicts a two-dimensional representation of a typical nano-LAMP along a section cut through its center.

While understanding the chemical enhancement mechanism has proven to be very important for quantitatively characterizing SEVS signals, designing the surface plasmon resonances ensures a minimum level of enhancement in the signal of reporter layers and, hence, of the probes. The effects due to chemical enhancement, other non-linear resonances and the unreliability of fabrication can all be modelled using a correction factor when translating the theoretically predicted signals to experimental observations. This factor can

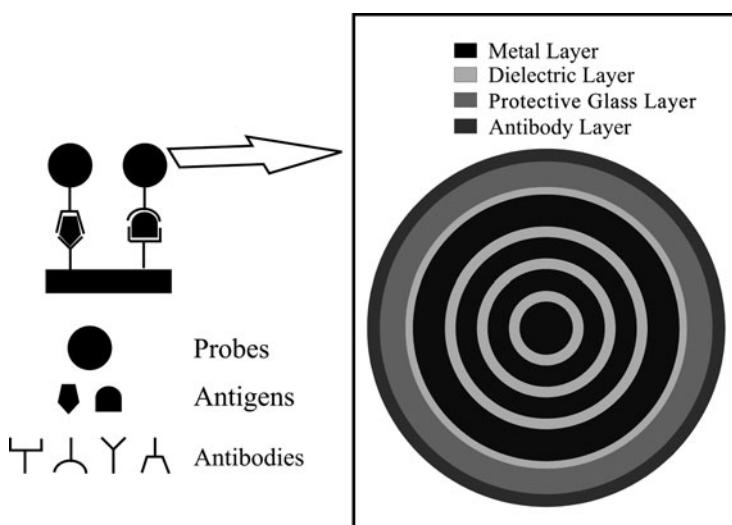


Fig. 15.1 Modality using nano-LAMPs, inset shows a 2D representation at the center.

be established by a series of validation experiments. The first important step then, is to theoretically design the probe geometry to take advantage of the surface plasmon resonances within the probe due to metal layers. Although the focus here is discussing such models for SERS, the principles essentially remain the same for SEIRA and, where needed, we refer to results pertaining to SEIRA.

We now consider the SERS enhancement in this nano-LAMP in terms of the Raman signal given by:

$$P_{\text{SERS}}(\omega_S) = N\sigma_{\text{ads}}G_{\text{EM}}I(\omega_L), \quad (15.1)$$

where  $N$  is the number of reporting molecules with enhanced Raman scattering (this generally is different from the actual number of reporting molecules or volume of reporting layer in the probes),  $\sigma$  is the increased Raman cross-section due to chemical effects,  $G_{\text{EM}}$  is the electromagnetic contribution to the enhancement and  $I(\omega_L)$  is the intensity at the excitation frequency. If the same amount of molecules/reporting dielectric volume, were free without the presence of the metal layers the Raman signal would have been given by:

$$P_{\text{RS}}(\omega_S) = N'\sigma'I(\omega_L), \quad (15.2)$$

where  $N'$  is the actual number of reporting molecules embedded in the probes and  $\sigma'$  is the Raman cross-section.

$G_{\text{EM}}$  in eqn (15.1) is the electromagnetic field enhancement factor and is given by:

$$G_{\text{EM}} \propto \sum M(\omega_S)M(\omega_L). \quad (15.3)$$

The factors  $M(\omega_L)$  and  $M(\omega_S)$  are the magnitudes of local electric field intensities. It should be noted here that the total signal of the probe is important rather than the average signal of the reporting molecules. For example, a local enhancement may be such that it allows for single-molecule measurements. Hence, it is important to consider the order of magnitude of the maximum value of the local E-fields and an integrated value over the whole volume in evaluating and maximizing the SERS signal. Also, the factor  $M$  is usually of the same order at the laser and shifted frequencies. Hence, it would be sufficient to consider the square of one of these intensities in the summation. One formulation of the problem of designing Raman enhancement then reduces to solving for the local electromagnetic fields within the probe when subject to a specific excitation laser. The electromagnetic enhancement factor for SEIRA would include only a single photon and not the emitted field. So it will include only the enhancement in the laser (excitation) field. To summarize, the electromagnetic field enhancement factor provides a measure of the probe signal and we seek to maximize it.

The electromagnetic fields due to an incident field on a probe can be evaluated by solving Maxwell's equations. The incident laser beam can be considered as a superposition of plane waves and hence it is sufficient to consider a plane-wave excitation. The solution for a multilayered sphere can be computed by using a well-established analytical scattering solu-

tion based on Mie theory. Here, we present a few details of the method we use that is based on the recursive formulation due to Wu *et al.* (1991).

### 15.3.1 EM scattering: Recursive formulation

Consider a plane EM wave of frequency  $\omega$  incident on an  $N$ -layered sphere and denote the relative refractive index of the  $j$ th layer with respect to the refractive index of the surrounding medium by  $m_j$ . Since we are interested in electric fields, consider the sphere and surrounding medium to have a magnetic permeability of 1. Without loss of generality the incident wave can be considered to be propagating along the positive  $z$ -axis and polarized along the positive  $x$ -axis. Such a field is expressed as:

$$\vec{E}_{\text{inc}} = E_0 e^{i\kappa z} \hat{e}_x, \quad (15.4)$$

where  $\kappa$  is the propagation constant in the surrounding medium.

To satisfy the continuity of fields at the spherical surface boundaries, the electric and magnetic fields need to be expanded in terms of geometrically similar basis functions, specifically, vector spherical harmonics (VSH). The incident plane-wave field can be expanded in terms of VSH as:

$$\begin{aligned} \vec{E}_{\text{inc}} &= E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( \vec{\mathbf{M}}_{o1n}^{(1)} - i \vec{\mathbf{N}}_{e1n}^{(1)} \right) \\ \vec{H}_{\text{inc}} &= -\frac{\kappa}{\omega} E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( \vec{\mathbf{M}}_{e1n}^{(1)} + i \vec{\mathbf{N}}_{o1n}^{(1)} \right). \end{aligned} \quad (15.5)$$

The scattered fields outside the sphere can be expanded in VSH as:

$$\begin{aligned} \vec{E}_{\text{sc}} &= E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( a_n^{(N+1)} \vec{\mathbf{M}}_{o1n}^{(3)} - i b_n^{(N+1)} \vec{\mathbf{N}}_{e1n}^{(3)} \right) \\ \vec{H}_{\text{sc}} &= -\frac{\kappa}{\omega} E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( b_n^{(N+1)} \vec{\mathbf{M}}_{e1n}^{(3)} + i a_n^{(N+1)} \vec{\mathbf{N}}_{o1n}^{(3)} \right). \end{aligned} \quad (15.6)$$

The fields in the  $j$ th layer in the sphere can be expanded as:

$$\begin{aligned} \vec{E}_j &= E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( A_n^j \vec{\mathbf{M}}_{o1n}^{(1)} - i B_n^j \vec{\mathbf{N}}_{e1n}^{(1)} + C_n^j \vec{\mathbf{M}}_{o1n}^{(3)} - i F_n^j \vec{\mathbf{N}}_{e1n}^{(3)} \right) \\ \vec{H}_j &= -\frac{\kappa_j}{\omega} E_0 \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left( B_n^j \vec{\mathbf{M}}_{e1n}^{(1)} + i A_n^j \vec{\mathbf{N}}_{o1n}^{(1)} + F_n^j \vec{\mathbf{M}}_{e1n}^{(3)} + i C_n^j \vec{\mathbf{N}}_{o1n}^{(3)} \right). \end{aligned} \quad (15.7)$$

In eqns (15.5–15.7)  $A_n$ ,  $B_n$ ,  $C_n$ ,  $F_n$ ,  $a_n^{(N+1)}$ , and  $b_n^{(N+1)}$  are the coefficients of expansion of VSH that need to be evaluated. The superscripts on 1 and 3 on VSH ( $\mathbf{M}$  and  $\mathbf{N}$ ) denote the usage of spherical Bessel and Hankel

functions, respectively, in their definition. The fields are expanded in appropriate VSH (Bessel or Hankel) based on the requirements of being mathematically well defined at certain locations. Specifically, the incident field has to be well defined and be zero at the origin. Similarly, the scattered field has to be finite at infinity. The internal fields on the other hand can be a combination of both inward and outward waves and accordingly both the Bessel functions of the first kind and Hankel functions are chosen to define them. The propagation constant in the  $j$ th layer appearing in eqns (15.7) is given by  $\kappa_j = 2\pi m_j/\lambda$ . The analytical expressions for VSH are given by:

$$\begin{aligned}\mathbf{M}_{o1n}(\kappa r, \theta, \phi) &= z_n(\kappa r)\pi_n(\cos \theta) \cos \phi \hat{e}_\theta - z_n(\kappa r)\tau_n(\cos \theta) \sin \phi \hat{e}_\phi, \\ \mathbf{M}_{e1n}(\kappa r, \theta, \phi) &= -z_n(\kappa r)\pi_n(\cos \theta) \sin \phi \hat{e}_\theta - z_n(\kappa r)\tau_n(\cos \theta) \cos \phi \hat{e}_\phi, \\ \mathbf{N}_{o1n}(\kappa r, \theta, \phi) &= n(n+1) \frac{z_n(\kappa r)}{\kappa r} \pi_n(\cos \theta) \sin \phi \hat{e}_r + \frac{d(\kappa r z_n(\kappa r))}{d(\kappa r)} \\ &\quad [\tau_n(\cos \theta) \sin \phi \hat{e}_\theta + \pi_n(\cos \theta) \cos \phi \hat{e}_\phi], \\ \mathbf{N}_{e1n}(\kappa r, \theta, \phi) &= n(n+1) \frac{z_n(\kappa r)}{\kappa r} \pi_n(\cos \theta) \cos \phi \hat{e}_r + \frac{d(\kappa r z_n(\kappa r))}{d(\kappa r)} \\ &\quad [\tau_n(\cos \theta) \cos \phi \hat{e}_\theta - \pi_n(\cos \theta) \sin \phi \hat{e}_\phi],\end{aligned}\quad (15.8)$$

where  $z_n$  are appropriate first-order Bessel functions ( $j_n$ ) or Hankel functions ( $h_n^{(1)}$ ). The angle-dependent functions are functions of associated Legendre polynomials and are given as:

$$\pi_n = \frac{P_n^{(1)}(\cos \theta)}{\sin \theta}, \quad \tau_n = \frac{d P_n^{(1)}(\cos \theta)}{d\theta}. \quad (15.9)$$

The coefficients of expansion are obtained by enforcing the continuity conditions at the interfaces. To accommodate the discontinuity between the  $j$ th and the  $(j+1)$ th layer we have:

$$\begin{aligned}(\mathbf{E}_{j+1} - \mathbf{E}_j) \times \hat{e}_r &= 0 \\ (\mathbf{H}_{j+1} - \mathbf{H}_j) \times \hat{e}_r &= 0.\end{aligned}\quad (15.10)$$

Similarly, the boundary conditions between the outermost layer and surrounding medium would lead to:

$$\begin{aligned}(\mathbf{E}_{sc} + \mathbf{E}_{inc} - \mathbf{E}_N) \times \hat{e}_r &= 0 \\ (\mathbf{H}_{sc} + \mathbf{H}_{inc} - \mathbf{H}_N) \times \hat{e}_r &= 0.\end{aligned}\quad (15.11)$$

The above equations can be reformulated using Riccati–Bessel functions (defined as  $\psi_n(\rho) = \rho j_n(\rho)$ ,  $\xi(\rho) = \rho h_n^{(1)}(\rho)$ ), logarithmic derivative functions ( $D_n^{(1)}(\rho) = \psi_n'(\rho)/\psi_n(\rho)$ ,  $D_n^{(3)}(\rho) = \xi_n'(\rho)/\xi_n(\rho)$ ) and ratio function ( $R_n(\rho) = \psi_n(\rho)/\xi_n(\rho)$ ). By defining:  $a_n^j = \frac{C_n^j}{A_n^j}$ ,  $b_n^j = \frac{F_n^j}{B_n^j}$ , a stable outward recursion could be worked out for coefficients  $a_n$  and  $b_n$ . In this formulation we start with the core, where  $a_n^1 = b_n^1 = 0$  and evaluate the rest of the coefficients  $a_n^j$  and  $b_n^j$  from the values  $a_n^{j-1}$  and  $b_n^{j-1}$ . Defining  $x$  as the size parameter given by  $x = 2\pi r/\lambda$ , for  $j = 2 \dots N$ , we have the following

equations:

$$\begin{aligned}
 U_n(m_{j-1}x_{j-1}) &= m_{j-1} \\
 &\quad \frac{R_n(m_{j-1}x_{j-1})D_n^{(1)}(m_{j-1}x_{j-1}) + a_n^{j-1}D_n^{(3)}(m_{j-1}x_{j-1})}{R_n(m_{j-1}x_{j-1}) + a_n^{j-1}}, \\
 V_n(m_{j-1}x_{j-1}) &= \frac{1}{m_{j-1}} \\
 &\quad \frac{R_n(m_{j-1}x_{j-1})D_n^{(1)}(m_{j-1}x_{j-1}) + b_n^{j-1}D_n^{(3)}(m_{j-1}x_{j-1})}{R_n(m_{j-1}x_{j-1}) + b_n^{j-1}}, \\
 a_n^j &= -R_n(m_jx_{j-1}) \frac{U_n(m_{j-1}x_{j-1}) - m_jD_n^{(1)}(m_jx_{j-1})}{U_n(m_{j-1}x_{j-1}) - m_jD_n^{(3)}(m_jx_{j-1})}, \\
 b_n^j &= -R_n(m_jx_{j-1}) \frac{m_jV_n(m_{j-1}x_{j-1}) - D_n^{(1)}(m_jx_{j-1})}{m_jV_n(m_{j-1}x_{j-1}) - D_n^{(3)}(m_jx_{j-1})}. \tag{15.12}
 \end{aligned}$$

The coefficients of the scattered field can be obtained as:

$$\begin{aligned}
 a_n^{N+1} &= -R_n(x_N) \frac{U_n(m_Nx_N) - D_n^{(1)}(x_N)}{U_n(m_Nx_N) - D_n^{(3)}(x_N)}, \\
 b_n^{N+1} &= -R_n(x_N) \frac{V_n(m_Nx_N) - D_n^{(1)}(x_N)}{V_n(m_Nx_N) - D_n^{(3)}(x_N)}. \tag{15.13}
 \end{aligned}$$

From this point, the remaining internal coefficients can be obtained by starting at the outer layer and using inward recursion. For the outer layer we have:

$$\begin{aligned}
 A_n^N &= m_N \frac{\xi_n(x_N)}{\xi_n(m_Nx_N)} \left( \frac{R_n(x_N) + a_n^{N+1}}{R_n(m_Nx_N) + a_n^N} \right), \\
 B_n^N &= \frac{\xi_n(x_N)}{\xi_n(m_Nx_N)} \left( \frac{R_n(x_N) + b_n^{N+1}}{R_n(m_Nx_N) + b_n^N} \right). \tag{15.14}
 \end{aligned}$$

For the inner layers  $j = N \dots 2$  we have:

$$\begin{aligned}
 A_n^{j-1} &= A_n^j \frac{m_{j-1}}{m_j} \frac{\xi_n(m_jx_{j-1})}{\xi_n(m_{j-1}x_{j-1})} \left( \frac{R_n(m_jx_{j-1}) + a_n^j}{R_n(m_{j-1}x_{j-1}) + a_n^{j-1}} \right), \\
 B_n^{j-1} &= B_n^j \frac{\xi_n(m_jx_{j-1})}{\xi_n(m_{j-1}x_{j-1})} \left( \frac{R_n(m_jx_{j-1}) + b_n^j}{R_n(m_{j-1}x_{j-1}) + b_n^{j-1}} \right). \tag{15.15}
 \end{aligned}$$

By using an appropriate cutoff for the number of VSH used (Wiscombe 1980) and stable recursive evaluation of the logarithmic derivative and ratio functions involved, an accurate mathematical solution for the EM fields can be obtained. The far-fields are dependent on the coefficients of scattered field and can be evaluated using the asymptotic expansions of the scattered fields. The

scattering and extinction efficiencies in the far-field are obtained using:

$$\begin{aligned} Q_{\text{ext}} &= \frac{2}{\kappa r_t^2} \operatorname{Re} \left\{ \sum_{n=1}^{\infty} (2n+1) \left( a_n^{(N+1)} + b_n^{(N+1)} \right) \right\}, \\ Q_{\text{sca}} &= \frac{2}{\kappa r_t^2} \left\{ \sum_{n=1}^{\infty} (2n+1) \left( |a_n^{(N+1)}|^2 + |b_n^{(N+1)}|^2 \right) \right\}, \end{aligned} \quad (15.16)$$

where  $r_t$  is the total size of the particle. The results we discuss here only correspond to coinage metals: gold, silver and copper. Their bulk optical constants are well reported in the literature (Johnson *et al.* 1972).

The complex refractive index  $n$  of a material is related to its dielectric constant  $\varepsilon$  by:

$$\operatorname{Re}(\varepsilon) = \operatorname{Re}(n)^2 - \operatorname{Im}(n)^2; \operatorname{Im}(\varepsilon) = 2\operatorname{Re}(n)\operatorname{Im}(n). \quad (15.17)$$

For particles smaller and metal shells thinner than the bulk electron mean-free path, the size-dependent dielectric function  $\varepsilon_s$  is related to the bulk experimental dielectric function  $\varepsilon_b$  by:

$$\varepsilon_s(r, \omega) = \varepsilon_b(\omega) + \frac{\omega_p^2}{\omega^2 + i\omega\gamma_b} - \frac{\omega_p^2}{\omega^2 + i\omega\Gamma}, \quad (15.18)$$

where  $\gamma_b$  is the bulk colloisonal frequency and  $r$  is the reduced electron mean-free path (Averitt *et al.* 1997).  $\Gamma$  is the modified collisional frequency given by:

$$\Gamma = \gamma_b + A \frac{v_F}{r}, \quad (15.19)$$

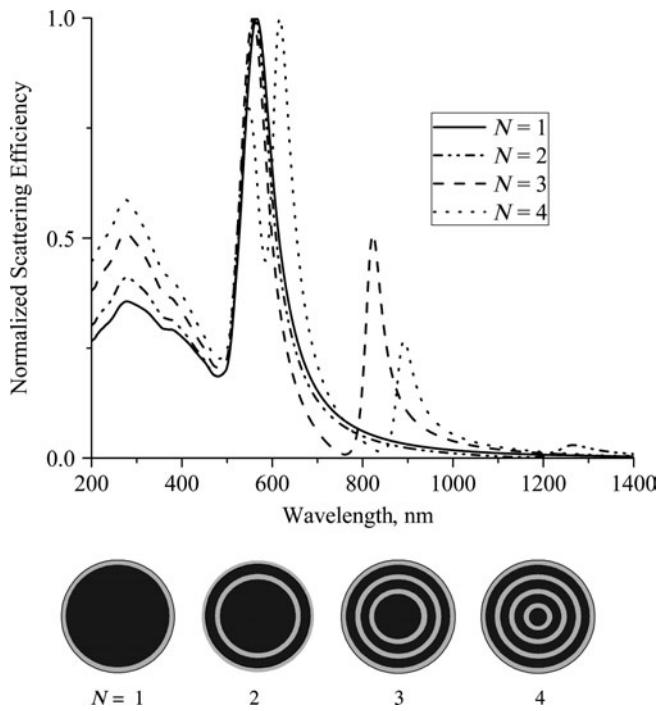
where  $v_F$  is the Fermi velocity  $A$  is a parameter dependent on geometry that is assumed as 1 for the case of isotropic scattering. For the results presented here we use a corrected refractive index for metal shells thinner than 5 nm using thickness of the shell as the reduced mean-free path.

### 15.3.2 Optical tunability

Using calculations discussed in the previous section, we can evaluate far-field and near-field properties of nano-LAMPs. Their optical tunability is with respect to different parameters: size (and number of layers or a combination), refractive indices of metal and dielectric. Before considering their near-field properties that are directly relevant to SEVS enhancement, let us consider the optical tunability of their extinction and scattering efficiency. Such effects are relevant in their applicability for tissue-imaging modalities based on elastic light scattering, low coherence tomography and light-scattering spectroscopy. In these applications, the dielectric layers within are designed to tune peak positions of scattering, while the metal layers determine the scattering.

#### 15.3.2.1 Far-field

We first consider the effects on scattering efficiencies for nano-LAMPs of a given total size and of a different numbers of layers, in visible and NIR excitation. Here, each layer of a nano-LAMP is considered to be synonymous



**Fig. 15.2** Scattering efficiency tunability with number of layers when total size is constant.

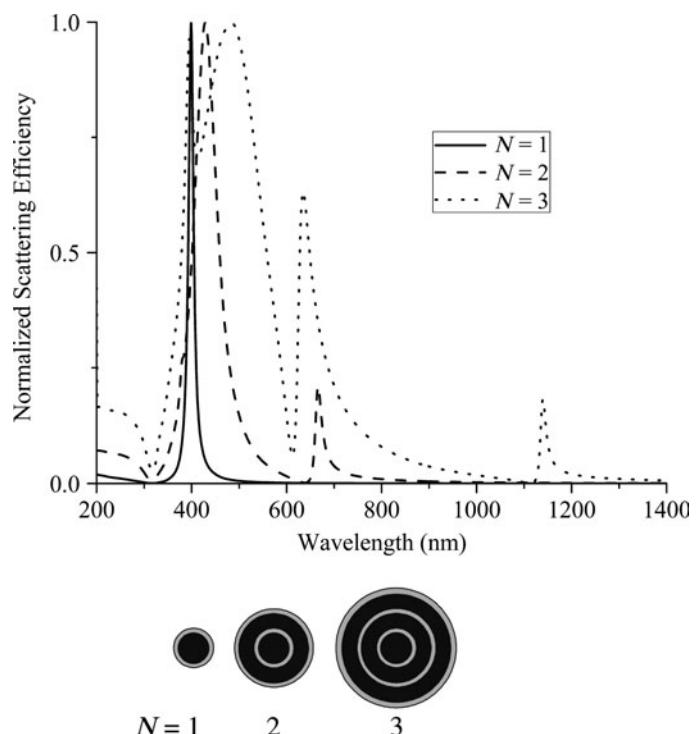
with one metal layer and one dielectric layer together. For example,  $N = 1$  corresponds to a metal core–dielectric shell particle and  $N = 2$  corresponds to this core–shell particle covered with a metal shell on the top of which a dielectric shell is coated. The metal layer is considered to be made of gold and the dielectric of silica, the refractive index of which is considered to be 1.45. The total size of nano-LAMPs is fixed to be 60 nm and the number of layers  $N$  is varied from 1 to 4. Further, each metal shell is of thickness 10 nm, whereas the thickness of each dielectric layer is 5 nm.

The size of the metal core is chosen according to the number of layers. In keeping with the theme of biomedical applications, these nano-LAMPs are assumed to be embedded in a water-like host medium whose refractive index is chosen to be 1.33. The normalized scattering efficiencies are plotted with excitation wavelength in Fig. 15.2.

In the range of excitation wavelengths considered, for  $N = 1$ , there are two dominant resonances in the spectral distribution of normalized scattering efficiency. The incorporation of an additional metal shell and dielectric shell, when  $N$  is increased to 2, shifts the existing resonances very slightly to the left. Also, the relative proportions of the resonances slightly change and a new smaller peak appears on the right. With further addition of metal and dielectric shells ( $N = 3$  and 4), new resonances appear and the existing resonances shift. This is in accordance with the existing understanding that, in alternating metal–dielectric multilayered spheres, the plasmonic coupling of metal shells results in hybridized multiple extinction resonances. The coupling strength, and in turn resonances, further depend on the extent and properties of the spacer dielectric (Sinzig *et al.* 1994; Prodan *et al.* 2004).

Such effects are pre-dominant in layered materials with metals showing narrow plasmon resonances and dielectrics with large refractive index in the chosen spectral region. In the systems considered here, which start with a gold core and shelled with alternate silica and gold layers, these shifting and splitting effects are perhaps not as obvious as in multilayered spheres considered by Sinzig *et al.* (1994). Moreover, lower-energy plasmon peaks might have superposed each other and are hence not spectrally resolved. Nevertheless, it is clear that changing the number of layers provides an opportunity to design layered probes for appropriate scattering resonance peaks.

Next, we consider a similar calculation on nano-LAMPs made of silica and silver, but with the core size kept constant at 20 nm. The number of layers  $N$  is varied from 1 to 3; the sizes of silver layers are fixed at 20 nm, the size of the dielectric layer is fixed at 5 nm and the embedding medium is chosen to be aqueous. The normalized scattering efficiency spectra obtained are plotted in Fig. 15.3. The effects due to increasing number of layers as well as size as observed in Fig. 15.3 are physically similar to those in Fig. 15.2. The peaks are much sharper for these nano-LAMPs, due to sharper plasmon resonances in the case of silver. Upon adding silver and silica shells to the initial  $N = 1$  nano-LAMP configuration, the sharp resonance peak broadens and a very slight splitting effect can be noticed. Also, a new resonance peak occurs on the right. Adding another set of silver and silica shells ( $N = 3$ ), further splits the left peak, shifts the right peak and adds a new peak on the

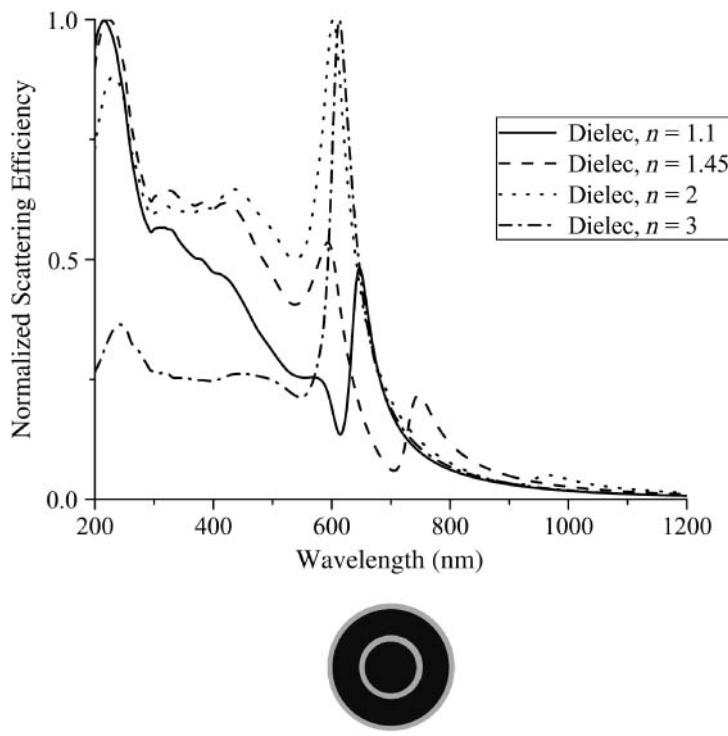


**Fig. 15.3** Scattering efficiency tunability with number of layers when core size is constant.

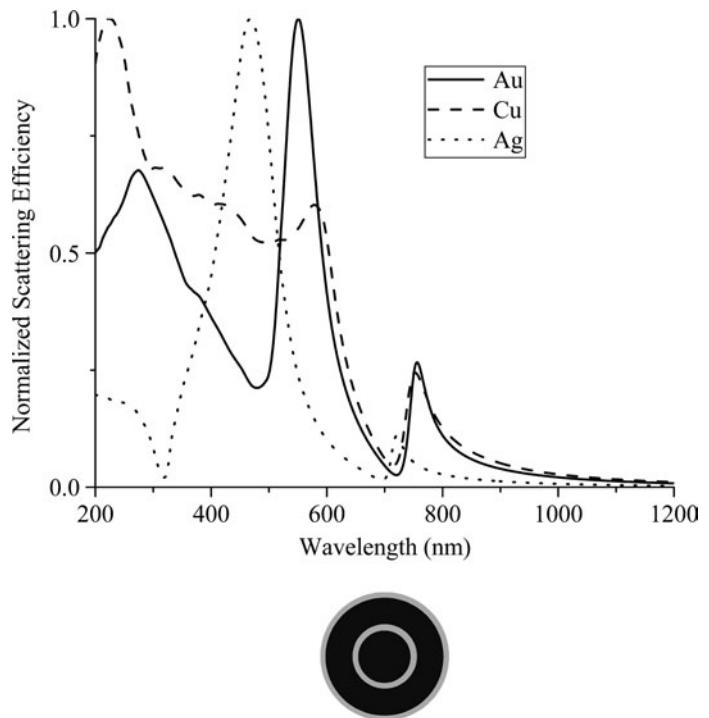
far right. Also here, in contrast to the previous example, the new plasmon peaks and shifting of the existing ones are demonstrated better. The total sizes and number of layers that could be accommodated are both dependent on experimental feasibility and size criteria. Based on such fabrication constraints, the geometrical configuration of nano-LAMPs can then be optimized for a given excitation frequency.

The other important parameter that determines the location, extent and height of far-field resonance peaks in nano-LAMPs, is the chemical composition. To demonstrate this, consider the scattering efficiency spectral distribution of nano-LAMPs of varying dielectric spacer in a fixed geometrical configuration of  $N = 2$  (with Cu core 20 nm, dielectric 5 nm, Cu shell 20 nm, dielectric 5 nm).

The dielectrics chosen are of refractive indices 1.1, 1.45, 2 and 3. An index of 1.1 represents a layer similar to biomolecules (but a separation of 5 nm is perhaps not practically achievable with biomolecular linkers along with producing a continuous shell of nanoparticles), 1.45 represents silica, while high refractive indices 2 and 3 are chosen as hypothetical values for comparison. The embedding host medium is again assumed to be water. From the scattering efficiency spectra plotted in Fig. 15.4, it is clear that the relative dominance of the resonances in the given excitation region vary with the dielectric. More importantly, the resonance peaks for  $n = 1.1$  and 1.45 are of interest because most biopolymers have refractive index values between 1 and 1.5. Also, the clear distinction between some of the peaks is relevant in creating multicolor probes.



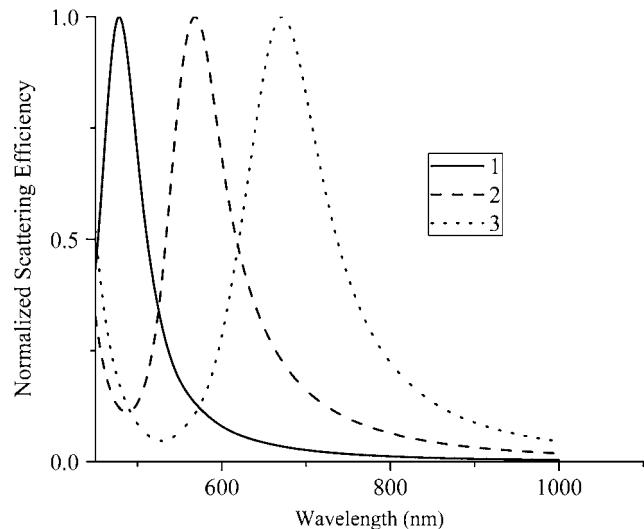
**Fig. 15.4** Scattering efficiency tunability in a fixed geometry with changing dielectric.



**Fig. 15.5** Scattering efficiency of bimetallic layered particles with same core, but different shells.

Next, we consider the same geometry with  $N = 2$ , but change the chemical composition of metal shell. The core of the particle is fixed to be Au and the metal shell is varied between Au, Cu and Ag. The dielectric layers are fixed to be silica and the embedding host medium is water. The scattering efficiencies for this case are presented in Fig. 15.5.

From Fig. 15.5, it can be observed that the scattering characteristics for a Cu shell are closer to those a Au shell. The greater differences between plasmon characteristics of Ag and Au than those between Au and Cu might result in such effects. The above-discussed tunability can be further used to design multicolor labels in the visible and NIR ranges of excitation. The controllability of the spectral positions of extinction resonances can be used in designing dielectric core metal shell particles (nanoshells) with different core/shell ratios (Jackson *et al.* 2001). Three layered particles with metal core–dielectric shell–metal shell, consisting of thin metal shells have been proposed to provide ultrasharp resonances (Chen *et al.* 2005). In this study the size effects on dielectric functions of metal layers thinner than the bulk electron mean-free path were not considered. In actuality, this dependence on size results in much wider resonances (Khlebtsov 2006). A representative example of the scattering efficiencies calculated using corrected refractive indices for probes proposed by these authors is provided in Fig. 15.6. The three configurations in the figure can be used as multicolor labels in the visible range. The caption for Fig. 15.6 provides the details of layer thicknesses in each configuration. The total size



**Fig. 15.6** Multicolor labels in the visible range. Size 10 nm; Geometry: Ag core, Silica shell, Ag shell. 1: 2.5 nm, 4 nm, 3.5 nm; 2: 4.25 nm, 2.75 nm, 3 nm; 3: 5 nm, 3 nm, 2 nm.

for these probes is kept constant to avoid diffusion disparities and hence the ability to bind to receptor sites.

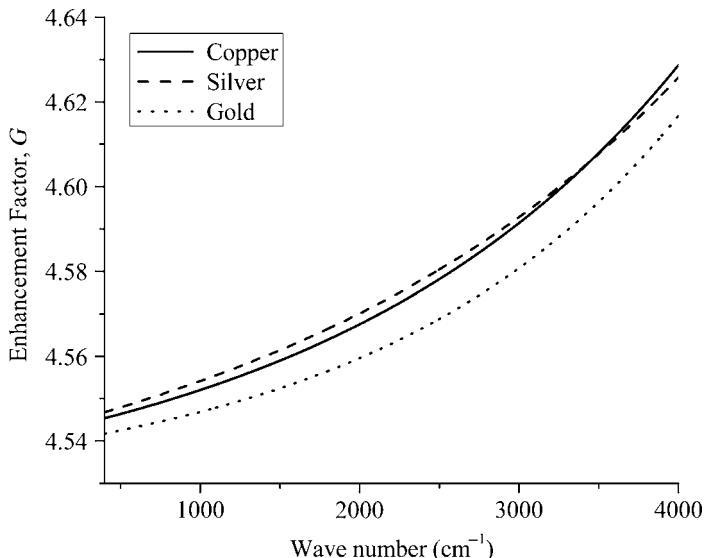
Such probes of multiplexing capability can be obtained, but the number of distinct probes that can be created is going to be limited by the widths of the resonance peaks; this is also the case for nanoshells. Nano-LAMPs with a higher number of layers, however, can provide a rich absorption spectral response. Their full distribution responses in the excitation regions might be used along with multivariate curve analysis for multiplexed detection of a greater number of nano-LAMPs. Although an unlimited multiplexing might not be possible for detecting elastic scattering of nano-LAMPs, the results above show that once they have been designed for optimum near-fields, their far-field properties can be used in multimodal techniques. The specificity of detecting Raman/IR responses of the dielectric layer is much higher due to the characteristics of vibrational transitions. Such a sensing mechanism for nano-LAMPs would, in theory, provide an opportunity for multiplexing an unlimited number of targets.

### 15.3.2.2 Near-field

The optical near-fields within nano-LAMPs play a very crucial part in designing optimized Raman/IR enhancement. Let us, for example, consider SEIRA for the molecules in the reporting layers of probes. For high-sensitivity applications the maximum probe size allowed is restricted to be less than 100 nm. In the case of mid-IR excitation, the nanoscale probes usually are under the small-particle limit and any size effects are negligible. Using the small-particle limit the surface enhancement in SEIRA for a molecule on a nanosphere is given by:

$$G = \left\{ |1 - g|^2 + \frac{3}{2} \left( 2\text{Re}(g) + |g|^2 \right) \right\}, \quad (15.20)$$

where  $g = (\varepsilon - \varepsilon_0)/(\varepsilon + 2\varepsilon_0)$ , in which  $\varepsilon$  and  $\varepsilon_0$  are the dielectric constants of the metal particle and surrounding medium, respectively. SEIRA enhancement

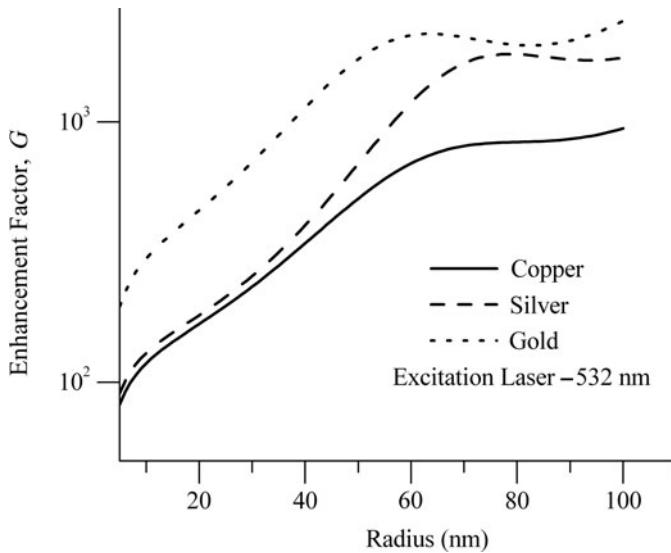


**Fig. 15.7** SEIRA enhancement on nanospheres of Au, Ag and Cu.

factors on single nanospheres of Au, Ag and Cu evaluated using eqn (15.20) are depicted in Fig. 15.7. From the figure, it can be observed that the differences in enhancement factors of SEIRA for the three different metals are minor. Also, the enhancement is not going to be benefited much by a careful design of layer thicknesses and size. So, we focus on optimizing Raman enhancement by studying near-fields with nano-LAMPs made of different chemical compositions and topology. Nevertheless, the principles remain the same for SEIRA and these calculations can be extended to SEIRA, if necessary.

The electromagnetic contribution to the Raman enhancement factor given in eqn (15.3) depends on enhanced intensities at both laser and Stokes frequencies. However, it is common practice to evaluate one of the intensities and approximate the EM contribution to be of the order of its square (Schatz *et al.* 2002). If we consider only the order of magnitude of the signal from the probe then it suffices to evaluate the maximum value of the field intensity within the probe. However, it is impossible to generate multilayered structures with uniform surface coverage for each shell. So sometimes, a more conservative approach would be to evaluate the order of surface average of intensities. Such a design criterion would ensure the maximization of the extent of intense fields called “hot-spots”. Here, we use both these approaches and mention the approach used when referring to a particular result. Another important aspect in the applicability of probes based on enhancement in interstices of metal features is the dependence on polarization of the incident beam due to orientation variability. Due to the spherical configuration of nano-LAMPs, such a dependence on polarization is absent.

Figure 15.8 depicts the Raman enhancement calculated from the maximum intensity on the surface of Au, Ag and Cu nanospheres of different sizes at an excitation wavelength of 532 nm. It can be observed that along with depending on plasmon characteristics of the metal, the enhancement is dependent on

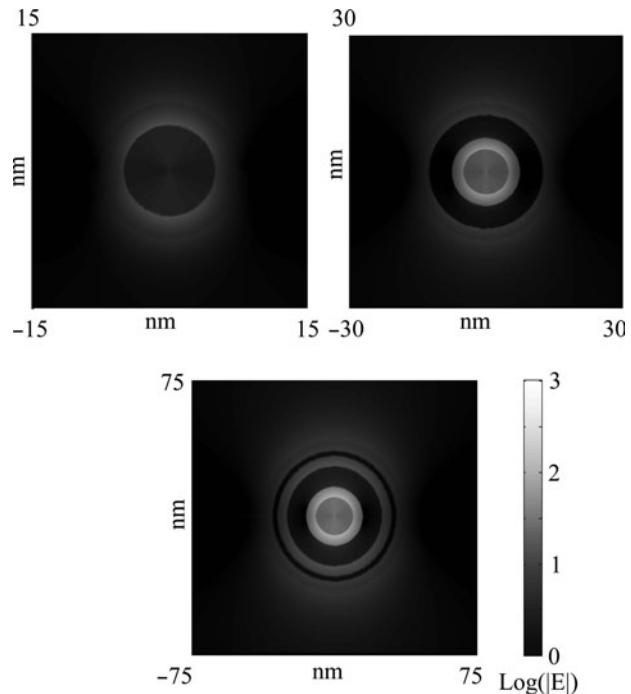


**Fig. 15.8** Dependence of Raman enhancement on size of metal sphere.

size. We will use this tunability further, to design Raman probes for a given excitation.

By plotting the near-fields of different configurations, the effect of geometrical parameters like relative sizes of the shells can be observed. Probing into near-fields of nano-LAMPs also provides an opportunity to understand the extent of “hot-spots” in their reporter dielectric layers. In such multilayered configurations, the outer layers seem to focus the field onto the surface of the core (Xu 2005). Figure 15.9 provides a plot of the logarithm of magnitudes of electric fields within a representative nano-LAMP made of gold layers and a dielectric of refractive index 1.1. The radii of the layers starting from the metal core are 20 nm, 21 nm, 26 nm, 27 nm, 32 nm, 33 nm, 53 nm, 54 nm. The excitation wavelength is taken to be 532 nm and the nano-LAMPs are considered to be embedded in vacuum. The polarization of the incident field and its propagation with respect to each plane are represented at the bottom of each subfigure. The subfigure at the bottom right represents the field at a spherical surface 0.5 nm above the core.

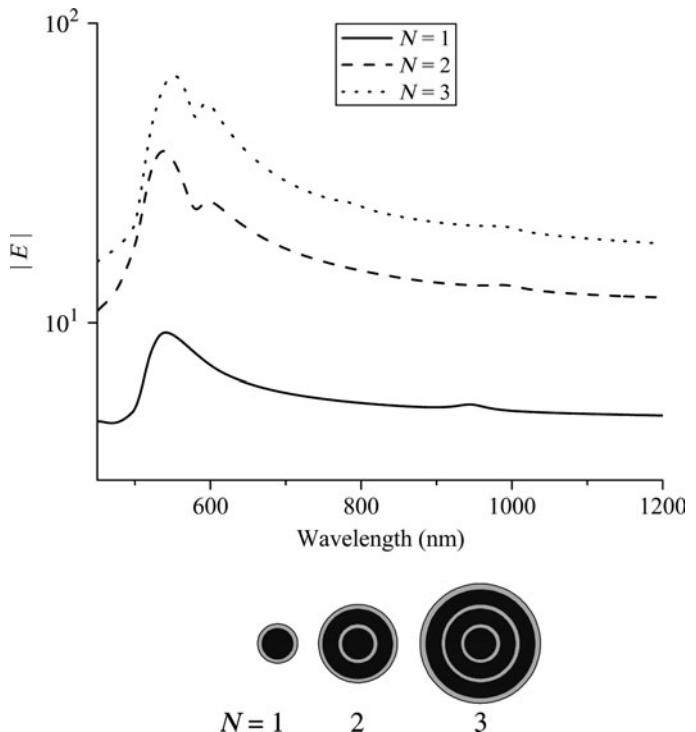
E-field magnitudes of  $\sim 10^2$ – $10^3$  are observed at the surface of the core, as shown in the figure, and this translates into a Raman enhancement factor of order  $\sim 10^8$ – $10^{12}$ . Moreover, it can be observed that all dielectric layers are bright, representing enhanced E-fields. Along with these “hot-spots”, dark spots or regions of no enhancement are also observed mostly in the metal regions. These regions of no enhancement also exist in dielectric layers; although they are not obvious in the representations chosen here. Mathematically, these can be interpreted to be due to cancellation of higher-order spherical harmonics and physically this is due to the retardation effects of higher-order multipoles. Hence, we needed to consider such higher-order effects in order to completely define enhancement and EM scattering solutions based on electrostatics would not be sufficient to design hot-spots.



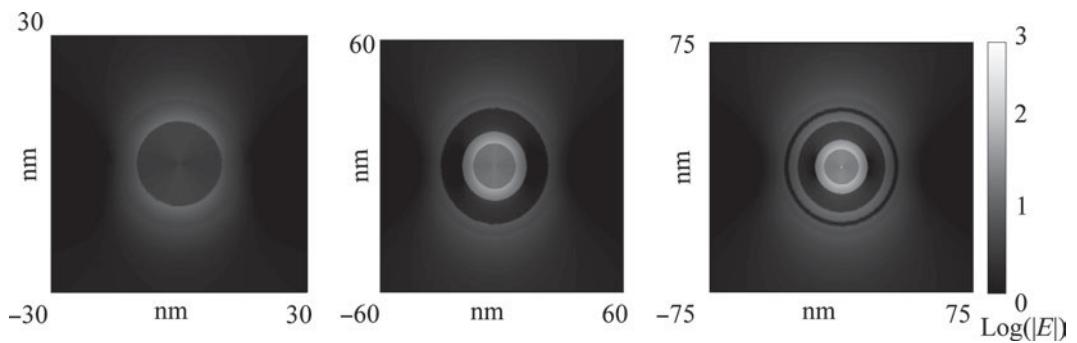
**Fig. 15.9** Enhanced electromagnetic fields of a multilayered sphere in different planes.

The enhancement predicted for nano-LAMPs under appropriate configurations are huge compared to what is achievable by single nanoparticles or nanoshells. Their near-field tunability can be demonstrated by considering the E-field magnitude near the core of particles with different numbers of layers and sizes but with core size kept constant. For the results here, a gold core of 10 nm is used and the refractive index of dielectric shells is considered to be 1.1. The geometrical configurations considered are given by:  $N = 1$ : 20 nm Au core + 5 nm dielectric shell;  $N = 2$ : 10 nm Au core + 5 nm dielectric shell + 10 nm Au shell + 5 nm dielectric shell;  $N = 3$ : 10 nm Au core + 5 nm dielectric shell + 10 nm Au shell + 5 nm dielectric shell + 2.5 nm Au + 5 nm dielectric. Figure 15.10 represents the spectral distribution of the E-field magnitude at a point 2 nm above the core in the direction of the incident field polarization. From the figure, it can be observed that as the number of layers increases the maximum E-field magnitude increases. This amplification is due to the additional metal layers interacting with the plasmon resonances of the core. The peak positions are located at 537, 542 and 552 nm, respectively. The shift towards lower energies is due to additional layers and spacing between the layers. This shows that by adding more layers of appropriate size one can keep the shift in the near-field resonances low, while increasing the maximum E-field possible.

Figure 15.11 presents the log values of E-fields in a plane at the symmetric section of these configurations. The propagation of the E-field in this case is into the paper and it is polarized in the upward direction. It can be noticed

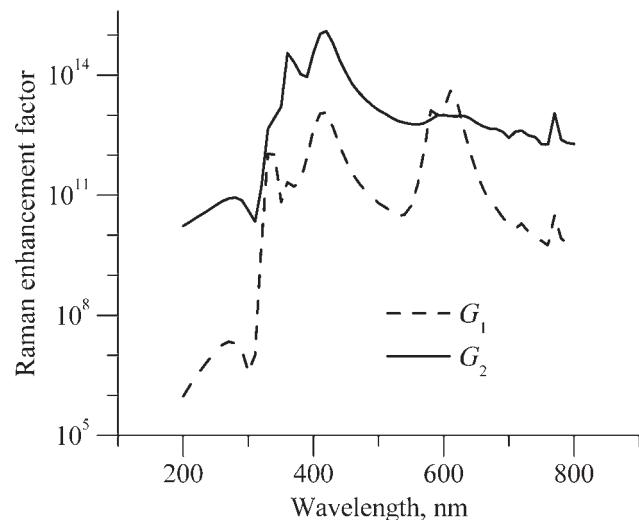


**Fig. 15.10** Near-field tunability of E-field magnitude of particles with constant core.



**Fig. 15.11** E-field distribution for nanostructured particles considered in Fig. 15.10 (from Kodali *et al.* 2008, with permission from SPIE).

from Fig. 15.11 that the E-field at the surface of the core increases with adding layers on the top of the initial nano-LAMP configuration. Further, it can be observed that some of the enhanced regions extend into metal layers and with proper selection of sizes and properties of different layers the probes can be designed to restrict and maximize the enhancement in dielectric layers.



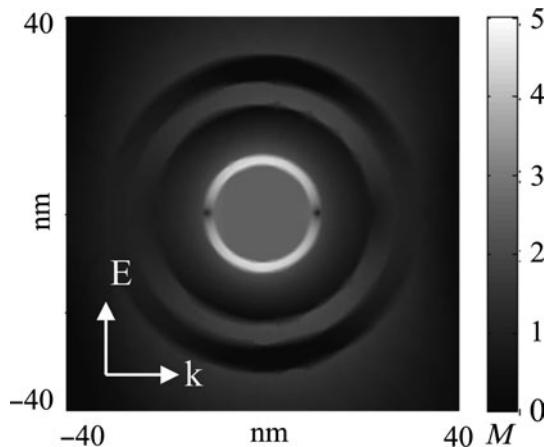
**Fig. 15.12** Multiple excitation capability of nano-LAMPs;  $G_1$  and  $G_2$  correspond to enhancement factors in dielectric layers 1 and 2 (from Kodali *et al.* 2008, with permission from SPIE).

## 15.4 Design

### 15.4.1 Design example

Using the properties discussed above, nano-LAMPs can be designed for maximum signals at given excitation laser characteristics. Further, the enhancement of the fields in each dielectric layer can be specifically designed for certain excitation frequencies. To demonstrate this, in Fig. 15.12, the enhancement of each dielectric layer for a probe with  $N = 2$  is presented. The metal layers considered are of silver, the dielectric is assumed to be of silica and the particle is assumed to be in vacuum. In this case, the enhancement at the Stokes frequency is considered by using a Raman shift of about 1010 cm<sup>-1</sup>. The geometry of configuration is defined as: 10 nm Au core + 2 nm silica shell + 10 nm Au shell + 5 nm silica shell + 5 nm Au shell. The probes are assumed to be embedded in vacuum.

Here, the enhancement factor is evaluated as a volume integral of the dielectric shell and represents the contribution of each dielectric layer to the total signal of the probe. Both layers exhibit resonant Raman responses at an excitation of about 470 nm, but layer 1 has a higher order of enhancement at an excitation of 664 nm. Due to the contribution to the signal from each layer being an integral of the values over the entire volume,  $G_2$  is usually higher than  $G_1$ . The spectral variation of the layers is slightly different but the total signal of the probe is high at multiple excitations throughout the distribution. In Fig. 15.13, the logarithm values of intensity ( $M$ ) distribution of this probe at laser frequencies is shown in the incident plane. It can be observed that the ‘hot-spot’ at the core in this case fully includes its surface except for the points on the equator. The outer dielectric is also completely enhanced at this excitation frequency. In this figure, the regions with retarded fields can be observed to a small extent in the outer dielectric layer.



**Fig. 15.13** E-field distribution for the probe considered in Fig. 15.11 at an excitation of 664 nm;  $M$  values are plotted on a log scale.

Using the above properties of layered particles one can optimize nano-LAMP structures subject to experimental constraints. An optimization algorithm can be used to devise a configuration with either maximum enhanced intensity or volume integral of enhanced intensity at given laser properties. The constraints that will be incorporated into such a design problem would be the fabrication and experimental constraints. The fabrication constraints limit the size of layer and chemical composition, while the experimental constraints would include the applicability and diffusion characteristics *in vivo* for nano-LAMPs of different sizes. In addition to designing the metal and dielectric layers, one needs to consider the effect of the protective layer and antibody layer (both act as dielectrics) over the Raman response of nano-LAMPs. The protective glass layer should be as thin as possible but with optimal shielding characteristics, whereas the antibody layer should be as thin as possible but with optimal binding efficiencies.

### 15.4.2 Fabrication and applicability

Here, we describe several possibilities of bottom-up fabrication procedures in creating nano-LAMPs and provide brief pointers to probable issues in their reliability. In these approaches, nano-LAMPs are built layer-by-layer from the core. The first approach is to start with a colloidal core and use seeding methods to add subsequent shells. For gold and silver nano-LAMPs the particle cores can be coated with silica by using silane linkers. The seed can be further grown using the *Stöber* method. Small colloidal Au/Ag particles can be attached again using silane linkers to silica shells. This attached colloid further acts as nucleation sites for further reduction of gold ions onto the surface. Synthesizing thin shells requires using carefully designed reaction conditions and reagents. The second approach is to self-assemble colloidal metal particles using biomolecular linkers like oligonucleotides, avidin, etc. In this technique, metal core particles and metal/dielectric shell particles are functionalized with two different molecules. They are linked by biomolecules partly complementary to each of these molecules and hence act as binding

agents. By using appropriate ratios of concentrations in this approach it is possible to generate core-shell structures. For a further control, this type of self-assembly can be guided with a substrate support and the probes can be detached from the surface and dispersed into a solution after their assembly. In addition to these approaches, nanometer-size platinum core-silica shell particles can be prepared by using a reverse micelle technique along with hydrolysis and condensation (Bae *et al.* 2002). Sol-gel techniques followed by reduction can also be used to create composite particles made of copper.

For nano-LAMP-based multiplexed detection, a quantitative expression obtained has to consider the effects of fabrication as well. For example, probes based on colloidal nanoparticles might possess a background signal due to enhancement of fluorescent signal of one of the stabilizing chemicals. Also, in a tissue-imaging modality a proper assessment on the binding kinetics of the antibodies on the probe and diffusion characteristics of the probe is needed. Apart from these, there might be issues pertaining to the theoretical model itself. A spherical continuous layered structure is far from what could be feasible experimentally. So the theoretical results obtained might not completely agree with validation experiments. SEVS, more specifically SERS, is predominantly a phenomenon of discontinuities. So it should be emphasized that the presence of such discontinuities, if not always, should at least aid in achieving greater enhancements.

Another important factor could be the effect of the protective layer and antibody layer in the outermost layers. These layers should only result in a slight shift of the resonance peaks, although such results were not presented here. While proposing theoretical configurations for the probes care should be exercised in determining whether the thickness of the shells proposed is experimentally feasible and precisely controllable. It is also worth mentioning that without considering appropriately sized dielectric functions for thin metal shells researchers might be misled into assuming very sharp resonances for such configurations.

## 15.5 Conclusion

Multilayered metal spheres offer a paradigm for creating nanoscale probes for multiplexed detection using surface-enhanced vibrational spectroscopy, especially for SERS. They also provide a ready template and wide tunability due to the diversity possible in the layered construct. For example, increasing the number of layers on core-shell particles amplifies the near-fields within the particle. Such characteristics of the near-field and the optical tunability of layered metal particles can be used in creating probes for specific applications. The enhancement in each dielectric/reporting layer can be designed to have its own spectral characteristic and the effective signal of a probe could always be high, providing the capability of responding to multiple excitations. Subject to experimental feasibility, size constraints and laser characteristics, the geometrical and chemical configuration of nano-LAMPs can be optimized for maximal SEVS responses.

## References

- Alabano, E.V.S., Daiser, S., Ertl, G., Mironda, R., Wandelt, K. *Phys. Rev. Lett.* **51**, 2314 (1983).
- Albrecht, M.G., Creighton, J.A. *J. Am. Chem. Soc.* **99**, 5215 (1977).
- Alivisatos, A.P., Johnsson, K.P., Peng, X., Wilson, T.E., Loweth, C.J., Bruchez, Jr., M.P., Schultz, P.G. *Nature* **382**, 609 (1996).
- Ansari, D.O., Stuart, D.A., Nie, S. *Proc. SPIE* (5699) (2005).
- Aravind, P.K., Nitzan, A., Metiu, H. *Surf. Sci.* **110**, 189 (1981).
- Aroca, R.F., Ross, D. *J. Appl. Spec.* **58**, 324 (2004).
- Aroca, R. *Surface-enhanced Vibrational Spectroscopy* (John Wiley & Sons Ltd Chichester, UK, 2006).
- Averitt, R.D., Sarkar, D., Halas, N.J. *Phys. Rev. Lett.* **78**, 4217 (1997).
- Avouris, Ph., Demuth, J.E. *J. Chem. Phys.* **75**, 4783 (1981).
- Bae, D.S., Han, K.S., Adair, J.H. *J. Am. Chem. Soc.* **85**, 1321 (2002).
- Barber, P.W., Hill, S.C. *Light Scattering by Particles: Computational Methods* (World Scientific Publishing Co. Singapore, 6, 1990).
- Baumberg, J.J., Kelf, T.A., Sugawara, Y., Cintra, S., Abdelsalam, M.E., Bartlett, P.N., Russell, A.E. *Nano Lett.* **5**, 2262 (2005).
- Cao, Y.W.C., Jin, R., Mirkin, C.A. *Science* **297**, 1536 (2002).
- Chen, K., Liu, Y., Ameer, G., Backman, V. *J. Biomed. Opt.* **10**, 024005 (2005).
- Creighton, J.A., Blatchford, C.G., Albrecht, M.G. *J. Am. Chem. Chem. Soc. Faraday Trans. II* **75**, 790 (1979).
- Deckman, H.W., Dunsmuir, J.H. *Appl. Phys. Lett.* **41**, 377 (1982).
- Doering, W.E., Nie, S. *Anal. Chem.* **75**, 6171 (2003).
- Dou, X.M., Ozaki, Y. *Rev. Anal. Chem.* **18**, 285 (1999).
- Efrima, S., Metiu, H. *J. Chem. Phys.* **15**, 1602 (1978).
- Fleischman, M., Hendra, P.J., McQuillan, A.J. *Chem. Phys. Lett.* **26**, 123 (1974).
- Gersten, J., Nitzan, A. *J. Chem. Phys.* **73**, 3023 (1980).
- Grow, A.E., Wood, L.L., Claycomb, J.L., Thompson, P.A. *J. Microbio. Meth.* **53**, 221 (2003).
- Hartstein, A., Kirtley, J.R., Tsang, J.C. *Phys. Rev. Lett.* **45**, 201 (1980).
- Hatta, A., Oshima, T., Suetaka, W. *Appl. Phys. A* **29**, 71 (1982).
- Hatta, A., Suzuki, Y., Suetaka, W. *Appl. Phys.* **135**, A35 (1984).
- Hatta, A., Chiba, Y., Suetaka, W. *Surf. Sci.* **158**, 616 (1985).
- Hatta, A., Suzuki, N., Suzuki, Y., Suetaka, W. *Appl. Surf. Sci.* **37**, 299 (1989).
- Hering, K., Cialla, D., Ackermann, K., Dorfer, T., Moller, R., Schneidewind, H., Mattheis, R., Fritsche, W., Rosch, P., Popp, J. *Anal. Bioanal. Chem.* **390**, 113 (2008).
- Hildebrandt, P., Stockburger, M. *J. Phys. Chem.* **88**, 5935 (1984).
- Hulteen, J.C., Van Duyne, R.P. *J. Vac. Sci. Tech.* **13**, 1553 (1995).
- Jackson, J.B., Halas, N.J. *Proc. Nat. Acad. Sci.* **101**, 17930 (2004).
- Jackson, J.B., Halas, N.J. *J. Phys. Chem. B* **105**, (2001).
- Jeanmarie, D.L., Duyne, R.P.V. *J. Electroanal. Chem.* **84**, 1 (1977).

- Jensen, T.R., Duval, M.L., Kelly, K.L., Lazarides, A.A., Schatz Van Duyne, R.P. *J. Chem. Phys.* **103**, 9846 (1999).
- Johnson, E., Aroca, R. *J. Phys. Chem.* **99**, 9325 (1995).
- Kamata, T., Kato, A., Umemura, J., Takenaka, T. *Langmuir* **3**, 1150 (1987).
- Kang, S.Y., Jeon, C., Kim, K. *Appl. Spectrosc.* **52**, 278 (1998).
- Kerker, M., Wang, D.S., Chew, H. *Appl. Opt.* **19**, 4159 *J. Chem. Phys.* **73**, 3023 (1980).
- Khlebtsov, B., Khlebtsov, N. *J. Biomed. Opt.* **11**, 044002 (2006).
- Kneipp, K., Wang, Y., Dasarai, R.R., Feld, M.S. *Appl. Opt.* **49**, 780 (1995).
- Kneipp, K., Wang, Y., Kneipp, H., Perelman, L.T., Itzkan, I., Dasari, R.R., Feld, M.S. *Phys. Rev. Lett.* **78**, 1667 (1997).
- Kneipp, K., Kneipp, H., Moskovits, *Surface-enhanced Raman Scattering: Physics and Applications* (Springer-Verlag, Heidelberg, 2006).
- Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R.R., Feld, M.S. *J. Phys. Condens. Matter.* **14**, R597 (2002).
- Kunz, K.S., Luebers, R.J. *The Finite Difference Time Domain Method for Electromagnetics* (CRC Press LLC, Boca Raton, 1993).
- Liao, P.F., Bergman, J.G., Chemla, D.S., Wokaun, A., Melangailis, J., Hawryluk, M., Economou, P. *Chem. Phys. Lett.* **82**, 355 (1981).
- Liver, N., Nitzan, A., Gersten, J.I. *Chem. Phys. Lett.* **111**, 449 (1984).
- Metiu, H., Das, P. *Ann. Rev. Phys. Chem.* **35**, 507 (1984).
- Michaels, A.M., Nirmal, M., Brus, L.E. *J. Am. Chem. Soc.* **121**, 9932 (1999).
- Mirkin, C.A., Lestinger, R.L., Mucic, R.C., Storhoff, J.J. *Nature* **382**, 607 (1996).
- Moskovits, M. *J. Chem. Phys.* **69**, 4159 (1978).
- Moskovits, M. *Rev. Mod. Phys.* **57**, 783 (1985).
- Mulvaney, S.P., Musick, M.D., Keating, C.D., Natan, M.J. *Langmuir* **19**, 4784 (2003).
- Muraki, H., Ito, T., Hiramatsu, M. *Thin Solid Films* **179**, 509 (1989).
- Murray, C.A., Allara, D.L., Rhinewine, M. *Phys. Rev. Lett.* **46**, 57 (1980).
- Murray, C.A., Bodoff, S. *Phys. Rev. Lett.* **52**, 2273 (1984).
- Nakao, Y., Yamada, H. *Surf. Sci.* **176**, 158 (1986).
- Neeves, A.E., Birnboim, M.H. *J. Opt. Soc. Am. B* **6**, 787 (1989).
- Nie, S., Emory, S.R. *Science* **275**, 1102 (1997).
- Nishikawa, Y., Fujiwara, K., Ataka, K., Osawa, M. *Anal. Chem.* **65**, 556 (1993).
- Oldenburg, S.J., Westcott, S.L., Averitt, R.D., Halas, N.J. *J. Chem. Phys.* **111**, 4729 (1999).
- Osawa, M., Kuramitsu, M., Hatta, A., Suetaka, W., Seki, H. *Surf. Sci.* **175**, L787 (1986).
- Osawa, M., Ikeda, M. *J. Phys. Chem.* **95**, 9914 (1991).
- Perney, N.M.B., Baumberg, J.J., Zoorob, M.E., Charlton, M.D.B., Mahnkopf, S., Netti, C.M. *Opt. Exp.* **14**, 847 (2006).
- Pinzarzu, S.C., Pavel, I., Leopold, N., Kiefer, W. *J. Raman Spectrosc.* **35**, 338 (2004).
- Porter, M.D., Lipert, R.J., Siperko, L.M., Wang, G., Narayan, R. *Chem. Soc. Rev.* **37**, 1001 (2008).
- Prodan, E., Radloff, C., Halas, N.J., Nordlander, P. *Science* **302**, 419 (2003).

- Qin, L., Zou, S., Xue, C., Atkinson, A., Schatz, G.C., Mirkin, C.A. *Proc. Nat. Acad. Sci.* **103**, 13300 (2006).
- Rowe, J.E., Shank, C.V., Zwemer, D.A., Murray, C.A. *Phys. Rev. Lett.* **44**, 1770 (1980).
- Schatz, G.C. *Acc. Chem. Res.* **17**, 370 (1984).
- Schatz, G.C., Van Duyne, R.P. in *Handbook of Vibrational Spectroscopy* (Wiley, New York, 2002).
- Schlukers, S. *Chem. Phys. Chem.* **10**, 1344 (2009).
- Sinzig, J., Quinten, M. *Appl. Phys. A* **58**, 157 (1994).
- Smardzewski, R.R., Colton, R.J., Murday, J.S. *Chem. Phys. Lett.* **68**, 53 (1979).
- Smith, W.E., White, P.C., Rodger, Dent, G. *Pract. Spectrosc.* **28**, 733 (2001).
- Stockman, M.I., Shalaev, V.M., Moskovits, M., Botet, R., George, T.F. *Phys. Rev. B* **46**, 2821 (1992).
- Tao, A.R., Yang, P. *J. Phys. Chem. B* **109**, 15687 (2005).
- Thaxton, C.S., Georganopoulos, Mirkin, C.A. *Clin. Chem. Act.* **363**, 120 (2006).
- Ueba, H., Ichimura, S. *Surf. Sci. Lett.* **118**, L273 (1982).
- Vo-Dinh, T., Stokes, D.L. *Biomed. Photonics Handbook* **64**, 1 (CRC Press, 2003).
- Wadayama, T., Sakurai, T., Ichikawa, S., Suetaka, W. *Surf. Sci.* **198**, 359 (1988).
- Weitz, D.A., Garoff, S., Gramila, T.J. *Opt. Lett.* **7**, 168 (1982).
- Wiscombe, W. *Appl. Opt.* **19**, 1505 (1980).
- Wu, Z.S., Wang, Y.P. *Radiat. Sci.* **26**, 1393 (1991).
- Xu, H., Bjerneld, E.J., Kall, M., Borjesson, L. *Phys. Rev. Lett.* **83**, 4357 (1999).
- Xu, H. *Phys. Rev. B* **72**, 074305 (2005).
- Xu, M., Dignam, M.J. *J. Chem. Phys.* **96**, 7758 (1993).
- Xu, M., Dignam, M.J. *J. Chem. Phys.* **99**, 2307 (1993).
- Xu, M., Dignam, M.J. *J. Chem. Phys.* **100**, 197 (1994).
- Yang, W.H., Schatz, G.C., Van Duyne, R.P. *J. Chem. Phys.* **103**, 869 (1995).
- Zeman, E.J., Schatz, G.C. *J. Phys. Chem.* **91**, 634 (1987).
- Zwemer, D.A., Shank, C.V., Rowe, J.E. *Chem. Phys. Lett.* **73**, 201 (1980).

# 16

# Protein-based nanodevices

P.P. Pompa and R. Rinaldi

16.1 Introduction	570
16.2 Protein fundamentals	572
16.3 Nanofabrication	574
16.4 Nanoelectronic devices based on proteins	580
16.5 Biophysical implications of protein-based nanobioelectronics	588
16.6 Nanodevices for biosensing	591
16.7 Conclusions	605
Acknowledgments	606
References	606

## 16.1 Introduction

Nanotechnology allows manufacturing and manipulation of matter on a scale, ranging from single atoms and molecules to micrometer-sized objects. This has already enabled the miniaturization of many current devices, resulting in faster operation or integration of several operations. Furthermore, at this scale, man-made structures match the typical sizes of natural functional units in living organisms, allowing them to interact with the biology of living organisms. Finally, nanometer-sized materials and devices often show novel properties.

Biotechnology is the technological application that uses biological systems, living organisms or their derivatives to produce or modify products or processes for a specific objective. The combination of these two disciplines and their applications in several fields has promoted growth in the interdisciplinary research at the interface between biology and nanotechnology (Lowe 2000; Whiteside 2003; Mohamadi *et al.* 2006).

DNA, proteins and cells, as natural building blocks, have captured significant attention of scientists due to their incomparable capability to self-assemble, selectively recognize and self-repair. Integration of these systems with inorganic micro- and nanostructures and devices will lead to noticeable progress in fields like biology/biotechnology, engineering, physics, medicine and open up the way to new non-conventional developments in them. From the fundamental point of view, it may be possible to fabricate new powerful tools to study, handle, control, and engineer these natural units on the one hand, and to design and produce novel nanoelectronic and nano-optical hybrid devices based on the properties and natural functions of the living matter on the other hand.

Hybrid nanodevices that encompass organic and inorganic components offer truly unique and flexible functionalities. The integration of biological elements, like active proteins, can provide a myriad of opportunities for custom-tailored devices with new potentialities to realize the envisioned applications (Willner 2002; Bashir 2004; Niemeyer and Mierkin 2004; Niemeyer 2007). Integration strategies of proteins with artificial man-made systems have been

promoting an unmatched progress in many fields by invoking broad interdisciplinary and transdisciplinary approaches in biology and materials science. Proteins have been proposed as materials for bioelectronics, thanks to their natural functionalities and the potentiality offered by protein engineering. In this scenario, new hybrid architectures, systems and devices can be realized down to the nanoscale by combining the top-down and bottom-up methodologies.

For a successful outcome, the research efforts have to overcome formidable obstacles arising from the large diversity between the two worlds, such as the totally different manipulation, treatment and assembling processes, and the need for a “humid” environment to maintain functional the biounits, which could be a detriment for devices. It is worth noting also that the charge carriers in inorganic and organic matter are typically different, being electrons and ions, respectively, and that the problem of interfacing and compatibility does not seem to have an easy solution. One of the most frequent questions in this matter relates to the viability and stability of biological units into a non-natural “solid-state” environment. The “communication” between them often demands the building up of more complicated and novel architectures.

In recent years rapid progress in the above direction has been made with the advent and development of more advanced alternative nanofabrication techniques to produce structures below 100 nm. Besides the extensively used electron-beam lithography, the host of other lithographic techniques employed are ion-beam lithography, X-ray lithography, scanning probe lithography, and the alternative techniques like soft lithography. The latter in particular has been demonstrated to be compatible with the handling and modification of organic and biological materials, and there exists already in the literature various examples of protein patterning realized by means of this technique (Xia and Whitesides 1998). New inspection tools, like scanning probe microscopy, developed in the last two decades have become ubiquitous systems to image nanoscale structure and to estimate certain structural, mechanical, and functional characteristics of biological entities, ranging from proteins and DNAs to cells and tissues. Various modes of imaging and SPM-based spectroscopies have been developed to correlate structure, properties, and chemomechanical interactions between biological units in different environments. This has induced a rapid improvement in handling, assembling and subsequently modifying these biological entities with the result that new insights into the properties of interfaces and binding mechanisms have been gained. In particular the detailed investigation of self-assembling processes at the base of protein and DNA formation and ligand–receptor interactions have opened new routes to the design and engineering of hybrid systems, comprising inorganic nanostructures and biological “smart” matter. The need for development of new strategies for the functional integration of biological units and electronic systems or nanostructured materials were also facilitated by the parallel progress in the biochemistry and molecular biology areas, namely the advances in protein engineering, with the ability to make “designer” proteins/peptides with specific functions or combinations of functions; and the establishment of surface display technologies, with the ability to generate and screen large repertoires of peptides/nucleic acids for high-affinity binding to potentially

any structure (organic or inorganic). New nanostructured sensors also based on nanowires or carbon nanotubes, electronic nanocircuits based on biomolecules and biomolecular templates are a few examples where biology meets nanoelectronics.

Here, we will review the exploitation of specific proteins for engineering nanoelectronic devices. We will also describe the recent development in the field of biosensing and proteomics and the related technological methods in realizing hybrid devices and biocompatible interfaces. This chapter is organized as follows: in Section 16.2 we briefly describe the basic characteristics of proteins, to make the chapter accessible also to the readers who are not experts in biology; in Section 16.3 we review the main procedures and strategies in nanomanufacturing suitable to interface inorganic matter to biological units; Section 16.4 is focused on the description of nanoelectronic devices having proteins as functional units; Section 16.5 deals with the major concerns about the possibility of integrating these devices with the present-day consumer electronics; Section 16.6 presents more recent examples of nanodevices for biosensing; and, finally, in Section 16.7 important conclusions are given.

## 16.2 Protein fundamentals

The term protein comes from the Greek word “*protas*”, whose meaning is “of primary importance”. A protein is a complex organic compound, typically of high molecular mass (up to 3 000 000 Da). The proteins can be assimilated to polymers built from 20 different L-alpha-amino acids, joined by peptide bonds. The details of their sequence are stored in the code of a gene: genes are transcribed into RNA, which is then subject to post-transcriptional modification and control, resulting in a mature mRNA. The mRNA is translated by ribosomes, originating in this way the protein. The amino-acid chain constituting a protein has two terminal points, the carboxy or C-terminus, and the amino or N-terminus.

The three-dimensional structure is a peculiar property of each protein. The shape into which a protein naturally folds is known as its native state, which is determined by its sequence of amino acids. In biochemistry there exists a peculiar terminology related to four distinct types of protein structure: the “*primary structure*”, which is exactly the amino acid sequence; the “*secondary structure*” related to the segments of the chain that assume particular shapes, like alpha helix and beta sheet, formed by hydrogen bonding; the “*tertiary structure*” is the overall shape of each protein molecule and is formed mainly by hydrophobic interactions, with also a contribution of hydrogen bonds, ionic interactions and disulfide bonds; finally, the “*quaternary*” structure is the shape resulting from the fusion of more than one protein molecule, called the protein complex. It is worth noting that proteins differing in their structure also differ in their biological functions. The functional rearrangements are usually referred to as “conformations”, and transitions between them are called conformational changes. The natural folding process of a protein occurs in less than a minute after they are formed: this self-assembling process brings the protein from

primary to higher structures. As a general rule, there exist more than one stable folded conformation corresponding to a precise biological activity and only one is considered to be the active one. Molecules and ions that are able to chemically bind to a specific site onto a protein, are called ligands and the sites are called binding sites. The strength of ligand–protein binding characterizes the chemical specificity of the binding and is called “affinity”. Proteins are involved in every function performed by a cell: therefore, by suitably controlling the protein’s activity it is possible to regulate these functions. The regulation can go through the so-called “allosteric modulation”, involving binding of ligands, or “covalent modulation”, involving protein’s modification.

As already outlined above, the activity of a protein is directly related to its structure. Small amino acid chains, lacking in tertiary structure, also occur in nature. They are referred to as “polypeptides”, like, for example, the hormones. Proteins are generally classified as soluble, filamentous, or membrane-associated. Enzymes are soluble proteins that catalyze biochemical reactions. Antibodies also belong to this class. Membrane–associated proteins usually transport mass across the cell membrane, but they do not modify the chemical structure of the shifted mass or substrate. Another kind of membrane proteins are the receptors, which may change their conformation upon interaction with specific ligands. Filamentous proteins constitute the cytoskeleton of cells and are the basic components of several animal tissues, like tubulin, actin, collagen, and keratin, which are the building blocks of skin, hair, and cartilage. Another special class of proteins consists of “motor proteins” such as myosin, kinesin, and dynein. These proteins are nanometer-sized molecular motors and are able to generate physical force that can move organelles, cells, and entire muscles. There is a considerable interest in the nanotechnological application of such proteins bound to a surface for the building of bioinspired nanomachines.

Proteins are involved in every function performed by a cell, including regulation of important cellular functions such as signal transduction and metabolism. It is also possible to group the proteins into functional classes like:

1. enzymes, which catalyze metabolism reactions;
2. structural proteins, such as tubulin, or collagen;
3. regulatory proteins, such as transcription factors or cyclins that regulate the cell cycle;
4. signalling molecules and their receptors, such as hormones and their receptors;
5. defensive proteins, like antibodies and toxins.

Proteins are sensitive to their environment. They are found to be active in their native state, over a small pH range, and in a solution with small quantities of electrolytes. A protein that is not in its native state is said to be “denatured” or “unfolded”, because they have no well-defined secondary structure. The sequence and hence the structure and properties of a protein can be modified by methods belonging to the “genetic engineering”, while the term “protein engineering” refers to a rather new research branch aimed to design and create proteins with entirely new properties or functions. As outlined clearly

in this brief description of protein characteristics, the self-assembly process resulting in the folding of a protein is greatly attractive towards potential uses in nanotechnology. Understanding mechanisms at the base of protein folding would allow the design and building of man-made machines at the nanoscale.

Proteins are considered to be of great interest due to their interesting properties and the range of functionalities they offer to bioelectronics. Almost all kinds of proteins listed above have been used in bioelectronics and biosensors at the microscale (Gopel 1995). However, only a few examples of nanoscale biological devices based on proteins have been developed in the last decade with the contemporary progress of nanofabrication and handling tools. The sizes and the potentiality offered by protein engineering makes them suitable candidates for the development of protein-based nanobioelectronics (Astier *et al.* 2005). Vice versa the progress in the present-day electronics and the related technological methods promoted the recent developments in the field of biosensing and proteomics (Mohamadi *et al.* 2006). In the following sections we will describe some representative examples of hybrid nanoarchitectures for bioelectronics.

## 16.3 Nanofabrication

To reach the nanoscale in a controlled manner, two very different paths are pursued. One is a top-down strategy of miniaturizing current technologies, while the other is a bottom-up strategy of building ever-more-complex molecular devices atom by atom. Top-down approaches are good for producing structures with long-range order and for making macroscopic connections, while bottom-up approaches are best suited for assembly and establishing short-range order at nanoscale dimensions. The integration of top-down and bottom-up techniques is expected to eventually provide the best combination of tools for nanofabrication.

### 16.3.1 Top-down methodologies

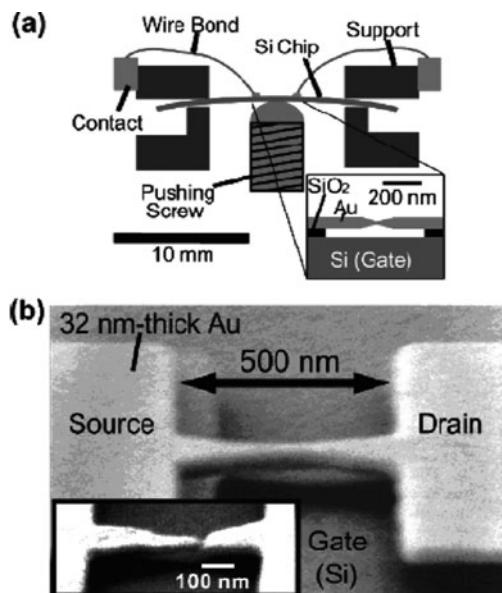
Microelectronics has been the primary beneficiary of technology pushing the limits of the small in manufacturing. Such complex, multistep processes of material removal and deposition become more arcane and expensive as the feature dimensions become smaller. This process of micromanufacture is referred to as “top-down” manufacturing since one starts on the macroscale and proceeds to create fine features by processing the bulk on a fine scale. Microelectronics engineers commonly employ photolithography to reduce the component size to submicrometer scale. The most common top-down approach to fabrication involves lithographic patterning techniques using short-wavelength optical sources. A key advantage of the top-down approach—as developed in the fabrication of integrated circuits—is that the parts are both patterned and built in place, so that no assembly step is needed. Optical lithography is a relatively mature field due to the high degree of refinement achieved in microelectronic chip manufacturing, with the current short-wavelength optical

lithography techniques reaching dimensions below 100 nanometers (the traditional threshold definition of the nanoscale). Short-wavelength sources, in the extreme ultraviolet and X-ray, are being developed to allow lithographic printing techniques to reach dimensions from 1 to 10 nanometers. The main drawback of standard photolithography lies in its resolution limit, ultimately determined by the radiation wavelength. The Rayleigh resolution criterion defines the critical dimension (smallest half-pitch) of a feature on an integrated circuit (IC):  $\approx \kappa_1 \lambda / N_A$ , where  $\lambda$  is the wavelength,  $N_A$  is the numerical aperture of the imaging system and  $\kappa_1$  is a value between 0.25 and 1 that depends on the configuration of the illumination system and the optical response of the photoresist. To go further down at the nanoscale other nanofabrication tools and methodologies are required. Among the numerous technologies available for patterning below 100 nm scale, electron-beam lithography (EBL) is of course the most widely employed, mainly because of its versatility that permits the fabrication of a variety of nanodevices, combining EBL with processes such as lift-off, etching and electrodeposition/electroplating. With EBL, the resist is exposed to a thin electron beam with a spot size of the order of 2 nm. The EBL resolution is limited by large-angle electron-scattering events leading to backscattering and additional exposure in the resist (proximity effect). As a consequence of this, in order to wire biomolecules, pairs of nanoelectrodes need to be fabricated with the separation in the few nanometer range. For this purpose, postprocessing techniques are typically employed post-EBL exposure, such as electrodeposition, electromigration and mechanical controllable compression to realize break junctions. In particular, nanogaps ranging from 20 to 3.5 nm were reproducibly fabricated by Kervennic and coworkers by monitoring the conductance between two free-standing EBL-fabricated electrodes during Pt electrodeposition (from an aqueous solution of 0.1 mol of  $K_2PtCl_4$  and 0.5 mol of  $H_2SO_4$ ) stopping the electrodeposition process at pre-defined conductance values (Kervennic *et al.* 2002).

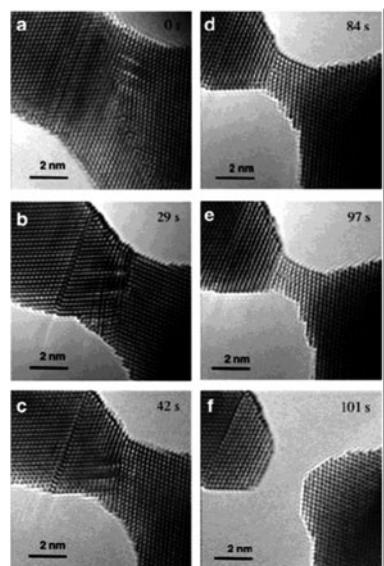
Electromigration, usually observed and controlled at cryogenic temperatures (Park *et al.* 2002), and mechanically controllable break junctions realized by the gentle fracture of an electrode due to an induced mechanical deformation in the supporting substrate ((Reed *et al.* 1997; Champagne *et al.* 2005), Fig. 16.1) have also been exploited to achieve nanometer-spaced electrodes. Since all these approaches do not completely ensure the reproducibility of the fabricated gap, the real separation has to be typically measured by SEM. Motivated by this, Zandbergen *et al.* (2005) developed an e-beam sculpting technique (Fig. 16.2) to fabricate nanojunctions in a TEM by reshaping a metallic bridge in a controllable manner until the creation of a nanogap. Finally, it is worth noting that the robustness of the electrodes and possible failure mechanisms during operation have also to be investigated (Maruccio *et al.* 2003; Coura *et al.* 2004). Besides the scattering-related limitations in the resolution, electron-beam lithography also suffers from high costs and comparatively long processing times.

Among the other lithographic techniques, we should mention:

- (1) Ion-beam lithography: a technique closely related to EBL which exploits lightweight high-energy ions, like protons, instead of electrons.

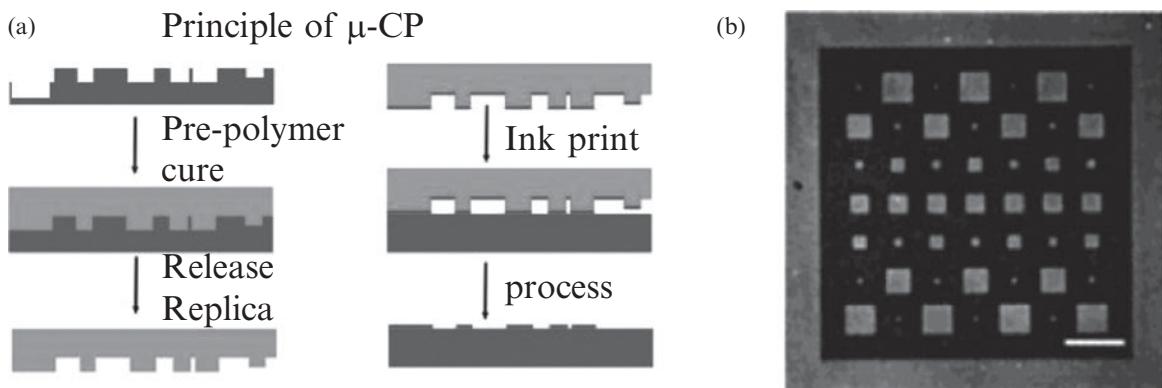


**Fig. 16.1** (a) Schematic of the mechanically adjustable and electrically gated nanojunction (by Champagne *et al.*), SEM-imaged in (b) at 78° tilt angle; in the inset of (b), an electromigration-broken nanojunction. Reproduced with permission from A.R. Champagne *et al.* *Nano Lett.* **5**, 305–308 (2005). Copyright 2008 American Chemical Society.



**Fig. 16.2** Successive moments in the fabrication of a nanogap by the TEM sculpting technique. Reproduced with permission from H.W. Zandbergen *et al.* *Nano Lett.* **5**, 549–553 (2008). Copyright 2008 American Chemical Society.

- (2) X-ray lithography (using 0.1–10 nm radiation, Silverman 1997): uses collimated X-rays to expose a resist in a parallel replication process. The increased lateral resolution is due to both the extremely short wavelength and the high penetration ability achieved. A mask is employed to define the pattern into a resist layer, such a mask is usually the most crucial element. X-ray lithography is expensive to perform because of the need for a synchrotron-radiation facility.
- (3) Interference lithography: simple and maskless. Exposure is obtained from the light of two or more interfering coherent laser beams. The interference patterns define the final geometry, and this is the major drawback of this technique, since only highly symmetrical structures can be obtained. Conversely, a large area can be exposed simultaneously, which results in higher speed.
- (4) Shadow-mask techniques: this exploits a very thin shadow-mask, whose nanometric holes play the role of the resist. The mask is placed very close to the substrate and the materials are then deposited to create the desired nanostructures on the substrate. Features in the 400-nm range can be fabricated, with resolution limited by the mask thickness.
- (5) Scanning probe lithography (SPL): exploits (a) a scanning probe for scratching, nanoindentation and local heating or (b) an intense electric field in the vicinity of the tip (when a bias is applied) for field-enhanced oxidation and electron exposure of resists. Its main disadvantages are the slow throughput and the small patternable area.
- (6) Soft lithography is attracting considerable attention, especially in bioinspired applications since it is biocompatible. Soft lithography (Xia and Whitesides 1998) is a collective term including different techniques that share the basic concept of using an elastomeric stamp or mold to transfer



**Fig. 16.3** (a)  $\mu$ CP printing; (b) Pattern of fluorescently labelled proteins by  $\mu$ CP. Reproduced with permission from J.L. Tan *et al.*, *Langmuir* **18**, 519–523 (2008) Copyright 2008 American Chemical Society.

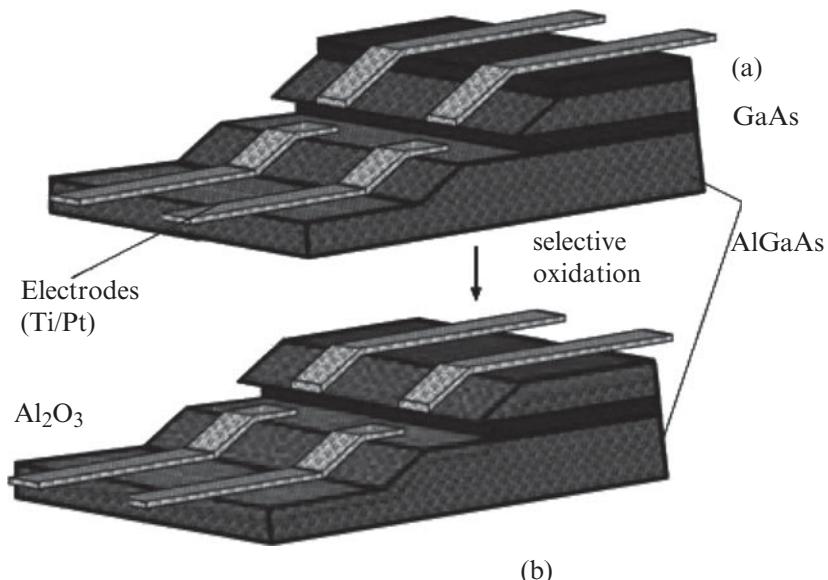
the pattern to the substrate. Among them we list here microcontact printing ( $\mu$ CP) (Fig. 16.3(a)), replica molding (REM), microtransfer molding ( $\mu$ TM), micromolding in capillaries (MIMIC), and solvent-assisted micromolding (SAMIM).

Their significance is due to the use of flexible organic molecules and materials relevant to our discussions rather than rigid inorganic materials. Soft lithography has the potential to fulfill the demands of a low-cost and high-resolution nanolithographic technique: recently a number of protein patterns were fabricated using this method. For example, Tan *et al.* (2002) reported protein patterns with micrometer resolution by  $\mu$ CP and examined the influence of substrate wettability on the process demonstrating that a minimum substrate wettability is required for successful  $\mu$ CP (Fig. 16.3(b)).

To a large extent, most of the proposed methods for patterning below the 100-nm scale are appropriate for contacting only single devices and not suitable for mass production, due to severe technological and economic limitations.

Large efforts are currently focused on developing techniques for economical fabrication of complex molecular electronic circuits on a large scale.

Among the strategies recently proposed, we want to mention in particular the approach of Krahne *et al.* (2002) (and recently improved by Maruccio *et al.* (2007a) to define networks of nanojunctions by only optical lithography and wet etching of an AlGaAs/GaAs quantum-well structure without the need for expensive e-beam systems. In this method, the thickness of the quantum well and of the deposited metal layer control the gap size with subnanometer precision. The leakage current through the semiconductor layer can be trimmed down by selectively oxidizing the two AlGaAs barriers, above and below the QW (Maruccio *et al.* 2007a) (Fig. 16.4). The approach is suitable for the simultaneous production of wafer-scale arrays of electrodes, mainly because the electrode pattern is defined only by photolithography; it thus represents an important step towards low-cost massive fabrication of nanodevices.



**Fig. 16.4** Mesa nanojunctions without (a) and with (b) selective oxidation of the AlGaAs barriers.

### 16.3.2 Bottom-up and self-assembly

The ability to pattern materials at the micro- and nanoscale has become of tremendous importance not only in the field of microelectronics industry (where the reduction of feature sizes is strongly related to the production of high-performance devices), but also in several important scientific areas such as photonics, sensing, advanced materials, biotechnology and nanomedicine. Basically, nanopatterning techniques can be summarized into two main categories, namely “top-down” and “bottom-up”. The top-down approach represents the most conventional way to produce spatially controlled nanostructures on different types of planar substrates. On the other hand, the bottom-up approach seeks to drive the natural self-assembly of single units into large structures, starting from single atoms, molecules or nanoparticles up to the realization of two- or three-dimensional (2D or 3D) structures exploiting chemical and/or physical interactions operating at the nanoscale. In the so-called bottom-up manufacturing technologies, processes are controlled by self-assembly, such as the manufacture of complex proteins using a DNA blueprint, followed by an ever-larger scale of hierarchical assembly as in cell walls, cells, tissues and organs.

The bottom-up approaches have become complementary and, in fact, of increasing importance to the top-down methodologies. Taking inspiration from nature, where biological systems have the ability to recognize specific signals coding for very selective interactions among molecules, such approaches, based on self-assembly interactions, provide the way for specific organizations of molecules on planar substrates. These methodologies enable easy production of bioinspired nanostructures, finely patterned both on conventional substrates (e.g. Si, SiO<sub>2</sub>, quartz, gold, etc.) and unconventional substrates, such as polymers and plastics. In order to produce nanostructures by means of the

bottom-up approach, biomolecules, such as nucleic acids, proteins, peptides and even cells, have been considered as the starting building blocks in the self-assembly process. Through several inter- and intramolecular interactions and/or interactions between biomolecules and substrates, the self-assembly processes may result in the formation of larger and spatially controlled structures. Seeman (1982) demonstrated the use of DNA structural differences and base-pair complementarities to produce well-ordered structures in assembled materials such as 2D lattices. Since then, structural DNA nanotechnology has led to fabrication of several examples of DNA-inspired nanostructures with different geometries (such as knots, cubes, octahedra, rings, etc.) including complex 2D DNA nanoarrays. Such nanostructures were produced by several approaches, all exploiting different ways to control the natural aggregation propensity of DNA and its self-assembling in complex architectures. For instance, some strategies, which were recently well reviewed (Stewart 2007), foresee single-step-based methods by the use of sequences with sticky ends encoding for tile-to-tile associations. Others require multistep methods utilizing a hierarchical self-assembly procedures. Noteworthy is the possibility to produce very complex shaped, sub-100-nm nanostructures (e.g. the smile acronym) as recently demonstrated by Rothemund *et al.* using the “DNA origami” technique (Rothenmund 2006). These techniques permit the realization of even more complex shaped nanostructures as compared to those obtainable by means of scanning probe lithographic techniques. Such applications have also opened a new scenario for diagnostic biosensors, based on the rapid biorecognition of specific biological events.

When the templates assisting the self-assembly processes are peptides or proteins, new interesting nanostructures, such as nanotubes, nanofibers and nanospheres, can also be obtained. Efforts to mimic nature for biological assembly are related to diseases, as in the case of the amyloidogenic diseases, where fibrils are produced by a natural self-assembly process, may provide challenging opportunities for the realization of novel nanostructures to be applied in material science, medicine and electronics. Cyclic peptides that are short repeated sequences of aromatic dipeptide and/or hydrophobic moieties, were used to drive the 3D reorganization of peptides into polymer nanotubes. Although the applications are numerous, of particular interest is the method developed by Gazit, enabling the formation and the alignment (both horizontal and vertical) of nanotubes to be exploited, for instance in the creation of metallic nanowires in electrochemical biosensors (Reches and Gazit 2006). 2D and 3D scaffolds formed by peptide nanostructures are also under investigation as new supports for advanced cell-culture studies of microdevices.

Another important and extensively explored research area is related to the investigation of the self-assembly properties of polymers, composites, ceramics and metals, on templates that can then be dissolved after the formation of the desired structure. In this case, the shape and dimensions of the resulting nanostructures are strictly related to the starting template features. Although several examples of such approaches have appeared in the literature, the use of nanoparticles as a sacrificial core appeared to be very promising. An important fabrication example is the layer-by-layer (LbL) tech-

nique, based on the consecutive assembly of oppositely charged synthetic polymer layers around a pre-formed nucleus, acquiring a specific shape conferred by the nanoparticles. The polymeric layer self-assembly is mainly mediated by electrostatic interactions and can occur on several cores, such as nanoparticles of polystyrene latex, melamine formaldehyde,  $\text{SiO}_2$ , carbonate particles ( $\text{MnCO}_3$ ,  $\text{CaCO}_3$ ,  $\text{CdCO}_3$ ), and even cells (Gil *et al.* 2008). After removing the core, hollow polymeric nanospheres or nanocapsules can be obtained. Such soluble nanostructures can be conjugated to several molecules, such as fluorescent and/or magnetic species, biological elements and can contain a cargo deliverable into several districts of the cells for therapeutic purposes.

Finally, the combination of the different approaches, i.e. top-down and bottom-up, with suitable chemical surface modifications and bioengineering of biological structures can lead to a deep integration of inorganic, organic, and even biological materials for the creation of tailored nanostructures, addressable at the nanoscale level by programmable chemical and physical methods of fabrication.

## 16.4 Nanoelectronic devices based on proteins

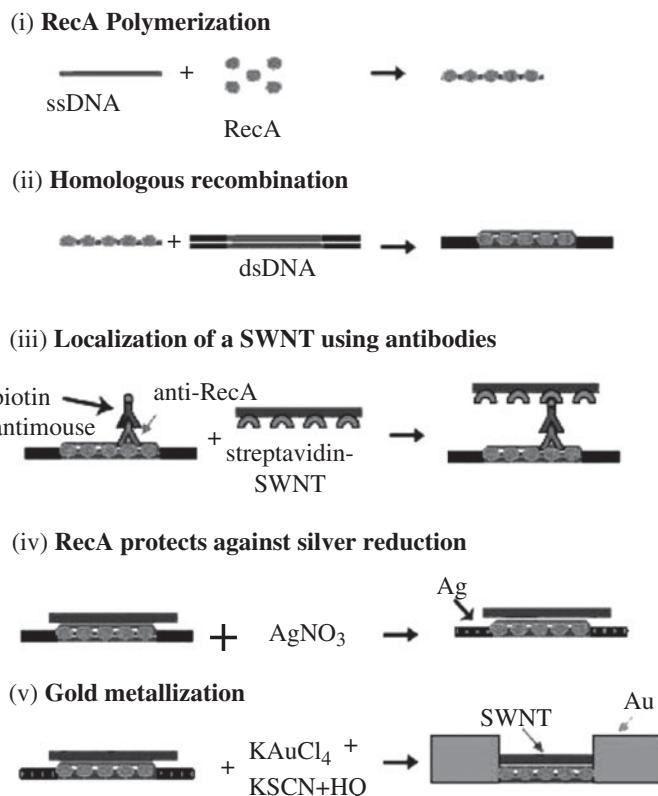
In recent decades, the enormous progresses in miniaturization and the advent of nanotechnology have been the key to open new doors at the interface between physics, chemistry and biology. In particular, these achievements have been favored by the invention of new, very powerful techniques capable of addressing single molecules/proteins, such as scanning probe microscopy and related techniques. As a result, new ways for a deeper understanding of biosystems have been opened up. In particular, great efforts worldwide were focused on biological systems like DNA, proteins, enzymes and cells, originating new fields such as bioelectronics and neuroelectronics.

The first two-terminal conductance measurements on the molecular scale were carried out in molecular tunnel junctions with the invention of the scanning tunnelling microscope (STM) (Datta *et al.* 1997). Only in the last decade were reproducible methods to interconnect and probe single molecules at the nanoscale in device geometry demonstrated using new lithographic techniques like extreme ultraviolet lithography, X-ray proximity lithography, electron- and ion-beam lithography.

The first biomacromolecule ever used to implement novel bioelectronic devices was DNA. A major issue then was to determine the DNA conductivity (Endres *et al.* 2004; Rinaldi *et al.* 2006), due to its crucial importance in understanding the significant life mechanisms (like DNA damage recognition and reparation) and the numerous applications in molecular electronics and biotechnology, such as the implementation of DNA chips with an electronic read-out. The main difficulties in addressing the DNA conductivity lie in the complexity of the system (controllable experiments are difficult, due to myriad differences in the DNA molecules sequence, length, character), the influence of the molecular environment, the detection scheme and the role of contacts between electrodes and molecules. As a result, DNA was reported

to be an insulator (Braun *et al.* 1998; De Pablo *et al.* 2000; Storm *et al.* 2001; Zhang *et al.* 2002), a semiconductor (Porath *et al.* 2000; Rakitin *et al.* 2001), an Ohmic conductor (Fink and Schönenberger 1999; Cai *et al.* 2000; Tran *et al.* 2000; Yoo *et al.* 2001; Hartzell *et al.* 2003) and even a superconductor (Endres *et al.* 2004; Rinaldi *et al.* 2006). Recently, an explanation in a comprehensive framework taking into account all the observed behaviors (excluding only the superconductor one) was proposed by Endres *et al.* (2004). They point out that near-Ohmic behaviors were reported only for DNA bundles, networks of bundles or supercoiled samples. However, single molecules and short molecules exhibited insulating and semiconducting behaviours, respectively. The observed near-Ohmic behavior was ascribed to the possible stabilization of floppy single DNA molecules by the bundles or condensed water and counterions trapped between the DNA molecules. This scenario would also explain the observed dependence of the conductance on humidity.

In a different approach, the self-assembling ability of DNA molecules was used to fabricate molecular nanodevices (Keren *et al.* 2002; Maruccio *et al.* 2003; Maruccio and Bramanti). Braun and coworkers developed a sequence-specific molecular lithography approach to produce large-scale functional circuits at the molecular level (Keren *et al.* 2002, 2003). The information encoded in the DNA molecules was employed as a mask to achieve the sequence-specific positioning of nano-objects by means of RecA protein acting as



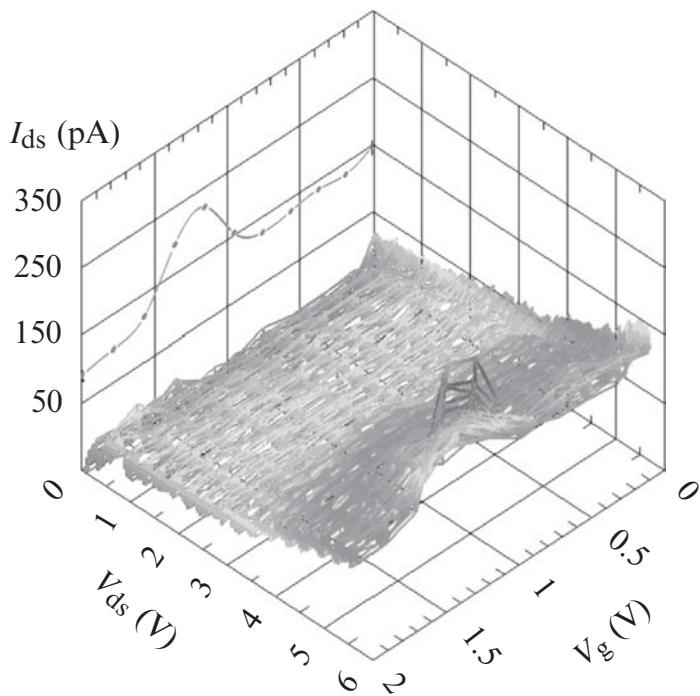
**Fig. 16.5** Assembly of a DNA-templated FET by molecular lithography: (i) Polymerization of RecA monomers on a ssDNA molecule, resulting in the formation of a nucleoprotein filament. (ii) Binding of the nucleoprotein filament at a desired address by homologous recombination. (iii) Delivery of a streptavidin-functionalized SWNT on the DNA-bound RecA by means of a primary antibody to RecA and a biotin-conjugated secondary antibody. (iv) Incubation in an  $\text{AgNO}_3$  solution and formation of silver clusters on segments unprotected by RecA. (v) Formation of two DNA-templated gold wires to contact the SWNT by electroless gold deposition (the silver clusters are used as nucleation centers). Reprinted from K. Keren *et al.* *Science* **302**, 1380–1382 (2003). Copyright 2003 American Association for the Advancement of Science.

the resist (Fig. 16.5). As difficulties in assembling reproducible and stable nanoelectronic devices based on DNA emerged, scientists started to explore proteins as functional units for nanodevices (Rinaldi *et al.* 2003; Rinaldi and Cingolani 2004; Maruccio *et al.* 2004a, 2005b; Rinaldi *et al.* 2004).

Due to their nature-tailored functions, proteins have also been extensively investigated for applications in molecular electronics, molecular motors and biosensors. In particular, their electron transfer (ET) capabilities (Rinaldi *et al.* 2006) are interesting for bioelectronics. In the past, spectroscopic and electrochemical methods were employed to obtain statistical information on ET in proteins, with special attention to metalloproteins, i.e. proteins containing a metal cofactor. The invention of the scanning probe microscopes, however, made it possible to probe single proteins (Tao 1996) by means of two-terminal conductance measurements in tunnel junctions. The electrochemical scanning tunneling microscopy (Sonnenfeld and Hansma 1986) was largely used to investigate redox proteins in self-assembled monolayers, mainly by the Ulstrup group, that demonstrated the capability of proteins immobilized in SAMs to retain their functional properties (Zhang *et al.* 2003). Analogous studies were carried out on the metalloprotein azurin by Alessandrini *et al.* (2003). More recently, conduction in metalloproteins was also studied by means of conductive AFM investigating the role of the metal site (Xu *et al.* 2005) and the influence of the compressional force employed to create the electrical contact (Zhao *et al.* 2004).

The ability of a protein in transferring electrons (Lee *et al.* 1997) was exploited in electronic devices. As an example, a prototype protein field effect transistor (pro-FET) (Maruccio *et al.* 2005) was recently demonstrated using the metalloprotein azurin immobilized in SAMs onto Si/SiO<sub>2</sub> substrates via the disulfide bridge (Chi *et al.* 2000; Alessandrini *et al.* 2003) (Cys-3-Cys-26). Azurin exists in two stable configurations—Cu(I) and Cu(II)—and its ET capability (*in vitro*, it is able to mediate electron transfer (ET) from cytochrome *c*<sub>551</sub> to nitrite reductase (Vijgenboom *et al.* 1997) depends on the equilibrium between these two oxidation states by means of the reversible redox reaction  $\text{Cu}^{2+} + \text{e}^- \rightleftharpoons \text{Cu}^{1+}$ . The pro-FET consisted of a planar nanojunction fabricated by EBL and a silver backgate with a protein monolayer connecting the source and drain electrodes. A pronounced resonance centered at  $V_g = 1.25$  V was observed in the transfer characteristic of the Pro-FET, enabling the device to switch from a n-MOS FET behavior to a p-MOS one (Fig. 16.6). In this device, electron transport occurs by hopping from one reduced protein to an adjacent oxidized one with the copper atom playing a central role in conduction. Lower currents were reported in devices based on two azurin variants where the Cu atom is absent (Apo-form) or replaced by a Zn atom.

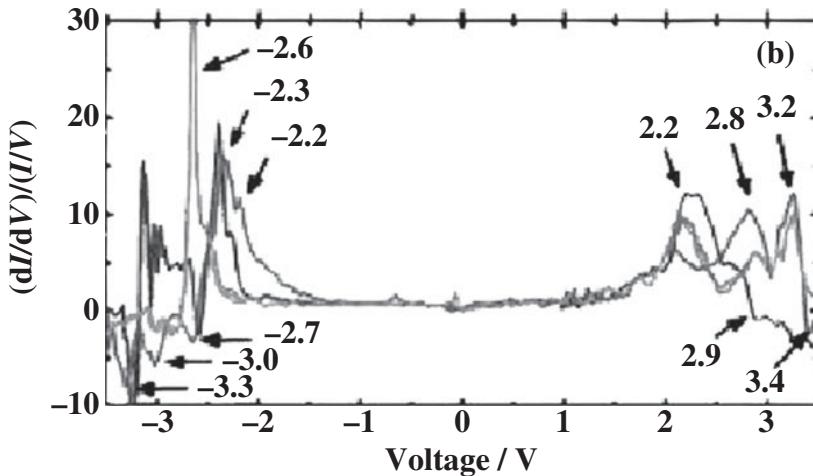
More recently, intramolecular ET in single/few Azurins was probed for the first time at the molecular level in a device geometry by exploiting an innovative technique to interconnect individual molecules and fabricate large-scale nanojunction arrays (Maruccio *et al.* 2007). In this experiment, mesa nanojunction structures described in the previous section were used to directly bind a single protein. Current–voltage characteristics of this device, measured in air and at room temperature, showed different maxima, ascribed to negative differential resistance (NDR) and an increased transparency of the molecular



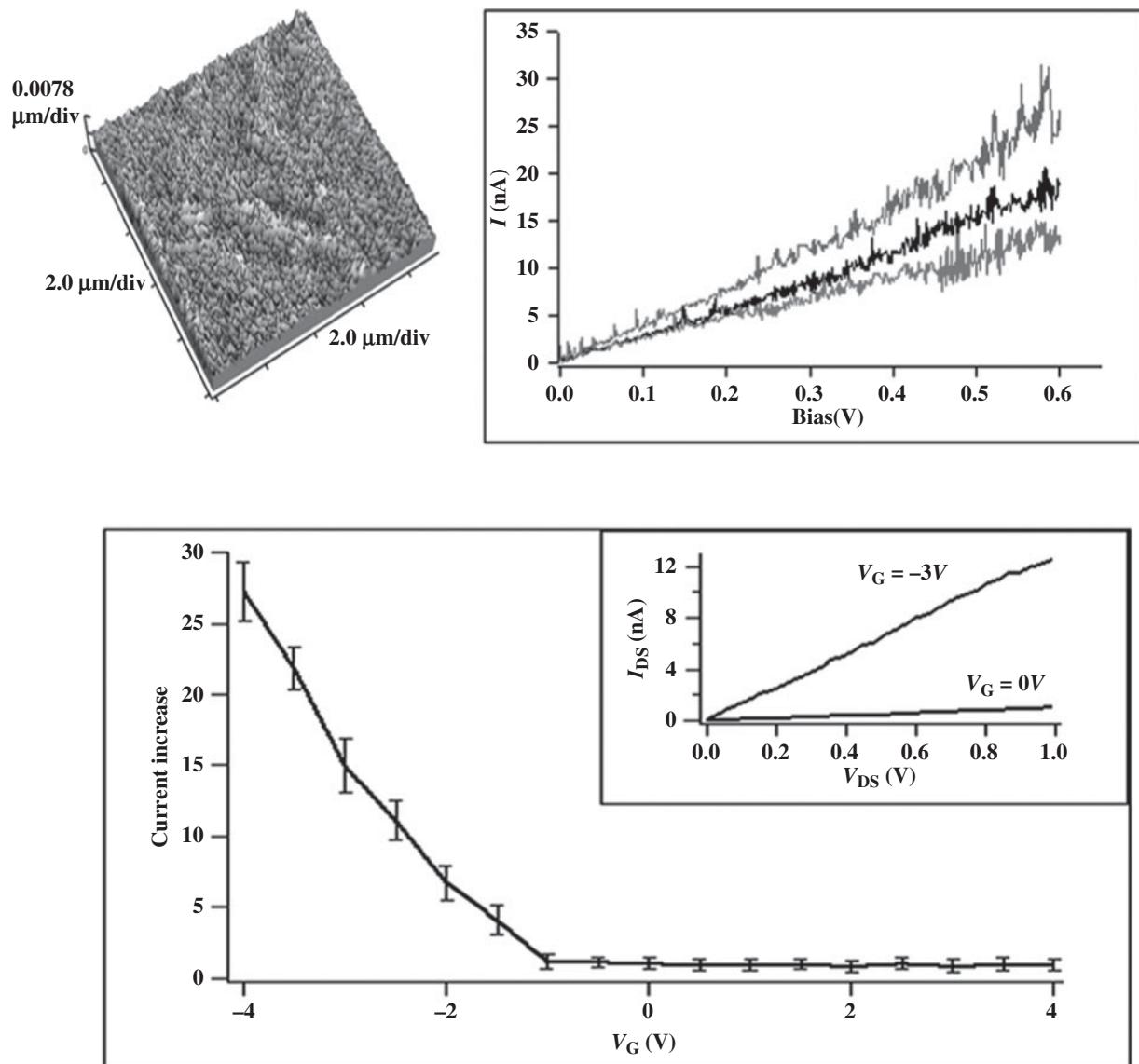
**Fig. 16.6** Protein FET: drain–source current ( $I_{ds}$ ) characteristic vs. the drain–source bias  $V_{ds}$  and gate bias  $V_g$ , in the dark, at room temperature. The pronounced resonance peak is centered at  $V_g = 1.25$  V. In the projection, the same resonance is visible in a transfer characteristic at  $V_{ds} = 5.5$  V. Reprinted with permission from G. Maruccio *et al.* *Adv. Mater.* **17**, 816–822 (2008). © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

tunnel junction under appropriate biasing conditions (Fig. 16.7). The occurrence of on/off resonance conditions in the conductance plot of this device was due to the tunnelling processes across the redox level of azurin and to the long-range ET (Maruccio *et al.* 2007a,b). The results for both current values and peak positions are in good agreement with the signatures of resonant tunnelling and NDR reported in azurin metalloprotein CP-AFM junctions at low applied voltage (Davis *et al.* 2004).

Electrical conduction has also been found in solid-state non-redox proteins, namely disordered multilayers of globular proteins (Pompa *et al.* 2006, 2007)

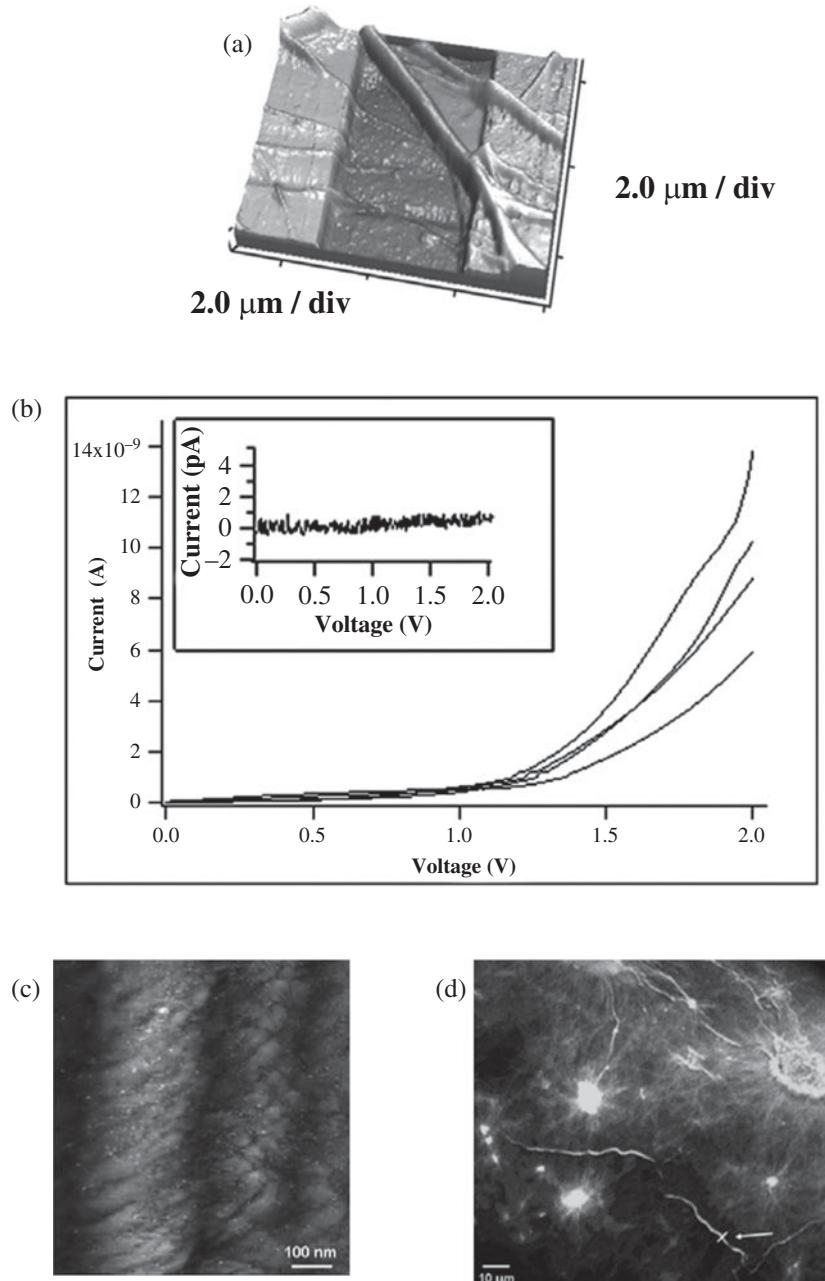


**Fig. 16.7** Normalized conductance of various mesa devices functionalized with Azurin by casting of a 20-mL drop of protein solution plotted as a function of bias voltage.



**Fig. 16.8** (top) Three-dimensional STM image of the protein film and STS characteristics obtained at different points of the biomolecular layer; (bottom) Dependence of the conductivity of the protein film on the gate voltage ( $V_G$ ): the current  $I_{DS}$  (at  $V_{DS} = 1$  V) is plotted as a function of  $V_G$ . The graph has been normalized so that the current value at  $V_G = 0$  V was fixed to one. Inset: Two typical  $I$ - $V$  characteristics, obtained for two different gate voltage values ( $V_G = 0$  and  $-3$  V).

and peptide nanofibrils (Del Mercato *et al.* 2007). In the first case, electrical conductivity was demonstrated both by scanning tunnelling microscopy STM/STS experiments (see Fig. 16.8, left), and by two- and three-terminal transport experiments at the nanoscale. Notably, the conduction of the protein multilayer was found to be strongly dependent on the application of a gate field (Fig. 16.8, right). As far as the fibrillar proteins are concerned, charge transport was experimentally observed by two-terminal transport experiments



**Fig. 16.9** (a) Three-dimensional AFM images of the protein fibrils across the gold electrodes used for transport experiments; (b) Typical current–voltage ( $I$ – $V$ ) characteristics of the nanofibrils. Inset: Control experiments carried out on empty devices (i.e. without fibrils), revealing very low current signals (always less than 1 pA); (c) High-resolution STM image of the nanofibrils deposited onto gold substrates; (d) Label-free imaging of protein fibrils by confocal microscopy.

using micrometer-sized interdigitated electrodes (Figs. 16.9(a) and (b)) and by high-resolution STM imaging experiments (Fig. 16.9(c)). In this latter study, a possible correlation of charge transport with intrinsic blue-green fluorescence processes in the nanofibrils (Fig. 16.9(d)) was also proposed. Both transport and fluorescence in nanofibrils typically displayed very low signals in vacuum conditions, as compared to ambient conditions (Del Mercato *et al.* 2007). A qualitative model for the transport mechanism suggested that charges could

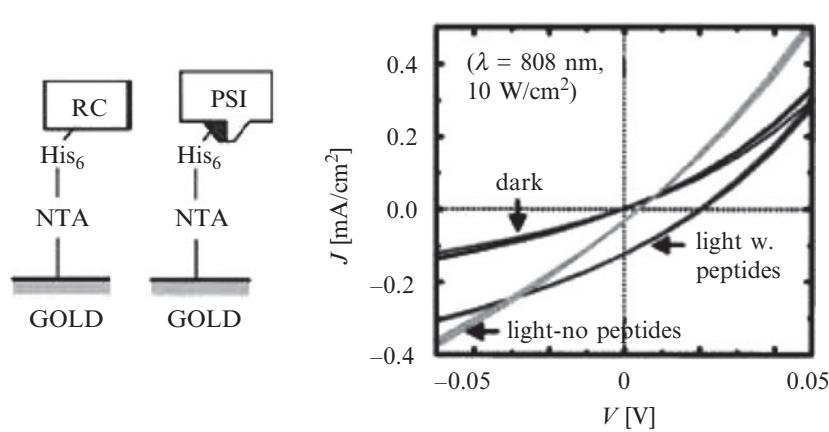
travel through the self-assembled polypeptides, due to the presence of efficient charge-transfer pathways in the protein/fibrillar structures. In such a model, a prominent role of water molecules in the transport mechanism was envisaged, since protein hydration shells are likely to be retained in the solid state and they can significantly support the intra/intermolecular charge-transport pathways along the peptide nanowires. The formation of continuous molecular chains in the solid-state proteins may therefore allow for long-range charge transfer between electrodes.

Once biomolecules of interest have been identified for their use in hybrid functional nanodevices, the major issue remains their assembling, wiring and interconnection to create molecular circuits along with the logic gates of interest. Biologically driven self-assembly could be exploited to reach this goal. As reported above, one promising area is DNA-driven self-assembly. But building complex ordered two- and three-dimensional structures by this technique is difficult, due to the one-dimensional nature of the scaffold. Instead, protein self-assembly can provide useful scaffolds and molds for nanostructures, two- and three-dimensional arrays and nanobiotechnology. A typical example of planar protein assembly is naturally found in bacterial surface layers. Crystalline bacterial surface layer (S-layer) proteins represent the outermost cell envelope component of a broad spectrum of bacteria and archaea (Fig. 16.1(a)) (Sára *et al.* 2005). S-layers are monomolecular arrays comprising of a single protein or glycoprotein species ( $M_w \sim 40\text{--}200\text{ kDa}$ ) and exhibit either oblique, square or hexagonal lattice symmetry with unit-cell dimensions in the range of 3 to 30 nm. Depending on the lattice symmetry one morphological unit (=unit cell) consists of one, two, three, four or six identical S-layer proteins. S-layers are highly porous protein lattices (30–70% porosity) with pores of uniform size and morphology in the 2 to 8 nm range (Fig. 16.1(b)). S-layers are generally 5 to 10 nm thick. Several applications have been demonstrated for S-layers, such as their use as templates for nanoscale patterning of inorganic materials or as ordered anchoring substrates for biomolecules in biosensing (Schuster *et al.* 2005). Colloidal metallic and semiconductor quantum dots (Györvary *et al.* 2004) and fusion proteins were assembled by using S-layers (Moll *et al.* 2002), opening the way to more complex hybrid nanostructured assemblies. Besides proteins, genetically engineered viruses have also been used to order nanostructures, acting as three-dimensional scaffolds (Lee *et al.* 2002; Blum *et al.* 2007). Blum and coworkers (2007) designed, fabricated and measured three-dimensional switchable nanoscale molecular networks on a 30-nm diameter cowpea mosaic virus particle. Photoinduced electron-transport processes in nature, such as photoelectric conversion and long-range electron transfer in photosynthetic organisms, are known to occur not only with high efficiency but also unidirectionally guided by biomolecular functional groups.

The concepts for the development of new functional bioelectronic devices can be inspired by the biological systems such as the electron-transfer chain or the photosynthetic reaction center. In the initial process of photosynthesis, a biological electron-transfer process, photoelectric conversion occurs followed by long-range electron transfer that takes place very efficiently in one direction through the biomolecules (Gust and Moore 1989; Birge *et al.* 1999). The

specific energy and electron transfer take place on a molecular scale due to the redox-potential difference as well as the electron-transfer property of functional molecules, especially the electron acceptor, sensitizer, and electron donor. Another interesting field is of course that of biomolecular photodiodes, recently reported by various groups worldwide (Lee *et al.* 2000; Choi *et al.* 2001; Manoj and Narayan 2003; Choi and Fujihira 2004; Frolov *et al.* 2005). In a typical approach, protein monolayers/heterolayers, like those made of cytochrome c and GFP (Choi *et al.* 2001) or bicomponent SAMs of helical peptides with different chromophores (Yasutomi *et al.* 2004) were used to produce a sizeable photocurrent that can be turned on and off or switched from anodic to cathodic by acting on the irradiating source. In some applications light was used to trigger or switch nanodevices based on pore proteins, like the  $\alpha$ -hemolysin ( $\alpha$ HL) (Ludwig and Bayley 2005). Kramer and colleagues made a light-switchable (as opposed to triggered) K channel (Banghart *et al.* 2004). A channel blocker was anchored to a specific site on the protein (Blaustein *et al.* 2000) through a photoisomerizable arm containing an azobenzene group. In the *trans* configuration, the azobenzene allows the appended blocker, a tetraalkyl ammonium group, to access its binding site at the external mouth of the channel. After irradiation at 380 nm, causing *trans*–*cis* isomerization, the blocker can no longer reach its binding site and the channel is activated causing a 10-fold increase in current flow. Irradiation at 500 nm causes *cis*–*trans* isomerization and the channel is blocked again. Presumably, the tetrameric channel carries four azobenzene blockers and therefore efficient conversion to the *cis* form is required for switching. Other examples of possible applications of proteins in optoelectronics devices are the integration of photosynthetic protein complexes in photodetectors and photovoltaic cells (Das *et al.* 2004, Fig. 10) and the implementation of nanobiotechnological components for memory storage and energy transduction.

For example, assemblies of bacteriorhodopsin, which in nature is a light-driven proton pump containing a photoisomerizable retinylidene chromophore,



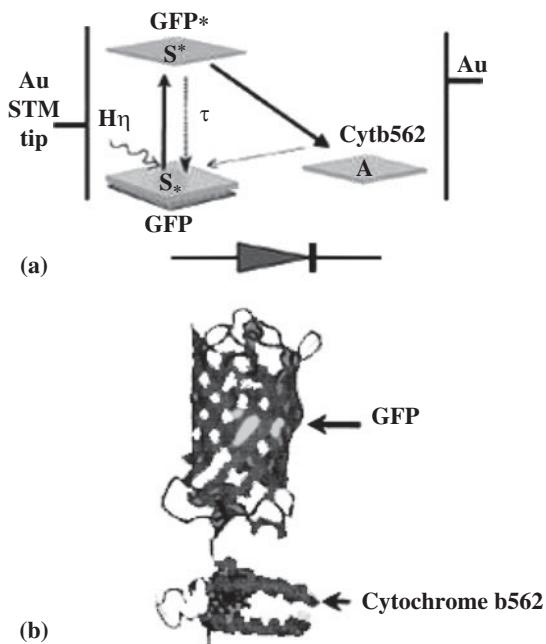
**Fig. 16.10** (a) and (b) Oriented immobilization of photosynthetic protein complexes (RC and PSI) on gold surfaces functionalized with DTSSP and (then)  $\text{Ni}^{2+}$ -NTA. (c) stabilization of RC complexes with A6K/V6D peptides improves the internal quantum efficiency of the devices to 12% under short-circuit conditions. Reproduced with permission from Rupa Das *et al.* *Nano Lett.* **4**, 1079–1083, (2004). Copyright 2008 American Chemical Society.

have been used as photorewriteable information storage devices or as photochromic ink for security cards (Xu *et al.* 2003). Different nanoscale devices were developed by using the bacteriorhodopsin, ranging from molecular motors, molecular phototransistors, molecular gates up to “artificial retina”-like sensors (Pandey 2006). Recently, a bacteriorhodopsin-based memory device was proposed with the possibility to overcome the actual limits of high-density optical recording by using near-field-based approaches (Renugopalakrishnan *et al.* 2007). The green fluorescent protein (GFP) is one of the most used proteins in fluorescence studies due to its very efficient fluorescence emission that also allows single-molecule detection. It exists in two distinct configurations (bright and dark) and recently, Beltram and coworkers succeeded in the optical control of transitions between them (Cinelli *et al.* 2001).

Specifically, they investigated two GFP mutants: EGFP (enhanced GFP) and E<sup>2</sup>GFP obtained by a single-point mutation T203Y of EGFP. A reversible transition between the bright and dark states in E<sup>2</sup>GFP by means of two laser beams with different wavelengths ( $\lambda$ ) = of 476 nm and 350 nm was demonstrated. Unlimited optically controllable cycles between the bright and dark configurations were also reported. The possibility to turn on and off the fluorescence in E<sup>2</sup>GFP could open the way to implementation of dense volumetric GFP-based optical memories exploiting photoconversion from the dark to bright state by irradiation at 350 nm (WRITE step), fluorescence emission following weak excitation at 476 nm (READ step) and photobleaching (switching off the emission, as ERASE step). Based on a similar approach Choi *et al.* (2006) produced a fusion protein, composed by Cytochrome b562 and green fluorescent protein. It was used as an electron acceptor and a sensitizer in the molecular layer by mimicking the photosynthesis. Self-assembled (SA) monolayer of fusion protein was formed on Au-coated glass, by means of micro-contact printing. Electric force microscopy and STM measurements were then performed to demonstrate the photoinduced one-way charge transfer in the SA layer and the subsequent generation of photocurrent (see the scheme reported in Fig. 16.11). Thus, the heterostructure fusion protein of GFP/cytochrome b562 can be used as a nanoscale optically driven biomolecular memory.

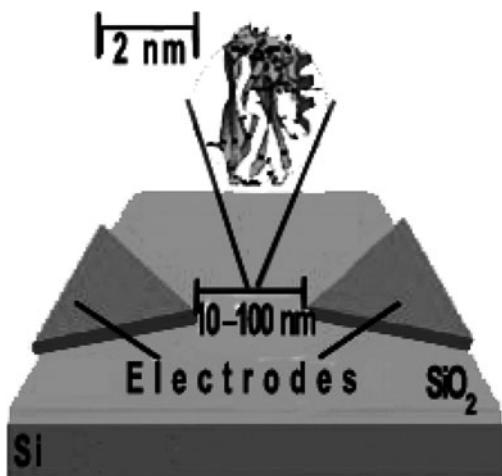
## 16.5 Biophysical implications of protein-based nanobioelectronics

The use of proteins as electron-conductive materials for bioelectronic/biosensing applications has recently attracted great attention, thanks to the possibility of exploiting their self-assembling capabilities and their specific functionalities at the nanoscale. However, a crucial point is the structural stability of biological macromolecules when used in the solid state and in device-like conditions as in biomolecular electronic devices. Resistance of biomolecules to high electric fields is the first important concern for nanobioelectronics/nanobiosensing applications. It is also relevant from a fundamental perspective to understand the dielectric properties and structural dynamics of



**Fig. 16.11** (a) Electron-transfer diagram of GFP/cytochrome b562 fusion protein and (b) structure of fusion protein. Reprinted from Choi J.-W., et al. *Current Applied Physics* **6**, 760–765 (2006). Copyright © 2008 Elsevier B.V.

proteins. In nanoscale devices, biomolecules may experience electric fields as high as  $10^7$  V/m in order to elicit charge transport/transfer (Fig. 16.12). Understanding the effects of such fields on their structural integrity is crucial to assess the reliability of biomolecular devices. A second issue is the structural stability of biological macromolecules when used under ambient conditions. This is, in fact, one of the most striking potential objections to “practical” bioelectronics, and turns out also to be important from a fundamental perspective. The behavior of biomolecules under non-physiological conditions is difficult to predict, and has been little investigated so far.

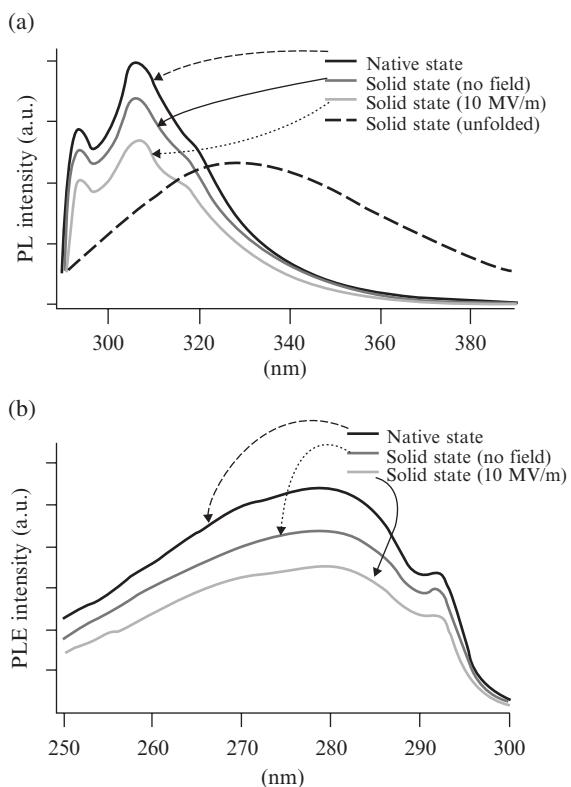


**Fig. 16.12** Schematic representation of a typical configuration for nanoscaled biomolecular devices, in which a protein is immobilized between two metallic tips. Generally, the interelectrode separation is in the 10–100 nm range, so that the application of a bias potential of a few volts results in the generation of electric fields up to  $10^7$  V/m. Investigations on the resistance of biomolecules to high electric fields are therefore of fundamental interest, because the high electric fields applied to nanometric devices are frequently indicated as potential sources of rapid molecular degradation.

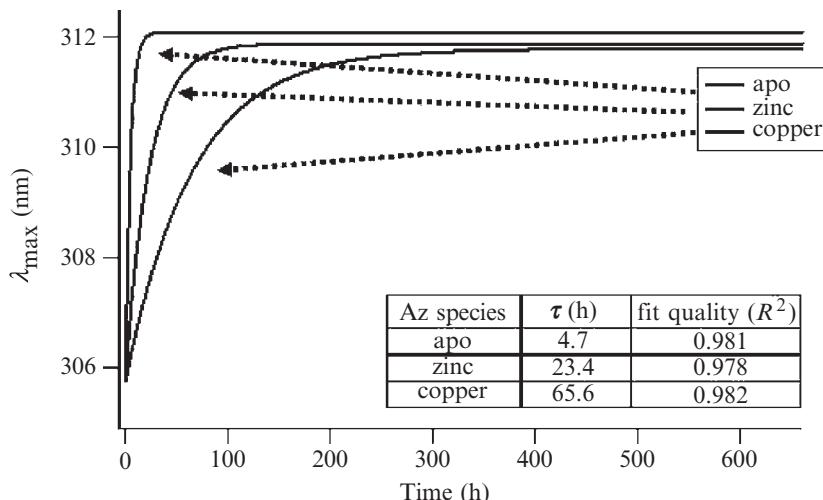
To this effect two main problems are faced, i.e. the interaction between biomolecules and solid substrates (the so-called surface effects), and the conformational transitions possibly induced by the removal of aqueous solvent.

As far as the resistance of “solid-state proteins” to high external electric fields is concerned, the metalloprotein azurin has been recently investigated, as a model system, to probe its response to external fields as high as  $10^7$  V/m (Pompa *et al.* 2005). Azurin was cast deposited onto mm-sized interdigitated electrodes. After the AFM assessment of protein deposition between electrodes, tryptophan fluorescence experiments were performed onto the protein films and revealed that no significant field-induced conformational alteration occurs, even in the case of extremely high electric fields (Fig. 16.13). Interestingly, such remarkable results were found to be supported by theoretical predictions based on X-ray structural data, since the calculated inner electric fields present in the protein had a mean value as high as  $2.45 \times 10^9$  V/m and peak intensities up to  $17 \times 10^9$  V/m. This means that the protein inner fields are orders of magnitude higher than those applied in the experiments, so that the fold pattern of the biased proteins is strongly unperturbed by external field intensities in the range of  $10^6$ – $10^7$  V/m.

As the general features of such inner fields are not peculiar to azurin, these conclusions should have a rather general validity (Pompa *et al.* 2005).



**Fig. 16.13** Intrinsic fluorescence spectra of apoazurin in the native state and in the film environment with and without an applied external electric field. (a) Emission spectra (PL); (b) Excitation spectra (PLE). A fluorescence spectrum of a heat-denatured sample of the solid-state protein is also reported for comparison.



**Fig. 16.14** Aging of the different protein derivatives at ambient conditions ( $20^\circ\text{C}$ , atmospheric pressure,  $\sim 50\%$  of humidity), over more than 600 h, as probed by intrinsic fluorescence experiments: emission maximum ( $\lambda_{\max}$ ) as a function of time from the solvent removal instant ( $\lambda_{\max}$ -vs.-time data were least-squares fitted with exponential curves (of the  $1 - e^{-t/\tau}$  type). The remarkable difference between the time constants ( $t_{\text{apo}}$ ,  $t_{\text{Zn}}$  and  $t_{\text{Cu}}$ ) is consistent with the increasing structural robustness of the three derivatives.

The biomolecules were deposited onto silicon dioxide surfaces and subsequently dehydrated, resulting in the formation of protein films characterized by good uniformity and an average thickness of 15–20 nm (approximately corresponding to 4–5 azurin layers), as probed by tapping mode AFM measurements. The dry molecular films were maintained at ambient conditions through several weeks, and their conformational changes were investigated by intrinsic fluorescence spectroscopy. The experimental evidence indicated a weak initial conformational rearrangement, followed by long-term stability (Fig. 16.14).

The aging effects were studied on wild-type azurin, as well as on two protein derivatives (apo- and zinc- forms). The results obtained in the solid state were consistent with the expected structural stability of the three species (Fig. 16.14). Moreover, upon rehydration of the biomolecular films at the end of the investigated period (approximately one month), azurin returned to exhibit a native-like conformation. Such a behavior suggests that, apart from the initial slight rearrangement, the proteins do not experience denaturation even after weeks in a non-physiological environment. It seems that proteins under ambient conditions move towards an equilibrium state, after which the new conformational pattern is no longer perturbed. This phenomenon may account for a somewhat unexpected stability of proteins in the solid state, which possibly stems from the retention of their hydration shells, also under these unusual experimental conditions. Such experimental results shed a somewhat unexpected positive light on the reliability in biomolecular electronics. The observed stability of azurin in non-physiological conditions confirmed its nature as a good candidate for electronic applications, making it worthy to take deeper investigations of other biological macromolecules.

## 16.6 Nanodevices for biosensing

The detection and transduction of signals from specific and single molecules is the basic requirement of a biosensing device. Different approaches have

been developed in the last decades in order to make biosensing devices more selective, stable and sensitive. The progress achieved in manufacturing micro- and nanosystems has given a new boost to this field, envisioning the possibility of integration, automation and parallelization in biosensing devices with the benefit of higher data quality, reproducibility and reagents saving. The need for a rapid and sensitive detection system in multiplexing channels is a common task of any biological analysis tool. Depending on the biological target and/or probe in the recognition and/or quantification analysis, different approaches and transduction systems were developed. In particular, for proteins, the recent developments are devoted to the integration of the standard biosensors on a micro- and nanoscale mimicking electronic, scalable and portable devices, like for example in standard CMOS technology. In this direction the devices developed using MEMS, and recently NEMS, technology (Bashir 2004) have many preferable advantages such as miniaturized size, mass production, low power consumption, and compatibility with IC manufacturing processes. The development of polymer lab-on-a-chip has evolved rapidly over the last decade and has shown numerous applications in biochemical analysis operations such as genomics analysis, proteomics analysis, and clinical analysis because of the high integration, low sample and reagent volume, short analysis time, less sample wastage, material cost effectiveness, and potential for making disposable devices (Oosterbroek and Van den Berg 2003; Li 2006).

### 16.6.1 FET-based biosensors

Interest in biosensors based on field effect transistors lies in the possibility to take advantage of the recent large-scale integration technologies for electronic circuits and provide economical and high-speed biodetection capabilities. Moreover, the advance in fabrication technologies reported in the previous section could further shrink the gate channel to the nanoscale, for few-molecule immobilization and detection. The electronic transduction based on these electronic devices is label free and is commonly based on two general strategies. In one method the degree of charging of the gate surface after the biorecognition events is used to control the gate potential, thus providing the detection of binding process (Bergveld 2003; Park *et al.* 2004; Dzyadevych *et al.* 2006; Schoning and Poghossian 2006).

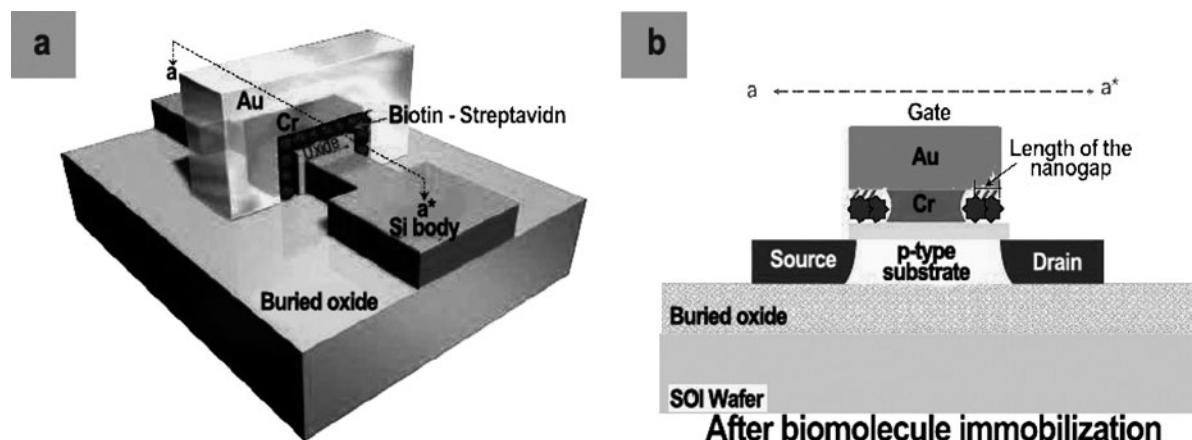
In the second one, the gate is modified by a redox-active unit, and the biocatalytic process alters the equilibrium of the reduced/oxidized state of the modifier, thus controlling the gate potential (Zayats *et al.* 2000; Freeman *et al.* 2007). These approaches are also suitable for on-chip integration of both the sensor and the measurement systems.

However, in order to increase the device sensitivity and reduce the number of biomolecules binding on the sensor, a dielectric-modulated FET nanogap device was proposed by Im *et al.* (2007). A three-dimensional picture (left) and cross-sectional scheme of the device is reported in Fig. 16.15. The device was realized by conventional metal-oxide semiconductor processes, and the nanogap was realized by depositing a thin layer of chromium underneath a

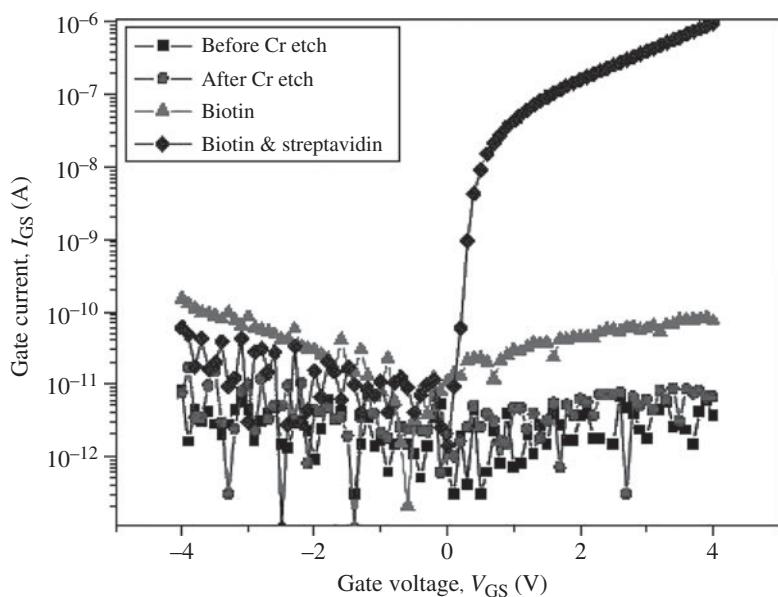
thicker gold layer, followed by etching. The realized gap was 200 nm wide and 15 nm high. The sensor was used to detect specific binding of streptavidin to biotin. The binding of the streptavidin changed the dielectric constant of the gate, resulting in a shift of the threshold voltage. In Fig. 16.16(a) a representative set of  $I_{GS}$ - $V_{GS}$  curves after the immobilization and binding events in one device is shown. Following a similar approach, an ion-sensitive field effect transistor (ISFET) was modified to follow a protein kinase activity (Freeman *et al.* 2007). The protein kinases are a large family of enzymes that modulate the activity of proteins by phosphorylation. They regulate the majority of cellular pathways, especially those involved in signal transduction. One of the most versatile protein kinases is casein kinase II (CK2), a serine/threonine-selective protein kinase, which can phosphorylate more than 160 proteins. It has been shown that CK2 is involved in signal transduction, transcriptional control, apoptosis, cell cycle and more. In biochemistry, to quantify the CK2 activity a radioactive or fluorescent labelling is required, while the electronic detection based on FET devices can allow label-free detection. The biodevice implementation is schematically depicted in Fig. 16.17. Peptides were covalently linked to the aminosylane functionalized gate interface and then rinsed with water and HEPES buffer solution. Figure 16.18(A) shows the measured change in the gate-to-source potential upon the treatment of the peptide-functionalized gate with CK2 and adenosine triphosphate as a function of time. These experiments were consistent with the fact that CK and ATP specifically phosphorylate the sequence-specific peptide. The resulting negatively charged phosphate units altered the gate potential and led to the observed potential changes. As the reaction with the enzyme/ATP was prolonged the content of the phosphate units increased, leading to higher  $\Delta V_{GS}$  values. The saturation of the surface groups by the phosphate units produced a saturation value of  $\Delta V_{GS} = 40$  mV. The authors also demonstrated the reusability of the device upon regeneration of the active sensing surface upon the application of a hydrolyzing enzyme.

Integration of nanowires or carbon nanotubes as sensing elements into devices could improve the device sensitivity, because of their reduced size. In particular, the large ratio between the surface extension and the volume increases the sensitivity and reduces the dissipation of current in the device. A self-assembling process is usually employed to realize these kind of sensors (Cui *et al.* 2001; Besteman *et al.* 2003), but this approach is scarcely compatible with modern micro- and nanoelectronic fabrication methods and becomes problematic for large-scale integration. To overcome this limitation Elibol *et al.* proposed a fabrication method based on top-down microelectronics processing techniques, to produce devices with silicon nanowires of a few tens of nanometer in diameter, onto which sensing biomolecules can be immobilized, as depicted in the schematic design of Fig. 16.19 (Elibol *et al.* 2003; Bashir 2004).

CNT-based transistor biosensors have emerged in recent years for efficient and label-free detection of a variety of biomolecules (Kim *et al.* 2007). Apart from their unique 1D quantum confinement properties, which have been exploited in different applications (Kong *et al.* 2000; Yao *et al.* 2000; Appenzeller *et al.* 2002; Heinze *et al.* 2002; Li *et al.* 2003; Kymakis *et al.* 2003; Wu *et al.* 2004; Gruener 2006), the main reasons for CNTs incorpora-

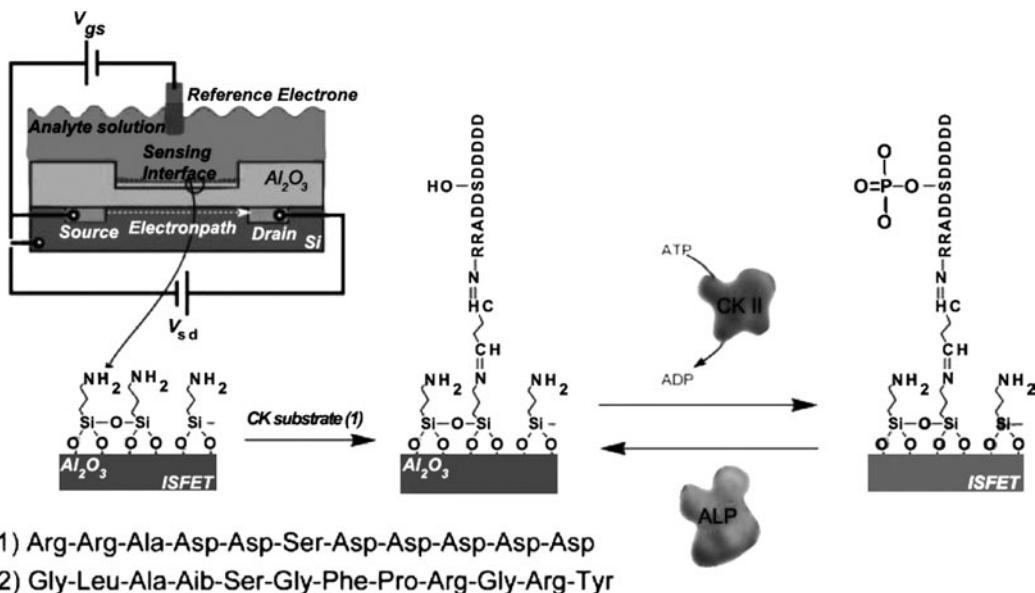


**Fig. 16.15** (a) Three-dimensional structure showing the silicon body, the gate oxide, the chromium and gold electrodes. The narrow region of the silicon body forms the active silicon region (the channel). The chromium layer is partially etched to form an air gap that can be filled with biomolecules. (b) A cross-section of the device along the direction of the dotted line in (a) shows the filled nanogap. Source and drain electrodes are formed by implanting phosphorous in the silicon body. Reprinted from Im H., et al. *Nature Nanotechnol.* **2**, 430 (2007), Copyright © 2008 Nature Publishing Group.



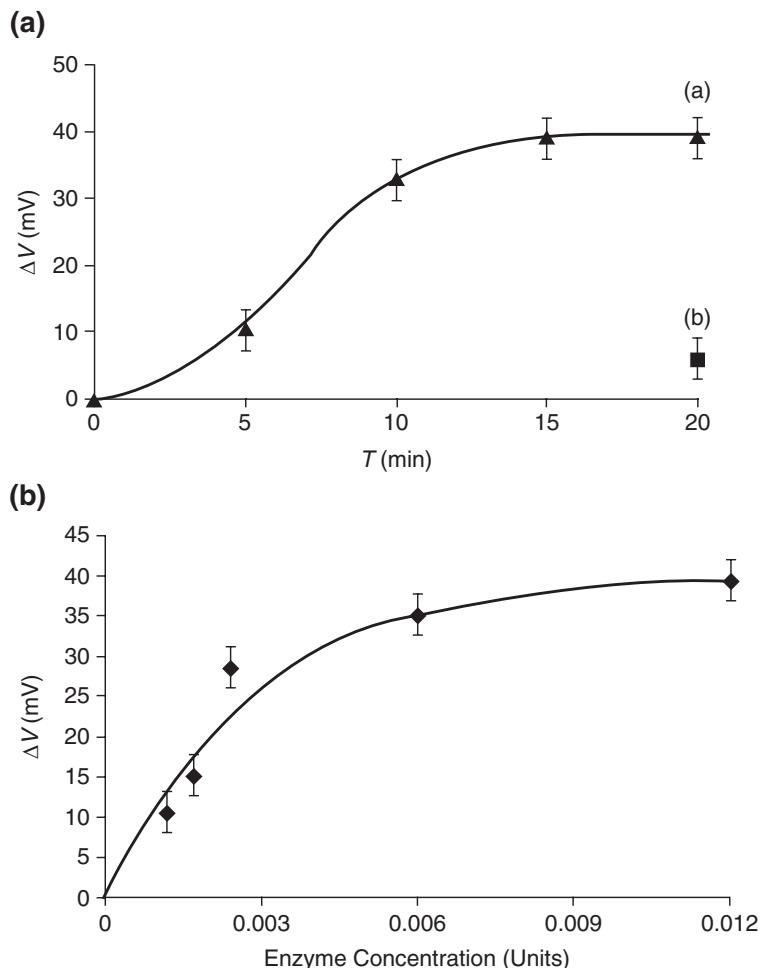
**Fig. 16.16** Electrical characteristics of the DMFET nanogap device before and after biomolecules are immobilized in the nanogap.  $I_{GS} - V_{GS}$  characteristics in one device at  $V_{DS} = 0.05$  V. Reprinted from Im H., et al. *Nature Nanotechnol.* **2**, 430 (2007), Copyright © 2008 Nature Publishing Group.

tion in biosensing devices rely on: (i) the high surface area of single-walled CNTs (SWNTs) that influence their electrical conductivity; (ii) diameter similar or smaller than those of individual proteins; (iii) the simple assembling methods based on bottom-up technologies and soft lithography; and (iv) the wide choice of covalent and non-covalent functionalization strategies to link a variety of biological entities (Chen *et al.* 2003; Balasubramanian and Burghard 2006). Two device architectures have been developed

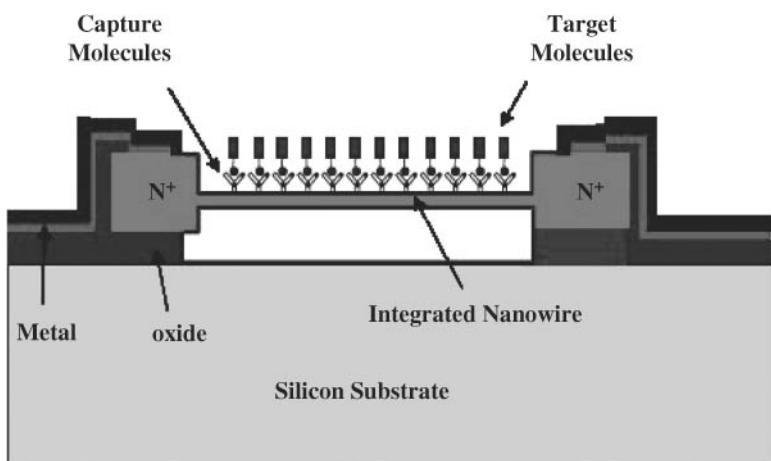


**Fig. 16.17** Modification of an ISFET device for the specific analysis of casein kinase II activity and for the recycling of the sensing interface by alkaline phosphatase. Reprinted from Freeman R. *et al.*, *Chem. Commun.* **33**, 3450 (2007), Copyright © 2008 The Royal Society of Chemistry.

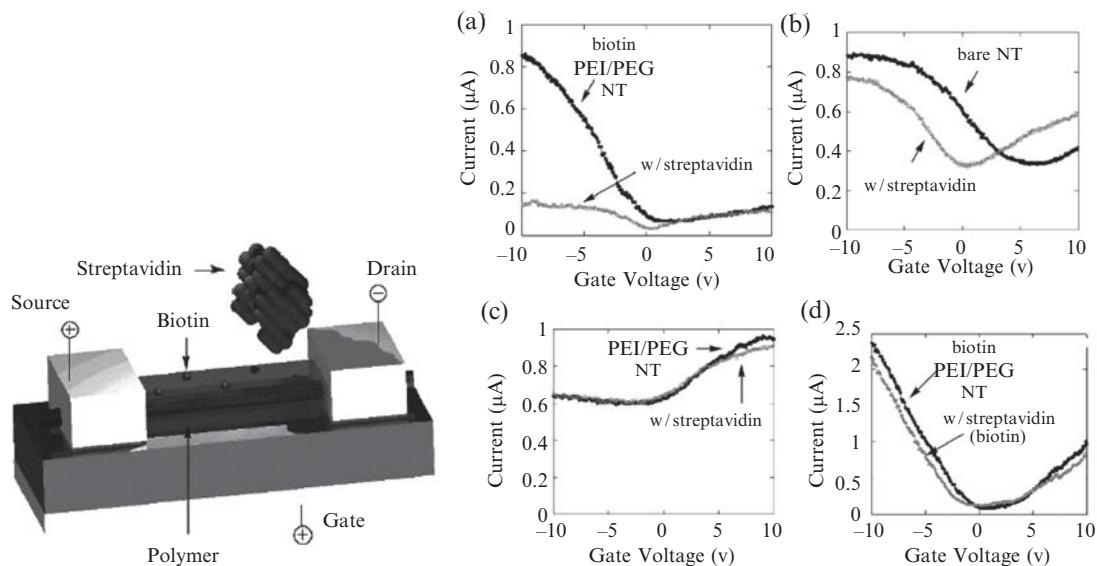
and explored by various groups working in this area. In one case a single CNT connects the source and the drain in a configuration that is shown in Fig. 16.20 (left). These devices exhibited very high sensitivity but were poorly reproducible in terms of device assembly and detection response. In the second case, a nanotube network was deposited in the conducting channel, showing less sensitivity but higher reproducibility and an easy integration procedure (Gruener 2006). Sensors based upon SWCNT-FETs can be operated in two different ways. One possibility is to monitor the conductance of an individual SWCNT or a network of SWCNTs during the introduction of the analyte solution. In this chemiresistor configuration, the resistance of the device is directly or inversely proportional to the concentration of the analyte. A second method is to measure the complete field effect modulation of conductance after introduction of the test solution. This latter methodology is referred to as chemFET, where the threshold voltage shift provides information about the analyte concentration. Different proteins and antibodies binding events, ligand–receptor interactions, enzyme activities and virus detection have been implemented onto these devices (Balasubramanian and Burghard 2006). A typical example of a modified carbon nanotube FET capable to detect biotin-streptavidin binding is presented in Fig. 16.20(a) and was demonstrated by Star *et al.* (2003). In response to biotin-streptavidin binding, they observed a clear change in the device characteristics (Fig. 16.20) with control experiments supporting the specificity of such biosensor. The biosensing mechanism of these carbon-nanotube biosensors has been extensively studied (Kuzmych and Star 2007). Recently, Heller and coworkers (2007) found that the sensing of biomolecules in these device contacts were passivated by



**Fig. 16.18** (A) (a) Time-dependent changes in the source-to-gate potential upon the treatment of the (1)-functionalized device with CK2, 0.012 U, and ATP, 100 mM. (b) Potential change of the (2)-modified device upon treatment with CK2, 0.012 U, and ATP, 100 mM, for 20 mins. (B) Source-to-gate potential changes upon phosphorylation of the (1)-functionalized device with different concentrations of CK2 in the presence of ATP, 100 mM, for 20 min. Readouts were made in 20 mMTris-HCl buffer solution, pH = 7.5 that included 50 mM KCl and 10 mM MgCl<sub>2</sub>. Reprinted from Freeman R., et al. *Chem. Commun.* **33**, 3450 (2007), Copyright © 2008 The Royal Society of Chemistry.



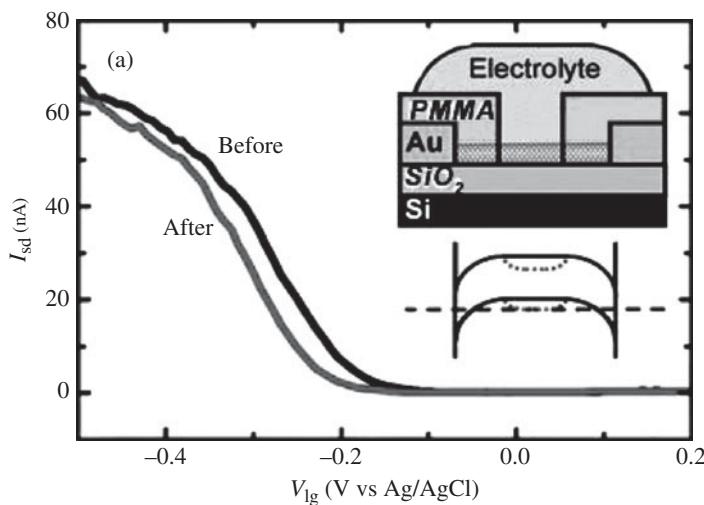
**Fig. 16.19** Schematic design of the device fabricated by Elibol *et al.* (2003). Reprinted from Bashir R. *Advanced Drug Delivery Reviews* **56**, 1565 (2004) Copyright © 2008 Elsevier B.V.



**Fig. 16.20** Left: Schematic of the modified carbon-nanotube (CNT) FET, functionalized with a molecular receptor (biotin) and successively exposed to streptavidin. Right: Change of the device characteristic due to biotin-streptavidin binding. (a) Response—in the absence and presence of streptavidin—of a biotinylated, polymer-coated CNT-FET, (b) of a bare CNT-FET and (c) of a polymer-coated CNT-FET device. (d) Response of a biotinylated, polymer-coated CNT-FET in the absence and presence of streptavidin previously incubated with biotin. Reproduced with permission from A. Star *et al.* *Nano Lett.* **3**, 459–463, (2003). Copyright 2008 American Chemical Society.

poly(methyl-metacrylate) (PMMA) and solution access was restricted to an exposed region along the bulk of the SWNT (see Fig. 16.21).

After addition of poly-L-lysine solution they observed a clear shift in the  $I-V_{lg}$  (liquid gate voltage) curves, concluding that work-function modulation at the metal-SWNT contact cannot be responsible for this change in  $I-V_{lg}$  curve.

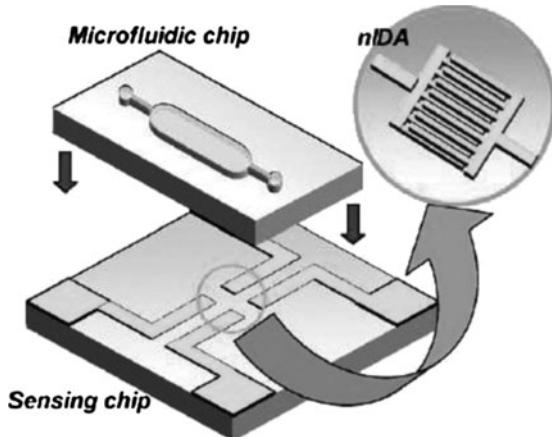


**Fig. 16.21**  $I-V_{lg}$  curves before and after PLL adsorption on modified SWNT devices. 200 nM PLL adsorption on a SWNT device where the electrodes and contact area are passivated by PMMA. Top and bottom insets illustrate device architecture and band-diagram changes upon protein adsorption, respectively. Reprinted from Heller, I., *et al.* *Nano Lett.* **8**, 591, (2007), Copyright © 2008 American Chemical Society.

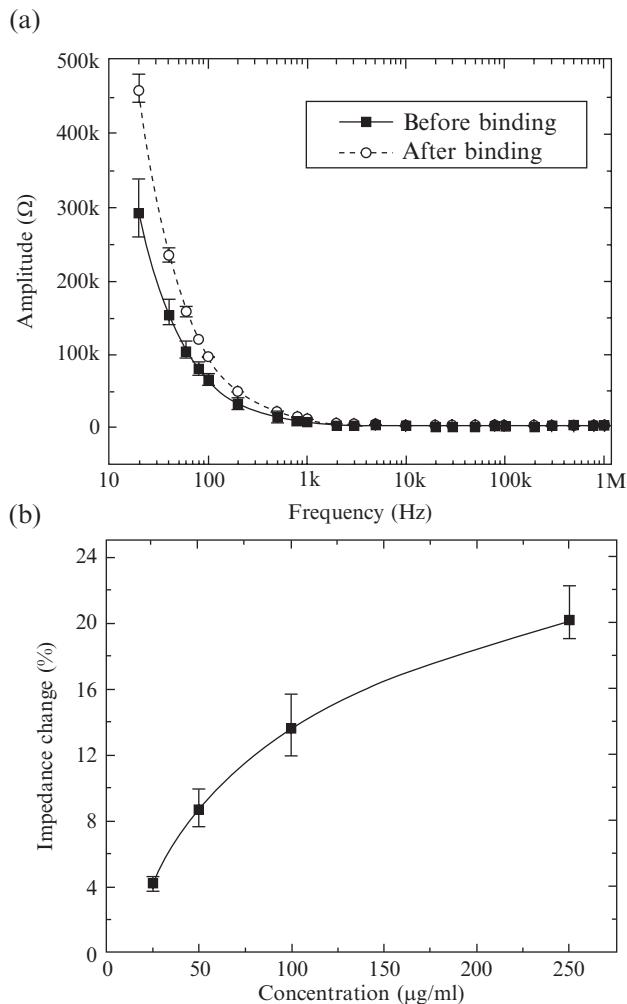
All these bioelectronic devices can be coupled to microfluidic systems (Jeon *et al.* 2000; Dertinger *et al.* 2001; Juncker *et al.* 2005; Malaquin *et al.* 2007) in order to promote automation, high-throughput analysis and screening. They are platforms comprising microchannel system and liquid reservoirs fully integrated and can combine sampling, sample enrichment, mixing, reaction modules (different heating zones), product separation, isolation and analysis. Usually polymers, plastic materials or glass are used to implement microfluidic networks.

Microfabricated networks of micrometer-sized channels have already proven useful for applications in chemical separations (e.g. capillary electrophoresis), biochemical assays, DNA analyses, medical diagnostics, drug delivery, cell manipulation and sorting (Haeberle and Zengerle 2007). Electrochemical detection schemes can be easily integrated with these systems on different scales, allowing fast and reproducible label-free biosensing. Electrochemical impedance spectroscopy for example is mainly based on interdigitated electrodes patterned on a solid substrate. While most of the reported integrated impedimetric biosensors have been fabricated on silicon or glass substrates (Athreya and Martin 1999; Ayliffe *et al.* 1999; Gomez-Sjoberg *et al.* 2005) polymers have recently been considered as the most suitable substrate material for micro-total analysis systems (microTAS) and lab-on-a-chip (Manz *et al.* 1990; Ohori *et al.* 1998) devices. Nanointerdigitated gold electrode arrays integrated in a microfluidic chip were used for direct bioaffinity sensing of protein binding (for example, mouse monoclonal antirabbit immunoglobulin, IgG), as reported in Fig. 16.22 (Zou *et al.* 2007).

Figure 16.23(a) shows that impedance changes were clearly detectable in the low-frequency range (20 Hz–1 kHz). Accordingly, the relative impedance changes at 100 Hz with respect to different mouse antirabbit IgG concentrations down to 25 µg/ml were tested (as reported in Fig. 16.23(b)). As control, pure PBS buffer solution without mouse antirabbit IgG was added into the sensing chamber and incubated for 10 min and no impedance change was observed. For the reproducibility, less than 15% chip-to-chip deviation was found in this study. In a more complex approach, Kim and coauthors developed



**Fig. 16.22** Schematic view of the gold nIDA on polymer substrate with a polymer microfluidic chip. Reprinted from Zou *et al.* *Sens. Actuators A* **136**, 518 (2007) Copyright © 2008 Elsevier B.V.

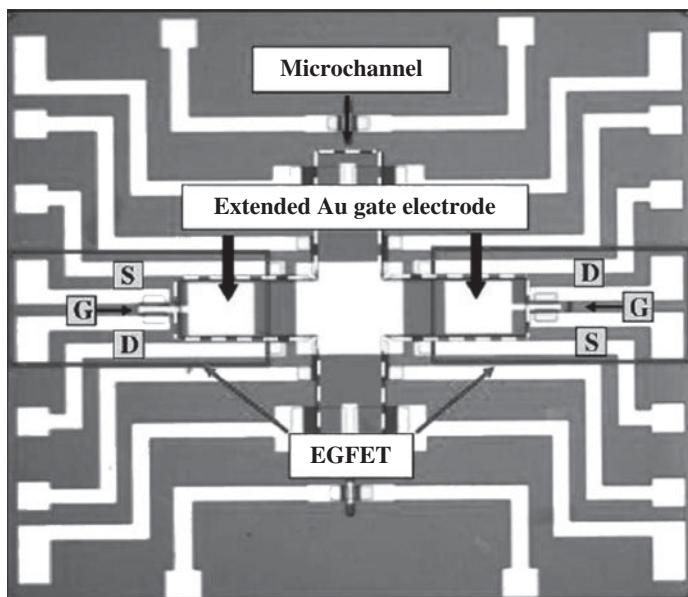


**Fig. 16.23** Impedimetric response: (a) before and after protein binding (0.25 mg/ml mouse antirabbit IgG) and (b) calibration curve of relative impedance change vs. different mouse antirabbit IgG concentrations at 100 Hz. Reprinted from Zou, *et al.*, *Sens. Actuators A* **136**, 518 (2007), Copyright © 2008 Elsevier B.V.

and fabricated an EGFET-based protein sensor in which an extended gate was formed at the bottom of a microfluidic channel across the (111) silicon wall (Fig. 16.24). This device was realized using both semiconductor IC process technology and MEMS technology in order to detect streptavidin–biotin protein complexes. Changes in the drain current were detected after the binding of the different molecular layers in the channel (Kim *et al.* 2006).

### 16.6.2 Biosensing with ion-channel membrane proteins

Transmembrane ion-channel proteins form pores through the highly resistive cell membrane and allow diffusion of ions and small molecules from one side to the other. One of the unique features of ion-channel proteins is that they show gating mechanisms superficially similar to transistors when a bias is applied to the gate (Sigworth and Klemic 2002). In the case of voltage-gated proteins, a potential applied across the membrane results in a change of the



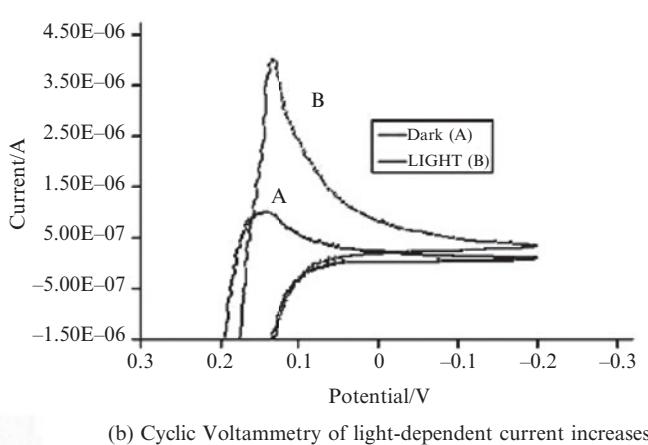
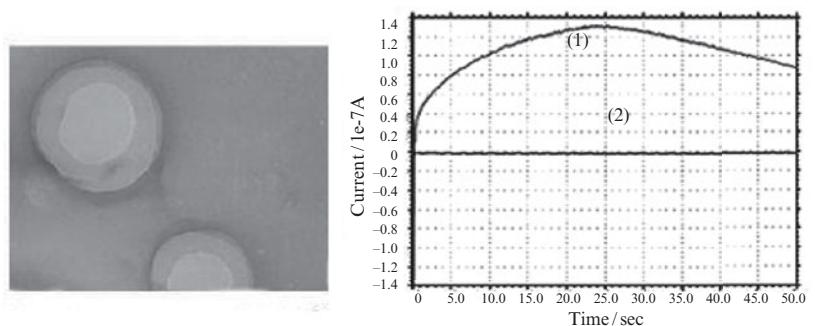
**Fig. 16.24** Photograph of the fabricated device. Reprinted from Kim *et al.* *Sens. Actuators B* **117**, 488 (2006), Copyright © 2008 Elsevier B.V.

conductance of channels embedded within. Ligand-gated channels exhibit similar conductance changes when a binding event occurs. Another unique feature of ion-channel proteins is that some are very specific towards ionic species passing through their pores. For example, voltage-gated potassium channels allow  $K^+$  ions to permeate at least 10 000 times more easily than  $Na^+$  (Doyle *et al.* 1998). The combination of signal transduction via gating mechanisms and high selectivity towards specific molecules makes ion-channel proteins a highly desirable candidate for biological sensors.

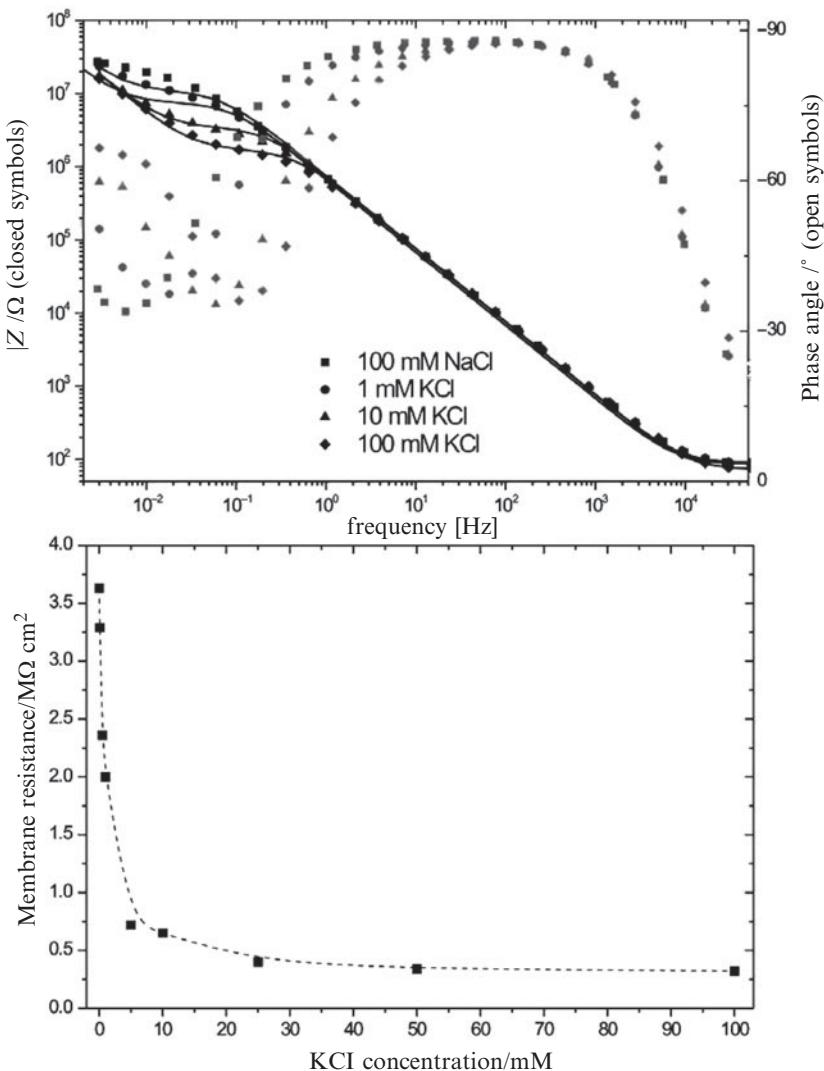
However, these proteins can work properly only if correctly embedded in the cell membrane. Recently, it has been shown that  $\alpha$ -hemolysin, a non-specific bacterial toxin channel protein, can be genetically engineered, modifying the inner pore regions so that specific analytes can be detected and identified (McGeoch *et al.* 2000; Bayley and Cremer, 2001; Bayley and Jayasinghe, 2004). This technique results in two types of information about the analyte molecule. The first is the actual magnitude of the conductance change due to particular molecules bound in the pore and is measured as a change in the current flowing through the protein. The second is the amount of time the molecule spends bound inside the pore. This information provides stochastic signatures of particular analytes using the same protein. The incorporation of engineered proteins into ion-channel measurement platforms will result in a highly selective sensor capable of providing stochastic information of particular analytes of interest. The natural biomembrane is a highly complex system characterized by high specificity and has a considerable number of functions available. The membrane is composed of a lipid bilayer with embedded membrane proteins. This highly engineered biological system is involved in both biological signalling processes between intracellular and extracellular

ambients and between the cells. The use of this natural architecture in biosensing would result in fast and efficient devices.

However, it is found difficult to connect the cell membrane in actual devices, for practical use and applications. Some groups realized synthetic membrane architectures that mimic some of the features of a natural membrane. The most common approaches to study protein–protein interactions use liposomes or vesicles as membrane models (Bernard *et al.* 2000; Apel *et al.* 2002; Sofou and Thomas 2003). However, these models are rather limited to electrical measurements or for practical applications. For example, Xi and coworkers (Xi *et al.* 2005) embedded two proteins, namely the bacteriorhodopsin (BR) and the cytochrome C oxidase (COX) in polymeric vesicles. These two proteins serve as “vectorial catalysts”, being able to produce unidirectional ion transport by converting light (BR) and chemical energy (COX) (Fig. 16.25). In this respect, solid-supported artificial membranes are promising systems that mimic the principal features of natural cell membranes (Cornell *et al.* 1997; Knoll *et al.* 2000; Krishna *et al.* 2003; Naumann *et al.* 2003; Atanasov *et al.* 2005; Terretaz and Vogel 2005). Transport across the membrane is only possible due to the function of incorporated proteins, e.g. ion channels, transporters or proton-pumping proteins. Additionally, the lipid bilayer is usually anchored via a spacer group to a solid support. This spacer serves as a stable anchoring moiety and provides an ion reservoir. In turn, the solid support provides the possibility



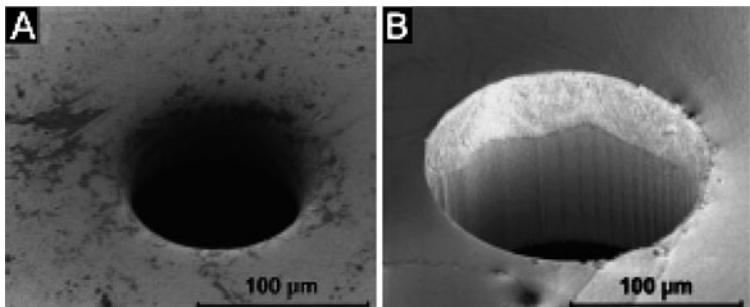
**Fig. 16.25** (a) Transmission electron microscopy (TEM) image of polymeric vesicles with diameters of approximately 250 nm. (b) Light-dependent current increase of the composite vesicles during cyclic voltammetry measurements. (c) Amperometric current detection as a function of time, (1) vesicles with proteins, (2) protein control. Reprinted from Xi *et al.* (*Adv. Funct. Mater.*) **15**, 1233, (2005) Copyright © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.



**Fig. 16.26** Incorporation of valinomycin into a DPTT-based tBLM. Top: Bode plot of the impedance and the phase shift as a function of frequency. Solid lines represent fits using an equivalent circuit. For better visibility, only the impedance fits are shown. Bottom: Bilayer resistance as a function of the potassium concentration in the electrolyte (ionic strength 100 mM). The dotted line is a guide to the eye only. Reprinted from Vockenroth *et al.* *Sensors* **608**, (2005) Copyright © 2008 IEEE.

to integrate the membrane architecture into micro- or nanoelectronic or optoelectronic read-out systems for biosensing.

A typical example is reported in Fig. 16.26, where electrochemical spectroscopy impedance measurements performed on a bilayer with Valinomycin are reported. Valinomycin is a small-ion-carrier peptide, which is known to diffuse into the membrane (Totu *et al.* 2001; Naumann *et al.* 2003). It is highly selective for potassium ions, which can be transported from one side of the bilayer to the other. If the peptide is functionally incorporated, the resistance of the membrane should thus decrease with increasing potassium concentration in the electrolyte. With increasing  $K^+$  concentration the resistance decreases as expected, first rapidly then slower until it approaches asymptotically a small fraction of the original bilayer resistance for high potassium concentrations.

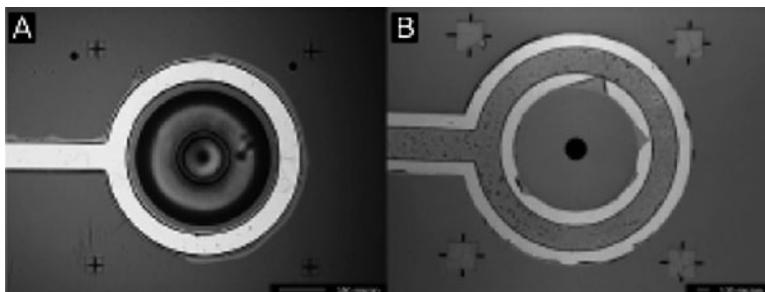


**Fig. 16.27** (A) Tilted plan-view SEM image of an aperture in a 25- $\mu\text{m}$  thick Teflon sheet created by electrostatic discharge. Due to the nature of the hole-formation process the sidewall is completely rounded and smooth. (B) SEM image of an aperture formed in silicon using deep reactive-ion etching. The sidewalls are vertical, exhibiting the roughness induced by the cyclic etch and deposition process. The small notches close to the edge of the aperture are due to plasma punch-through effects before the material in the center is removed. Reprinted from Wilk *et al.* *Biosens. Bioelectron.* **23**, 183 (2007), Copyright © 2008 Elsevier B.V.

At the same time, the membrane capacitance remains almost constant. After rinsing with 100 mM NaCl solution, the membrane resistance increases again to a value of  $2.6 \text{ M}\Omega \text{ cm}^2$ . Similar experiments were conducted with other ion-channel proteins, for example Gramicidin (Alonso-Romanowski *et al.* 1995; Jing *et al.* 1998; Andersen *et al.* 1999) or M2  $\delta$  (Oblatt-Montal *et al.* 1993; Opella *et al.* 1999; Montal and Opella 2002). These devices can be further engineered and miniaturized by using standard microelectronic fabrication materials and methods.

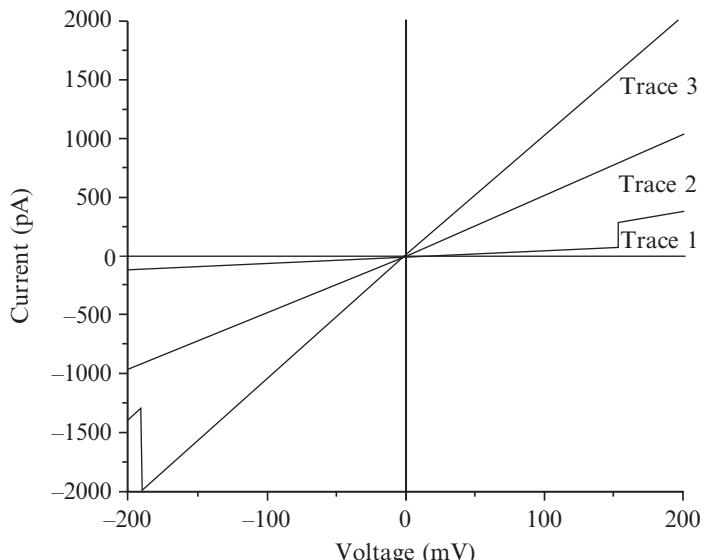
Wilk *et al.* (2007) developed a microfabricated silicon system with integrated patterned electrodes for low measurements of suspended bilayers with transmembrane ion-channel proteins (Figs. 16.27 and 16.28). As reported above, some other class of membrane proteins are sensitive to small molecules. The Vertebrate olfactory system belongs to this family and is capable of distinguishing among hundreds of thousands of different volatile molecules exhibiting diverse functional groups, branching patterns, charges or unsaturated bonds, size and shape. After device fabrication, lipid bilayers were formed using painting techniques such that a small amount of lipid in decane solvent was brushed across the microfabricated aperture. Following bilayer formation, a voltage ramp was applied across the membrane and the slope of the resultant current–voltage ( $I$ – $V$ ) trace was used to determine the resistance of the membrane.

Figure 16.29 shows the insertion and measurement of OmpF porin using the microfabricated device with Ag/AgCl electrodes. The ORs mediating the perception of odorants belong to the large superfamily of G-protein coupled receptors (GPCRs) (Vidic *et al.* 2006). These proteins have attracted great attention for the possibility to implement a bioelectronic olfaction

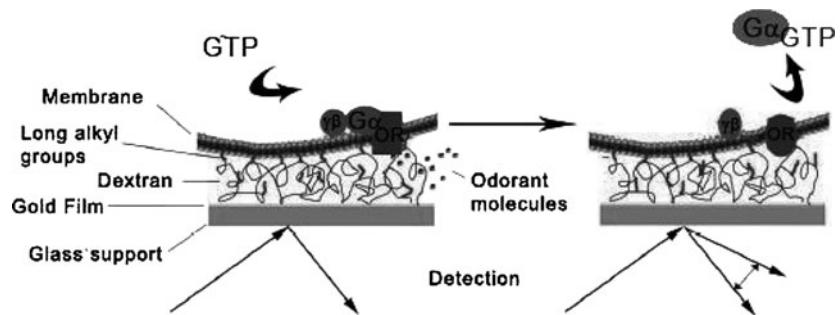


**Fig. 16.28** (A) Micrograph of the silicon integrated device from the backside with an additional 10- $\mu\text{m}$  thick layer of SU-8 2010 patterned 15  $\mu\text{m}$  over the edge of the Ag/AgCl electrodes. (B) Micrograph viewed from the front-side of the device showing the lifted-off PTFE layer around the silver electrode. The 150- $\mu\text{m}$  aperture in the middle is completely coated with a 30-nm PTFE layer. Reprinted from Wilk *et al.* *Biosens. Bioelectron.* **23**, 183 (2007), Copyright © 2008 Elsevier B.V.

**Fig. 16.29** Measurement of the insertion of OmpF porin using the microfabricated integrated electrodes on the surface of the silicon device in 0.50 M KCl solution. Trace one shows the baseline resistance with the insertion of an OmpF porin protein and the open state of two channels. Trace two shows the measurement of two complete proteins (6 open channels,  $t = 19$  s) while trace three shows the measurement of additional open channels (13 total open channels,  $t = 172$  s). The time with reference to the first baseline plot shows typical rates of porin insertion. The number of channels was determined using a single-channel conductance of 0.7 nS, which is typical for OmpF porin in 0.5M KCl. Reprinted from Wilk *et al.* *Biosens. Bioelectron.* **23**, 183 (2007), © 2008 Elsevier B.V.

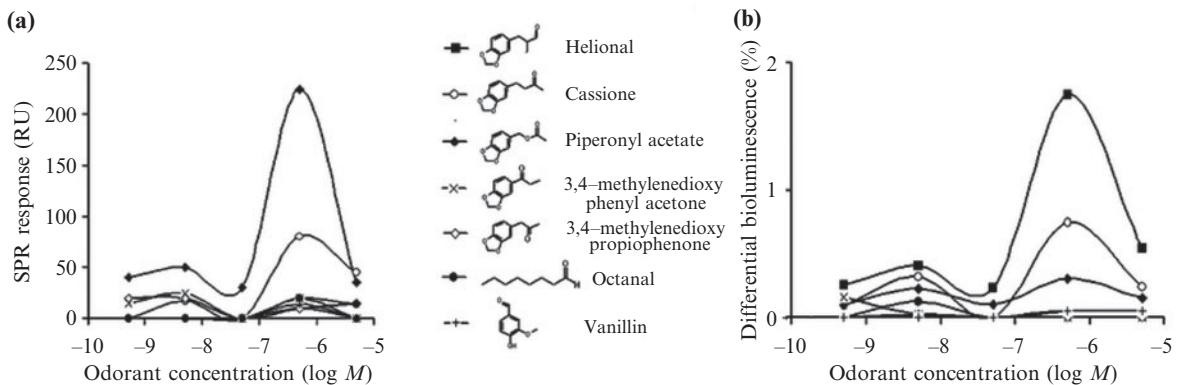


**Fig. 16.30** Principle of the sensor system. The L1 sensor chip (BIAcore) is composed of a gold film deposited onto a glass support, on which a carboxylated dextran grafted with long alkyl groups is attached. Nanosomes are immobilized via interaction of their lipidic surface with the long alkyl chains, leaving membrane receptors fully available for ligands and G protein interactions. The scheme presents only a restricted area of the lipidic bilayer. Upon stimulation of an OR with an odorant ligand, the Ga subunit is activated and desorbed in the presence of GTP. This results in a shift of the SPR response level expressed as arbitrary resonance units (RU). No shift of the SPR response level is observed when nanosomes are stimulated either with odorant alone, or GTP alone (c), as compared to the control stimulated with water. Reprinted from Vidic *et al.* *Lab on a Chip* **6**, 1026, (2006), © 2008 The Royal Society of Chemistry.



device, which could overcome the actual drawbacks of the corresponding inorganic electronic nose, they can be scaled down in dimension, due to the very high detection sensitivity of these natural systems. The nanosomes were immobilized on a Biacore sensor chip L1 whose surface consists of a covalently linked carboxymethyl-modified dextran polymer hydrogel on which an important part of the glucose moiety is grafted with lipophilic alkyl chains.

Nanosomes were efficiently hooked by those alkyl anchors resulting in very stable deposition, yielding a SPR response proportional to the total protein concentration of loaded material. Functional tests were performed with odorants at various concentrations and 10 mM GTP (or GTPcS) in HBS, at a  $5 \mu\text{l min}^{-1}$  flow rate (Fig. 16.30). The robustness of the biosensor was tested by exposing the chip surface to different concentrations of various odorants and its response was compared to the response of the whole yeast cells expressing the receptor (see Fig. 16.31), in order to assess if the receptor retains its sensing characteristics when immobilized onto a chip. Olfactory biosensors were produced



**Fig. 16.31** Detection of odorants by cmyc-OR1740 in immobilized nanosomes and by the receptor in whole yeast cells. (a) Differential SPR responses obtained upon stimulation of the ORs in nanosomes with various odorants in the presence of 10 mM GTPcS. (b) Differential bioluminescence responses upon stimulation of the receptor in whole yeast. Concentration-dependent curves are plotted as a difference of response to odorants relative to controls obtained by replacing odorants with water. Odorants tested are shown in the middle panel. Reprinted from Vidic *et al.* *Lab on a Chip* **6**, 1026 (2006), © 2008 The Royal Society of Chemistry.

on a nanoscale by trapping a few olfactory receptors onto nanoelectrodes or nanoelectromechanical systems (Gomila *et al.* 2006). Micro- and nanoelectrochemical cells for impedance spectroscopy measurements were produced by a combination of standard microfabrication processes and focused ion-beam milling. The size of the working electrode of the microelectrochemical cells fabricated ranged from  $80\text{ }\mu\text{m}$  down to 40 nm in diameter, allowing the microdevices to probe the response from thousands of receptors down to a single receptor. Different configurations and metals could be used to implement the nanosensor devices, and perform measurements in physiological medium.

## 16.7 Conclusions

In this chapter we have reviewed the recent progress in the field of protein-based nanodevices. From current studies, it is clear that the fabrication of synthetic nanometer-sized elements from biomolecular building blocks represents a well-established approach to synthesize devices for nanobiotechnology and nanobioelectronics.

The highlights of the results presented here demonstrate a promising trend towards the implementation of new technological methods and devices, at the nanoscale, based on interaction and communication between the biological matter and the inorganic devices. The future will reserve the design and demonstration of systems and architectures with increasing improvements in terms of efficiencies, robustness and capabilities. The hope is that the entire human life will continue to benefit from this new multicomponent nanotechnology.

The fast-paced growth and exciting progress in these fields makes it difficult to give a complete and real-time overview of all the happenings in the “nanobio” field. The purpose of the present review was to provide interested students and researchers the current status of the frontiers and advances in the area of protein nanodevices.

## Acknowledgments

We very much appreciated the help of Giuseppe Maruccio and Imma Cortese for valuable contributions and support to the preparation of this chapter. We are thankful for exciting collaboration with various colleagues: Roberto Cingolani, Valentina Arima, Antonio Della Torre, Loretta Del Mercato, Stefania Sabella, Roman Krahne, Eliana D'Amone, at NNL in Lecce (Italy), Alessandro Bramanti at ST-Microelectronics (Lecce-Italy). Financial support by NNLoF CNR-INFM, Italian Institute of Technology (IIT), EC through SAMBA and SpiDME project, Italian MIUR (FIRB MOLECULAR DEVICES and FIRB NANOGENOMICS) is gratefully acknowledged.

## References

- Alessandrini, A., Gerunda, M., Canters, G.W., Verbeet, M.Ph., Facci, P. *Chem. Phys. Lett.* **376**, 625 (2003).
- Alonso-Romanowski, S., Gassa, L.M., Vilche, J.R. *Electrochimica Acta* **40**, 1561 (1995).
- Andersen, O.S., Apell, H.J., Bamberg, E., Busath, D.D., Koeppe, R.E., Sigworth, F.J., Szabo, G., Urry, D.W., Wooley, A. *Nature Struct. Biol.* **6**, 609 (1999).
- Apel, C.L., Deamer, D.W., Mautner, M.N. *Biochimica et Biophysica Acta* **1559**, 19 (2002).
- Appenzeller, J., Knoch, J., Derycke, V., Martel, R., Wind, S., Avouris, P. *Phys. Rev. Lett.* **89**, 126 801/1 (2002).
- Astier, Y., Bayley, H., Howorka, S. *Curr. Opin. Chem. Biol.* **9**, 576 (2005).
- Atanasov, V., Knorr, N., Duran, R.S., Ingebrandt, S., Offenhäuser, A., Knoll, W., Köper, I. *Biophys. J.* **89**, 1780 (2005).
- Athreya, S.A., Martin, D.C. *Sens. Actuators A* **72**, 203 (1999).
- Ayliffe, H.E., Frazier, A.B., Rabbitt, R.D. *IEEE J. MEMS* **8**, 50 (1999).
- Balasubramanian, K., Burghard, M. *Anal. Bioanal. Chem.* **385**, 452 (2006).
- Banghart, M., Borges, K., Isacoff, E., Trauner, D., Kramer, R.H. *Nature Neurosci.* **7**, 1381 (2004).
- Bashir, R. *Adv. Drug Delivery Rev.* **56**, 1565 (2004).
- Bayley, H., Cremer, P.S. *Nature* **413**, (6852) 226 (2001).
- Bayley, H., Jayasinghe, L. *Mol. Membr. Biol.* **21**, (4) 209 (2004).
- Bergveld, P. *Sens. Actuators B* **88**, 1 (2003).
- Bernard, A.L., Guedea-Boudeville, M.A., Sandre, O., Palacin, S., Di Meglio, L.-M., Jullien, L. *Langmuir* **16**, 6801 (2000).
- Besteman, K., Lee, J.L., Wiertz, F.G.M., Heering, H.A., Dekker, C. *Nano Lett.* **3**, 727 (2003).
- Birge, R.R., Gillespie, N.B., Izaguirre, E.W., Kusnetzow, A., Lawrence, A.F., Singh, D., Song, Q.W., Schmidt, E., Stuart, J.A., Seetharaman, S., Wise, K.J. *J. Phys. Chem. B* **103**, 10746 (1999).
- Blaustein, R.O., Cole, P.A., Williams, C., Miller, C. *Nature Struct. Biol.* **7**, 309 (2000).

- Blum, A.S., Soto, C.M., Wilson, C.D., Amsinck, C., Franzon, P., Ratna, B.R. *IEEE Trans. Nanobiosci.* **6**, 270 (2007).
- Braun, E., Eichen, Y., Sivan, U., Ben-Yoseph, G. *Nature* **391**, 775 (1998).
- Cai, L., Tabata, H., Kawai, T. *Appl. Phys. Lett.* **77**, 3105 (2000).
- Champagne, A.R., Pasupathy, A.N., Ralph, D.C. *Nano Lett.* **5**(2), 305 (2005).
- Chen, R.J., Bangsaruntip, S., Drouvalakis, K.A., Kam, N.W.S., Shim, M., Li, Y., Kim, W., Utz, P.J., Dai, H. *Proc. Natl. Acad. Sci. USA* **100**, 4984 (2003).
- Chi, Q., Zhang, J., Nielsen, J.U., Friis, E.P., Chorkendorff, I., Canters, G.W., Andersen, J.E.T., Ulstrup, J. *J. Am. Chem. Soc.* **122**, 4047 (2000).
- Choi, J., Fujihira, M. *Appl. Phys. Lett.* **84**, 2187 (2004).
- Choi, J., Nam, Y., Lee, W., Kim, D., Fujihira, M. *Appl. Phys. Lett.* **79**, 1570 (2001).
- Choi, J.-W., Nam, Y.-S., Lee, B.H., Ahn, D.J., Nagamune, T. *Curr. Appl. Phys.* **6**, 760 (2006).
- Cinelli, R.A.G., Pellegrini, V., Ferrari, A., Faraci, P., Nifosi, R., Tyagi, M., Giacca, M., Beltram, F. *Appl. Phys. Lett.* **79**, 3353 (2001).
- Cornell, B.A., Braach-Maksvytis, V.L.B., King, L.G., Osman, P.D.J., Raguse, B., Wieczorek, L., Pace, R.J. *Nature* **387**, 580 (1997).
- Coura, P.Z., Legoas, S.B. *et al.* *Nano Lett.* **4**(7), 1187 (2004).
- Cui, Y., Wei, Q., Park, H., Lieber, C.M. *Science* **293**, 1289 (2001).
- Das, R., Kiley, P.J., Segal, M., Norville, J., Amy Yu, A., Wang, L., Trammell, S.A., Evan Reddick, L., Kumar, R., Stellacci, F., Lebedev, N., Schnur, J., Bruce, B.D., Zhang, S., Baldo, M. *Nano Lett.* **4**, 1079 (2004).
- Datta, S., Tian, W.D. *et al.* *Phys. Rev. Lett.* **79**(13), 2530 (1997).
- Davis, J.J., Morgan, D.A., Wrathmell, C.L., Axford, D.N., Zhao, J., Wang, N. *J. Mater. Chem.* **15**, 2160 (2004).
- Del Mercato, L.L., Pompa, P.P., Maruccio, G., Della Torre, A., Sabella, S., Tamburro, A.M., Cingolani, R., Rinaldi, R. *Proc. Natl. Acad. Sci. USA* **104**(46), 18019–24 (2007).
- De Pablo, P.J., Moreno-Herrero, F., Colchero, J., Gomez Herrero, J., Herrero, P., Bar, A.M., Ordejon, P., Soler, J.M., Artacho, E. *Phys. Rev. Lett.* **85**, 4992 (2000).
- Dertinger, S.K.W., Chiu, D.T., Jeon, N.L., Whitesides, G.M. *Anal. Chem.* **73**, 1240 (2001).
- Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A.L., Gulbis, J.M., Cohen, S.L., Chait, B.T., MacKinnon, R. *Science* **280**(5360), 69 (1998).
- Dzyadevych, S.V., Soldakin, A.P., El'skaya, A.V., Martelet, C., Jaffrezic-Renault, N. *Anal. Chim. Acta* **568**, 248 (2006).
- Endres, R.G., Cox, D.L., Singh, R.R.P. *Rev. Mod. Phys.* **76**, 195 (2004).
- Elibol, O.H., Morissette, D., Akin, D., Denton, J.P., Bashir, R. *Appl. Phys. Lett.* **83**, 4613 (2003).
- Fink, H.W., Schönenberger, C. *Nature* **398**, 407 (1999).
- Freeman, R., Elbaz, J., Gill, R., Zayats, M., Willner, I. *Chem. Eur. J.* **13**, 10317 (2007).
- Freeman, R., Gill, R., Willner, I. *Chem. Commun.* **33**, 3450 (2007).
- Frolov, L., Rosenwaks, Y., Carmeli, C., Carmeli, I. *Adv. Mater.* **17**, 2434 (2005).
- Gil, P.R., Del Mercato, L.L., Del Pino, P., Javier, A.M., Parak, W.J. *NanoToday* **3**, 12 (2008).

- Gopel, W. *Biosens. Bioelect.* **10**, 35 (1995).
- Gomez-Sjoberg, R., Morisette, D.T., Bashir, R. *IEEE J. MEMS* **14**, 829 (2005).
- Gomila, G., Casuso, I., Errachid, A., Ruiz, O., Pajot, E., Minic, J., Gorojankina, T., Persuy, M.A., Aioun, J., Salesse, R., Bausells J., Villanueva, G., Rius, G., Hou, Y., Jaffrezic, N., Pennetta, C., Alfinito, E., Akimov, V., Reggiani, L., Ferrari, G., Fumagalli, L., Sampietro, M., Samitier, J. *Sens. Actuators B* **116**, 66 (2006).
- Gruener, G. *Anal. Bioanal. Chem.* **384**, 322 (2006).
- Gust, D., Moore, T.A. *Science* **244**, 35 (1989).
- Gyorvary, E., Schroedter, A., Talapin, D.V., Weller, H., Pum, D., Sleytr, U.B. *J. Nanosci. Nanotechnol.* **4**, 115 (2004).
- Haeberle, S., Zengerle, R. *Lab on a Chip* **7**, 1094 (2007).
- Hartzell, B.B., McCord, D., Asare, H., Chen, J., Heremans, J., Soghomonian, V. *Appl. Phys. Lett.* **82**, 4800 (2003).
- Hartzell, B.B., McCord, D., Asare, H., Chen, J., Heremans, J., Soghomonian, V. *J. Appl. Phys.* **94**, 2764 (2003).
- Heinze, S., Tersoff, J., Martel, R., Derycke, V., Appenzeller, J., Avouris, P. *Phys. Rev. Lett.* **89**, 106, 801/1 (2002).
- Heller, I., Janssens, A.M., Mannik, J., Minot, E.D., Lemay, S.G., Dekker, C. *Nano Lett.* **8**, 591 (2007).
- Im, H., Huang, X.J., Gu, B., Choi, Y.K. *Nature Nanotechnol.* **2**, 430 (2007).
- Jeon, N.L., Dertinger, S.K.W., Chiu, D.T., Choi, I.S., Stroock, A.D., Whitesides, G.M. *Langmuir* **16**, 8311 (2000).
- Jing, W., Wu, Z., Wang, E. *Electrochimica Acta* **44**, 99 (1998).
- Junker, D., Schmid, H., Delamarche, E. *Nature Mater.* (2005).
- Keren, K., Krueger, M., Gilad, R., Ben-Yoseph, G., Sivan, U., Braun, E. *Science* **297**, 72 (2002).
- Keren, K., Berman, R.S., Buchstab, R.S., Sivan, U., Braun, E. *Science* **302**, 1380 (2003).
- Kervennic, Y.V., Van der Zant, H.S.J. et al. *Appl. Phys. Lett.* **80**(2), 321 (2002).
- Kim, D.S., Park, J.E., Shin, J.K., Kim, P.K., Lim, G., Shoji, S. *Sens. Actuators B* **117**, 488 (2006).
- Kim, S.N., Rusling, J.F., Papadimitrakopoulos, F. *Adv. Mater.* **19**, 3214 (2007).
- Kong, J., Franklin, N.R., Zhou, C., Chapline, M.G., Peng, S., Cho, K., Dai, H. *Science* **287**, 622 (2000).
- Knoll, W., Frank, C.W., Heibel, C., Naumann, R., Offenhäusser, A., Rühe, J., Schmidt, E.K., Shen, W.W., Sinner, A. *Rev. Molecu. Biotechnol.* **74**, 137 (2000).
- Krahne, R., Yacoby, A., Shtrikman, H., Bar-Joseph, I., Dadosh, T., Sperling, J., *Appl. Phys. Lett.* **81**, 730 (2002).
- Krishna, G., Schulte, J., Cornell, B.A., Pace, R.J., Osman, P.D. *Langmuir* **19**, 2294 (2003).
- Kuzmych, D.R., Star, A. *J. Phys. Chem. C* **111**, 3539 (2007).
- Kymakis, E., Alexandrou, I., Amarantunga, G.A.J. *J. Appl. Phys.* **93**, 1764 (2003).
- Lee, S., Mao, C., Flynn, C.E., Belcher, A.M. *Science* **296**, 892 (2002).
- Lee, I., Lee, J.W., Greenbaum, E. *Phys. Rev. Lett.* **79**, 3294, 97 (1997).

- Lee, I., Lee, J.W., Stubna, A., Greenbaum, E. *J. Phys. Chem. B* **104**, 2439 (2000).
- Li, J., Lu, Y., Ye, Q., Cinke, M., Han, J., Meyyappan, M. *Nano Lett.* **3**, 929 (2003).
- Li, C.H. *Chromatographic Science Series*, (eds) Klans Anton and Claire Berger, 1st edn. (CRC Press, Boca Raton, Fla., USA, 2006).
- Liu, H., Fan, F.F., Lin, C.W., Bard, A.J. *J. Am. Chem. Soc.* **108**, 3838 (1986).
- Ludwig, S., Bayley, H., (ed.) Goeldner, M., Givens, R. Wiley-VCH Verlag **253**, 304 (2005).
- Lowe, C.R. *Curr. Opin. Struct. Biol.* **10**, 428 (2000).
- Malaquin, L., Kraus, T., Schmid, H., Delamarche, E., Wolf, H. *Langmuir* **23**, 11513 (2007).
- Manz, A., Gruber, N., Widmer, H.M. *Sens. Actuators B* **1**, 244 (1990).
- Manoj, A.G., Narayan, K.S. *Appl. Phys. Lett.* **83**, 3614 (2003).
- Maruccio, G., Visconti, P. *et al.* *Microelectron. Eng.* **67–8**, 838 (2003).
- Maruccio, G., Marzo, P., Krahne, R., Passaseo, A., Cingolani, R., Rinaldi, R. *Small* **3**, 1184 (2007a).
- Maruccio, G., Marzo, P., Krahne, R., Della Torre, A., Passaseo, A., Cingolani, R., Rinaldi, R. *Microelectron. Eng.* **84**, 1585 (2007b).
- Maruccio, G., Visconti, P., Arima, V., D'Amico, S., Biasco, A., D'Amone, E., Cingolani, R., Rinaldi, R., Masiero, S., Giorgi, T., Gottarelli, G. *Nano Lett.* **3**, 479 (2003).
- Maruccio, G., Bramanti, A. Nanoelectric devices based on proteins, in *Nanobioelectronics – for electronics, Biology, and Medicine*, (ed.) A. Offenhäusser, R. Rinaldi, Springer, (2009).
- Maruccio, G., Biasco, A., Visconti, P., Bramanti, A., Pompa, P.P., Calabi, F., Cingolani, R., Rinaldi, R., Corni, S., Di Felice, R., Molinari, E., Verbeet, M.R., Canters, G.W. *Adv. Mater.* **17**, 816 (2005a).
- Maruccio, G., Cingolani, R., Rinaldi, R. *J. Mater. Chem.* **14**, 542 (2004a).
- Maruccio, G., Visconti, P., Biasco, A., Bramanti, A., Della Torre, A., Pompa, P.P., Frascerra, V., Arima, V., D'Amone, E., Cingolani, R., Rinaldi, R. *Electroanalysis* **16**(22), 1853 (2004b).
- Maruccio, G., Biasco, A., Visconti, P., Bramanti, A., Pompa, P.P., Calabi, F., Cingolani, R., Rinaldi, R., Corni, S., Di Felice, R., Molinari, E., Verbeet, M.R., Canters, G.W. *Adv. Mater.* **17**(7), 816 (2005b).
- McGeoch, J.E.M., McGeoch, M.W., Carter, D.J.D., Shuman, R.F., Guidotti, G. *Med. Biol. Eng. Comput.* **38**, 113 (2000).
- Mohamad, R.M., Mahmoudian, L., Kaji, N., Tokeshi, M., Chuman, H., Baba, Y., *Nanotoday* **1**, 38 (2006).
- Moll, D., Huber, C., Schlegel, B., Pum, D., Sleytr, U.B., Sra, M. *Proc. Natl. Acad. Sci. USA* **99**, 14646 (2002).
- Montal, M., Opella, S.J. *Biochim. Biophys. Acta* **1565**, 287 (2002).
- Niemeyer, C.M. *Nanotoday* **2**, 42 (2007).
- Niemeyer, C.M., Mierkin, C.A. *Nanobiotechnology*, Wiley-VCH (2004).
- Naumann, R., Schiller, S.M., Giess, F., Grohe, B., Hartman, K.B., Kärcher, I., Köper, I., Lübben, J., Vasilev, K., Knoll, W. *Langmuir* **19**, 5435 (2003).

- Naumann, R., Walz, D., Schiller, S.M., Knoll, W. *J. Electroanal. Chem.* **550–551**, 241–252 (2003).
- Oblatt-Montal, M., Bühler, L.K., Iwamoto, T., Tomich, J.M., Montal, M. *J. Biol. Chem.* **268**, 14601 (1993).
- Ohori, T., Shoji, S., Miura, K., Yotsumoto, A. *Sens. Actuators A* **64**, 57 (1998).
- Oosterbroek, R.E., Van den Berg, A. *Lab-on-a-chip, Miniaturized System for (Bio) Chemical Analysis and Synthesis*, 1st edn (Elsevier, Amsterdam, 2003).
- Opella, S.J., Marassi, F.M., Gesell, J.J., Valente, A.P., Kim, Y., Oblatt-Montal, M., Montal, M. *Nat. Struct.* **6**(4), 374 (1999).
- Pandey, P.C. *Anal. Chimica Acta* **568**, 47 (2006).
- Park, J., Pasupathy, A.N. *et al. Nature* **417**, 722 (2002).
- Park, K.-Y., Han, S.-W., Kim, M.-S., Choi, S.-Y. *Electron Lett.* **40**, n. 3 (2004).
- Pompa, P.P., Bramanti, A., Maruccio, G., Del Mercato, L.L., Cingolani, R., Rinaldi, R. *Chem. Phys. Lett.* **404**, 59 (2005).
- Pompa, P.P., Bramanti, A., Maruccio, G., Cingolani, R., De Rienzo, F., Corni, S., Di Felice, R., Rinaldi, R. *J. Chem. Phys.* **122**, 181102 (2005).
- Pompa, P.P., Della Torre, A., Del Mercato, L.L., Chiuri, R., Bramanti, A., Calabi, F., Maruccio, G., Cingolani, R., Rinaldi, R. *J. Chem. Phys.* **125**, 021103 (2006).
- Pompa, P.P., Del Mercato, L.L., Della Torre, A., Chiuri, R., Calabi, F., Maruccio, G., Cingolani, R., Rinaldi, R. *PCT Int. Appl.* WO 2007060632 (2007).
- Porath, D., Bezryadin, A., De Vries, S., Dekker, C. *Nature* **403**, 635 (2000).
- Rakitin, A., Aich, P., Papadopoulos, C., Kobzar, Yu., Vedeneev, A.S., Lee, J.S., Xu, J.M. *Phys. Rev. Lett.* **86**, 3670 (2001).
- Reches, M., Gazit, E. *Nature Nanotech.* **1**, 195 (2006).
- Reed, M.A., Zhou, C., *et al. Science* **278**(5336), 252 (1997).
- Renugopalakrishnan, V., Khizroev, S., Anand, H., Pingzuo Li, Lindvold, L. *IEEE Trans. Magn.* **43**, 773 (2007).
- Rinaldi, R., Cingolani, R. *Physica E; Low-dimen. Syst. Nanostruct.* **21**(1), 45 (2004).
- Rinaldi, R., Pompa, P.P., Maruccio, G., Biasco, A., Visconti, P., Pisignano, D., Blasi, L., Sgarbi, N., Krebs, B., Cingolani, R. *IEE Proc. Nanobiotech.* **151**, 101 (2004).
- Rinaldi, R., Biasco, A., Maruccio, G., Arima, V., Visconti, P., Cingolani, R., Facci, P., De Rienzo, F., Di Felice, R., Molinari, E., Verbeet, M.P., Canters, G.W. *Appl. Phys. Lett.* **82**(3), 472 (2003).
- Rinaldi, R., Maruccio, G. *Nano-bio Electronics, Nanotechnology for Life Sciences – Vol. 4 (Nanodevices for life sciences)*, (ed.) Challa Kumar (Wiley VCH, 2006).
- Rinaldi, R., Maruccio, G. *Molecular Electronics*, in *Wiley Encyclopedia of Biomedical Engineering*, (ed.) M. Akay (Wiley Inc., 2006).
- Rothemund, P.W.K. *Nature* **440**, 297 (2006).
- Sara, M., Pum, D., Schuster, B., Sleytr, U.B. *J. Nanosci. Nanotech.* **5**(12), 1939 (2005).
- Schonning, M.J., Poghossian, A. *Electroanalysis* **18**, 189 (2006).
- Schuster, B., Györvary, E., Pum, D., Sleytr, U.B. *Protein Nanotech.* **300**, 101 (2005).
- Seeman, N.C. *J. Theor. Biol.* **99**(2), 237 (1982).
- Sigworth, F.J., Klemic, K.G. *Biophys. J.* **82**, 2831 (2002).

- Silverman, J.P. *J. Vac. Sci. Technol. B* **15**, 2117 (1997).
- Sofou, S., Thomas, J.L. *Biosens. Bioelectron.* **18**, 445 (2003).
- Sonnenfeld, R., Hansma, P.K. *Science* **232**, 211 (1986).
- Star, A., Gabriel, J.C.P., Bradley, K., Grüner, G. *Nano Lett.* **3**, 459 (2003).
- Stewart, M.E. *J. Nanoeng. Nanosyst.* **220**, 81 (2007).
- Storm, A.J., Van Noort, J., De Vries, S., Dekker, C. *Appl. Phys. Lett.* **79**, 3881 (2001).
- Tan, J.L., Tien, J., et al. *Langmuir* **18**(2), 519 (2002).
- Tao, N.J. *Phys. Rev. Lett.* **76**, 4066 (1996).
- Terretaz, S., Vogel, H. *MRS Bull.* **30**, 207 (2005).
- Totu, E., Josceanu, A.M., Covington, A.K. *Mater. Sci. Eng. C* **18**, 87 (2001).
- Tran, P., Alavi, B., Gruner, G. *Phys. Rev. Lett.* **85**, 1564 (2000).
- Vijgenboom, E., Busch, J.E., Canters, G.W. *Microbiology* **143**, 2853 (1997).
- Vidic J.M., Grosclaude J., Persuy M.A., Aioun J., Salesse R., Pajot-Augy E. *Lab on a Chip* **6**, 1026 (2006).
- Whitesides, G.M. *Nature Biotechnol.* **21**, 1161 (2003).
- Wilk, S.J., Petrossian L., Goryll, M., Thornton, T.J., Goodnick, S.M., Tang, J.M., Eisenberg, R.S. *Biosens. Bioelectron.* **23**, 183 (2007).
- Willner, I. *Science* **298**, 2407 (2002).
- Wu, Z., Chen, Z., Du, X., Logan, J.M., Sippel, J., Nikolou, M., Kamaras, K., Reynolds, J.R., Tanner, D.B., Hebard, A.F., Rinzler, A.G. *Science* **305**, 1273 (2004).
- Xi, J., Ho, D., Chu, B., Montemagno, C.D. *Adv. Func. Mater.* **15**, 1233 (2005).
- Xia, Y.N., Whitesides, G.M. *Angew. Chem. Int. Ed.* **37**(5), 551 (1998).
- Xu, D., Watt, G.D., Narb, J.N., Davis, R.C. *Nano Lett.* **5**(4), 571 (2005).
- Xu, J., Stickrath, A.B., Bhattacharya, P., Nees, J., Varo, G., Hillebrecht, J.R., Ren, L., Birge, R.R. *Biophys. J.* **85**, 1128 (2003).
- Yasutomi, S., Morita, T., Imanishi, Y., Kimura, S. *Science* **304**, 1944 (2004).
- Yao, Z., Kane, C.L., Dekker, C. *Phys. Rev. Lett.* **84**, 2941 (2000).
- Yoo, K.H., Ha, D.H., Lee, J.O., Park, J.W., Kim, J., Kim, J.J., Lee, H.Y., Kawai, T., Choi, H.Y. *Phys. Rev. Lett.* **87**, 198102 (2001).
- Zandbergen, H.W., Van Duuren, R., et al. *Nano Lett.* **5**(3), 549 (2005).
- Zhang, Y., Austin, R.H., Kraeft, J., Cox, E.C., Ong, N.P. *Phys. Rev. Lett.* **89**, 198102 (2002).
- Zhang, J., Grubb, M., Hansen, A.G., Kuznetsov, A.M., Boisen, A., Wackerbarth, H., Ulstrup, J. *J. Phys. Condens. Matter.* **15**, S1873–1890 and references therein (2003).
- Zhao, J., Davis, J.J., Sansom, M.S.P., Hung, A. *J. Am. Chem. Soc.* **126**, 5601 (2004).
- Zayats, M., Kharitonov, A.B., Katz, E., Buckmann, A.F., Willner, I. *Biosens. Bioelectron.* **15**, 671 (2000).
- Zou, Z., Kai, J., Rust, M.J., Han, J., Ahn, C.H. *Sens. Actuators A* **136**, 518 (2007).

# 17

# Bioconjugated quantum dots for tumor molecular imaging and profiling

17.1 Introduction	612
17.2 Photophysical properties of quantum dots	615
17.3 Engineering of QD-based probes for biomedical applications	621
17.4 Tumor molecular imaging and profiling	629
17.5 Conclusions	636
Acknowledgments	636
References	637

P. Zrazhevskiy and X. Gao

## 17.1 Introduction

Medicine, as a practice of treating common disease symptoms utilizing generalized therapeutic approaches and tools, is quickly transforming to a qualitatively new stage—personalized medicine—which addresses individual diseases in a pathology-specific and patient-specific manner. This transition is being made possible by uncovering and utilization of physiologically relevant molecular information and incorporation of highly specific and sensitive tools into biomedical research and clinical practice. Scientists working in the field of biomedical research are constantly exploring new elaborate ways for obtaining comprehensive molecular information in order to better understand and, eventually, control normal and pathologic processes underlying complex physiological phenomena. In addition to higher-level (i.e. tissue to whole organism) symptoms widely used for disease diagnostics and therapy today, subcellular and molecular-level information is indispensable in the design of specific targeted medications and development of personalized treatment schemes (Geho *et al.* 2007). However, the conventional biomedical toolbox is not sufficient to address the increasing complexity of molecular diagnostics and therapeutics. Nanotechnology, on the other hand, is already introducing novel nanoparticle-based tools for molecular imaging, therapy, and targeted drug delivery and opening doors to highly sensitive and quantitative diagnostics as well as targeted and personalized treatment (Alivisatos *et al.* 2005; Geho *et al.* 2007; True and Gao 2007).

The need for personalized diagnostics and therapy is becoming apparent in all areas of medicine; however, it is especially urgent and sought after in treating cancer. Despite considerable efforts in the scientific and medical community to combat this disease, cancer still accounts for 25% of all deaths in the United States and remains one of the leading causes of death in the twenty-first century (Jemal *et al.* 2005). Mechanisms of cancerogenesis and cancer

response to therapy remain poorly understood, thus precluding accurate cancer diagnosis, prognosis, and effective treatment. Elucidation of such mechanisms will undoubtedly provide insights not only for novel approaches to detection and treatment of cancers in the early stages, but also for development of effective prophylaxis. Yet, implementation of this task is quite challenging, as each cancer appears to be as unique as a fingerprint (True and Gao 2007). It is evident that different types of cancers have different biomarker expression schemes. However, even when tumors appear histologically identical between patients of the same ethnic and demographic group, cancer cells still sometimes exhibit drastically different genotypes and phenotypes (Liu *et al.* 2004), resulting in different (and, quite often, unpredictable) prognoses and responses to therapy. For example, Paik and colleagues have identified 21 genes that predict 10-year metastasis-free survival in patients with estrogen receptor-positive node-negative tumor, and have demonstrated that two-thirds of such patients get no benefit from chemotherapy (Paik *et al.* 2004). Therefore, accurate molecular profiling of individual tumors is one key to effective treatment. Tumor-specific molecular information might identify cellular markers for targeted and effective anticancer therapy with minimal adverse side-effects. As such, expression levels of hormone receptors have been directly correlated with the benefit of endocrine treatment, while over-expression of Her2 protein or amplification of its gene has been identified as a requirement for effective treatment with monoclonal antibodies (such as trastuzumab) (Elledge *et al.* 2000; Vogel *et al.* 2002; Konecny *et al.* 2003). Successful targeted therapy has been implemented for the treatment of breast cancer (Paik *et al.* 2004), lung carcinoma, chronic myelogenous leukemia, and gastrointestinal stromal tumor (Arora and Scholar 2005). Therefore, the ability to thoroughly and quantitatively analyze complex panels of cancer biomarkers of individual tumors is strongly needed for comprehensive understanding of cancer pathophysiology—a pre-requisite for accurate cancer diagnostics and effective therapy.

Non-invasive medical imaging techniques (e.g. magnetic resonance imaging and positron emission tomography) are routinely used for cancer screening and diagnostics. However, despite considerable achievements in the area of *in-vivo* diagnostics, these techniques are not yet sensitive and/or specific enough to assess biomarker expression profiles of cancer cells. The invasive but highly informative and reliable technique of *ex-vivo* pathological evaluation of biopsies of primary tumors and their distal metastases remains the basis for addressing cancer diagnostics on the molecular level. In clinics, abnormal expression of cellular markers characteristic to cancer cells is commonly evaluated using standard semi-quantitative immunohistochemistry methods (Umemura and Osamura 2004). For example, the immunoperoxidase method reveals biomarker location via deposition of colored product resulting from enzymatic conversion of diaminobenzidine by horseradish peroxidase. Since the density of staining is proportional to the number and distribution of biomarkers in the specimen, visual analysis of the colorimetric stains is routinely used to “quantify” biomarker expression levels. Another common method—immunofluorescence—utilizes antibodies linked to fluorescent proteins or organic dyes for labelling of biomarkers, providing higher contrast and sensitivity of staining. Signal quantification can be achieved to a certain degree

of accuracy by measuring fluorescence intensity of the specimen. However, both techniques are most well suited for single-color imaging and have limited use in multiplexed analysis of biomarkers (True 1988). Besides being single color in nature, immunoperoxidase analysis is prone to inconsistent staining results, as slight deviations in staining protocol (e.g. activity of enzyme, duration of reaction, and amount of substrate added) lead to significant differences in staining density. Additionally, visual evaluation of stains is subjected to inter- and intraobserver variability, further limiting accurate quantification (Pedersen *et al.* 1989; Thomson *et al.* 2001). The immunofluorescence method provides more room for detection of multiple markers with careful choice of fluorescent probes, but quantitative analysis is heavily compromised by relatively quick photobleaching of organic fluorophores and significant signal interference due to tissue autofluorescence (Ghazani *et al.* 2006). In order to overcome these limitations, more sensitive and quantitative techniques, such as reverse transcription polymerase chain reaction, gene chips, protein chips, and biomolecular mass spectrometry, can be used to search for multiple proteins as well as RNA and DNA in cancer cells and tumor-tissue specimens (Liotta and Petricoin 2000; Chaurand *et al.* 2004). However, due to the destructive nature of these procedures, potentially valuable structural cellular and tissue information is lost during sample processing. This is particularly undesirable considering the substantial heterogeneity of tumor composition (True and Gao 2007). Therefore, the inability to accurately detect and quantify expression of multiple biomarkers in tissue samples with preserved morphology precludes obtaining a comprehensive “molecular portrait” of cancer cells from analysis of a limited number of sections available from a biopsy (in most cases less than 5–6 sections). Development of new molecular-profiling technologies capable of quantitative analysis of molecular signatures of individual patients’ tumors would greatly enhance the quality and predictive power of cancer diagnostics.

Nanotechnology offers a handful of novel solutions to the above-stated issues—namely metallic, semiconductor, and magnetic nanoparticles—based on their unique physical properties. Among these, quantum dots (QDs) have emerged as a promising new tool for detection and quantification of multiple biomarkers in cells and tissue samples *ex vivo*, and even whole organisms *in vivo* (Gao *et al.* 2005). Being semiconductor nanoparticles 2 to 10 nm in diameter, QDs possess unique photophysical properties drastically different from single atoms or bulk materials. The nanometer-scale size of QDs, comparable with the size of large proteins, allows for integration of nanoparticles and biomolecules yielding biologically functional nanomaterials (Chan *et al.* 2002; Gao *et al.* 2005; Yezhelyev *et al.* 2006). Narrow size-tunable light emission and effective light absorption throughout a wide spectrum make QDs much brighter and, thus, more sensitive detection tools than organic fluorophores and provide massive multiplexing capabilities. Accurate and consistent quantitative analysis of biomarker expression is readily achievable due to exceptional QD resistance to photobleaching and photodegradation. Additionally, the large Stokes shift and long fluorescence lifetime allow separation of the QD signal from tissue autofluorescence and, therefore, substantial improvement of signal-to-background ratio in both *ex-vivo* and *in-vivo* studies

(Ness *et al.* 2003; Tokumasu and Dvorak 2003; Wu *et al.* 2003; Nisman *et al.* 2004; Giepmans *et al.* 2005; Matsuno *et al.* 2005; Ghazani *et al.* 2006; Xing *et al.* 2007). Therefore, QDs are well suited for sensitive quantitative molecular profiling of cancer cells and tissues, holding tremendous promise for unravelling the complex gene-expression profiles of cancers, accurate clinical diagnosis and personalized treatment of patients (Jain 2005; True and Gao 2007).

Currently, QDs have found use in most of the conventional biomedical tools where fluorescence or colorimetric imaging of target biomarker is utilized (e.g. cell and tissue staining, Western blot, ELISA, etc.) and have launched novel applications (e.g. *in-vivo* fluorescence imaging, single-molecule tracking, combined drug delivery and imaging, etc.) utilizing their unique photophysical properties. The number of biomedical applications of QDs continues to grow, ranging from ultrasensitive detection *in vitro* to targeted drug delivery and imaging *in vivo*. Yet, we believe that the most profound impact of QD technology on cancer research with immediate clinical benefits will come from utilization of the QD unique features for multiplexed and quantitative analysis of biomarker expression in tumor-tissue biopsies, as QD toxicity is not a major concern in these studies. Therefore, the scope of this review is limited to discussion of QD applications for *ex-vivo* tumor molecular imaging and profiling.

## 17.2 Photophysical properties of quantum dots

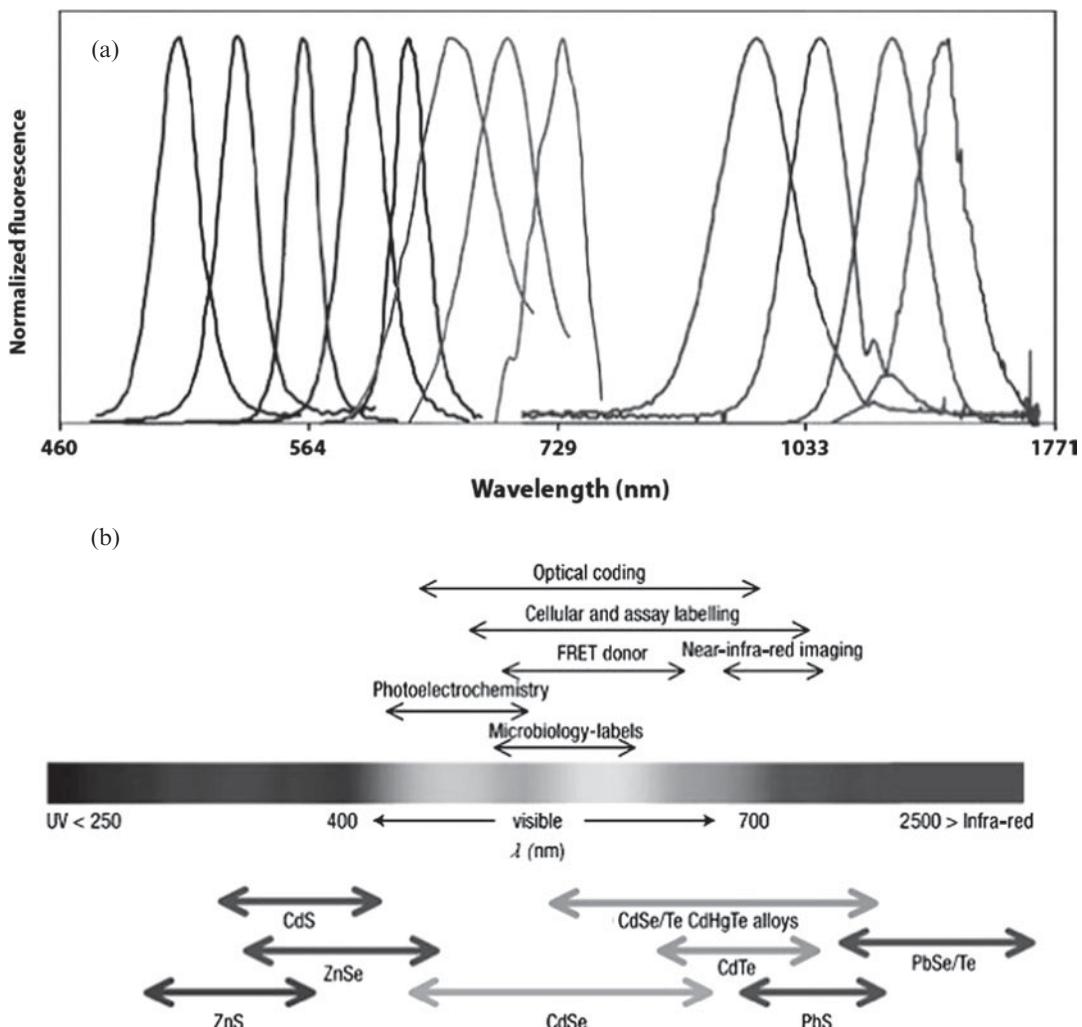
QDs represent a class of spherical (or nearly spherical) nanoparticles made of semiconductor materials. Their name, as well as unique photophysical properties, come from the quantum confinement of charge carriers within the nanoparticle. Electrons in individual atoms occupy discrete energy levels; therefore, transitions from one energy level to another are quantized. Bulk materials, on the other hand, are composed of a great number of tightly interacting atoms. As a consequence of these interactions, discrete energy levels occupied by individual electrons overlap and fuse into nearly continuous energy bands—conduction band and valence band. Valence-band electrons are tightly bound to atoms, whereas electrons in the conduction band have weak association with individual atoms and are relatively free to move within the material. In metals, valence and conduction bands overlap, thus readily providing free electrons for conduction of electricity. Insulators, on the other hand, have a large energy bandgap between valence and conduction bands. At room temperature, valence-band electrons cannot acquire enough energy to “jump” over the bandgap into the conduction band. As a result, insulators in general cannot conduct electricity. Bulk semiconductors have electrical conductivity in between that of a conductor (e.g. metals) and that of an insulator (e.g. rubber-like polymers). This property comes from the existence of a relatively small bandgap (less than 4 eV) between semiconductors’ valence and conduction bands. At ambient conditions, the conduction band is poorly populated with electrons, yielding low conductivity. However, upon excitation (i.e. absorption of extra energy from outside sources) by application of an external electric

field, increasing temperature, or illumination, significantly more valence-band electrons can acquire enough energy to jump to the conduction band, improving the conductivity of the material. Unlike bulk materials, semiconductor nanoparticles are made of only hundreds to thousands of atoms and do not have a sufficient number of electrons to form continuous energy bands. As a result, discrete energy levels appear on the edges of the conduction and valence bands, with spacing between levels dependent on the size of the nanoparticle (Alivisatos 1996a). Therefore, quantized size-dependent electronic transitions can be observed in QDs.

### 17.2.1 Tuneable and size-dependent light emission

Promotion of an electron to the conduction band leaves a positively charged “hole” in the valence band, creating a physically separated electron–hole pair termed an exciton. When the physical dimensions of a semiconductor become smaller than the exciton Bohr radius, a quantum confinement effect is observed. In spherical semiconductor nanoparticles (QDs) an electron and hole are confined in all three dimensions, and the degree of confinement increases with decreasing size of the nanoparticle. The quantum confinement effect can be understood in analogy with the “particle-in-a-box” concept: when the mobility of a particle is limited by an energy barrier, uncertainty in its position becomes significantly smaller, and, according to the Heisenberg uncertainty principle, the momentum of the particle cannot be well defined, thus yielding a non-zero kinetic energy (Smith *et al.* 2004). Since confined particles can occupy only certain discrete energy levels, oscillator strength is concentrated into a few discrete transitions. In the case of QDs, as the size of the nanoparticle decreases, the degree of confinement of the charge carriers increases, thus increasing the kinetic energy of these particles, expanding the spacing between allowed electronic states, and shifting electronic excitation to higher energy (Alivisatos 1996a). As a consequence, smaller QDs require higher-energy photons for electron excitations, while larger QDs with smaller bandgap can be excited with photons of lower energy. The newly created exciton is not stable, and tends to relax back to its ground state through electron–hole recombination. In a direct semiconductor, such as CdSe, this transition is electronically allowed and is usually accompanied by the conversion of bandgap energy into an emitted photon. The size of the bandgap (and, therefore, light emission color) is determined by both chemical composition and physical dimensions (degree of electron–hole confinement) of the QD (Fig. 17.1) and can be tuned by adjusting these parameters (Alivisatos 1996b; Bruchez *et al.* 1998; Medintz *et al.* 2005b). Recent advances in QD synthesis have allowed precise control of particle size, shape, and internal structure, producing nanoparticles emitting light from the UV, throughout the visible, and into the infrared region (400–4000 nm) (Qu and Peng 2002; Kim *et al.* 2003; Zhong *et al.* 2003a; Zhong *et al.* 2003b; Pietryga *et al.* 2004; Smith *et al.* 2004).

The capability to precisely tune the QD size and, therefore, light emission profile enables utilization of multicolor QDs in a highly powerful application of simultaneous imaging and tracking of multiple molecular targets. This is



**Fig. 17.1** Emission color of QDs depends on the size and chemical composition of the nanocrystal. (A) Emission spectra of CdSe (left), InP (middle), and InAs (right) QDs are shown. CdSe QDs emit light at shorter wavelengths, and as the size increases from 2.1 to 4.6 nm in diameter, emission spectrum shifts to longer wavelengths (from left to right). The same tendency is observed for InP (3.0, 3.5, and 4.6 nm) and InAs (2.8, 3.6, 4.6, and 6.0 nm) QDs. (Bruchez, *et al.*, *Semiconductor nanocrystals as fluorescent biological labels*. *Science*, **281**, (5385): p. 2013–6 1998. Copyright AAAS.) (B) Emission ranges of several QD core materials are superimposed over the light spectrum. Also, potential areas of biological applications of QDs are presented. (Medintz, *et al.*, *Quantum dot bioconjugates for imaging, labelling and sensing*. *Nature Mater*, **4**(6), p. 435–46, 2005. Copyright Macmillan Publishers Ltd.)

especially useful in molecular profiling of tumor-tissue specimens obtained from biopsies considering a very limited amount of material is available and a large panel of biomarkers are to be tested for. The multicolor imaging capability of QDs can be primarily attributed to the presence of very narrow and symmetric light emission bands (20–30 nm in the visible range, and can be as narrow as 14 nm at full width at half-maximum or FWHM) (Bruchez *et al.* 1998; Zhong *et al.* 2003a; Lin *et al.* 2004), which produce no or little cross-talk between adjacent colors and allow simultaneous detection and

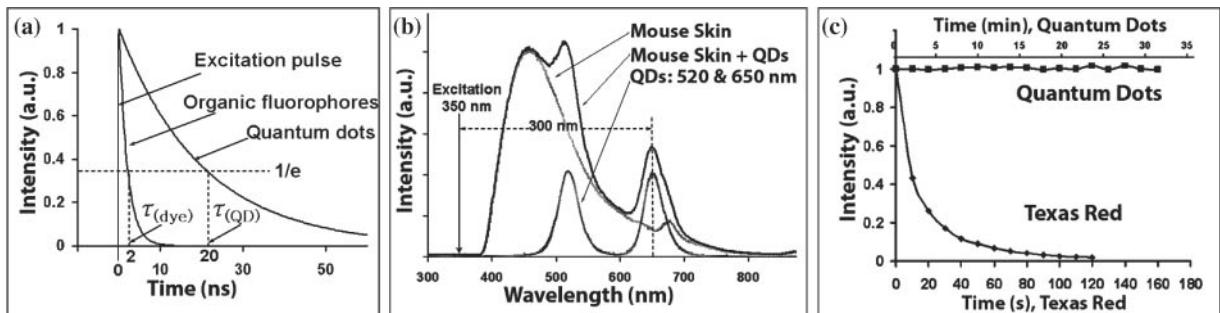
quantification of multiple signals. In addition, the QD core size can be measured and used for differentiation between different targets in multiplexed high-resolution samples imaging in an electron microscope (EM). Since there is a direct correlation between the size of a nanocrystal and the wavelength of emitted light, high-resolution EM images can be readily correlated with lower-resolution fluorescence images of the sample (Nisman *et al.* 2004; Giepmans *et al.* 2005).

### 17.2.2 Effective light absorption over a wide spectrum

The multiplexing capability of QDs is complemented by efficient light absorption over a broad spectral range (hundreds of nanometers). Even though lifetime-limited emission rates of QDs are 5–10 times lower than those of organic dyes due to QD longer excited-state lifetimes (20–50 ns), QDs absorb light 10–50 times faster at the same excitation photon flux since their molar extinction coefficients ( $0.5\text{--}5 \times 10^6 \text{ M}^{-1} \text{ cm}^{-1}$  (Leatherdale *et al.* 2002)) are 10–50 times larger than those of organic dyes ( $5\text{--}10 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ ). Therefore, in fluorescence imaging operating under absorption-limited conditions (in which the rate of absorption is the main limiting factor of fluorescence emission) QDs appear to be 10–20 times brighter than organic dyes, allowing highly sensitive detection of biomarkers present in low copy numbers (Bruchez *et al.* 1998; Chan and Nie 1998; Gao *et al.* 2005). Moreover, the molar extinction coefficient of QDs gradually increases towards shorter wavelength, thus allowing for simultaneous excitation of various QDs by a single high-energy light source (e.g. UV lamp) (Chan *et al.* 2002), eliminating the need for multiple excitation sources, reducing the cost of imaging instrumentation, and simplifying data analysis.

### 17.2.3 Long excited-state lifetime

The long excited-state lifetime of QDs (20–50 ns), though detrimental for light emission dynamics, is very useful in separation of the QD signal from the short-lived autofluorescence background (2 ns) and organic fluorophore signals (1–4 ns) via a technique known as time-domain imaging (Pinaud *et al.* 2006). The exponential model of fluorescence decay after a pulse excitation is presented in Fig. 17.2(a). Since the QD excited-state lifetime is one order of magnitude longer compared to organic dyes, the fluorescence intensity of QDs drops significantly slower, thus producing a QD-to-dye intensity ratio of  $\sim 100$  in only 10 ns. Therefore, the image contrast (measured by the signal-to-noise or signal-to-background ratio) can be dramatically improved by time-delayed data acquisition (Gao *et al.* 2005). This is especially helpful in fluorescence imaging of formalin-fixed-paraffin-embedded (FFPE) tissue samples, where autofluorescence from endogenous proteins and fixation agents can be significant. For example, Dahan *et al.* have demonstrated that time-gated imaging significantly and selectively reduces the autofluorescence contribution, achieving enhancement of the signal-to-background ratio by more than an order of magnitude (Dahan *et al.* 2001). Time-gated imaging can also be utilized in



**Fig. 17.2** Unique optical properties of QDs provide enhanced sensitivity and multiplexing ability to molecular-profiling techniques. (a) A comparison of the excited-state decay curves (monoexponential model) between QDs and common organic dyes. The longer excited-state lifetimes of QD probes allow the use of time-domain imaging to discriminate against the background fluorescence (short lifetimes).  $(\text{dye})t$  and  $(\text{QD})t$  are the delay times for the fluorescence signals to decrease to  $1/e$  of their original values, where  $e$  is the natural log constant and is equal to 2.718. (b) Comparison of mouse skin and QD emission spectra obtained under the same excitation conditions, demonstrating that the QD signals can be shifted to a spectral region where autofluorescence is reduced. (c) Photobleaching curves show that QDs are several thousand times more photostable than organic dyes (e.g. Texas red) under the same excitation conditions. (Gao, *et al.*, *In vivo molecular and cellular imaging with quantum dots*. *Curr. Opin. Biotechnol.* **16**(1), p. 63–72, 2005. Copyright Elsevier.)

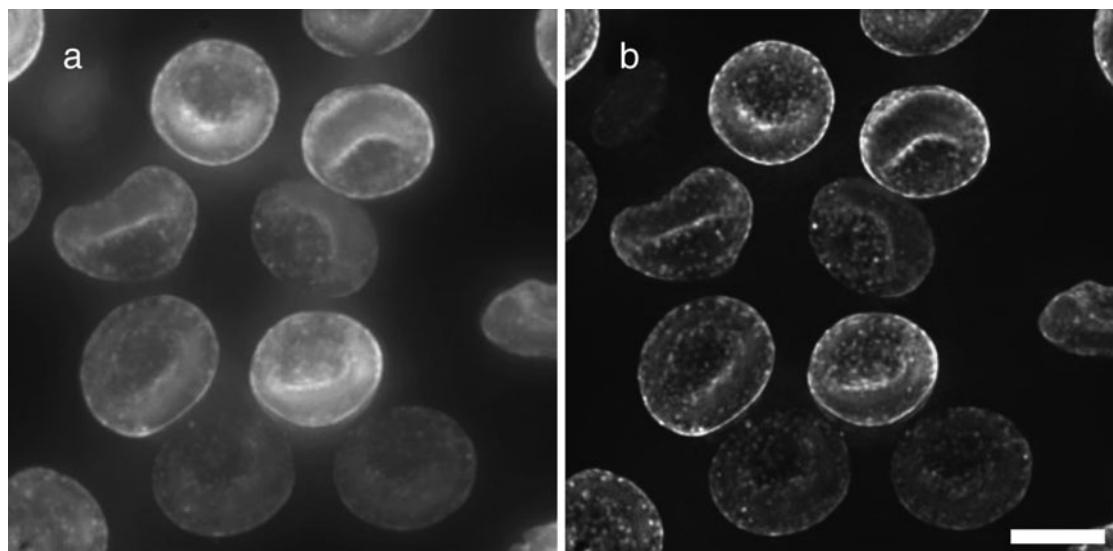
multiplexed biomarker imaging when both QDs and conventional organic fluorophores are used. In this case, the short-lived fluorophore signal is detected first, and then the QD signal is measured (Grecco *et al.* 2004).

#### 17.2.4 Large redshift between absorbed and emitted photons

Excitation of an electron to the conduction band and consecutive relaxation of an exciton to its ground state is accompanied by some energy loss. As a consequence, the emitted photon has a longer wavelength (i.e. less energy) than the absorbed photon. This phenomenon is known as the Stokes (or red) shift and is measured by the distance between the excitation and emission peaks. Redshift produced by quantum dots is significantly larger than that of organic fluorophores, allowing for further improvement in detection sensitivity (Mews *et al.* 1996; Klostranec and Chan 2006). The Stokes shift of semiconductor QDs can be as large as 300–400 nm, depending on the wavelength of the excitation light. Organic dye signals with a small Stokes shift are often buried by strong tissue autofluorescence, whereas QD signals are clearly recognizable above the background (Fig. 17.2(b)) (Gao *et al.* 2005). For example, Gao *et al.* have demonstrated the utility of red QDs (emission peak around 640 nm) conjugated to antibodies against prostate-specific membrane antigen (PSMA) for *in-vivo* tumor imaging in mice (Gao *et al.* 2004).

#### 17.2.5 Photostability

QDs present a unique tool for high-resolution imaging, real-time tracking, and quantitative analysis of biomarkers due to their resistance to photobleaching and photodegradation. It has been shown that QDs resist photobleaching



**Fig. 17.3** QDs exceptional resistance to photobleaching allows collection of z-stack data for high-quality 3D reconstruction of images. (a) Maximum projection image of QD-labelled erythrocytes without signal processing is blurred by out-of-focus signals and by haze due to refractive-index mismatch between mounting media and cells. (b) Additional image processing produces a high-contrast image. Scale bar = 5 µm. (Tokumasu, and Dvorak, *Development and application of quantum dots for immunocytochemistry of human erythrocytes*. *J. Microsc.* **211**(Pt. 3): p. 256–61, 2003. Copyright Wiley-Blackwell Publishing Ltd.)

for more than 30 min, while organic dyes fade by more than 90% in less than one minute under identical experimental conditions (Fig. 17.2(c)) (Gao and Nie 2003). Unprecedented photostability renders QDs well suited for imaging when long exposure to excitation source is required (for example, when taking high-resolution confocal images or monitoring target movement in cell cultures), while keeping signal intensity constant and allowing for consistent analysis of samples. Tokumasu and Dvorak have reported the use of this property in immunocytochemical studies of human erythrocytes with a high-magnification, three-dimensional reconstruction technique, where utilization of QD-bioprobes allowed oversampling of the fluorescent images and extension of the useful magnification range for documentation and reliable collection of z-stack image data for 3D reconstruction without loss of image intensity (Fig. 17.3) (Tokumasu and Dvorak 2003). Cui *et al.* have utilized pseudo-TIRF (total internal reflection fluorescence) microscopy for real-time tracking of intracellular transport of QD-labelled nerve growth factor along axons of rat dorsal root ganglion neurons (Cui *et al.* 2007). Additionally, prolonged light exposure of tissue samples effectively photobleaches the autofluorescence resulting from organic molecules and fixation agents, while keeping the QD signal constant.

It should be noted that bare QDs are not stable under UV illumination for extended periods of time due to photolysis (since the energy of UV irradiation is close to that of covalent bonds in nanoparticles) (Derfus *et al.* 2004). However, core/shell QDs with a stable polymer coating (which are currently used in most biomedical applications) do not exhibit degradation under standard

imaging and biologically relevant conditions (Dubertret *et al.* 2002; Gao *et al.* 2004).

## 17.3 Engineering of QD-based probes for biomedical applications

Having developed robust techniques for synthesis of nanoparticles with precisely controlled size and geometry, internal structure, and surface chemistry, scientists have placed increasing effort towards the design of novel coating and bioconjugation methods for production of small biocompatible and multifunctional QD-based nanoprobes. A brief overview of most common QD synthesis techniques along with water-solubilization and bioconjugation approaches is presented in this section.

### 17.3.1 QD synthesis

Quantum dots are generally composed of atoms from Groups II–VI (e.g. CdSe and CdTe) and III–V (e.g. InP and InAs) elements of the Periodic Table. Both types of nanocrystals have been synthesized and studied extensively in the past. The first synthesis protocols involved preparation of QDs in aqueous solutions with addition of stabilizing agents, such as thioglycerol and polyphosphate. Even though such methods provided a relatively easy way for synthesis of water-soluble QDs, poor control over the nanoparticle size (with relative standard deviation, RSD, greater than 15%) resulted in production of a polydisperse QD population, which emitted light over a broad spectral range, thus greatly compromising the multiplexing capabilities of this technology (Chan *et al.* 2002). Moreover, QDs produced in this manner had poor fluorescence efficiencies, resulting in reduced sensitivity of QD-based fluorescent imaging and detection.

Realization of QD potential for use in biomedical applications came in the 1990s, when a high-temperature organometallic procedure for synthesis of highly uniform colloidal CdSe QDs was introduced by Bawendi and coworkers (Murray *et al.* 1993). In this procedure a mixture of trioctyl phosphine/trioctyl phosphine oxide (TOP/TOPO) is combined with pyrolysis of organometallic precursors at high temperature (Medintz *et al.* 2005b). TOPO, being a high boiling point base with long hydrocarbon chains, coordinates with unsaturated metal atoms on the QD surface, coating the QD with a monolayer of hydrophobic alkyl chains and preventing the formation of bulk semiconductor (Alivisatos 1996a; Smith *et al.* 2006). The conical shape of TOPO improves the packing density on the curved surface of nanocrystals—while only half of the Cd sites can be covered on a flat crystal surface, almost all Cd sites are capped on a highly curved surface of a 2-nm diameter QD (Alivisatos 1996a). Although the overall procedure yields highly crystalline CdSe QDs with narrow size distribution (RSD < 5%), the fluorescence quantum yield of such nanoparticles remains low (~10%) (Chan *et al.* 2002). Moreover,

TOPO-coated Group II–VI QDs are unstable with respect to photooxidation. For example, in CdSe nanocrystals, Se is oxidized to selenate upon exposure to visible light. The resulting oxide evaporates from the surface as a molecular species and leaves behind reduced Cd and a freshly exposed layer of chalcogenide (Alivisatos 1996a; Derfus *et al.* 2004). This process leads to effective destruction of nanocrystals and to potential QD toxicity due to release of free Cd ions.

Deposition of a surface-capping layer of a material with larger bandgap has been found to dramatically increase the quantum yield of QDs, effectively passivate the core surface, protect it from oxidation, and prevent leaching of core atoms into surrounding solution. The best available QDs with highly refined chemistry are made of CdSe cores overcoated with a layer of ZnS (Medintz *et al.* 2005b). Due to the similarity between the lengths of Zn–S and Cd–Se bonds, the lattice mismatch between the two materials is reasonable, thus allowing formation of a stable core/shell structure (Alivisatos 1996a; Chan *et al.* 2002). ZnS has a wider bandgap (3.8 eV) than CdSe (1.75 eV) and, therefore, can efficiently confine the exciton to the core, eliminating non-radiative relaxation pathways and increasing quantum yield (sometimes up to 80%, but typically around 40–50% at room temperature) (Michalet *et al.* 2001; Chan *et al.* 2002). Although thin ZnS shells (1–2 monolayers) often produce the highest fluorescence yields, thicker shells (4–6 monolayers) provide more core protection from photooxidation and degradation in biological media (Medintz *et al.* 2005b). Also, a thicker layer of ZnS might significantly reduce QD blinking (Nirmal *et al.* 1996). Blinking (alteration of bright and dark states) is associated with charge trapping and untrapping at surface defects of a nanocrystals and with the charge ejection from the dot (Auger ionization process) followed by a recombination process (Efros and Rosen 1997; Michalet *et al.* 2001; Pinaud *et al.* 2006). Even though this effect can be used to differentiate single QDs from aggregates, it might be detrimental to some QD applications. For example, blinking causes signal fluctuations in imaging of fast biological processes (Lidke *et al.* 2004), loss of distance information when movement of a single molecule is observed, and spectral jumping (change in the emission peak position) that affects the FRET efficiency (Michalet *et al.* 2001; Pinaud *et al.* 2006). Therefore, elimination of blinking is desirable, and a compromise between quantum yield and blinking may need to be found for individual applications. Alternative capping methods, such as QD surface passivation with thiol groups of beta-mercaptoproethanol ligands, have been reported to yield near-complete elimination of blinking (Hohng and Ha 2004).

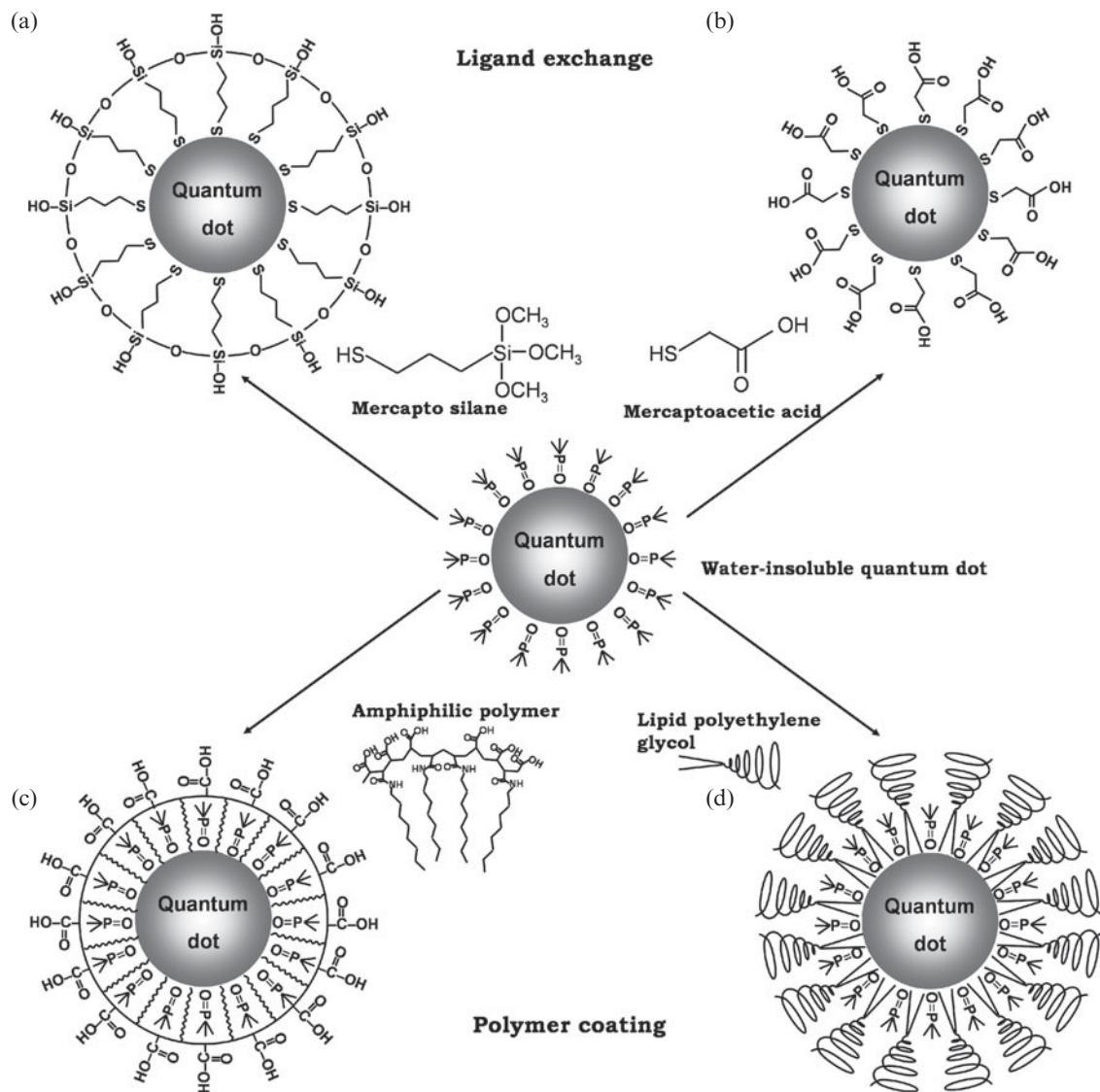
Recently, a leap towards large-scale synthesis of high-quality quantum dots was made by Peng *et al.*, who have utilized alternative cheap precursor materials (such as CdO) for production of highly crystalline QDs that show excellent quantum yields (85% at room temperature) without an inorganic capping layer (Peng and Peng 2001, 2002; Li *et al.* 2003). However, utilization of such QDs in biological applications still requires capping and further surface modification to yield biocompatible probes stable in physiologically relevant conditions.

### 17.3.2 QD water solubilization

So far, synthesis of quantum dots in organic solvents with TOPO stabilization has provided the best control over nanoparticle size and resulted in production of stable nanocrystals with high fluorescence efficiencies. However, such QDs are soluble only in organic solvents, such as chloroform or hexane, and, therefore, are not readily applicable for exploration of water-based biological systems. A variety of surface-coating techniques have been implemented (and more are still under development) with the ultimate goal of making QDs soluble and stable in biological buffers, while preserving their original photophysical properties (Pinaud *et al.* 2006). As new QD-based applications are being explored, more stringent and demanding requirements for QD surface coating arise: the size of QDs should stay small after coating, the surface should be biocompatible and non-immunogenic, reactive groups should be available for conjugation of biomolecules and targeting ligands, and yet QDs should show minimal non-specific interactions with the biological environment. Therefore, improvement of existing coating techniques and design of novel application-specific water-solubilization and bioconjugation approaches remains an active area of research.

One of the first effective water-solubilization strategies via encapsulation of QDs into polymerized silanol shells was developed by Bruchez and coworkers (Bruchez *et al.* 1998). In this procedure 3-(mercaptopropyl)trimethoxysilane (MPS) is directly absorbed onto the nanocrystals, displacing TOPO molecules (Fig. 17.4(a)). Upon addition of base, silanol groups are hydrolyzed producing silica/siloxane shell and rendering particles soluble in intermediate polar solvents (e.g. methanol or dimethyl sulfoxide). Further reaction with bifunctional methoxy compounds provides solubility in aqueous buffers. Nanoparticles produced in such a manner maintain a relatively small size and are highly stable against flocculation. However, low reaction yield, poor fluorescence efficiency (~21%), and the tendency to precipitation and gel formation at neutral pH preclude utilization of this approach for large-scale production and widespread use of QD-based bioprobes (Chan *et al.* 2002). In a similar approach, Guo *et al.* have obtained highly stable nanocrystals soluble in polar solvents by encapsulation of QDs in cross-linked dendron boxes (Guo *et al.* 2003). Further surface modifications are required to render such QDs water soluble and capable of bioconjugation.

Another route for water solubilization of QDs is substitution of the native TOPO coating with small bifunctional ligands presenting both a surface-anchoring group (e.g. thiol) and a hydrophilic end group (e.g. carboxyl or hydroxyl). Chan and Nie first implemented a ligand-exchange procedure by coating CdSe/ZnS core/shell QDs with mercaptoacetic acid and demonstrated the utility of such nanoparticles for *in-vitro* live-cell staining and sensitive immunoassays (Chan and Nie 1998). In their procedure, when reacted with TOPO-coated nanoparticles at the water chloroform interface in the presence of organic base, mercapto groups displaced TOPO and bound to Zn atoms of the ZnS shell. Polar carboxylic acid groups rendered QDs soluble and stable in aqueous buffers. Additionally, carboxylic groups provide binding sites for cross-linking to proteins, peptides, and nucleic acids (Fig. 17.4(b)) (Chan and



**Fig. 17.4** Two general strategies for water solubilization of TOPO-coated QDs are presented. Ligands are drawn disproportionately large for detail; ligand-polymer coatings are usually only 1–2 nm in thickness. (a & b) In a ligand-exchanged approach TOPO coating is replaced by heterobifunctional ligands, such as mercapto silanes (a) or mercaptoacetic acid (b) generating hydrophilic QDs with a shell of silica or carboxylic acids on the surface. (c & d) In the polymer-coating procedure, TOPO is retained on the QD surface and nanocrystals are rendered water soluble through micelle-like interactions with amphiphilic polymers (c) or lipids (d). (Smith, *et al.*, *Multicolor quantum dots for molecular diagnostics of cancer. Expert. Rev. Mol. Diagn.* **6**(2), p. 231–44, 2006. Copyright Future Drugs Ltd.)

Nie 1998; Chan *et al.* 2002). Although this procedure is quick, simple, and efficient in producing small water-soluble nanoparticles, the quantum yield of QDs decreases significantly upon water solubilization. The shelf life of such nanoparticles is also very short due to poor stability of mono-thio compounds. Once QDs are purified and excess mercaptoacetic acid is removed from the solution, ligands start to detach from the surface, causing QD precipitation.

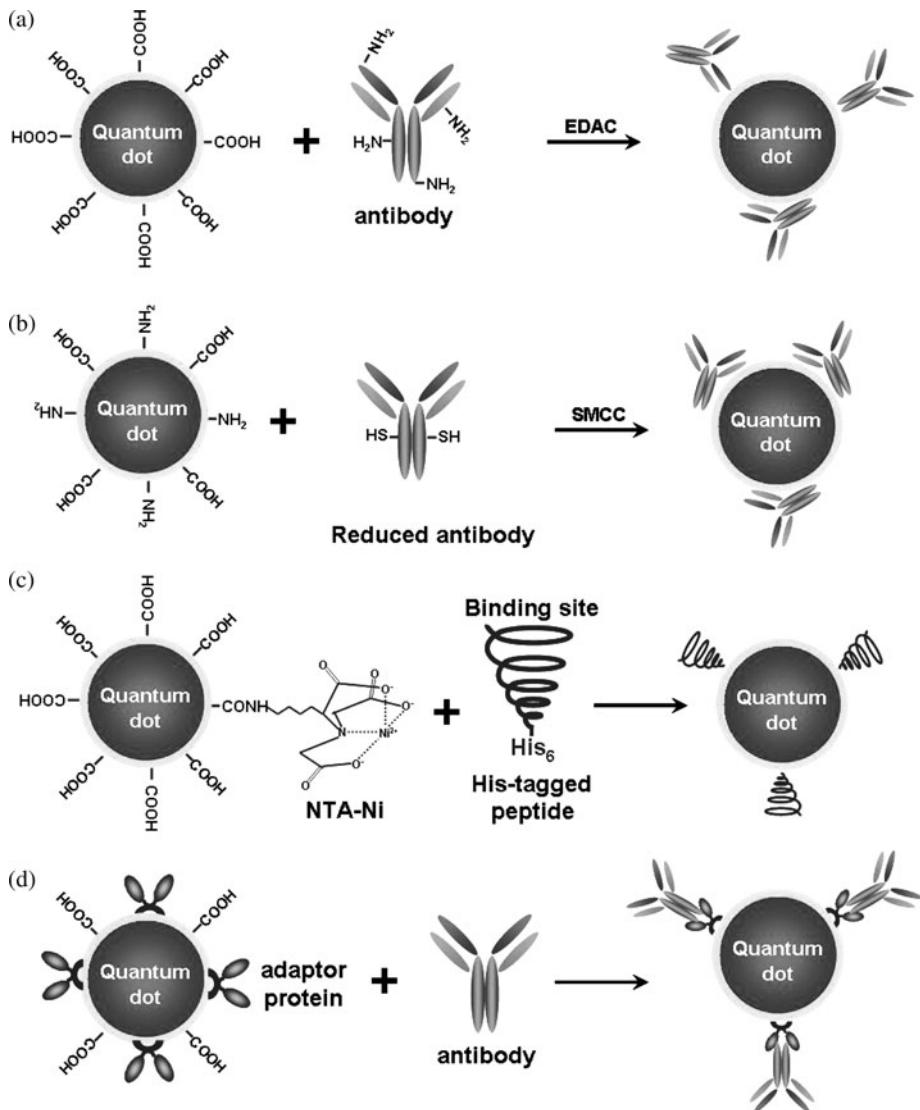
Substitution from mono-thio to di-thio ligands can improve the storage stability (Medintz *et al.* 2005b; Pons *et al.* 2006). For example, Liu *et al.* have utilized di-thiol ligand dihydrolipoic acid conjugated to poly(ethylene glycol) to prepare small (hydrodynamic diameter of 11.4 nm) and stable QDs with minimal loss of fluorescence efficiency (drop in quantum yield from 65% to 43%) (Liu *et al.* 2008). In an alternative approach, Jiang *et al.* have improved the stability of mercapto-acid shell by covalently cross-linking neighboring molecules with lysine (Jiang *et al.* 2006). However, the dramatic increase in nanoparticle size (from 8.7 to 20.3 nm hydrodynamic diameter) induced by shell cross-linking is undesirable, and further optimization of this procedure is required. Weiss and colleagues have developed a method for CdSe/ZnS QD surface passivation and water solubilization using phytochelatin-related peptides (Pinaud *et al.* 2004). Besides yielding relatively small water-soluble nanoparticles, this approach also provides points of chemical modification and biological functionality on QD surfaces. Belcher *et al.* have demonstrated that the phage-display libraries can be used to identify, develop, and amplify binding between peptides and inorganic semiconductor substrates (Whaley *et al.* 2000). With the use of such libraries and accelerated evolution techniques it might be possible to select peptide sequences that can specifically bind to any type of nanomaterial, thus making peptide coating and surface functionalization more general (Pinaud *et al.* 2006).

The third strategy preserves the native TOPO coating on the QDs and involves overcoating with variants of amphiphilic “diblock” and “triblock” copolymers and phospholipids (Dubertret *et al.* 2002; Wu *et al.* 2003; Gao *et al.* 2004). The reaction is mainly driven by absorption of amphiphilic polymer coating onto the TOPO-coated nanoparticles via hydrophobic interactions. In one method QDs are solubilized with an octylamine-modified polyacrylic polymer. The hydrophobic alkyl sidechains interweave alkylphosphine chains of TOPO on the QD surface exposing hydrophilic carboxylic acid groups to the solution, thus rendering QDs water soluble (Fig. 17.4(c)). Another method utilizes coating with polyethylene glycol (PEG)-lipid layer (Dubertret *et al.* 2002), which has an amphiphilic surfactant structure (Fig. 17.4(d)). Both methods keep the original TOPO layer, which is important for maintaining the optical properties of nanoparticles and for shielding the core from contact with the outside environment (Gao and Nie 2003). In fact, polymer-coated QDs linked to PEG molecules are protected to such a degree that their optical properties do not change in a broad range of pH and salt concentrations (Gao *et al.* 2004). Linking to PEG also leads to improved biocompatibility and reduced non-specific binding. The drawback of this approach is that the resulting particles have final sizes three to four times larger than the original particles. For example, in some cases block copolymer coating increases the diameter of CdSe/ZnS QDs from ~4–8 nm before encapsulation to ~20–30 nm (Medintz *et al.* 2005b). This might be detrimental for quantitative biomarker detection in a crowded biological environment as bulky QD probes might introduce steric hindrance and block access to densely localized targets, which, in turn, might lead to a drop in detection sensitivity and non-linearity in quantitative analysis. Therefore, polymer-coated QDs are most suited for sensitive detection and quantitative analysis of low copy number and/or sparsely distributed markers

(Dahan *et al.* 2003; Pinaud *et al.* 2006; Cui *et al.* 2007). The thickness of the polymer coating might also preclude QDs from fluorescence resonance energy transfer (FRET)-based investigation (Clapp *et al.* 2004; Medintz *et al.* 2005b; Pinaud *et al.* 2006). Novel encapsulation and bioconjugation approaches are being developed to provide higher nanocrystal protection with thin, biocompatible, and functional coatings.

### 17.3.3 QD bioconjugation

Quantum dots are unique in that they combine small protein-scale size and large surface area. Due to this feature, QDs represent versatile nanoscaffolds for attachment of multiple proteins, peptides, and nucleic acids, thus enabling design of a variety of multifunctional nanoparticle–biological hybrids (Kim *et al.* 2008; Nehilla *et al.* 2008; Qi and Gao 2008). Such hybrids combine properties of both materials involved (i.e. optical properties of the nanocrystals and biological functions of ligands attached) and are therefore a very powerful tool in biomedical applications. Several approaches can be used for conjugation of QDs and biological molecules. For example, covalent-bond formation between reactive functional groups (e.g. primary amines, carboxylic acids, alcohols, and thiols) is one of the most popular bioconjugation methods. Many proteins contain primary amine groups that can be linked to negatively charged QDs, exposing carboxylic acid groups on the surface, via carbodiimide-mediated amide formation (i.e. EDC, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide, condensation reaction). This reaction does not require additional chemical modification of proteins, preserving their natural structure. However, molecular orientation of the attached proteins cannot be manipulated using EDC conjugation, as proteins can attach to QDs at any point where an amine group is available, including the active sites (such as F(ab) domains of 4antibodies) (Fig. 17.5(a)). Attachment at a point close to a ligand’s active site might result in partial or complete loss of biological functionality of that ligand. Moreover, since a large excess of ligand molecules is required to achieve efficient conjugation, EDC reaction provides poor control over the number of biomolecules attached to a single nanocrystal and usually results in QD aggregation due to cross-linking between multiple reactive sites on QDs and proteins. Nonetheless, this approach was successfully used in preparation of commercial QD-streptavidin conjugates with ~5–10 proteins attached to single nanocrystals. In such an approach, biotinylated biomolecules can be attached to QDs via a biotin–streptavidin bond (Medintz *et al.* 2005b). Even though this procedure has been shown easy to use and highly effective for staining of cellular proteins (Wu *et al.* 2003; Howarth *et al.* 2008), it is not optimal for multiplexed protein detection, since streptavidin binds to all biotinylated proteins indiscriminately. Recently, Bagalkot *et al.* have utilized EDC coupling to decorate the surface of polymer-encapsulated QDs with double-stranded RNA aptamers targeting PSMA on prostate-cancer cells. Efficient targeted drug delivery and simultaneous live-cell imaging was demonstrated using this construct (Bagalkot *et al.* 2007). Yet, the utility of such QD bioconjugates for staining of fixed cells and



**Fig. 17.5** Methods for conjugating QDs to biomolecules. (a) Traditional covalent cross-linking chemistry using EDAC (ethyl-3-dimethyl amino propyl carbodiimide) as a catalyst produces bioconjugates with random orientation of ligands. (b) Conjugation of antibody fragments to QDs via reduced sulphydryl-amine coupling mediated by heterobifunctional cross-linker SMCC, succinimidyl-4-Nmaleimidomethyl-cyclohexane carboxylate controls ligand orientation, but destroys its natural structure. (c) Conjugation of histidine-tagged peptides and proteins to Ni-NTA-modified QDs provides potential control of the attachment site and QD:ligand molar ratios. (d) Conjugation of antibodies to QDs via an adaptor protein preserves the natural structure of the ligand and controls its orientation. (Gao, *et al.*, *In vivo molecular and cellular imaging with quantum dots*. *Curr. Opin. Biotechnol.* **16**(1), p. 63–72, 2005. Copyright Elsevier.)

tissues or their applicability for *in-vivo* drug delivery and imaging applications requires further evaluation.

Another common covalent-bonding procedure involves active ester maleimide-mediated amine and sulphydryl coupling. Since free sulphydryl groups are rare in native biomolecules, additional treatment of the ligands

is often required. For example, in a procedure used in Invitrogen's antibody-QD conjugation kit, an antibody is first reduced by DTT (dithiothreitol) to expose free sulfhydryls, while QDs are activated using a hetero-bifunctional cross-linker SMCC (4-(maleimidomethyl)-1-cyclohexanecarboxylic acid N-hydroxysuccinimide ester) providing sylfhydryl-reactive maleimide groups. This reaction yields stable QD–antibody complexes and provides more control over antibody orientation on the surface of nanocrystals (Fig. 17.5(b)). However, reduction of antibodies with DTT substantially decreases their affinity to binding sites and, therefore, decreases the overall sensitivity of the bio-probe (Pathak *et al.* 2007). Despite these limitations Parak *et al.* have successfully conjugated amine-modified double-stranded DNA oligonucleotides to silanized QDs via SMCC reaction (Parak *et al.* 2002). Yezhelyev and colleagues have utilized a SMCC procedure for QD–antibody conjugation and have successfully developed a QD-based assay for quantitative detection of three breast cancer markers—estrogen receptor, progesterone receptor, and ERBB2—in paraffin-embedded human breast-cancer cells (Yezhelyev *et al.* 2005).

As an alternative to covalent bonding to functionalized QD coating, biomolecules can be linked directly to the nanocrystal surface via a thiol-exchange reaction. Here, mercapto-coated QDs are mixed with thiolated biomolecules or biomolecules containing polyhistidine (HIS) residues, and a small fraction of the surface ligands is replaced by the molecule of interest (Medintz *et al.* 2005b). A thiol-exchange reaction provides more control over the final bioconjugate assembly due to the ability of rational design of poly-HIS tags or targeted incorporation of thiolated anchors. In one example, Lao *et al.* have designed a tripartite fusion protein consisting of an N-terminal HIS-tag, a stimulus-responsive elastin-like peptide (ELP), and a C-terminal IgG-binding protein L (Lao *et al.* 2006). In this construct, HIS-tag effectively binds to QD surface, orienting the remainder of the ligand away from the nanoparticle surface; ELP causes reversible aggregation in high-salt buffers, aiding QD–bioconjugate purification; and protein L binds to IgG light chains with high affinity, allowing preparation of aggregate-free QD–antibody conjugates. Yet, the major downside of a thiol-exchange approach is the necessity for unstable surface coatings (such as mercapto compounds), which significantly reduce the brightness and stability of such bioconjugates in aqueous solutions. Gao *et al.* have developed an indirect HIS-tag coupling method by linking QDs to a chelating compound Ni-NTA (nickel-nitrilotriacetic acid), which binds to hexahistidine-tagged biomolecules in a quantitative and controlled manner. In contrast to direct HIS-tag coupling described above, this procedure is well suited for conjugation with stable polymer-coated nanocrystals (Fig. 17.5(c)) (Gao *et al.* 2005).

Another conjugation approach involves non-covalent self-assembly of engineered proteins on the surface of QDs via electrostatic interactions. For example, in a simple and straightforward procedure Goldman *et al.* deposited avidin—a highly positively charged glycoprotein—on the surface of negatively charged QDs for further conjugation to biotinylated antibodies (Goldman *et al.* 2002), while Medintz *et al.* have utilized this procedure to decorate biotinylated cowpea mosaic virus with QDs (Medintz *et al.* 2005a). Similarly,

Mattossi and colleagues have described the use of a chimeric fusion protein for indirect coupling of native unmodified immunoglobulin G antibodies to QDs (Mattossi *et al.* 2000). In this approach, the positively charged leucine zipper domain of an adapter protein electrostatically interacts with negatively charged QDs while the protein G domain binds to the antibody Fc region. As a result, the target-specific F(ab')<sub>2</sub> domain of IgG faces outwards (Fig. 17.5(d)). Since primary antibodies can be directly conjugated to QDs, a multistep immunostaining procedure becomes straightforward, and multiple biomarkers can be labelled during one staining round. However, the size of such bioconjugates is large due to a number of thick biomolecule layers on the QD surface. Our experience has shown that the large size of QD probes is usually associated with a number of problems in biomarker detection, such as increased non-specific binding, slower diffusion to a target, decreased ability to penetrate biological membranes, and steric hindrance between QD probes.

Recent achievements in merging nanoparticle encapsulation and bioconjugation and design of pre-functionalized surface coatings promise to provide more compact, stable, and biocompatible nanoparticles with controlled density and orientation of ligands attached. Amphiphilic polymers with maleic anhydride backbone are being actively explored for this purpose. In organic anhydrous solvents, such polymers encapsulate TOPO-coated QDs and introduce reactive anhydride groups on the surface. In basic aqueous buffers anhydride rings are quickly hydrolyzed, thus yielding negatively charged carboxylic acid groups and rendering QDs water soluble (Pellegrino *et al.* 2004). More importantly for conjugation, anhydride groups are highly reactive towards amine-containing molecules. By utilizing this feature, the Parak group have conjugated a variety of biomolecules to polymer chains, thus producing a pre-functionalized coating for QD encapsulation (Fernandez-Arguelles *et al.* 2007; Lin *et al.* 2008).

A great number of different surface coating and bioconjugation techniques have been developed. The complexity of these procedures as well as the quality of resulting QD bioconjugates and degree of reaction control range widely. Some methods, such as direct covalent conjugation of ligands via SMCC or EDC reaction and indirect coupling via the streptavidin–biotin bond, have found wide use in many biomedical applications and especially in the area of multicolor cell and tissue imaging (Xing *et al.* 2007). Yet, presently there is no universal method best suited for all biomedical applications.

## 17.4 Tumor molecular imaging and profiling

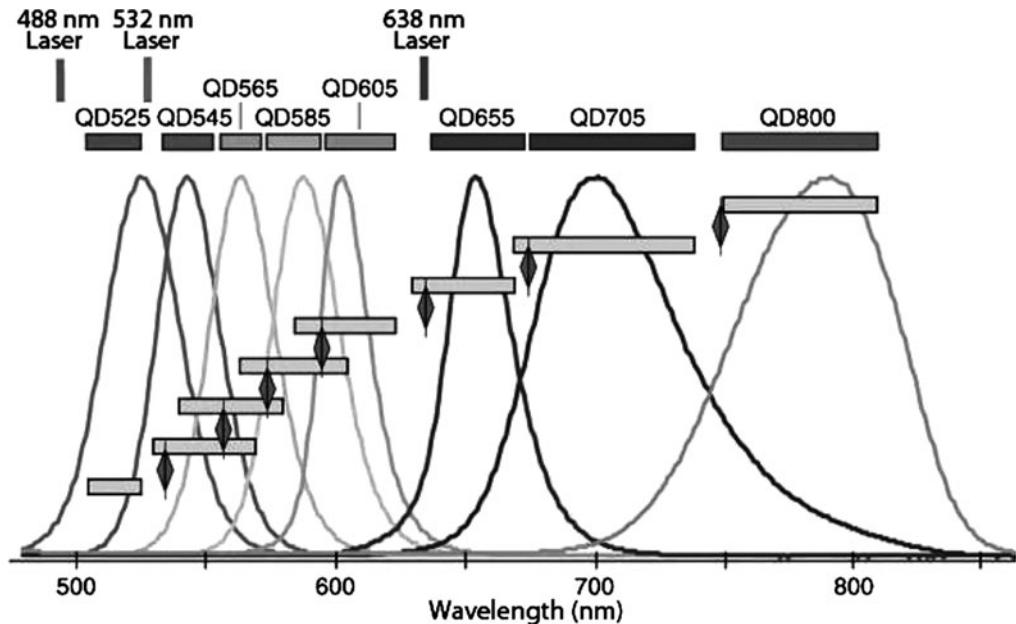
Various labelling techniques for detection of DNA, mRNA, proteins, and other biomolecules in cells and tissue samples are currently used in clinical practice and research. All methods are based on specific interaction between the target and its ligand (e.g. antigen and antibody, or DNA and complementary oligonucleotide) with visualization of the target position via fluorescence (e.g. organic dyes), bright-field (e.g. peroxidase-mediated color development), or electron (e.g. gold nanoparticles) microscopy. Conventional immunohistochemistry

methods have been successfully used for reliable and consistent staining of clinical tissue samples. However, their largely single-color and qualitative nature greatly limits their application in molecular profiling of tumor-tissue specimens. At the same time, QD-based probes with unique optical properties, stable and biocompatible coatings, and functionalized surfaces have already shown their outstanding performance in multiplexed fluorescent detection and quantitative analysis of proteins and nucleic acids in cells and tissue sections.

### 17.4.1 QD-based multicolor flow cytometry

Polychromatic flow cytometry is a great tool for quick and accurate identification, separation, counting, and characterization of many cell types by fluorescence immunophenotyping. In a basic setup, a beam of monochromatic light (usually produced by a laser) is directed onto a focused stream of fluid containing fluorescently labelled cells. Excitation of fluorescent tags by a laser is followed by emission of light of a longer wavelength, which is quantified by detectors. Many organic fluorophores have been successfully used in flow-cytometry applications. However, broad overlapping emission spectra and narrow excitation bands of organic dyes require the use of quite complicated instrumentation, consisting of multiple excitation sources and detectors. This significantly hampers simultaneous detailed characterization of multiple cellular proteins. Currently, the practical limit of 12-color flow cytometry has been achieved with organic fluorescent dyes (Chattopadhyay *et al.* 2006).

Utilization of the unique optical properties of QDs can help to overcome these limitations, improving the multiplexing capability and sensitivity of flow cytometry and bringing down the cost of experimental setups. Chattopadhyay *et al.* have shown that at least 17 fluorophores can be independently quantified on a cell-by-cell basis using polychromatic flow cytometry (Chattopadhyay *et al.* 2006), thus achieving accurate immunophenotyping of antigen-specific T-cells using a panel of 8 QD probes and 9 organic fluorophores. In this setup, all QDs can be simultaneously excited by a single 408-nm laser source, while three lasers of longer wavelengths (488, 532, and 633nm) have to be used to separately excite organic dyes (Fig. 17.6). As most organic fluorophores cannot be efficiently excited by a blue light, substantial reduction of emission overlap from organic dyes in the QD channel is readily achieved. The ability to multiplex quantum dots with other fluorophores in this manner significantly increases the number of parameters that can be measured simultaneously on single cells (Chattopadhyay *et al.* 2006). Application of bright multicolor QDs in polychromatic flow cytometry empowers this method with unprecedented sensitivity and multiplexing capabilities, thus allowing for comprehensive analysis of the cellular composition of complex biological fluids (such as blood) and extraction of single cells based on their biomarker profiles. Even though this technology has been primarily used for phenotyping of T-cells and for the study of their differentiation, it is reasonable to expect that QD-based high-throughput multicolor flow cytometry will become an indispensable tool



**Fig. 17.6** Multicolor QDs can be simultaneously detected and resolved in polychromatic flow cytometry. Emission spectra of various QDs are narrow and symmetrical and can be easily separated using a set of dichroic and bandpass filters. Wavelengths for long-pass dichroic filters are shown as diamonds and for bandpass filters—as bars. The bars above the spectra represent the fraction of light collected by each of the detectors. All QDs can be excited by a single 408-nm laser (not shown here), whereas a set of longer-wavelength lasers is used for excitation of organic dyes. (Chattopadhyay, et al., *Quantum dot semiconductor nanocrystals for immunophenotyping by polychromatic flow cytometry*. *Nature Med.* **12**(8), p. 972–7, 2006. Copyright Macmillan Publishers Ltd.)

for molecular profiling of cancer and stem cells, study of tumor proliferation and metastasis, accurate cancer diagnostics and prognosis, and monitoring of therapeutic efficacy.

#### 17.4.2 Applications of QDs in high-resolution correlated fluorescence/electron microscopy

Electron microscopy is an imaging technique that utilizes electrons rather than photons to probe a sample and has a much higher magnification and resolving power than fluorescence microscopy. It provides access to localization of proteins on a subcellular level. In order to be visible in a transmission electron microscope (TEM), a probe must scatter electrons efficiently, i.e. it must be electron dense. QDs are electron-dense moieties since they represent nanocrystals with tightly packed metal atoms in their cores and shells. Current synthetic techniques provide a good control over nanoparticle sizes and shapes. TEM, in turn, provides high resolution to discriminate between these slight changes, thus making QDs suitable for multiplexed target detection with electron microscopy. Since the size of QDs directly determines the emission color, low-resolution fluorescence images can be readily correlated with high-resolution TEM images, thus obtaining subcellular, cellular, and tissue-level information from a single specimen.

Giepmans *et al.* have evaluated the performance of antibody-conjugated QDs in correlated fluorescence and electron microscopy (Giepmans *et al.* 2005). In the first step of tissue staining, a standard two-step immunohistochemistry procedure using QD conjugates is performed, and fluorescence microscopy is used for staining quality assessment and optimization of various staining parameters, such as fixation strength, permeabilization time, and concentration of primary and secondary antibodies. In the second step, samples are analyzed by TEM. Preservation of the QD fluorescence is essential in order to be able to perform correlated microscopy on light and electron microscopic levels. However, preparation for the TEM requires postfixation in osmium tetroxide, which quenches QD fluorescence. Skipping the osmication step might negatively affect the ultrastructure of a sample and decrease TEM resolution. Therefore, there is a tradeoff between image resolution and correlated fluorescence/TEM imaging on the same sample until new QD-compatible protocols for TEM sample preparation become available. Nonetheless, Giepmans and colleagues have reported successful detection of at least three QD-labelled biomarkers distinguishable at both fluorescence and TEM levels with good resolution (Giepmans *et al.* 2005). Additional multiplexing functionality of this technique can be obtained from discrimination of QDs based on their elemental composition. Nisman *et al.* have proposed the use of electron spectroscopic imaging (ESI, a technique for generating elemental maps of materials with high resolution and detection sensitivity) for mapping the distribution of quantum dots in cell and tissue components based on QD internal chemistry (Nisman *et al.* 2004). Fluorescence/TEM correlated imaging is not meant to be used for high-throughput quantitative screening of clinical samples, but rather as a tool for ultrasensitive detection of target markers and for the detailed study of distribution and relationships between different biomarkers on a subcellular level. This technique will certainly play a major role in uncovering complex molecular pathways underlying development of cancer and other pathological processes.

#### 17.4.3 QD bioprobes for molecular profiling of tumor-tissue sections and microarrays

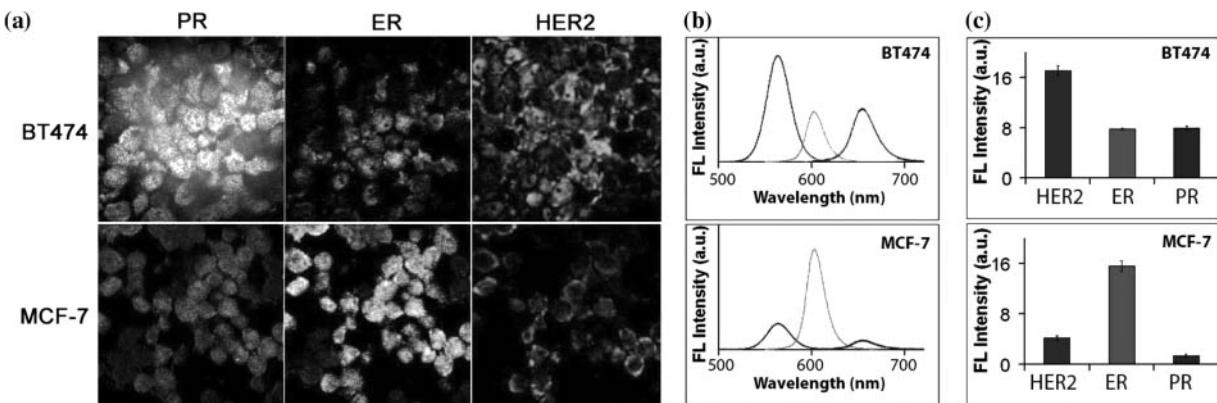
Tissue microarrays have been extensively used for high-throughput histological analysis of tumor biopsies. While standard immunohistochemistry techniques are mostly limited to qualitative one-color imaging and are subjected to inter- and intraobserver variability (Pedersen *et al.* 1989), QDs represent a promising tool for quantification and multicolor labelling of antigen-derived markers in FFPE tissue samples. Wu *et al.* have investigated the use of QDs conjugated to streptavidin and IgGs for simultaneous labelling of membrane-associated Her2 receptor and of a nuclear antigen in SK-BR-3 breast-cancer cells (Wu *et al.* 2003). In comparison to Alexa dyes, QDs have proven to be much more photostable and have produced higher signal intensity. However, while staining of cell-surface antigens was reliable and effective, staining of cytoplasmic and nuclear markers was more variable. This issue is directly associated with the large size of QD probes and can be potentially

resolved by stronger sample permeabilization and by adjusting staining conditions.

Further improvement in detection sensitivity can be achieved through the use of the tyramide signal amplification (TSA) technique reported by Ness *et al.* (2003). In this procedure, the target biomarker is identified by horseradish peroxidase (HRP)-labelled antibody. Biotinylated tyramide is converted by HRP into a highly reactive oxidized intermediate that covalently binds to cell-associated proteins in the vicinity of the HRP-labelled antibody generating a high-density labelling of target protein. Upon subsequent addition of QD-streptavidin probes, nanocrystals attach to the deposited tyramide residues via biotin–streptavidin bonds producing labelling of one biomarker with multiple QDs. Therefore, this technique can be used for ultrasensitive detection of low copy number biomarkers, but it is inherently single colored and not quantitative.

Ghazani and coworkers have integrated an optical spectroscopy technique and QDs for quantitative analysis of tumor biopsies in tissue microarrays (Ghazani *et al.* 2006). Three-color staining of lung carcinoma xenografts for epidermal growth factor receptor (EGFR), E-cadherin, and cytokeratin has been achieved by utilization of 655, 605, and 565 nm QD-based assays, and the specificity of staining has been confirmed by standard immunofluorescence imaging with Alexa 488 dye. QDs have shown superior signal intensity and photostability, enabling effective removal of autofluorescence background and reliable quantification of signal intensity using denoising and normalizing algorithms. Ghazani *et al.* have noted that quantitative comparison of different biomarkers in multiplexed staining of one tissue section could be severely compromised by the strong signal enhancement of red QDs and reduction of the 565-nm signal (Ghazani *et al.* 2006). In general, signal intensity reduction becomes more pronounced as emission wavelength decreases across the spectrum, since smaller QDs possess lower molar extinction coefficients and, thus, absorb light less efficiently compared to larger QDs. Discordance between the fluorescence properties of individual probes should not be an issue when statistical analysis of biomarkers between large numbers of samples is attempted. Yet, in order to obtain a uniform staining of multiple targets within one sample it is advisable that brighter red QDs be used for labelling of low copy number biomarkers, while dimmer blue QD conjugates be used for targets present in large quantities. Adjustment of antibody concentrations and other staining conditions can also be used to prevent saturation of a detector by one highly bright signal.

Recently, Yezhelyev *et al.* have demonstrated the use of QD bioconjugates for multiplexed labelling and quantification of five clinically significant breast-cancer markers—HER2, ER, PR, EGFR, and mTOR (Yezhelyev *et al.* 2007). The performance of QD–antibody conjugates was carefully evaluated on FFPE breast cancer cell lines as a model system. The specificity of staining and accuracy of quantitative analysis was validated by conventional techniques (IHC, western blot, and FISH) for three representative biomarkers—ER, PR, and HER2 (Fig. 17.7). In order to account for signal enhancement of red QDs and compare expression levels of biomarkers within one sample, the acquired data was adjusted according to the relative intensities of QDs



**Fig. 17.7** Detection and quantification of expression of three representative breast-cancer markers—PR, ER, and HER2 in formalin-fixed paraffin-embedded breast-cancer cell lines. (a) Fluorescence images are deconvoluted into three individual components. (b) Representative spectra obtained with single-cell spectroscopy are used in quantitative analysis of biomarker expression levels. The 565-nm and 605-nm peaks are enhanced by a factor of 8 and 4, respectively, to compensate the differential brightness of multicolor QDs. (c) Statistical analysis of biomarker expression obtained by averaging spectra of 100 single cells. (Yezhelyev, *et al.*, *In situ molecular profiling of breast cancer biomarkers with multicolor quantum dots*. *Adv. Mater.* **19**(20), p. 3146–51, 2007. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

(QD655:QD605:QD565 = 8:4:1 as measured in a separate experiment for equal QD concentrations). The utility of this technology for clinical evaluation of tumor-tissue specimens has been demonstrated by detection and quantification of the panel of all five biomarkers on FFPE breast-cancer tissue biopsies.

#### 17.4.4 QD-oligonucleotide bioconjugates for *in-situ* hybridization

Fluorescence *in-situ* hybridization (FISH) is a common method of determining gene-expression level or messenger RNA (mRNA) distribution by the use of fluorescent-labelled DNA or RNA probes. This technique is based on specific binding of oligonucleotide probes to complimentary mRNA molecules with a high degree of sequence similarity. Ideally, 1:1 probe-to-target binding should occur, yielding direct quantitative correlation between gene amplification (i.e. number of mRNA molecules present) and signal intensity, and providing accurate information about mRNA localization within the cell. However, the low copy number of target mRNA and quick photobleaching of organic fluorophores heavily compromise the quantitative potential and sensitivity of FISH. QDs conjugated with oligonucleotides could become brighter FISH probes that are easy to detect and quantify (Yezhelyev *et al.* 2006). Pathak and colleagues have evaluated the performance of QD-oligonucleotide-based probes in FISH assays in human sperm cells (Pathak *et al.* 2001). While the high negative charge of QDs carrying carboxylic acid groups on their surface precludes efficient oligonucleotide loading and resulted in significant non-specific nanoparticle binding to target cell, hydroxylated CdSe/ZnS core-shell QDs produced by coating with dithiothreitol (DTT) showed good stability and substantially lower non-specific binding under FISH conditions (pH 7.0–8.5).

Subsequent coupling of such nanoparticles with 5'-aminated oligonucleotides via carbamate linkages has yielded fully functional and stable bioconjugates. In a test assay on sperm cells, QD conjugates identified 41% of sperm cells to carry the Y chromosome (against a theoretical total of 50%), while showing very little (less than ~5%) non-specific binding and suggesting that the QD bioconjugates are stable and specific (Pathak *et al.* 2001). A direct comparison and quantification of the performance of QDs and organic fluorophores under the same experimental conditions were not available through these studies.

Xiao and Barker have used highly stable PEGylated QD–Streptavidin bioconjugates (commercially available from QD Corporation) for visualization of biotinylated oligonucleotide probes in FISH detection of human metaphase chromosomes and assessment of amplification of clinically important ERBB2 gene (Xiao and Barker 2004). In this 2-step hybridization procedure, biotinylated human DNA probes are first hybridized to cells, and then QD–streptavidin, Texas Red–streptavidin or fluorescein–streptavidin conjugates are added. Substantially higher signal-to-noise ratios have been observed for QD bioconjugates compared to organic dyes, while specificity of hybridization has been demonstrated for all three probes. Unfortunately, due to utilization of a universal biotin–streptavidin linkage, the 2-step experimental procedure used implies that only one target can be detected per sample, rendering multiplexed FISH impossible.

In a similar approach, Tholouli *et al.* have modified the hybridization protocol to allow for simultaneous detection of at least 3 mRNA molecules, while still using biotin–streptavidin linkage for QD–oligonucleotide conjugation (Tholouli *et al.* 2006). Pre-incubation of biotinylated DNA probes with QD–streptavidin conjugates allowed detection of 3 mRNA targets in a 1-step FISH procedure by simultaneous incubation of all three QD–oligo probes with a sample. Naturally, pre-conjugation of multiple oligonucleotides to QDs significantly increases the overall size of the probe, thus requiring stronger specimen permeabilization with enzymes (e.g. proteinase K), which necessarily degrades cell and tissue architecture and destroys most of the protein-based biomarkers useful for immunohistochemical studies.

Chan *et al.* have developed a more controlled procedure for pre-conjugation of exactly one oligonucleotide probe per QD via biotin–streptavidin linkage (Chan *et al.* 2005). Starting with commercial QD–streptavidin conjugates (from Quantum Dot Corp.), excess streptavidin sites are blocked with biocytin (water-soluble biotin derivative), and only a few biotinylated oligonucleotides are allowed to bind. Further purification of QD–oligo conjugates in agarose gel electrophoresis yields relatively small mono-oligonucleotide FISH probes suitable for multiplexed mRNA detection under mild specimen permeabilization. High-resolution multiplexed FISH has been demonstrated in simultaneous detection of dopamine D2 receptor mRNA,  $\varepsilon$ -sarcoglycan mRNA, tyrosine hydroxylase (TH) mRNA, and vesicular monoamine transporter (Vmat2) mRNA using two different QD probes and two different organic fluorophore probes within a single mouse midbrain neuron. The ability to use a protein-compatible specimen permeabilization techniques has allowed Chan and colleagues to successfully combine QD-based FISH and QD-based immunohistochemistry to compare cellular distribution patterns of the Vmat2

mRNA and immunoreactivity of TH in dopaminergic neurons (Chan *et al.* 2005). These results offer the possibility of correlating gene expression at the mRNA level with the number of corresponding protein copies simultaneously in tumor cells as well (Yezhelyev *et al.* 2006).

Recently, the combined QD-based FISH/immunohistochemistry technique has been improved by Matsuno *et al.* who demonstrated the use of confocal laser scanning microscopy for three-dimensional imaging of the intracellular localization of growth hormone (GH), prolactin (PRL), and of their mRNAs (Matsuno *et al.* 2005). The exceptional photostability and signal intensity of QDs have been utilized for reconstruction of high-resolution three-dimensional images of tissue samples. However, the use of a 2-step FISH procedure and universal biotin–streptavidin linker for QD–oligonucleotide conjugation limits this technique to qualitative detection of only one mRNA plus a few proteins (detected by QD–antibody conjugates) per section. Therefore, with incorporation of new probes suitable for multiplexed FISH and immunohistochemistry, Matsuno’s technique should allow three-dimensional mapping of the relative position of biomarkers and corresponding mRNAs inside cells and tissues with high resolution and sensitivity.

## 17.5 Conclusions

Quantum dots have emerged as a new class of molecular imaging agents and have already fulfilled some of their promises in quantitative molecular profiling of cells and tissues. Water solubilization of QDs with stable biocompatible coatings and functionalization with multiple biomolecules have converted QDs into multifunctional nanodevices that find their use in various applications in research and clinical diagnostics. However, further improvements are needed for this technology to receive widespread adaptation. For example, the relatively large size of QD bioconjugates hampers their deep-tissue penetration and requires strong permeabilization treatment, which sometimes destroys subcellular structures. Development of compact QD bioconjugates would resolve this problem. The QD–ligand conjugation chemistry is still under development and currently cannot effectively control the number and orientation of bioligands, which directly affects the staining stoichiometry. The lack of standardization procedures makes large-scale quantitative analysis of cancer biopsies problematic due to possible variations in data interpretation by different labs. With these improvements, QD-based molecular profiling technology will allow researchers to correlate a panel of cancer biomarkers with cancer behavior and clinical outcome and will provide an opportunity for development of personalized treatment schemes targeted specifically against biomarkers expressed.

## Acknowledgments

We thank the Department of Bioengineering at the University of Washington for the new faculty startup fund, the National Science Foundation for a Faculty Early Career Development award (CAREER), and the Congressionally

Directed Medical Research Programs for a new investigator award. We are also grateful to Prof. Larry True for fruitful discussion and technical assistance in pathology studies of prostate cancer, and to Jesse Burk-Rafel and Shivang Dave for critical reading of the manuscript.

## References

- Alivisatos, A.P. *J. Phys. Chem.* **100**, 13226 (1996a).
- Alivisatos, A.P. *Science* **271**, 933 (1996b).
- Alivisatos, A.P., Gu, W., Larabell, C. *Ann. Rev. Biomed. Eng.* **7**, 55 (2005).
- Arora, A., Scholar, E.M. *J. Pharmacol. Exp. Ther.* **315**, 971 (2005).
- Bagalkot, V., Zhang, L., Levy-Nissenbaum, E., Jon, S., Kantoff, P.W., Langer, R., Farokhzad, O.C. *Nano Lett.* **7**, 3065 (2007).
- Bruchez, Jr., M., Moronne, M., Gin, P., Weiss, S., Alivisatos, A.P. *Science* **281**, 2013 (1998).
- Chan, P., Yuen, T., Ruf, F., Gonzalez-Maeso, J., Sealfon, S.C. *Nucleic Acids Res.* **33**, e161 (2005).
- Chan, W.C., Maxwell, D.J., Gao, X., Bailey, R.E., Han, M., Nie, S. *Curr. Opin. Biotechnol.* **13**, 40 (2002).
- Chan, W.C., Nie, S. *Science* **281**, 2016 (1998).
- Chattopadhyay, P.K., Price, D.A., Harper, T.F., Betts, M.R., Yu, J., Gostick, E., Perfetto, S.P., Goepfert, P., Koup, R.A., De Rosa, S.C., et al. *Nature Med.* **12**, 972 (2006).
- Chaurand, P., Sanders, M.E., Jensen, R.A., Caprioli, R.M. *Am. J. Pathol.* **165**, 1057 (2004).
- Clapp, A.R., Medintz, I.L., Mauro, J.M., Fisher, B.R., Bawendi, M.G., Matoussi, H. *J. Am. Chem. Soc.* **126**, 301 (2004).
- Cui, B., Wu, C., Chen, L., Ramirez, A., Bearer, E.L., Li, W.P., Mobley, W.C., Chu, S. *Proc. Natl. Acad. Sci. USA* **104**, 13666 (2007).
- Dahan, M., Laurence, T., Pinaud, F., Chemla, D.S., Alivisatos, A.P., Sauer, M., Weiss, S. *Opt. Lett.* **26**, 825 (2001).
- Dahan, M., Levi, S., Luccardini, C., Rostaing, P., Riveau, B., Triller, A. *Science* **302**, 442 (2003).
- Derfus, A.M., Chan, W.C.W., Bhatia, S.N. *Nano Lett.* **4**, 11 (2004).
- Dubertret, B., Skourides, P., Norris, D.J., Noireaux, V., Brivanlou, A.H., Libchaber, A. *Science* **298**, 1759 (2002).
- Efros, A.L., Rosen, M. *Phys. Rev. Lett.* **78**, 1110 (1997).
- Elledge, R.M., Green, S., Pugh, R., Allred, D.C., Clark, G.M., Hill, J., Ravdin, P., Martino, S., Osborne, C.K. *Int. J. Cancer* **89**, 111 (2000).
- Fernandez-Arguelles, M.T., Yakovlev, A., Sperling, R.A., Luccardini, C., Gaillard, S., Medel, A.S., Mallet, J.M., Brochon, J.C., Feltz, A., Oheim, M., et al. *Nano Lett.* **7**, 2613 (2007).
- Gao, X., Cui, Y., Levenson, R.M., Chung, L.W., Nie, S. *Nature Biotechnol.* **22**, 969 (2004).
- Gao, X., Nie, S. *Trends Biotechnol.* **21**, 371 (2003).

- Gao, X., Yang, L., Petros, J.A., Marshall, F.F., Simons, J.W., Nie, S. *Curr. Opin. Biotechnol.* **16**, 63 (2005).
- Geho, D.H., Luchini, A., Garaci, E., Belluco, C., Petricoin, E., Liotta, L.A. *Nanomedicine* **2**, 1 (2007).
- Ghazani, A.A., Lee, J.A., Klostranec, J., Xiang, Q., Dacosta, R.S., Wilson, B.C., Tsao, M.S., Chan, W.C. *Nano Lett.* **6**, 2881 (2006).
- Giepmans, B.N., Deerinck, T.J., Smarr, B.L., Jones, Y.Z., Ellisman, M.H. *Nature Methods* **2**, 743 (2005).
- Goldman, E.R., Anderson, G.P., Tran, P.T., Mattossi, H., Charles, P.T., Mauro, J.M. *Anal. Chem.* **74**, 841 (2002).
- Grecco, H.E., Lidke, K.A., Heintzmann, R., Lidke, D.S., Spagnuolo, C., Martinez, O.E., Jares-Erijman, E.A., Jovin, T.M. *Microsc. Res. Tech.* **65**, 169 (2004).
- Guo, W., Li, J.J., Wang, Y.A., Peng, X. *J. Am. Chem. Soc.* **125**, 3901 (2003).
- Hohng, S., Ha, T. *J. Am. Chem. Soc.* **126**, 1324 (2004).
- Howarth, M., Liu, W., Puthenveetil, S., Zheng, Y., Marshall, L.F., Schmidt, M.M., Wittrup, K.D., Bawendi, M.G., Ting, A.Y. *Nature Methods* **5**, 397 (2008).
- Jain, K.K. *Expert Opinion on Pharmacotherapy* **6**, 1463 (2005).
- Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R.C., Ghafoor, A., Feuer, E.J., Thun, M.J. *Ca-a Cancer J. Clin.* **55**, 10 (2005).
- Jiang, W., Mardyani, S., Fischer, H., Chan, W.C.W. *Chem. Mater.* **18**, 872 (2006).
- Kim, J., Lee, J.E., Lee, S.H., Yu, J.H., Lee, J.H., Park, T.G., Hyeon, T. *Adv. Mater.* **20**, 478 (2008).
- Kim, S., Fisher, B., Eisler, H.J., Bawendi, M. *J. Am. Chem. Soc.* **125**, 11466 (2003).
- Klostranec, J.M., Chan, W.C.W. *Adv. Mater.* **18**, 1953 (2006).
- Konecny, G., Pauletti, G., Pegram, M., Untch, M., Dandekar, S., Aguilar, Z., Wilson, C., Rong, H.M., Bauerfeind, I., Felber, M., et al. *J. Nat. Cancer Inst.* **95**, 142 (2003).
- Lao, U.L., Mulchandani, A., Chen, W. *J. Am. Chem. Soc.* **128**, 14756 (2006).
- Leatherdale, C.A., Woo, W.K., Mikulec, F.V., Bawendi, M.G. *J. Phys. Chem. B* **106**, 7619 (2002).
- Li, J.J., Wang, Y.A., Guo, W.Z., Keay, J.C., Mishima, T.D., Johnson, M.B., Peng, X.G. *J. Am. Chem. Soc.* **125**, 12567 (2003).
- Lidke, D.S., Nagy, P., Heintzmann, R., Arndt-Jovin, D.J., Post, J.N., Grecco, H.E., Jares-Erijman, E.A., Jovin, T.M. *Nature Biotechnol.* **22**, 198 (2004).
- Lin, C.A., Sperling, R.A., Li, J.K., Yang, T.Y., Li, P.Y., Zanella, M., Chang, W.H., Parak, W.J. *Small* **4**, 334 (2008).
- Lin, S.L., Pradhan, N., Wang, Y.J., Peng, X.G. *Nano Lett.* **4**, 2261 (2004).
- Liotta, L., Petricoin, E. *Nat. Rev. Genet.* **1**, 48 (2000).
- Liu, A.Y., Roudier, M.P., True, L.D. *Am. J. Pathol.* **165**, 1543 (2004).
- Liu, W., Howarth, M., Greytak, A.B., Zheng, Y., Nocera, D.G., Ting, A.Y., Bawendi, M.G. *J. Am. Chem. Soc.* **130**, 1274 (2008).
- Matsuno, A., Itoh, J., Takekoshi, S., Nagashima, T., Osamura, R.Y. *J. Histochem. Cytochem.* **53**, 833 (2005).
- Mattossi, H., Mauro, J.M., Goldman, E.R., Anderson, G.P., Sundar, V.C., Mikulec, F.V., Bawendi, M.G. *J. Am. Chem. Soc.* **122**, 12142 (2000).

- Medintz, I.L., Sapsford, K.E., Konnert, J.H., Chatterji, A., Lin, T., Johnson, J.E., Mattoucci, H. *Langmuir* **21**, 5501 (2005a).
- Medintz, I.L., Uyeda, H.T., Goldman, E.R., Mattoucci, H. *Nature Mater.* **4**, 435 (2005b).
- Mews, A., Kadavanich, A.V., Banin, U., Alivisatos, A.P. *Phys. Rev. B* **53**, 13242 (1996).
- Michalet, X., Pinaud, F., Lacoste, T.D., Dahan, M., Bruchez, M.P., Alivisatos, A.P., Weiss, S. *Single Molecules* **2**, 261 (2001).
- Murray, C.B., Norris, D.J., Bawendi, M.G. *J. Am. Chem. Soc.* **115**, 8706 (1993).
- Nehilla, B.J., Bergkvist, M., Popat, K.C., Desai, T.A. *Int. J. Pharmaceut.* **348**, 107 (2008).
- Ness, J.M., Akhtar, R.S., Latham, C.B., Roth, K.A. *J. Histochem. Cytochem.* **51**, 981 (2003).
- Nirmal, M., Dabbousi, B.O., Bawendi, M.G., Macklin, J.J., Trautman, J.K., Harris, T.D., Brus, L.E. *Nature* **383**, 802 (1996).
- Nisman, R., Dellaire, G., Ren, Y., Li, R., Bazett-Jones, D.P. *J. Histochem. Cytochem.* **52**, 13 (2004).
- Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. *N. Engl. J. Med.* **351**, 2817 (2004).
- Parak, W.J., Gerion, D., Zanchet, D., Woerz, A.S., Pellegrino, T., Micheel, C., Williams, S.C., Seitz, M., Bruehl, R.E., Bryant, Z., et al. *Chem. Materials* **14**, 2113 (2002).
- Pathak, S., Choi, S.K., Arnheim, N., Thompson, M.E. *J. Am. Chem. Soc.* **123**, 4103 (2001).
- Pathak, S., Davidson, M.C., Silva, G.A. *Nano Lett.* **7**, 1839 (2007).
- Pedersen, L., Holck, S., Schiodt, T., Zedeler, K., Mouridsen, H.T. *Breast Cancer Research and Treatment* **14**, 91 (1989).
- Pellegrino, T., Manna, L., Kudera, S., Liedl, T., Koktysh, D., Rogach, A.L., Keller, S., Radler, J., Natile, G., Parak, W.J. *Nano Lett.* **4**, 703 (2004).
- Peng, Z.A., Peng, X.G. *J. Am. Chem. Soc.* **123**, 183 (2001).
- Peng, Z.A., Peng, X.G. *J. Am. Chem. Soc.* **124**, 3343 (2002).
- Pietryga, J.M., Schaller, R.D., Werder, D., Stewart, M.H., Klimov, V.I., Hollingsworth, J.A. *J. Am. Chem. Soc.* **126**, 11752 (2004).
- Pinaud, F., King, D., Moore, H.P., Weiss, S. *J. Am. Chem. Soc.* **126**, 6115 (2004).
- Pinaud, F., Michalet, X., Bentolila, L.A., Tsay, J.M., Doose, S., Li, J.J., Iyer, G., Weiss, S. *Biomaterials* **27**, 1679 (2006).
- Pons, T., Uyeda, H.T., Medintz, I.L., Mattoucci, H. *J. Phys. Chem.* **110**, 20308 (2006).
- Qi, L., Gao, X. *Expert Opin. Drug Delivery* **5**, 263 (2008).
- Qu, L.H., Peng, X.G. *J. Am. Chem. Soc.* **124**, 2049 (2002).
- Smith, A.M., Dave, S., Nie, S., True, L., Gao, X. *Expert Rev. Mol. Diagn.* **6**, 231 (2006).
- Smith, A.M., Gao, X., Nie, S. *Photochem. Photobiol.* **80**, 377 (2004).
- Tholouli, E., Hoyland, J.A., Di Vizio, D., O'Connell, F., Macdermott, S.A., Twomey, D., Levenson, R., Yin, J.A., Golub, T.R., Loda, M., et al. *Biochem. Biophys. Res. commun.* **348**, 628 (2006).
- Thomson, T.A., Hayes, M.M., Spinelli, J.J., Hilland, E., Sawrenko, C., Phillips, D., Dupuis, B., Parker, R.L. *Mod. Pathol.* **14**, 1079 (2001).
- Tokumasu, F., Dvorak, J. *J. Microsc.* **211**, 256 (2003).

- True, L.D. *Am. J. Clin. Pathol.* **90**, 324 (1988).
- True, L.D., Gao, X. *J. Mol. Diagn.* **9**, 7 (2007).
- Umemura, S., Osamura, R.Y. *Breast Cancer* **11**, 334 (2004).
- Vogel, C.L., Cobleigh, M.A., Tripathy, D., Gutheil, J.C., Harris, L.N., Fehrenbacher, L., Slamon, D.J., Murphy, M., Novotny, W.F., Burchmore, M., *et al.* *J. Clin. Oncol.* **20**, 719 (2002).
- Whaley, S.R., English, D.S., Hu, E.L., Barbara, P.F., Belcher, A.M. *Nature* **405**, 665 (2000).
- Wu, X., Liu, H., Liu, J., Haley, K.N., Treadway, J.A., Larson, J.P., Ge, N., Peale, F., Bruchez, M.P. *Nature Biotechnol.* **21**, 41 (2003).
- Xiao, Y., Barker, P.E. *Nucleic Acids Res.* **32**, e28 (2004).
- Xing, Y., Chaudry, Q., Shen, C., Kong, K.Y., Zhai, H.E., Chung, L.W., Petros, J.A., O'Regan, R.M., Yezhelyev, M.V., Simons, J.W., *et al.* *Nature Protoc.* **2**, 1152 (2007).
- Yezhelyev, M., Morris, C., Gao, X., Lewis, M., Cohen, C., Nie, S., O'Regan, R.M. *J. Clin. Oncol.* **23**, 843s (2005).
- Yezhelyev, M.V., Al-Hajj, A., Morris, C., Marcus, A.I., Liu, T., Lewis, M., Cohen, C., Razhevskiy, P., Simons, J.W., Rogatko, A., *et al.* *Adv. Mater.* **19**, 3146 (2007).
- Yezhelyev, M.V., Gao, X., Xing, Y., Al-Hajj, A., Nie, S., O'Regan, R.M. *Lancet Oncol.* **7**, 657 (2006).
- Zhong, X.H., Feng, Y.Y., Knoll, W., Han, M.Y. *J. Am. Chem. Soc.* **125**, 13559 (2003a).
- Zhong, X.H., Han, M.Y., Dong, Z.L., White, T.J., Knoll, W. *J. Am. Chem. Soc.* **125**, 8589 (2003b).

# Modulation design of plasmonics for diagnostic and drug screening

C.-W. Lin, N.-F. Chiu, and C.-C. Chang

# 18

18.1 Introduction	641
18.2 Theoretical insights	642
18.3 Substrate effect (prism coupler, Ge-doped Si waveguide, grating, plasmonic)	649
18.4 Metallic effect (LRSPR, CMO adhesive layer)	657
18.5 Microfluidic parts	659
18.6 Biomolecular layer effect	661
18.7 Conclusions	669
Acknowledgments	670
References	670

## 18.1 Introduction

The behavior of free electrons or plasma near a metallic dielectric interface is a special type of light–matter interaction. A high-density electron gas, when subjected to collective longitudinal excitation or oscillation, will manifest particle-like behavior. Analogous to phonons, magnons and excitons, the phenomenon has been called a “plasmon” (Raether 1988). It occurs in the forms of bulk, surface, and particle plasmons, and research on the phenomenon has led to many important applications in optoelectronics and biomedicine. Recently, with increasing capabilities in micro- and nanofabrication, the original planar configuration has been extended into three dimensions and the nanoscale to explore the effects of selected material and structure design. Accordingly, the term “plasmonics” has been used to describe the resultant devices for applications in optics, communications, storages, and sensors (Andrew and Barnes 2004; Piliarik *et al.* 2005; Bravo-Abad *et al.* 2006). The unique characteristics of non-labelling and real-time monitoring have opened up a wide spectrum of novel developments in biochemical and biophysical applications, especially those involving the dynamic interactions of biomolecules at the solid/liquid or solid/gas interface. However, there is still room to further enhance its performance to make it an analytical useful modality (Homola 2003). This calls for optimization of multiple governing entities, e.g. a prism for coupling the photon to plasmon, metals supporting the resonance, biomolecules interacting on the sensor surface, light sources tuning for proper modes of detection, and microfluidic parts for delivering the precise quantity of samples for possible total integration of analytical processes and high throughput. In this chapter, we report the advances made in the following aspects of the present area:

- theoretical insights (origins, admittance loci);
- substrate effect (prism, Ge-doped silica waveguide, grating, plasmonic);

- metallic effect (long-range surface plasmon resonance (LSPR), conducting metal oxides (CMO) adhesive layer);
- light source (laser, light-emitting diode (LED), white light);
- microfluidic parts (switch, lateral flow);
- biomolecular layer effect.

## 18.2 Theoretical insights

### 18.2.1 Origins of surface plasmon wave and manipulation

Plasmon resonance from light–matter interactions and the behavior of free electrons in a metal can be described by the harmonic oscillator model with external applied force or EM field. A surface plasmon (SP) wave can interact with a photon if the momentum or energy is matched under pertinent conditions. However, incident light cannot excite an SP wave directly from the dielectric medium. For a given frequency  $\omega$ , light travelling in a dielectric medium has a wave vector

$$k = \left( \frac{\omega}{c} \right) \sqrt{\varepsilon_1}. \quad (18.1)$$

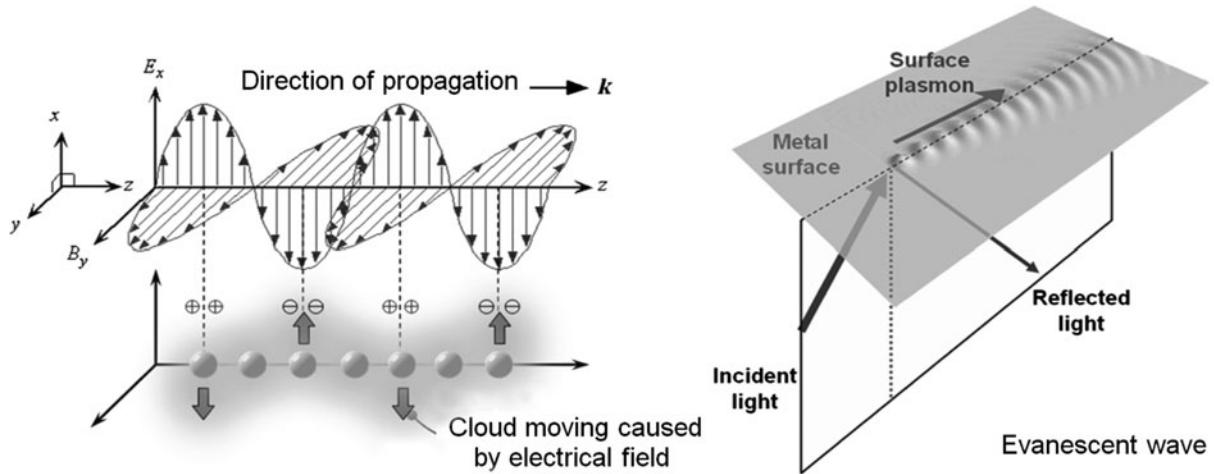
The parallel component of this wave vector in the plane of the incident surface is

$$k_{\parallel} = \left( \frac{\omega}{c} \right) \sqrt{\varepsilon_1} \sin \theta. \quad (18.2)$$

Because the propagation constant of an SP wave is always higher than the wave propagation in the dielectric medium, an SP wave cannot be excited directly. The requisite momentum change can be achieved by prism coupling, diffraction gratings, optical waveguides, or active plasmonics. Among these configurations, it was found that prism coupling produces the highest resolution. Using a coupling prism, the inplane component of the propagation constant can have a cross-point with the SP dispersion relation curve. The  $k_{\parallel}$  varies with the incident angle ( $\theta$ ). If  $\omega_L$  is the frequency of incident light then its momentum will match the SP propagation constant at the metal/air interface:

$$k_{sp} = k_{\parallel} \Rightarrow \sqrt{\varepsilon_p} \sin \theta = \sqrt{\frac{\varepsilon_a \cdot \varepsilon_m}{\varepsilon_a + \varepsilon_m}}. \quad (18.3)$$

Since a metal has a negative dielectric constant ( $\varepsilon_2$ ) and a dielectric material has a positive dielectric constant ( $\varepsilon_1$ ), the discontinuous electric field will have two opposite propagating directions at the interface. This results in surface charge-density oscillations, and the macroscopic oscillation of the electric field in the two media with components along the  $x$ - and  $z$ -directions. The phenomenon occurs at a frequency different from the bulk oscillations, is confined to the surface, and vanishes towards both sides of the metal surface. Since the surface density alternates in sign, the spatial summation in the  $z$ -direction gives an exponential decay in the magnitude of the electric field. The surface plasmons wave vector  $k_x$  is a complex function;



**Fig. 18.1** Illustrations of the origin of free-electron oscillation in an applied EM field and the excitation of an SP wave on the metallic surface with nm scale on the  $z$ -axis and  $\mu\text{m}$  scale on the  $x$ -axis.

$k_x = k_x' + ik_x''$ , hence the surface plasmons propagate in the  $x$ -direction and decay in amplitude as  $\exp(-2k_x''x)$ . Thus, one can define a propagation length  $L_x$ ,

$$L_x = (2k_x'')^{-1}. \quad (18.4)$$

Since for surface plasmon  $k_x > k_0$ , the surface plasmon wave amplitude decays in the  $z$ -direction as  $\exp(-|k_{z1,2}||z|)$ . One can define the skin depth of the surface plasmon wave as,

$$Z = \frac{1}{|k_{z1,2}|}. \quad (18.5)$$

As shown in Fig. 18.1, the exponential decay of the electrical field on the metal (Au)/air interface has both  $z$ - and  $x$ -directions with different space constants on the nm and  $\mu\text{m}$  scale. It can be shown that this surface plasmon resonant (SPR) wave on the surface is subjected to the effects of wavelength, incident angle, metal and dielectric medium. It thus opens up a wide spectrum of possible engineering designs of SPR devices for targeted applications with specific sensing requirement in refractive index, molecular sizes, sensitivity, dynamic range, and instrumentation (Andrew and Barnes 2004; Lin *et al.* 2005).

### 18.2.2 Admittance loci design method

The admittance diagram is a class of phase diagrams or loci that have been used in the design of multilayer optical coatings (Macleod 1992). Its application has been used to visualize the reflectivity behavior of optical thin films with the optical constants and thickness of selected materials and the optimal design of layer structure. Combined with the optical parameter database, particularly for biomolecules, the admittance diagram can be a unique and powerful design

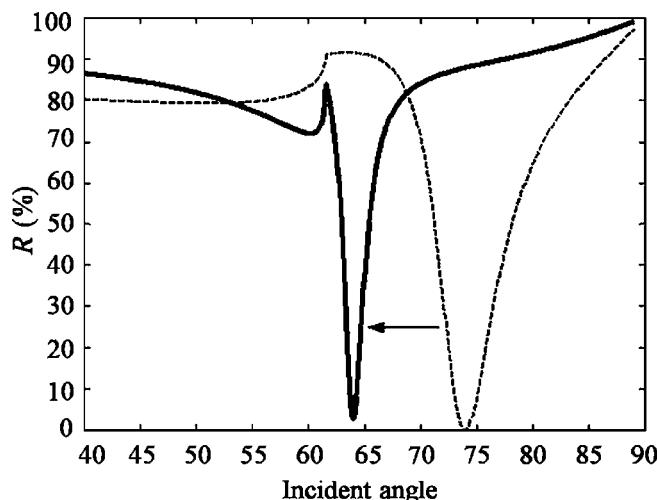
tool for multilayer optical biosensor/biochips (Lin *et al.* 2006b), especially for the design of phase-matching or bistable plasmonics in the future. Admittance loci can be used to visualize the optical behavior of thin film, as we move a virtual reference plane (equivalent admittance  $Y_e$ ) from the substrate admittance ( $Y_s$ ) to the front surface of the multilayer, with the incident light emitting from the surrounding medium onto a multilayer system with substrate  $Y_s$ . The equivalent admittance beyond the reference plane will change due to the thin-film thickness change. The plot of  $Y_e$  on the complex plane gives a trace of these changes, which is called the admittance diagram or admittance loci. In the loci, we can know how close  $Y_e$  is to  $Y_0$  (incident medium) and the reflectance. For example, we have used a coupling prism (dense medium)  $n = 1.5$ , gold film ( $d = 50$  nm), less dense medium (water)  $n = 1.33$ , incident light wavelength (633 nm), and found the critical angle to be around 62 degrees and the reflectance dip at 74.5 degrees (the SPR angle). Figure 18.2 shows the goal of our proposed design, where we try to reduce the SPR angle from 74 degrees to 64 degrees for easier optical alignment and better optical performance. In the following sections, we will discuss four different design cases that can be used to elucidate the approach.

#### 18.2.2.1 Single layer of dielectric thin film

Under fixed incident wavelength and incident angle, a single-layer structure with admittance,  $y$ , on substrate,  $y_s$ , in the ambient medium,  $y_0$ , can be represented as  $y_0|y|y_s$ . The equivalent admittance,  $Y_e$ , of different layer thicknesses is a complex number, which can be expressed as  $\alpha + i\beta$  for its real part and imaginary part or as H/E for optical admittance. It can be derived from the Maxwell equations with oblique incidence as:

$$Y_e = \alpha + i\beta = H/E = \frac{y_s \cos \delta + iy \sin \beta}{\cos \delta + iy_s/y \sin \delta}, \quad (18.6)$$

where  $\delta$  is the phase thickness.



**Fig. 18.2** The goal is to design a device with lower SPR angle for practical applications in solution or biological fluids. Simulated SPR curves show normalized reflectance from a prism ( $n = 1.5$ ), Au, and water ( $n = 1.33$ ) at 633 nm 74 degrees (dashed line) and designated 64 degrees (solid line) for better optical alignment, higher sensitivity, and larger dynamic range. Reprinted with permission from Lin *et al.*, *Sens. Actuators B* **117**, 219 (2006). Copyright 2006, Elsevier B.V.

For non-trivial solution of  $\cos \delta$  and  $\sin \delta$ , one can have

$$\left( \alpha - \frac{1}{2} \left( y_s + \frac{y^2}{y_s} \right) \right)^2 + \beta^2 = \frac{1}{4} \left( y_s - \frac{y^2}{y_s} \right)^2. \quad (18.7)$$

It is a circle starting from  $(y_s, 0)$  with a clockwise spiral due to the changes of film thickness and crosses the real axis again at  $(y^2/y_s, 0)$  for every quarter-wavelength thickness.

#### 18.2.2.2 Single layer of metallic film

For the special case of SPR, we will examine the admittance loci of an ideal metal with absorbance ( $y = -ik$ ) or high-performance metal (high  $k/n$  value for  $y = n - ik$ ), as in a Kretschmann-type SPR device where the prism is coated with a single high-performance metal layer. We are treating the glass prism as the incident medium and the surrounding air as the substrate. Thus, the starting admittance for the film is on the imaginary axis, negative part for s-polarization and the positive part for p-polarization. It will show a near-zero reflectance at a particular angle of incidence and metal thickness. The condition is very sensitive to the angle of incidence. Under this situation, the phase thickness ( $\delta$ ) will become a pure imaginary number ( $i\gamma$ ), i.e.  $\delta = \frac{2\pi}{\lambda}(-ik)d = -i\gamma$ , where  $k$  is the extinction coefficient of the metal.

For a non-trivial solution of  $\gamma$ , we can get the locus of  $\alpha$  and  $\beta$ ,

$$\alpha^2 - 2 \frac{y_s^2 - k^2}{2y_s} \alpha + \beta^2 = k^2. \quad (18.8)$$

It is also a circular locus with a center on  $\left( \frac{y_s^2 - k^2}{2y_s}, 0 \right)$  and a radius of  $\frac{y_s^2 + k^2}{2y_s}$ . It goes from  $(0, k)$  to  $(0, -k)$  in the ideal metal and from  $(-n, k)$  to  $(n, -k)$  in the high-performance metal.

#### 18.2.2.3 Multilayer with symmetrical structure

A symmetrical multilayer SPR device has been proposed and realized by using a dielectric mirror structure with repetitive high/low refractive-index materials to modulate the SPR resonance conditions (Lin *et al.* 2006b). In this case, we proposed to shift the SPR angle from 74 to 64 degrees with a 633 nm light source on a BK7 coupling prism under water, which is equivalent to the use of a SF2 prism. With a single layer of 50 nm gold, its resonant angle is close to 74 degrees, where  $Y_e = 1.4593 - i0.0225$  is close to  $Y_o$  and its reflectance is near zero. However, at the designated 64 degrees, the starting point  $Y_e = 0.0686 - i1.5465$ , is close to the imaginary axis and its reflectance is higher than 80%. By using the proposed symmetrical structure of glass/Au(40 nm)/[TiO<sub>2</sub>(20 nm)/SiO<sub>2</sub>(20 nm)]<sub>4</sub>/Au(30 nm), its admittance loci will leave the imaginary axis and get close to  $Y_o$ . The final value is  $1.3664 + i0.4685$  and within the 5% isoreflectance region (2.83%). Experimental results confirmed the expected 10 degrees shift in resonant peak position towards the critical angle and 3.25 times decrease in FWHM (full width half-maximum). This device also resulted in a wider dynamic range for up to 1.50 refractive index unit (RIU), compared to 1.38 RIU in a conventional single-layer Au film.

#### 18.2.2.4 Multilayer with asymmetrical structure

Silver has been used extensively in the SPR devices. However, it does tarnish when exposed to the sensing environment due to surface contaminations of silver sulfide or silver chloride. It thus often requires an overcoating of dielectric layer to protect it from corrosion or a high-performance multilayer coating to enhance its optical performance. It is thus reasonable to replace the inner metal layer of Au with Ag to have further enhancement of SPR performance with an asymmetrical structure. The admittance locus of the dielectric layer between the exit medium (e.g. air) and the silver layer starts on the imaginary axis and it must remain on the axis if there is no absorption in the layer. Provided the incident angle in the dielectric layer remains below the critical, the locus simply climbs to the top of the imaginary axis and disappears to re-enter the diagram at the very foot of the imaginary axis to climb upward and repeat the pattern, if the layer is thick enough. In this case, we also used a 633 nm wavelength light source on a BK7 coupling prism under water. The only difference in the structure design is the replacement of the bottom Au layer with Ag layer to study its enhancement effect on SPR. With a single layer of 55 nm silver, its resonant angle is close to 68 degrees where  $Y_e = 1.4595 - i0.0226$  is close to  $Y_0$  and its reflectance is near zero. However, at 61.52 degrees, the starting point (0.00000, 71.43949) is close to the imaginary axis and its reflectance is around 100%. By using the asymmetrical structure, glass-Ag(50 nm)/[TiO<sub>2</sub>(20 nm)/SiO<sub>2</sub>(20 nm)]<sub>4</sub>/Au(20 nm), its admittance loci will leave the imaginary axis and get close to  $Y_0$ . The final value is (1.33905, -0.57627) and within the 5% isoreflectance region (4.28%). The resultant SPR spectrum showed a resonant angle at 61.52 degrees and narrower HMBW of 0.25 degrees. Table 18.1 summarizes the performance of three fabricated SPR devices according to the above-mentioned designs. It shows that the design of a multilayer structure does have significant improvements in the water, which is essential for biological applications.

#### 18.2.3 Surface plasmon grating coupled emission (SPGCE)

The radiative decay, in the form of light emission near the metal/organic interface following excitation, can enhance the luminescence efficiency. The pathways for the excited states to undergo various types of radiative decay

**Table 18.1** Performance comparisons of different SPR configurations. Reprinted with permission from Lin *et al.*, *Sens. Actuators B* **117**, 219 (2006). Copyright 2006, Elsevier B.V.

	SPR angle	HMBW	Dynamic range (RIU) <sup>a</sup>	Resolution (RIU) <sup>b</sup>	Intensity slope <sup>c</sup>
Single layer Au	73.9°	6.5°	1.331–1.38	$8.33 \times 10^{-5}$	58.65
Symmetric multilayer	62.54°	2°	1.331–1.50	$1.25 \times 10^{-5}$	127.5
Asymmetric multilayer	61.52°	0.25°	1.33–1.48	$8.13 \times 10^{-6}$	215.19

Note: <sup>a</sup> RIU is the refractive-index unit; <sup>b</sup> Angular measurement of  $10^{-3}$  angular resolution; <sup>c</sup> Intensity change (0–255)/1 degree shift.

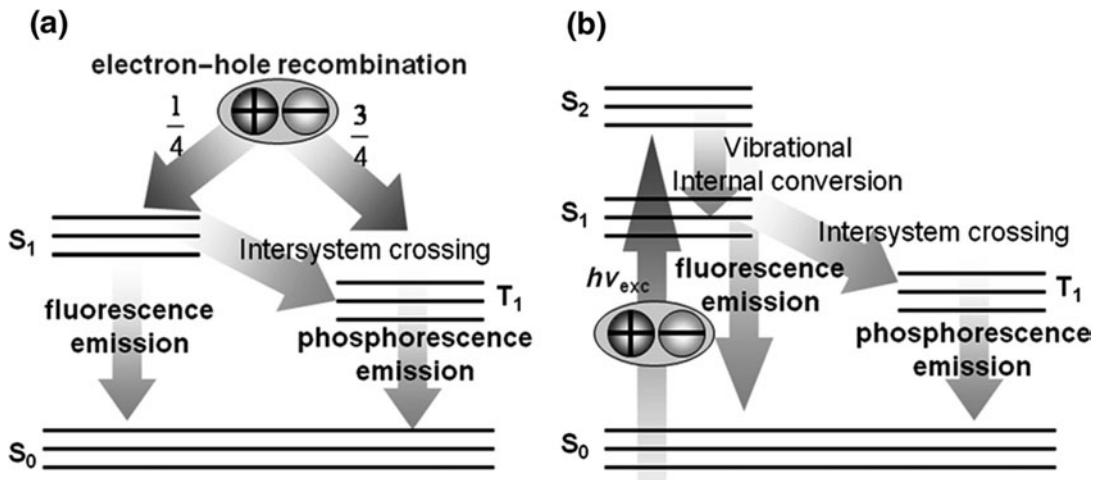
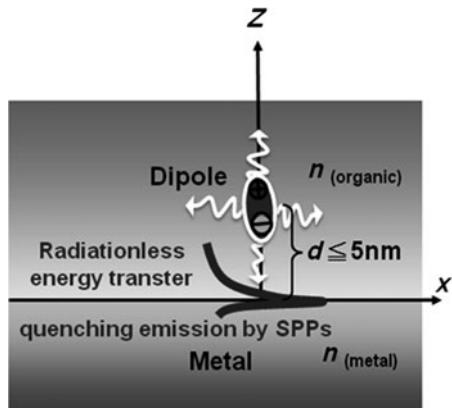


Fig. 18.3 Schematic diagrams show the pathways of radiative decays for (a) electroluminescence (EL), and (b) photoluminescence (PL).

are depicted by the Jablonski diagram shown in Figs. 18.3(a) and (b), for electroluminescence (EL) and photoluminescence (PL), respectively. For electroluminescence, light of well-defined color can be produced upon applying a potential across the electrodes with organic layer(s) sandwiched in-between (Kalinowski 2005). This happens because molecular excitons that decay radiatively are formed once electrons and holes are injected into the organic layer. In the case of photoluminescence, a fluorophore is usually excited to the higher energy level of  $S_2$ . With a few rare exceptions, molecules in condensed phases rapidly relax to the lower energy level of  $S_1$ . This process is called internal conversion (non-radiative) and generally occurs in  $10^{-12}$  s or less (Lakowicz 1999). These excited molecules then return to the highest level in the ground state and quickly reach thermal equilibrium. An interesting consequence of the emission from the highest level in the ground state is that the emission spectrum is typically a mirror image of the absorption spectrum for the  $S_0 \rightarrow S_1$  transition. Molecules in the  $S_1$  state can also undergo a spin conversion to the first triplet state,  $T_1$ . Emission from  $T_1$  is termed phosphorescence and is generally shifted to lower energy relative to the fluorescence. Conversion of  $S_1$  to  $T_1$  is called intersystem crossing. In general, fluorescence results from the lowest singlet excited state and phosphorescence resulted from the lowest triplet excited state. The ground state is always a singlet.

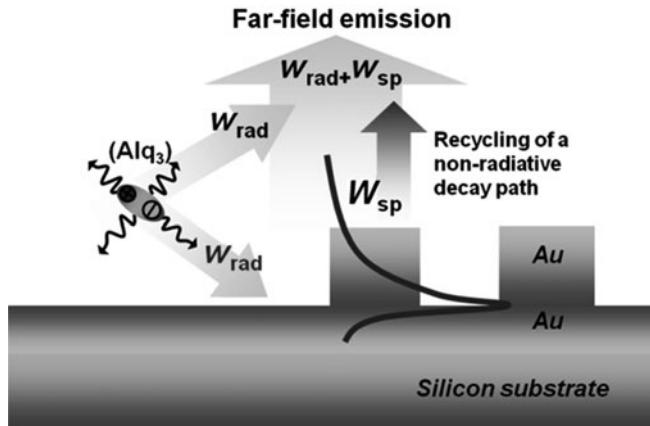
The energy-transfer process is similar to the near-field coupling of the dipole to SP modes in the organic/metal interface. Such a dipole can form an exciton and contribute to its optical properties that are subjected to a quenching process near the metal/organic interface. This irreversible energy transfer from the excited organic molecules to the metal (Förster transfer) is due to the dipole-dipole interaction (Chance *et al.* 1975; Choong *et al.* 1998; Burin and Ratner 2000). In addition, via the diffusion of excitons to the metal/organic interface it is possible to have higher non-radiation energy transfer as compared to localized excitons without such an interaction. The interface dipole field is coupled to the surface modes causing excitons to oscillate and thus absorb



**Fig. 18.4** Schematic representation showing metal/organic interface energy transfer that is equivalent to the near-field coupling of the interface dipole to surface plasmons in the plane metal. The energy transfer across the interface could cause quenching of the interface dipole at the metal/organic interface.

energy. Normally, surface plasmons are not coupled to the radiation field. It does not violate the rule in this case because the short-range interface dipole field is not regarded as a radiation field without a proper matching condition in the energy-transfer processes (Chance *et al.* 1975). The excitation of SPPs is one of the possible radiationless routes by which the energy transfer of the organic semiconductor molecules can be absorbed by the metal. Quenching by the surface plasmon absorption of the metal due to Förster energy transfer occurs as shown in Fig. 18.4, where the dipole layer interacts between the metal and organic thin film.

Furthermore, an appropriate organic/metal interface can be used to manipulate the radiative decay properties. This approach is sometimes referred to as radiative decay engineering (RDE) (Lakowicz *et al.* 2001, 2002, 2003). Recently, enhanced transmission and emission of organic semiconductor on metallic thin film mediated by energy transfer of coupled SPPs has been used in the design of various optoelectronic devices. Initially, coupled SPPs transmission through a thin metal film was investigated on a surface having a few hundred nanometers of roughness (Fang *et al.* 2003) or corrugated microstructure (Gruhlke *et al.* 1986; Gifford *et al.* 2002; Wedge and Barnes 2004; Feng *et al.* 2007). Others have made SPPs coupled emissions from excited fluorescent molecules by evanescent field near the metallic surface (Pockrand and Brillante 1980; Kitson *et al.* 1996; Hung *et al.* 2006). The coupled emission around the SPR angle has been observed by using an attenuated total reflection (ATR) prism in the Kretschmann and reverse Kretschmann configurations (Nakano *et al.* 2001; Enderlein and Ruckstuhl 2005; Shinbo *et al.* 2005; Winter and Barnes 2006), grating coupler (Kitson *et al.* 1996; Kalkman *et al.* 2003; Hung *et al.* 2006) and nanoparticles (Maier *et al.* 2003). One can thus achieve enhanced quantum yield or increased radiative rate of fluorescent emission. The existence of non-radiative decay near a metallic surface can have effects on fluorescence intensity and lifetimes. The further interactions of coupled resonance energy by the so-called surface plasmon coupled emission (SPCE) can have interesting optical properties for many applications (Pockrand and Brillante 1980; Lakowicz *et al.* 2002).



**Fig. 18.5** SPs can radiate into the far-field by using a grating structure to resolve the momentum mismatch and result in recycling of a non-radiative decay path with  $W = W_{\text{rad}} + W_{\text{sp}}$ , which can turn the “non-radiative” plasmon into a far-field photon. In this case, we use on organic light emitter,  $\text{Alq}_3$ , as a localized light source that can interact with a metallic structure and change its lifetime from a slow emitter to a fast emitter.

The emission of an organic semiconductor layer with a thin metal layer greater than 5 nm in thickness is almost completely quenched (Chance *et al.* 1975; Lakowicz *et al.* 2003). Additionally, the dipole of organic semiconductor can induce a field in the metal, which consequently increases the intrinsic radiative decay rates and quantum yields by the nearby metal surfaces; this results in decreasing the lifetime of the organic semiconductor. In summary, the quantum yields, intensity, resonance energy transfer and lifetimes of the emission light are governed by the magnitudes of the radiative decay rate and the sum of the non-radiative decay rates. The effects of transferred energy on a plane occurred over distances are comparable to the excitation and emission wavelengths. For a planar metallic film system, all the energy non-radiatively transferred to the metal is ultimately lost as heat and the emission from the device is strongly quenched. For a metallic grating system, the energy transferred to SPPs can be recovered as photons, as shown in Fig. 18.5. A grating is then incorporated into this basic sample structure. The combined effects of the multilayered substrate and the grating-induced directional emission further lead to an effective enhancement in a narrow angular range. Enhanced energy transfer can be combined with directional emission to result in higher sensitivity. It is known that fluorescence can be excited by the evanescent wave due to surface plasmons (Lakowicz *et al.* 2003). In this case, the metal-induced increase in the transfer rate will result in transfer over longer distances and the light emission will become detectable with an increase in the transfer efficiency.

### 18.3 Substrate effect (prism coupler, Ge-doped Si waveguide, grating, plasmonic)

It is well known that SPR can be excited by three different coupling mechanisms, i.e. prism coupler, waveguide, and grating coupling. It is now better understood that we can use engineering design to fine tune coupling conditions for different applications. We will now discuss these excitation methods in detail.

### 18.3.1 Prism coupler

In 1968, Otto introduced the prism-coupled method to excite surface plasmons by light radiation (Otto 1968). By total internal reflection, an evanescent wave is set up at the prism air interface. The inplane propagation constant of the evanescent wave is

$$k_{\parallel} = (\omega_c) * \sqrt{\varepsilon_p} * \sin \theta. \quad (18.9)$$

If a metal surface is now brought into the evanescent field with an air gap between the prism and metal, the evanescent wave can excite the surface plasmon at the metal/air interface. With a coupling prism, the inplane component of the propagation constant ( $k_{\parallel}$ ) can have a cross-point with the SP dispersion relation curve. The  $k_{\parallel}$  varies with the incident angle ( $\theta$ ). If the incident light,  $\omega_L$ , can match the SP propagation constant at the metal/air interface by changing its incident angle,  $\theta$ , or dielectric constant of the metal layer,  $\varepsilon_m$ , as follows:

$$k_{sp} = k_{\parallel} \Rightarrow \sqrt{\varepsilon_p} * \sin \theta = \sqrt{\frac{\varepsilon_a \cdot \varepsilon_m}{\varepsilon_a + \varepsilon_m}}. \quad (18.10)$$

When this occurs, it will cause a reflectance dip at a resonant angle  $\theta$  of the p-polarized light. We call this the surface plasmon resonance angle (SPR angle). Such a behavior can be described by using Fresnel's equations. Practically, it is difficult to make the air gap with hundreds of nanometers between the prism and metal surface precisely. In 1971, Kretschmann published another prism coupling configuration, with a thin layer of metal directly deposited on the prism, to excite surface plasmons (Kretschmann and Raether 1971). At the same incident angle as in Otto's configuration, there is also a reflectance dip. The differences between these two devices include the thickness of the metallic film (in Kretschmann-type  $\sim 50$  nm and in Otto-type  $\sim$ hundreds of nanometers) and the practical realization of the required thickness in nanometers with the precise air gap between the prism and the metal surface. So nowadays, it is the most popular configuration widely adapted by commercial systems to excite surface plasmons. In either case, the refractive index of the prism is a critical factor for the coupling of SPR as an optical dense medium to work in air, water or glycerol for different biological preparations. For example, we can use SF10 ( $n = 2.1$ ) to replace BK7 ( $n = 1.5$ ) to measure glycerol-preserved proteins with smaller SPR angle at 633 nm within the operable range of a mechanical scanning stage. Alternatively, one can use the near-infrared region of wavelength to excite SPR with a BK7 prism and result in a smaller SPR angle for the same mechanical constraints. Recently, this has been extended to use on a non-transparent Si wafer as a coupling prism (Patskovsky *et al.* 2004). This resulted in a 2–5 times sensitivity enhancement and better surface preparation due to less surface roughness of silicon than glass.

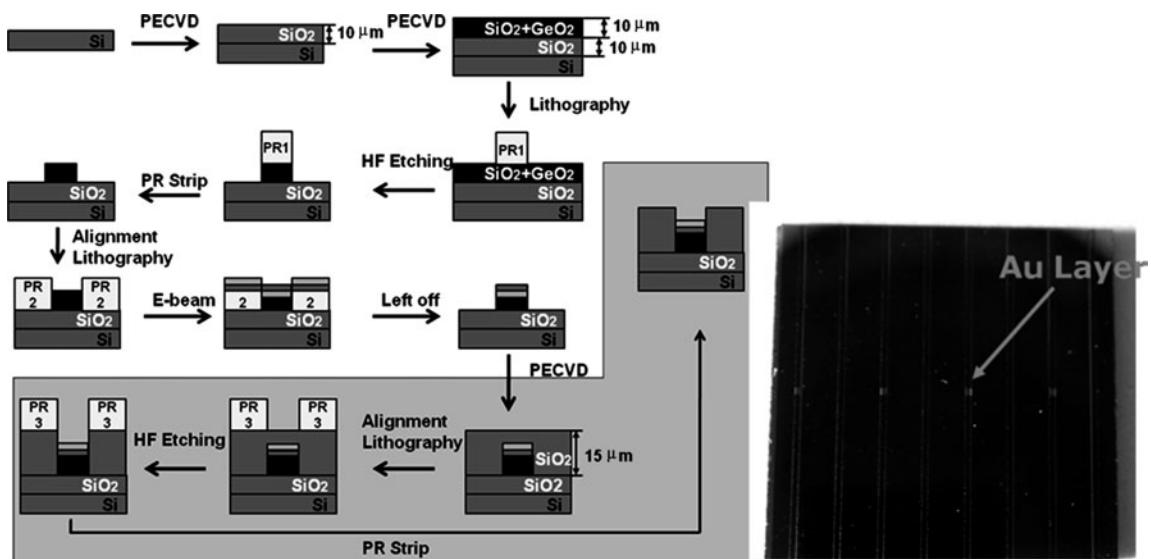
### 18.3.2 Ge-doped silica waveguide

For waveguide coupling (Chu *et al.* 2006; Huang *et al.* 2006; Lin *et al.* 2006a), it is important to control the difference between the refractive index of the

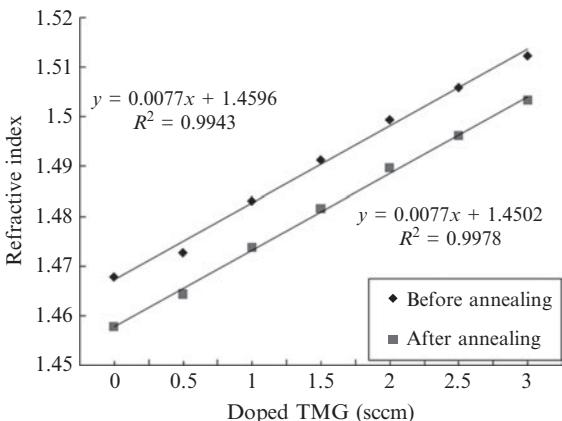
core layer and the cladding layer to induce SPR in the visible light region. One can use a simulation package, for example BeamPROP (RSOFT Design Group, Inc.), to visualize the optimal electromagnetic field profile within the waveguide structure for the effects on SPR response due to the optical properties and the defined boundary of core and cladding materials. Silica-on-silicon has been used extensively as planar substrate to fabricate an integrated waveguide due to its low light loss in the visible-near-infrared (NIR) region, which is suitable for the excitation of SPR. Plasma-enhanced chemical vapor deposition (PECVD) can be used for the fabrication of high-quality silicon dioxide ( $\text{SiO}_2$ ) as the cladding layer at lower temperature ( $400^\circ\text{C}$ ) than the traditional thermal-oxidation method according to the eqn (18.11). Our optimal working parameters for tetra ethyl ortho silicate (TEOS) and  $\text{O}_2$  flow rate are 7 SCCM and 500 SCCM under 90 Pa vacuum pressure with 300 W for RF power. A precision Ge-doped silica waveguide process can be used to change the core layer refractive index by adding tri-methyl-germanium (TMG) simultaneously according to eqn (18.12):



Figure 18.6 shows the fabrication processes of the designed waveguide sensors and final device according to the above-mentioned design principles. First, a single-side polished silicon wafer was washed with piranha solution ( $\text{H}_2\text{SO}_4 : \text{H}_2\text{O}_2 = 3 : 1$ ) at  $80^\circ\text{C}$  for 5 min and then with diluted buffered oxide etchant (BOE:  $\text{H}_2\text{O}$  1:3) for 30 s. Subsequently,  $\text{SiO}_2$  ( $n = 1.469$ ) and  $\text{SiO}_2\text{-Ge}$  ( $n = 1.491$ ) were deposited on the wafer by PECVD with a thickness



**Fig. 18.6** Fabrication processes of waveguide SPR sensor and photograph of waveguide sensor. The gold spot on the waveguide is the SPR sensing area. Reprinted with permission from Huang *et al.*, *Biosen. Bioelect.* **22**, 519 (2006). Copyright 2006, Elsevier B.V.

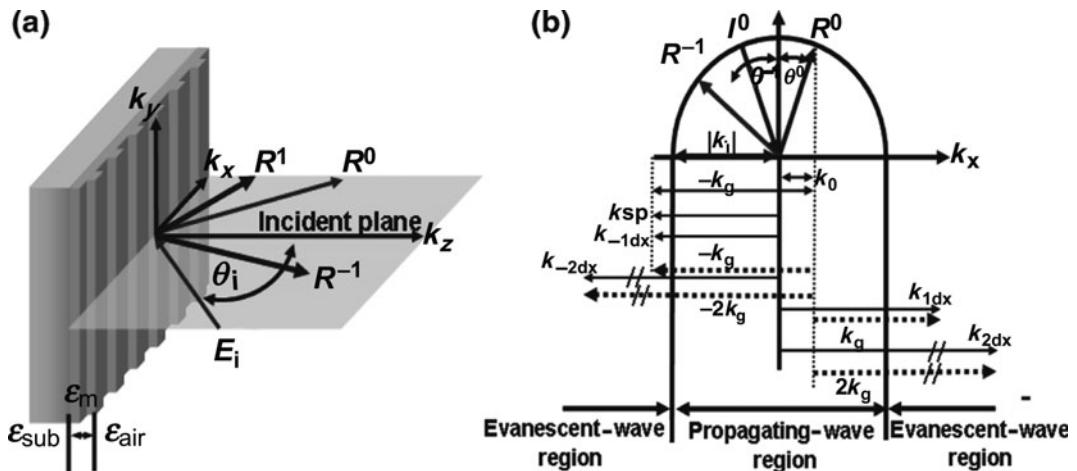


**Fig. 18.7** The controllability of Ge-doped silicon refractive indices by PECVD shows a linear range from 1.46–1.51 with doped TMG in different flow rates before and after annealing processes.

of  $10\text{ }\mu\text{m}$  each. Figure 18.7 shows the experimental verification of the final  $\text{SiO}_2$ -Ge core refractive index with different TMG flow rate. To define the waveguide pattern on the Ge-doped  $\text{SiO}_2$ , the waveguide pattern was first defined onto the substrate by lithography, and then the substrate was etched by BOE. Finally, to fabricate the SPR detection area on the waveguide, another lithography process was applied to the substrate and was subsequently evaporated with 1 nm Cr and 50 nm Au, subsequently. After the lift-off process and chip dicing the fabrication of a waveguide chip was accomplished. The dimensions of the waveguide were  $\sim 1\text{ cm}$  length,  $200\text{ }\mu\text{m}$  width. For the purpose of improving the light-coupling efficiency with the waveguide, a curved structure at the end of the waveguide was designed and fabricated to facilitate the focusing of incident light onto the waveguide.

### 18.3.3 Nanograting

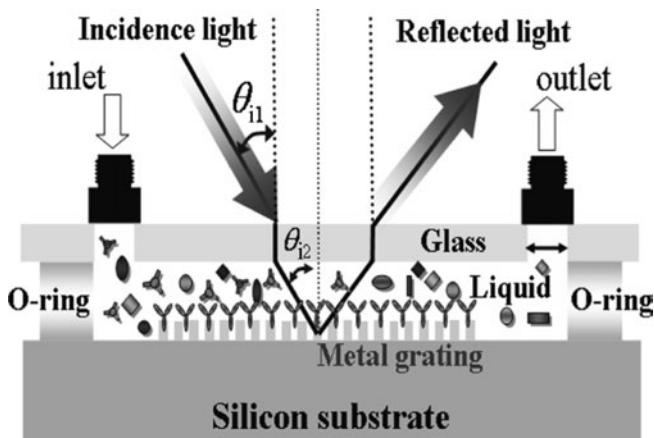
For a grating coupler, it is well known that one can control the angular and intensity radiation fields by the pitches of the grating and the incident angle. It will have enhanced the localized EM field due to surface plasmon wave (SPW) interactions when the metallic grating scale decreases to a few hundred nanometers. Recent efforts in enhanced transmission of metallic nanohole array have triggered a lot of activities in this surface-plasmon-related phenomena. This field corresponds to non-radiative diffraction or to an evanescent field near the grating with a wavelength equal to the period of the grating. When the grating period is shorter than the wavelength of light, the grating vector  $k_g$  is longer than the wave vector of the light and is able to meet with the SPR condition. Figure 18.8 shows the experimental configuration for a planar light wave incident on a periodically modulated surface with angle of incidence and various diffractive orders. Its corresponding vector diagram of the various wave vectors associated with the incident beam ( $k_i$ ), the grating ( $k_g$ ), the diffracted light ( $k_{\pm\text{ndx}}$ ), and the surface plasmon ( $k_{\text{SP}}$ ). In this situation, the surface plasmon is excited by 1st-order diffracted light. It demonstrated that



**Fig. 18.8** Non-radiative diffraction or evanescent diffraction is generated by an ultrafine grating with a period shorter than the wavelength of light. (a) Experimental configuration for a planar light wave incident on a periodically modulated surface with angle of incidence. (b) A vector diagram of the various wave vectors associated with the incident beam, the grating, the diffracted light, and the surface plasmon. In this situation, the surface plasmon is excited by 1st-order diffracted light.

this 1D grating pattern creates a photon forbidden zone, which is named as the photonic bandgap (Chiu *et al.* 2006a).

For its application as immunosensors, the grating surface will serve as a sensing layer that requires further functionalized modification for the detection of a given biomolecule as antigen or antibody. Based on a specific molecular-recognition hypothesis, the interactions of different molecules can be differentiated by the changes in reflectance or resonant angles due to changes in effective refractive index on the surface. Every molecule, for which a selective counterpart is found, may in principle be measured with this sensor system. Figure 18.9 shows the illustration of the recognition process on the chip surface along with the use of a flow-through microchannel. The incoming light beam passes through the air/glass interface first and then the glass/fluid interface before reaching the grating surface. Refraction takes place at both interfaces.



**Fig. 18.9** Metallic grating surface with biochemical sensing layer as an integrated sensor system. Reprinted with permission from Chiu *et al.*, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, 6521 (2006b). Copyright 2006, IEEE.

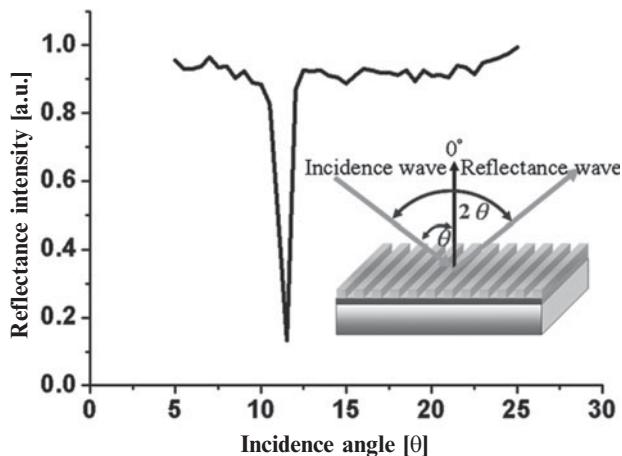
In the case of an empty flow cell, both refractions make up for each other and the beam suffers only a small lateral displacement that can be neglected. Figure 18.10 shows the changes of angular-dependent reflectivity around resonant angle from metallic grating by using a TM-polarized 530 nm light source. The resonant peak for the grating coupler was measured on a rotary table at zero position (angle  $\varphi = 0$  degree) with different angles of reflected light ( $\theta = 5\text{--}25$  degrees). The differences in analytes are measured as a change in the peak positions with reference to the SPR dip in the water, as shown in Fig. 18.10.

Other than the uses of the fixed plane of incidence and changes of the incident angle,  $\theta$ , we can also record the reflectivity as a function of the azimuthal angle ( $\varphi$ ) for a fixed angle of incidence. The advantage of this procedure is that only the grating has to be rotated whilst the source and detector remain fixed. The experimental setup used to record azimuthal-angle-dependent reflectivity data from a diffraction grating is shown in Fig. 18.11. The grating is placed on the rotating table with the incident light at a fixed angle (Lin *et al.* 2007). The incident wave on a grating with pitch  $\Lambda$  permits momentum match such that we may couple to surface plasmon resonances,  $k_{\text{SP}}$ . It can be expressed as the linear summation of azimuthal-angle-dependent coupled resonance at angle  $\varphi$  and the incident-angle-dependent coupled resonance at angle  $\theta$ , in the following equation, where the angle and sign refer to the diffracted order,  $m$ , and the refractive index of ambient medium,  $n_a$  (Hibbins *et al.* 1998):

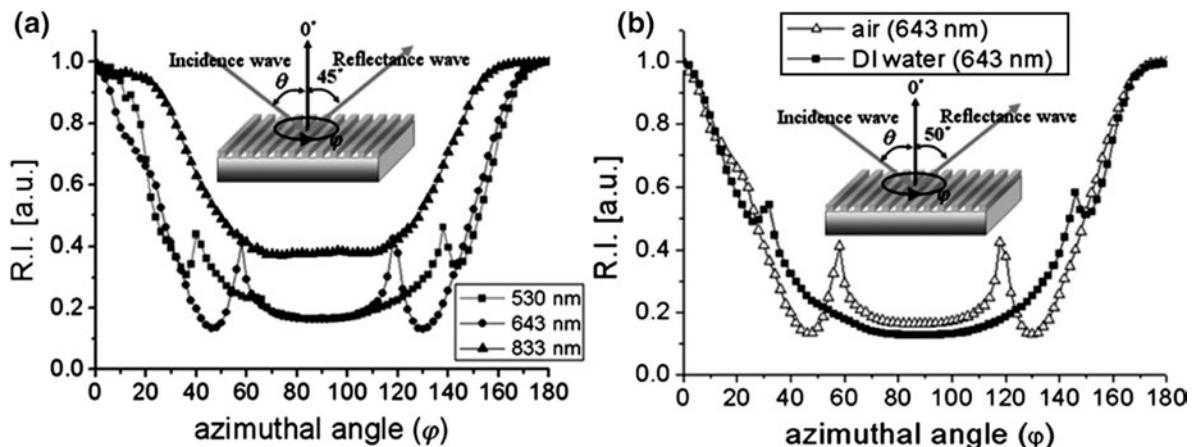
$$k_{\text{SP}}^2 = n_a^2 k_0^2 \sin^2 \theta + \left( m \frac{2\pi}{\Lambda} \right)^2 \pm 2n_a m \frac{2\pi}{\Lambda} k_0 \sin \theta \cos \varphi. \quad (18.13)$$

The special case of azimuthal angle  $\varphi = 0$  can be shown as:

$$n_a \sin \theta \pm m \frac{\lambda}{\Lambda} = \pm \sqrt{\frac{\varepsilon_m n_a^2}{\varepsilon_m + n_a^2}}. \quad (18.14)$$



**Fig. 18.10** The measured reflected light angle from the metallic grating for TM-polarized light with a wavelength of 530 nm. Reprinted with permission from Chiu *et al.*, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, 6521 (2006b). Copyright 2006, IEEE.



**Fig. 18.11** Illustration of the azimuthal-angle-dependent scans at (a)  $\theta = 45$  degree with 530, 643, and 833 nm in air and (b)  $\theta = 50$  degree in air and DI water at 643 nm. Two minima in reflectivity can be observed that are attributed to the coupling to a surface plasmon in the first order. The minimal reflectance at azimuthal angles smaller than 90 degrees can be used for quantitative comparison due to its symmetry. Reprinted with permission from Chiu *et al.*, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, 6521 (2006b). Copyright 2006, IEEE.

Figure 18.11(a) shows the scanning azimuthal angular SPR spectra at three different wavelengths (530 nm, 643 nm, and 833 nm) and the reflectance measurement setup at 45 degrees. We then compared the spectral changes due to immersive medium of air (filled square) and deionized (DI water) open triangle at 643 nm as shown in Fig. 18.11(b). Further quantitative comparison of different experimental conditions can depend on the shifting of the minimal reflectance at the azimuthal angle smaller than 90 degrees for its symmetry at 90 degrees. Table 18.2 shows the summary of experimental results that are obtained from the scanning spectra.

### 18.3.4 Active plasmonics

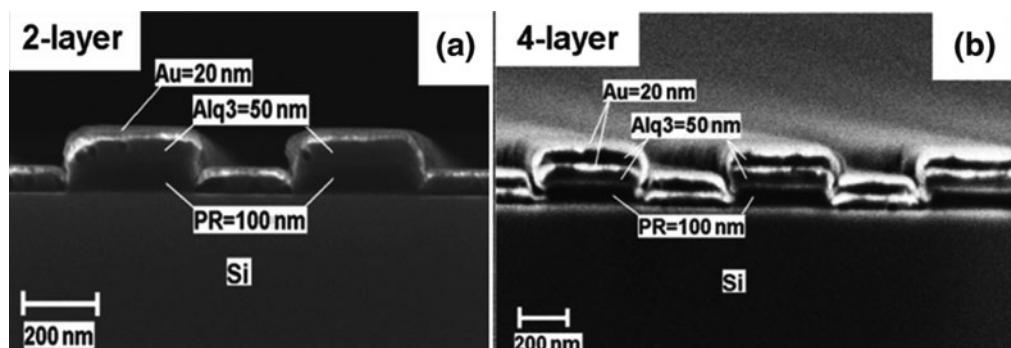
The properties of the light source used in a SPR system play an important role in signal quality. It is known that SPR can only be induced by P-polarized (TM) light that is parallel to the incident plane. Such a P-polarized light can come from many different sources by placing a polarizer between either light source and sample or sample and detector. It can be a point source with finite diameter for collective measurement in spatial average, or in imaging mode

**Table 18.2** Summary of grating 400 nm for SPR resonance angles.

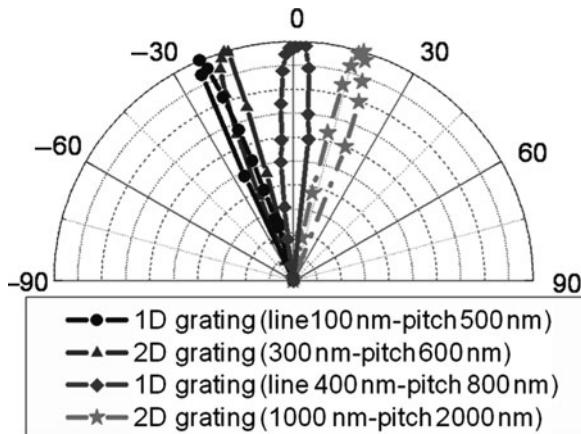
Air ( $n = 1$ )	Water ( $n = 1.3319$ )	NaCl (10 mM) ( $n = 1.3322$ )	NaCl (50 mM) (1.3324)	NaCl (0.1 M) (1.333)	Glucose (0.125 M) (1.3353)	Glucose (0.25 M) (1.3386)	Alcohol (50%) ( $n = 1.3563$ )	Alcohol (95%) (1.3585)
530 nm	12.8°							
643 nm	33.6°	21.525°	21.52°	21.515°	21.48°	21.418°	21.315°	20.72°
833 nm		64.4°	64.38°	64.35°	64.3°	64.18°	63.96°	62.98°

with a spatial resolution of  $\mu\text{m}$ . Replacing the bulky external light source with a smaller light source for portable applications is a challenge. One successful example is the use of TI Spreeta, a small footprint LED light source with a polarizer film.

Recently, Lin's group has reported the measurement results of biochemical solutions by using a fabricated 1D grating plasmonics device (Chiu *et al.* 2007a). In this device, a thin layer of light-emitting molecules, Alq<sub>3</sub> has been deposited on top of a nanograting structure for the modulation of directional emission by the interaction of surface polaritons. They have tried to combine a planar light source of organic LED (OLED) with the above-mentioned nanograting patterns and a microlens for the excitation of surface plasmon wave. It was found that the resultant PL emission from OLED molecules when interacting with a Au nanostructure is subject to the modulation effect of local SP (Chiu *et al.* 2007a,b). The samples were prepared by an electron-beam lithography system. The exposure was carried out for 2  $\mu\text{s}$  by a pixel map of  $60\,000 \times 60\,000$  dots to give a total exposure area of  $1.2 \times 1.2 \text{ mm}^2$ . Next, 50 nm of organic Alq<sub>3</sub> and subsequently 20 nm of gold were deposited on the grating by a thermal evaporator with vacuum level and evaporation rate of approximately  $2 \times 10^{-6}$  Torr and 0.2 Å/s, respectively. They have designed, fabricated, and characterized two SPR configurations, a 2-layer one of [Si/grating (PR)/Alq<sub>3</sub>(50 nm)/Au(20 nm)] and a 4-layer device of [Si/grating (PR)/Alq<sub>3</sub>(50 nm)/Au(20 nm)/Alq<sub>3</sub>(50 nm)/Au(20 nm)] symmetrical organic dielectric films, with grating linewidth and pitch size of 400 nm and 800 nm. Figure 18.12(a) and (b) shows the cross-sections of a 2-layer and a 4-layer device, respectively. They use the Alq<sub>3</sub> emission to excite SPPs on multilayer grating coupled emission. The emissions correspond to the resonant condition of SP modes on the Alq<sub>3</sub>/Au interface and grating couple to the Au/air interface for the emission of light. This technique has surface plasmon grating coupled emission (SPGCE) of light passing through the metal and is a multilayer grating approach for the excitation of SPPs. The experimental results showed that these devices have directional emission with enhanced intensity and reduced full width half-maximum (FWHM). They have



**Fig. 18.12** SEM images of the gratings cross-section, which show the arrangement of a periodically lamellar layer (a) 2-layer structure of Alq<sub>3</sub>/Au and (b) 4-layer structure of (Alq<sub>3</sub>/Au/Alq<sub>3</sub>/Au) on top of a 100 nm PR structure. Reprinted with permission from Chiu *et al.*, *Opt. Exp.* **15**, 11608 (2007). Copyright 2007, Optical Society of America.



**Fig. 18.13** Directional control of scattering reflectance by 1D and 2D grating patterns.

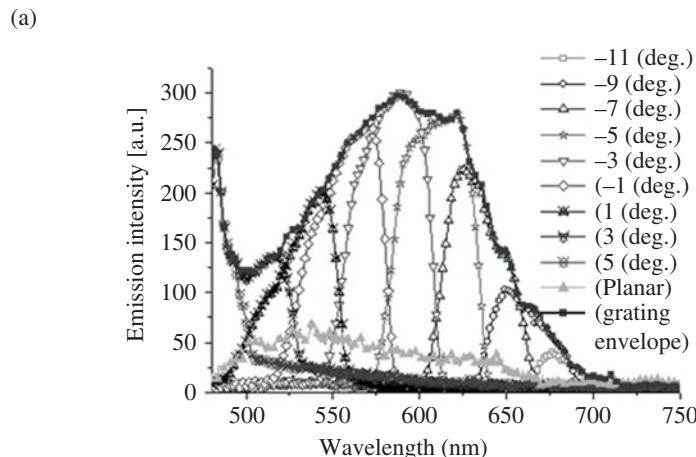
also tried to fit the measurement results to a revised theoretical model of a grating coupler on a thin layer of metal, which confirms the existence of SPP mode with the momentum-matching condition of a surface-plasmon resonance in the experimental configuration.

The PL spectra and corresponding luminescence changes at different emission angles are measured and shown in Fig. 18.13. The enhanced PL and directional emission is due to the SPR effect as mentioned above. The resultant angular emission spectra of enhanced luminescence from a metal/organic grating are shown in Fig. 18.14(a). Figure 18.14(b) shows the grayscale-coded images for 4-layer devices. It was found that the 4-layer device has higher intensity and smaller FWHM. The average shift in peak wavelength is 14 nm/degree for the grating with 800 nm pitch size. The emission spectra can shift from 750 nm to 480 nm by changing the emission angle for measurement. Emission intensity at different viewing angles of the planar, two-layer, and four-layer devices are shown in Fig. 18.15. This results in an angular-dependent tuneable color device with specific structural parameters, e.g., pitch constant ( $\Lambda$ ), the thickness of each layer of the grating and the optical indices of used materials, to satisfy the SPP resonant conditions. One can thus design the geometrical features of a grating structure to optimally control the directional scattering and associated field intensity.

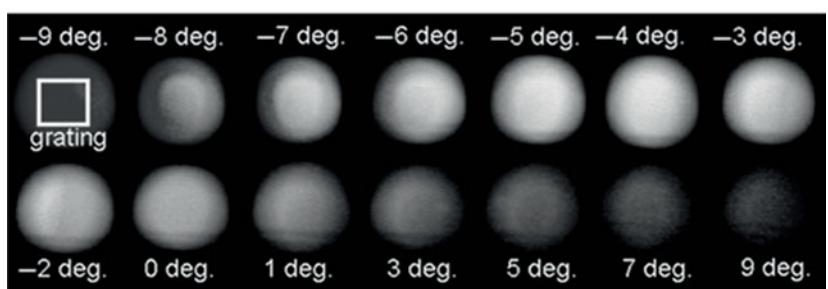
## 18.4 Metallic effect (LRSPR, CMO adhesive layer)

### 18.4.1 Long-range surface plasmon resonance (LRSPR)

The interactions of SPW can be further enhanced by using a multilayer of metals or dielectric materials to create the so-called “long-range SPR condition”. In SPR devices, different metallic thin-film materials can be used for supporting free-electron oscillations. Other than the well-known Au or Ag, different metals (e.g. Al) can be used for controlling SP wave propagation with



(b)

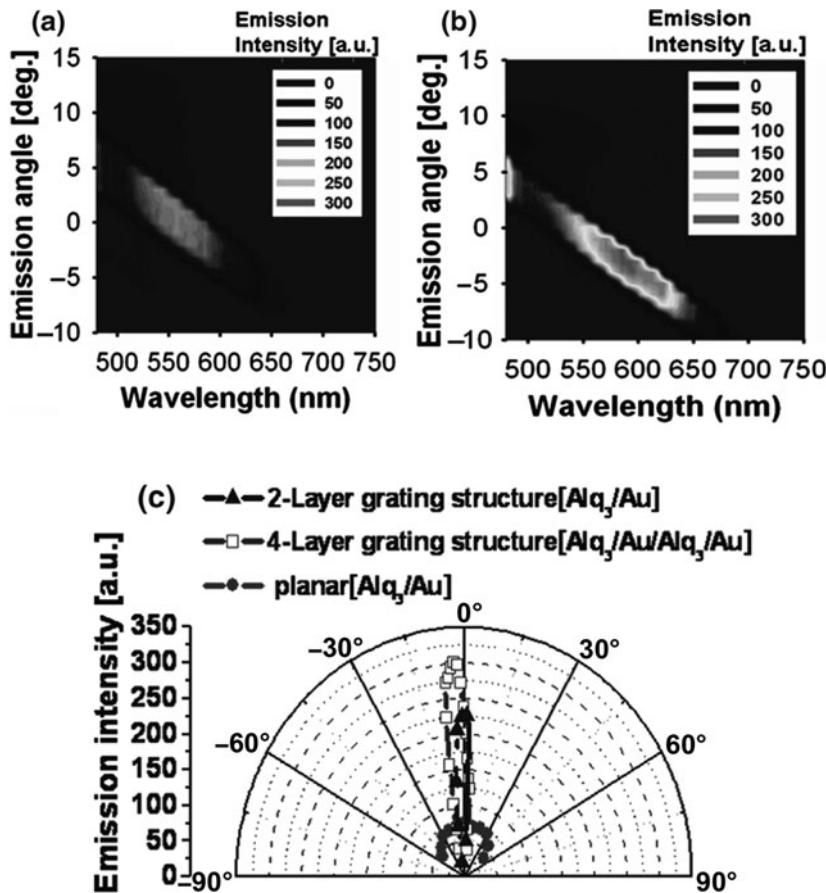


**Fig. 18.14** The PL emission obtained from a grating of a 4-layer sample, grating size (line 400 nm, pitch 800 nm, area size  $1.2 \times 1.2$  mm). (a) The PL of grating and non-grating (planar) samples as well as the integration of overall emission angles (envelope). The PL is measured on the organic/metal grating. (b) The changes in SPR were measured from the grayscale images of different angular spectra. Reprinted with permission from Chiu *et al.*, *Appl. Phys. Lett.* **91**, 083114 (2007). Copyright 2007, American Institute of Physics.

different length constant at the metallic/dielectric interface for plasmonics. For possible application in microfluidic, it is feasible to use longer length constant designs for SP waves on the surface. Due to possible complications resulting from surface reactions, the choice of metallic surface for biological samples is quite limited. Alternatively, one combines the principle of LRSPPR to accomplish the goal. By using  $MgF_2$  as the buffer layer between the prism and Au, the propagation length can be dramatically enhanced up to 250 times the Kretschmann configuration, where  $L_x$  is about  $3.3 \mu m$  at  $633 nm$ . A longer propagation constant can be achieved by using NIR with critical thickness of Au layer, as shown in Table 18.3.

**Table 18.3** Summary of design parameters for enhanced SP wave at different excitation wavelengths with  $MgF_2$  buffer layer and critical thickness of Au ( $d_{cut}$ ).

Wavelength	633 nm	790 nm	980 nm
$n_{MgF_2}$ (1000 nm)	1.37895	1.37600	1.37600
$n_{Au}$	$0.16195 + i3.21$	$0.1472 + i4.5764$	$0.177 + i5.9028$
$n_{H_2O}$	1.33169	1.32640	1.32200
$d_{cut}$	17 nm	23 nm	30 nm
$L_x$ (at cutoff d)	$839.78 \mu m$	5.7637 mm	158.68 mm



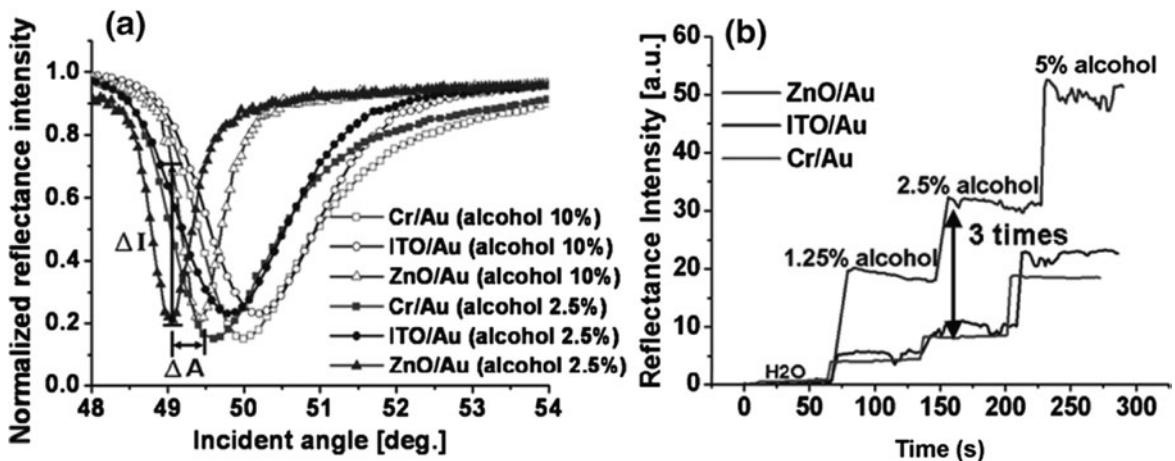
**Fig. 18.15** PL emission obtained from a grating sample having a 2-layer and 4-layer structure (grating size: line 400 nm, pitch 800 nm, area size  $1.2 \times 1.2 \text{ mm}^2$ ). In (a) and (b) are shown a PL 3D emission image obtained from a grating sample. The dependence of the emission spectra on observation angle ( $\theta$ ) is shown in (a) and (b) for the 2-layer and 4-layer structure, respectively. (c) Shows the planar, 2-layer and 4-layer. The emission maximum was about  $0^\circ$  and  $-3^\circ$  for 2-layer, 4-layer devices, respectively. Reprinted with permission from Chiu *et al.*, *Appl. Phys. Lett.* **91**, 083114 (2007). Copyright 2007, American Institute of Physics.

#### 18.4.2 Conducting metal oxide as adhesive layer

There are increasing interests in using conducting metal oxide (CMO) such as indium tin oxides (ITO), with its free electrons, as resonant medium. We have found that both ITO and zinc oxide (ZnO) can be used to replace chromium (Cr) as an adhesive layer necessary for the deposition of Au film onto glass substrate. As shown in Fig. 18.16, replacing Cr with ZnO or ITO shows (a) the largest slope sensitivity and (b) the highest reflectance under various concentrations of alcohol solutions, which can be very useful in the low-concentration region. The enhancement mechanism has been proposed and attributed to its compatible work function of ZnO to Au ( $\Phi_{\text{ZnO}} = 5.4 \text{ eV}$ ,  $\Phi_{\text{ITO}} = 4.7 \text{ eV}$ ,  $\Phi_{\text{Cr}} = 4.5 \text{ eV}$ ,  $\Phi_{\text{Au}} = 5.7 \text{ eV}$ ) and low optical loss due to absorption in the vis-NIR region ( $n_{\text{ZnO}} = 2.2$ ,  $n_{\text{ITO}} = 1.626 + i0.01$ ,  $n_{\text{Cr}} = 4.276 + i4.328$ ,  $n_{\text{Au}} = 0.189 + i5.419$ , @ 833 nm).

## 18.5 Microfluidic parts

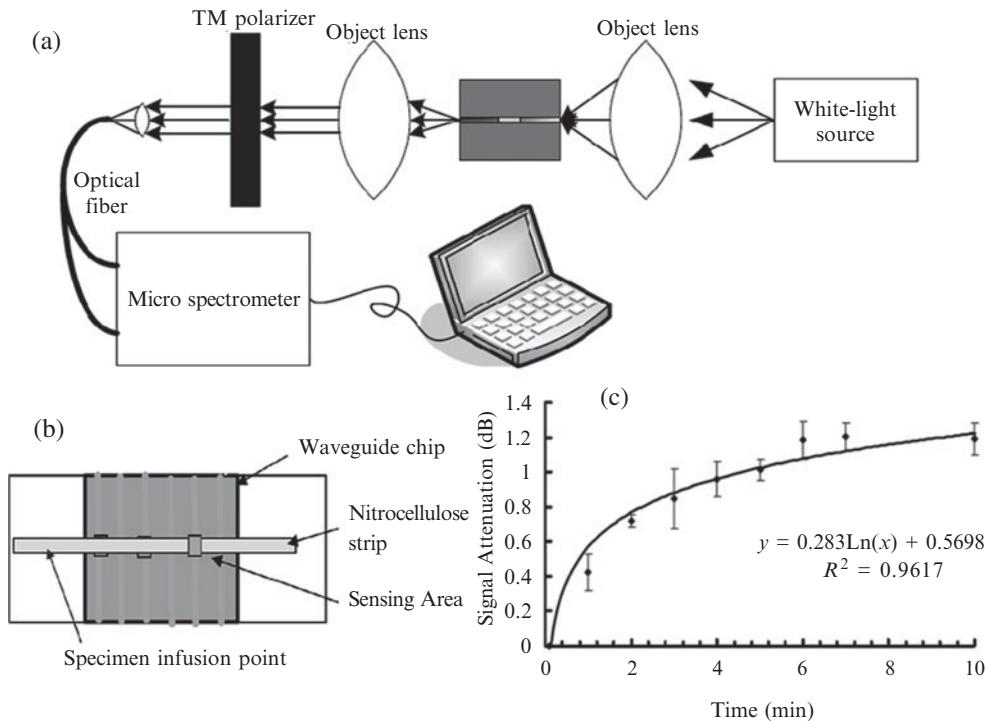
Most of the biomolecular interactions on the sensor surface are introduced by using a microchannel with pre-defined geometrical features. It has the



**Fig. 18.16** The uses of conducting metal oxide (ITO and ZnO) as adhesive material demonstrated better performance than Cr. (a) Slope sensitivity and (b) the reflectance under different alcohol concentrations.

advantages of delivering precise amounts of sample to the reaction site for possible automation and parallel processing with better accuracy and reproducibility (Yager *et al.* 2006). However, it is still a challenge to adapt the existing protocols in micro/nanoscale environments for continuous flow (Myszka 1997). The biomolecular interactions in the microchannel are diffusion limited and may require a specially designed structure to improve its performance in the passive mode, e.g. mixer, separation, or directional control. In some field applications, one would prefer powerless fluidic manipulation, which can be used for sample loading and even buffer-washing processes. It will also require surface-modification techniques to control the surface wettability for a fluidic switch, ratio mixing, and molecular attachment. The driving force is to design a working device with low operating voltage (<1 V) in comparison to current electrowetting technologies that require a much higher potential of several hundred volts.

For some field applications, a quantitative fast lateral flow device can be designed by using nitrocellulose membrane as both filtering and supporting substrate for flow carriers, e.g. buffer or plasma, to interact with surface-immobilized biomolecules on the active region of the SPR sensor area(s). A simple photospectrometer with a white-light source can be used to measure the absorption spectra for the specific recognition of antigen–antibody on the active site of a fabricated SPR waveguide sensor as shown in the upper part of Fig. 18.17. A strip of nitrocellulose membrane is then applied onto the surface that acts as a fluidic carrier and filter for the analyte. The efficacy of such a lateral flow device depends on many engineering factors for precision, e.g. pore size, length, thickness, added surfactant, etc. By monitoring the changes in absorption at a specific wavelength, the binding of virus particles (Avian Leukosis Virus, ALV) to the immobilized monoclonal antibodies (mAb22) can be detected in real time. Figure 18.17(c) shows the average values of three measurements that reached steady state in about 5 min.



**Fig. 18.17** (a) The optical setup for a SPR waveguide measurement system. A white-light source is used to provide wavelength modulation for the SPR detection in the visible region by a computer-controlled spectrophotometer through an objective lens for focusing and TM polarizer. (b) A strip of nitrocellulose membrane is applied across the waveguide device, where in the active region mAbs have been immobilized before the experiment. (c) A time-series measurement of light intensity at a specific wavelength (570 nm) shows the dynamics of molecular interactions in 10 min ( $n = 3$ ). Reprinted with permission from Huang *et al.*, *Biosens. Bioelect.* **22**, 519 (2006). Copyright 2006, Elsevier B.V.

## 18.6 Biomolecular layer effect

An affinity-based biosensor is one of the technologies that holds great promise in biomedical research due to its specificity in molecular recognition. Its application has been widely used for the detection of pathogens in the food industry, environment, and biomedical science, including virus (Baac *et al.* 2006), bacteria (Waswa *et al.* 2007), cancer marker (Teramura and Iwata 2007) and signal transduction (Li *et al.* 2007). To achieve the high sensitivity and specificity of immunoassays, it is critical for the immobilization process of antibodies to have corrected orientation to access their active sites for the intended functions on the sensor surface. Therefore, it is imperative to develop biochemical methods that can efficiently increase the binding capacity for antigens and have good stability in different sensing conditions. Antibodies in a highly oriented manner have been shown to maintain higher binding capacity than those of a random manner (Babacan *et al.* 2000). Generally, antibodies can be captured onto the surface by three methods: physical (Joshi *et al.* 2007), chemical (Sai *et al.* 2006), and biological (Kang *et al.* 2006) immobilizations. The physical methods are the earliest and simplest approach for biomolecular immobilization, which can be done by using entrapment of porous matrix or

using van der Waals or weak electrostatic forces to adsorb onto the sensor surface. In general, the physical method based on the adsorption of antibodies to the metal surface is not stable enough for commercial biosensors. Although it shows advantages of simplicity and does not require complicated protein modification, it does tend to reduce binding activities and bind non-specific molecules in random orientation. The most widely used approach, which has the advantage of the strong protein attachment, is the immobilization of antibodies to a self-assembled monolayer (SAM) onto gold films (Kim *et al.* 2004; Sai *et al.* 2006). This approach is based on the covalent interaction between the carboxyl group (-COOH) on the chemically activated surface and primary amines (-NH<sub>2</sub>) of lysine residues on antibodies. Unlike adsorption, the coupling bond is strong and the biomolecular coverage is in a high density. However, there is a major problem on antibody immobilization through lysine residues that are usually present on the outside of protein. Due to the several lysine residues on an antibody, random coupling of antibodies through the distribution of their exposed amine groups always leads to an increase in steric-hindrance problems and reduces antigen-binding capacities. As an alternative strategy for immobilization, the biological capture method is quickly evolving, too. It can provide proper orientation of the antibody on the sensing surface for antigen binding. The most commonly oriented immobilizations include the use of protein A and protein G, both are surface proteins from bacterium *Staphylococcus aureus* and streptococcal strain, respectively. These proteins can lead to bacteria immune escape because they are generally expected to bind to the Fc part of the antibody specifically (Muñoz *et al.* 1998). Owing to the strong interaction between these proteins and antibody, protein A and protein G have led to their widespread use in powerful bioaffinity assays for antibody purification and immunoprecipitation. While these proteins are involved in the binding of Fc region, the antibody can be immobilized in the order orientation and the Fab region is free to interact with the antigen. For the reasons mentioned above, the use of biological immobilization can improve sensitivity rather than random amine coupling. Thus, this bioaffinity immobilization has been extensively reported for higher binding efficiency, contributing to the enhancement of biosensor performance (Babacan *et al.* 2000; Lee *et al.* 2003; Oh *et al.* 2004; Bae *et al.* 2005). On account of the high affinity of bioimmobilization for the Fc region of monoclonal antibody, we would like to perform a comparative evaluation of the binding efficiency between protein G and protein A with monoclonal antibody of avian leukosis virus (ALV), which causes neoplasia in chickens and becomes a widespread contagious disease affecting the poultry industry.

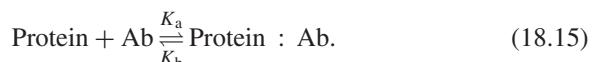
By utilizing surface plasmon resonance (SPR), the protein kinetics can be measured on the SAM-modified surface. By coating the surface with receptor, the presence of analyte in the sample can be detected by the effective refractive-index change as analyte binds to the SPR sensing surface (Homola *et al.* 1999). SPR has been used to provide the possibility of developing faster diagnosis methods compared to traditional detection approaches. Due to it being an accurate, sensitive and fast method for a wide range of application, the number of studies using SPR is increasingly growing (Myszka 1997; Hsieh *et al.* 2004; Oh *et al.* 2005; Natarajan *et al.* 2008). However, it is very important

to optimize the conditions of immobilized biomolecules, which include buffer and wash solutions to prevent artifacts due to the bulk effect of refractive index, especially for SPR-based biosensors.

All chemical reagents used were of analytical grade. 8-Mercaptooctanoic acid ( $C_8H_{16}O_2S$ ) was synthesized with the help of Prof. Adam Shih-Yuan Lee's group at Tamkang University, and was dissolved with absolute alcohol. 1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC), N-hydroxysuccinimide (NHS), bovine serum albumin (BSA) and human serum albumin (HSA) were purchased from Sigma. Binding reactions were analyzed by using SPRimager instrumentation (GWC Technologies, Madison, USA), where changes in surface-plasmon resonance were monitored in real time as analyte was passed over the sensor chip. The gold-coated slide was first immersed into 8-mercaptopoctanoic acid solution at room temperature for 20 min. Then the chemical immobilized surface was activated using 400 mM EDC/100 mM NHS for covalent-bond formation in 10 min at room temperature. The gold-film chip was then soaked in 50  $\mu$ g/ml protein A (protein G) solution that fixed on the gold surface for 10 min. The antibody at a concentration of 50  $\mu$ g/ml was immobilized onto the surface for 20 min and the surface was blocked with 1% BSA for 5 min. The modified chip was placed on the holder and matching oil dripped on the glass surface of the chip. Using match oil made a prism and the chip is fully aligned with the planes on the holder. At the starting step, the gold surface was exposed to PBS buffer at a flow rate of 40  $\mu$ L/min with the syringe pump in order to get the stable baseline. The chip was measured by using SPR at a fixed angle of 50 degrees under 790 nm wavelength. Monoclonal antibodies at different molar concentration (17, 33, 67, 167, 333 and 667 nM) in PBS were injected into the immobilized surface at the above flow rate for 20 min. The molar concentration, also called molarity, (in units of mol/L, or M) usually expressed as the number of moles of solute per liter of solution. PBS was finally added to the immobilized surface followed by buffer rinse. For detection using a biosensor, it is important to block the unreacted sites on the sensing surface to reduce the amount of non-specific binding of proteins during subsequent steps in the detection. Many blocking agents, e.g. milk powders, BSA or other highly purified proteins, have been used to block unreacted sites on the sensing surface. In this work, we chose BSA, which is a commonly used agent for all immunoassay applications as a blocking agent. The chip immobilized with a protein layer was immersed in monoclonal antibody at a concentration of 25  $\mu$ g/ml for 20 min. The gold surface was dipped into 1 mg/ml of BSA that had been used for blocking non-specific protein adsorption for 10 min. The modified chip was placed in the holder and matching oil dripped on the glass surface of the chip. The gold film was exposed to PBS buffer at a flow rate of 10  $\mu$ L/min in order to get the stable baseline. The chip was measured at a fixed angle of 50 degrees under 790 nm wavelength. ALV antigen at 10-fold dilution in PBS was added into the device surface at the above flow rate for 80 min. PBS was finally added to the immobilized surface followed by buffer rinse. We have followed the above-mentioned preparation procedures to study both the steric effect in antibody–antigen interaction and the effect of flow rate.

### 18.6.1 Kinetics analysis of antibody binding

Binding of analytes to the ligands on the sensor chip is in principle a two-step event. At the beginning, the analytes will diffuse into the unstirred layer on the sensor surface following the concentration gradient from the center of the flow to the side-wall. Then the analytes would interact with the immobilized ligands on the surface. The unstirred layer is also known as the Nernst diffusion layer, which is formed by diffusion and is mainly concerned with mass-transport effects. Only diffusion can control the transfer of analyte to the sensing surface (Sikavitsas *et al.* 2002). Outside this layer, the solution is assumed to be well mixed and the concentration is nearly constant with respect to distance. In other words, the binding rate is entirely dependent on the analyte concentration in the unstirred layer. Therefore, when performing concentration measurements, such a condition can be achieved by using a low flow rate (Myszka *et al.* 1998). As mentioned above, it is important to control the fluidic flow rate in the protocol design. When using a different flow rate to deliver the same amount of sample volume, one would normally prefer a slower flow rate that can maximize the total amount of binding analyte to the surface. It is reasonable to ignore mass-transfer limits in this situation so that the measured kinetic constant can represent the reaction rate between molecular pairs. This model also assumes that the binding reaction is reversible. We can then describe the binding kinetics by the following equation:



The rate of association,  $K_a$ , is:

$$\frac{d[\text{Protein : Ab}]}{dt} = K_a[\text{Protein}][\text{Ab}]. \quad (18.16)$$

And the rate of dissociation,  $K_d$ , is:

$$-\frac{d[\text{Protein : Ab}]}{dt} = K_d[\text{Protein : Ab}]. \quad (18.17)$$

When the reaction reaches equilibrium, the rates of association and dissociation are equal. Rearrange the equation to define the equilibrium dissociation constant,  $K_D$ , as:

$$K_D = \frac{K_d}{K_a} = \frac{[\text{Protein}][\text{Ab}]}{[\text{Protein : Ab}]}, \quad (18.18)$$

where Ab acts as the ligand and protein is either protein G or A and serves as the receptor.

$$\text{Total available site} = [\text{Protein}] + [\text{Protein : Ab}] = B_{\max}. \quad (18.19)$$

The total available site is considered as the greatest binding capacity,  $B_{\max}$ , which is equal to the sum of the bounded and unbounded protein concentration on the surface.

From the above-mentioned equations, we find:

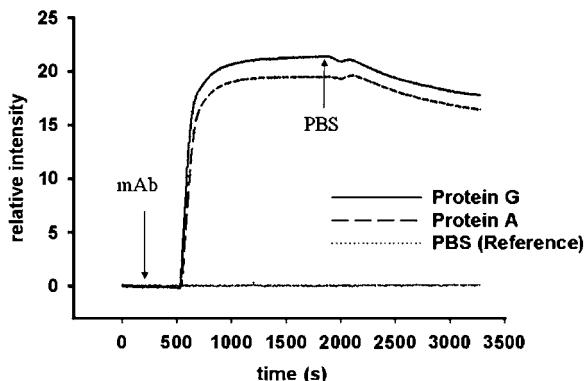
$$[\text{Protein : Ab}] = \frac{B_{\max}[\text{Ab}]}{K_D + [\text{Ab}]} \quad (18.20)$$

For the SPR device, the optical intensity response is treated as being proportional to the surface-attached mass. In these experiments, optical intensity response is proportional to [Protein: Ab] and therefore the equation becomes

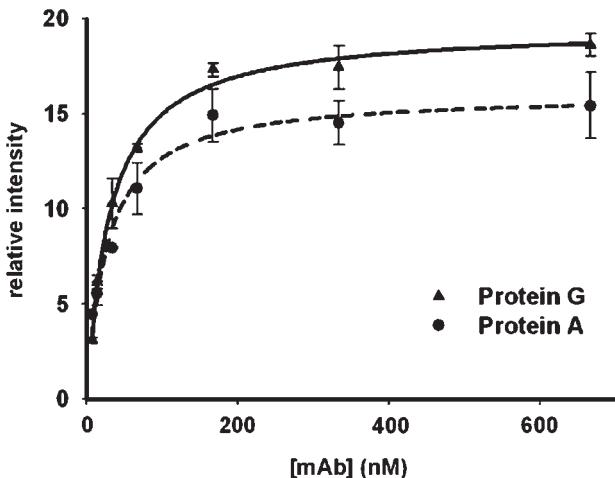
$$\text{relative intensity} = \frac{B_{\max}[\text{Ab}]}{K_D + [\text{Ab}]} \quad (18.21)$$

By setting [Protein] equal to  $K_D$  in the above equation, we will get that  $[\text{Ab}]/[\text{Protein:Ab}] = 1$ , so  $[\text{Ab}]$  equals  $[\text{Ab:Protein}]$ . This means that half of the antibodies are free and the other half are bound to the protein layer. In other words, when the concentration of antibody equals  $K_D$ , half the protein will be occupied at equilibrium. If the antibodies have a higher affinity for the protein layer, the  $K_D$  value will be lower, as it will take a lower concentration of protein to bind half the antibodies.

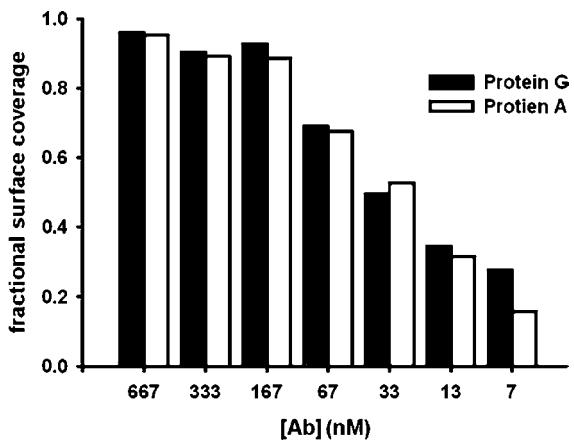
First, we will try to evaluate the capturing efficiency of mAbs (IgG) by the immobilized protein A and protein G layer. The reactions can be monitored with relative intensity changes on the sensing surface in real time. The chip immobilized with protein G layer was immersed in monoclonal antibody (mAb22) at a concentration of 17, 33, 67, 167, 333, and 667 nM for 20 min. The gold surface was dipped into 1 mg/ml of BSA that had been used for blocking non-specific protein adsorption for 10 min. The modified chip was then measured by placing on the holder and matching oil dripped on the glass surface of the chip. The gold film was exposed to PBS buffer in order to get the stable baseline. The chip was measured at an angle of incidence of 50 degrees under 790 nm wavelength. ALV antigen at 10-fold diluted concentration in PBS was injected into the immobilized surface at a flow rate of 10  $\mu\text{L}/\text{min}$  for 80 min. PBS was finally added to wash the surface followed by buffer rinse. Figure 18.18 showed two overlapped data traces of adding 167 nM to the immobilized protein A and G, followed by rinsing with PBS buffer as indicated by the arrows. It was found that mAbs-protein G has higher relative intensity change than mAbs-protein A. By using eqn (18.21), we can plot the relative intensity versus different concentrations of antibodies to make the best regression analysis, as shown in Fig. 18.19. The Langmuir fit of the data yields  $K_D$  of protein G-Ab =  $3.03 \times 10^{-8} \text{ M}$  with saturation of relative intensity (maximum binding;  $B_{\max}$ ) =  $19.54 \pm 1.08$ ,  $R^2 = 0.995$ ,



**Fig. 18.18** Real-time relative intensity changes of 167 nM monoclonal antibodies (mAb22) show greater response to immobilized protein G layer than protein A with reference to the PBS buffer.



**Fig. 18.19** Binding plots (relative intensity versus bulk solution concentration) used to determine the kinetic constants for mAb22 adsorption to protein layer with Langmuir fit.



**Fig. 18.20** Plot of the fractional surface coverage ( $\theta$ ) computed by relative intensity divided by saturation of relative intensity for the determination of optimal concentration.

and  $K_D$  of protein A-Ab =  $2.65 \times 10^{-8}$  M with saturation of relative intensity =  $16.06 \pm 1.83$ ,  $R^2 = 0.969$ .

In order to estimate the coverage of antibodies on the gold surface, we can calculate the surface coverage according to the following equation:

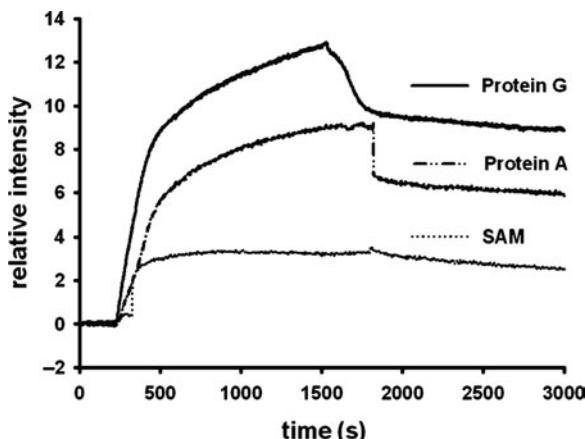
$$\text{fraction surface coverage}(\theta) = \frac{\text{sites occupied}}{\text{total available sites}} = \frac{\text{relative intensity}}{B_{\max}}, \quad (18.22)$$

where  $\theta$  is the fractional surface coverage ( $0 < \theta < 1$ ). It defines the saturation surface coverage of a particular binding molecule on the surface. Figure 18.20 shows the plot of the fractional surface coverage ( $\theta$ ) as a function of mAb22 concentration on protein G and protein A, respectively.

Staphylococcal protein A and Streptococcal protein G are bacterial Ab-binding proteins that are widely used as immunological tools. Formerly, it was believed that protein G and protein A both interacted mainly with the Fc part of

Ab. Therefore, they widely offered the optimal spatial orientation of antibodies and maximum antigen-binding efficiency for the biosensor development. We have investigated which one is suited for the antibodies of avian leukosis virus to improve biosensor sensitivity. Accordingly, we can see that the greater binding affinity between the protein layer and antibody is due to the smaller equilibrium dissociation constant. From the Langmuir fit of the data, the  $K_D$  of the protein A layer to antibody is smaller than the  $K_D$  of the protein G layer. This indicates that protein A has the better binding force to the avian types of antibodies than protein G. Basically, the relative intensity based on the SPR principle is proportional to the amount of analytic molecules attached to the recognized receptor. So the concentration of the antibody/protein A complex should be higher than the concentration of the antibody/protein G complex. But this is not the case, as shown in Fig. 18.19 and Fig. 18.20, and obviously not only do the protein G monolayers make a better intensity response to interaction with antibody but they also express a higher degree of the surface coverage. Therefore, from Fig. 18.20, these results suggest that the effective binding capacity between protein G and antibodies is superior to protein A. Protein G (molecular weight of 22 kDa) has two binding domains that can bind to the Fc region of IgG. While protein A (molecular weight of 42 kDa) has four binding subunits that can bind to the residues in the CH<sub>2</sub> and CH<sub>3</sub> domains of IgG heavy chains. From this experimental result, it implies that protein A binding subunits don't bind with antibody completely, possibly due to steric hindrance. On the other hand, protein G encounters a smaller steric effect in spite of only two binding subunits.

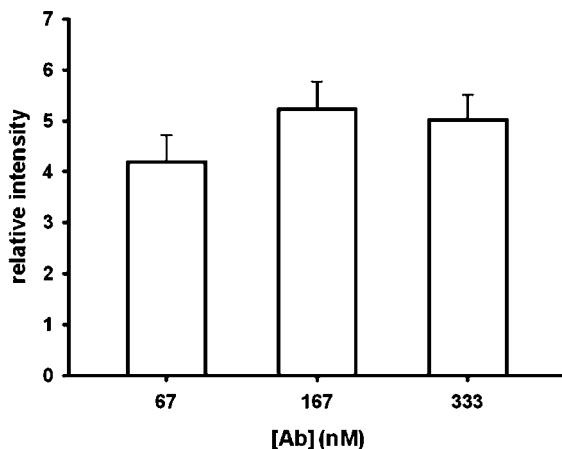
Figure 18.21 shows the change of SPR intensity after adding ALV virus particles at 10-fold dilution in PBS to the prepared mAb22 on top of self-assembled protein G and protein A layers. It is clear that the binding of ALV to the immobilized antibodies on the protein G layer has a higher SPR intensity change than protein A layer and the control layer of SAM only. The relative intensity values are about 8.8, 5.8 and 2.5, respectively. After three repeated experiments, the changes of SPR intensity on protein G, protein A, and SAM layer are found with relative intensity values about  $7.90 \pm 0.76$ ,  $5.59 \pm 2.01$ , and  $2.63 \pm 1.61$ , respectively. Antibody immobilization is an important tech-



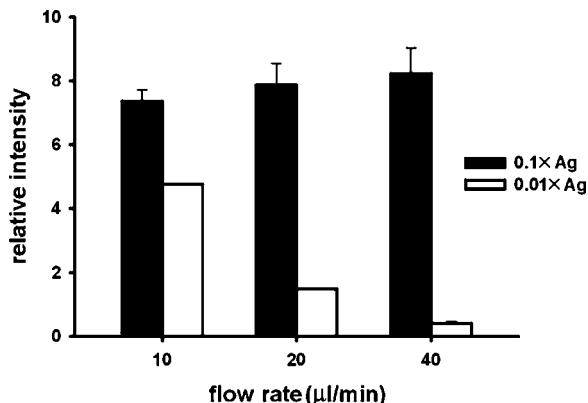
**Fig. 18.21** Measurement of SPR intensity with injected  $0.1\times$  ALV virus particles to the mAb22 layer on top of different protein layers and SAM layer as control.

nique with a variety of applications such as in diagnostic biosensors. The orientation of the antibodies immobilized on the surface is considered to be a determinant of their effectiveness. To confirm the selectivity of the protein layer to antibody, we compared the SPR binding intensity of antigen to immobilized Ab on the protein G, protein A, and SAM layer. By comparison between different protein layers and SAM layers, the intensity of antigen binding to the antibody on the protein layer increased at least twofold more than the control SAM layer. In other words, immobilization using a protein layer resulted in a sensitivity two times larger than that of random immobilization in any case. And with the immobilized protein G or protein A, the variation in intensity response to the experiments was smaller than only with the SAM layer. It had been studied previously that antibody orientation plays a significant role in antibody activity and function. In our work, the results are consistent with previous reports and support the use of oriented immobilization methods rather than random couplings. There is about 40% more in SPR intensity change by using protein G than protein A. These results showed that using protein G as an immobilized layer provided the better avidity to the avian antibody and improved the binding efficiency of avian antigen to the antibody so that the optimal method of immobilization determined was to make use of protein G.

Steric effects are a widely known phenomenon in biological studies of biosensor development. On the biosensor development, the higher ligand density is not better because the conformation of ligand may be formed as a barrier between ligand and receptor. Therefore, it may reduce the opportunity to access the active sites and decrease the sensitivity of the biosensor. To monitor the effect of steric hindrance on the SPR intensity response of the immobilized antibody, three different concentrations of antibody monolayer (67 nM, 167 nM, 333 nM) were used to measure the relative intensity signal after binding to antigens. As the concentration of antibody increased from 67 nM to 333 nM, the antibody surface loading was saturated at the concentration of 167 nM. At a concentration of the immobilized antibody layer of



**Fig. 18.22** Averages and standard deviations of SPR intensity measurement for the adsorption of antigen to antibody at the different concentration on the protein G layer ( $N = 3$ ).



**Fig. 18.23** Averages and standard deviations of SPR intensity measurement for the adsorption of antigen to antibody at the different flow rates on the protein G layer.

333 nM may cause steric effect which hinder the attachment of antigen. In addition, it showed that the fraction surface coverage is over 90% from 167 nM concentration. From the above results, the optimal immobilized concentration of antibody for antigen was 167 nM. The changes of SPR intensity by binding of  $0.01 \times$  ALV to the antibody on the self-assembled protein G layer is shown in Fig. 18.22. The SPR intensity changes were about  $4.19 \pm 0.52$ ,  $5.23 \pm 0.54$ , and  $5.02 \pm 0.48$  on the protein G layer, respectively. Finally, Fig. 18.23 showed the effect of flow rate on the relative intensity shift for a fixed volume of  $800 \mu\text{L}$  of  $0.1 \times$  and  $0.01 \times$  diluted ALV solutions that flow through the surface. It showed that the intensity changes are not significantly different under different flow rates of  $10$ ,  $20$ ,  $40 \mu\text{l}/\text{min}$ . However, it does have significant changes with lower concentration at  $0.01 \times$  dilution in higher flow rate.

## 18.7 Conclusions

The fundamental theory of SPR has allowed numerous approaches to realize miniature high-performance biosensors. Nano/microtechnologies have been used to fabricate several novel SPR devices for the applications in biomolecular sensing. The uses of microfluidic, e.g. micro/nanochannel, pump, mixer, valve, are now well realized for integrated clinical or home-care applications. We have reported a novel design of alternative dielectric layers to enhance the SPR signal quality and modulate its resonant position. The admittance loci method of thin-film design has been used in such a device. Revisiting Maxwell's equations also leads to the development of a novel SPR device, which is based on the lateral propagation of a surface plasmon wave along the metal–dielectric surface. For the realization of a hand-held SPR system, we have developed three SPR waveguide sensors. A lateral-flow microfluidic channel based on a nitrocellulose membrane has been integrated with a SPR waveguide sensor to achieve dynamic detection. We found that replacing the traditional Cr layer with ZnO resulted in better performance. An integrated SPR device, which couples an OLED light source with a nanograting structure and a top sensing layer, has been tested and its angular-dependent emission has been studied. Analysis of biological interaction in life science research,

drug development and clinical testing are very important. Proper orientation of antibodies on surfaces is critical to the performance of biosensors. In this chapter, we showed and compared the difference in bioimmobilization of protein G and protein A and indicated how to establish the reaction kinetics. Compared with a conventional method, the oriented immobilized antibody demonstrated a higher binding efficiency. The SPR analysis showed that the IgG layer was immobilized more evenly on the protein G layer than on the protein A layer. Using protein G as an immobilized layer provided the better avidity to the avian antibody and improved the binding efficiency of avian antigen to the antibody. In addition, using a SPR biosensor, we found the surface coverage of over 90% at 167 nM antibody on the immobilized surface and the steric effect contributed to the intensity phase of SPR response decreasing at 333 nM antibody. Moreover, by surface-coverage analysis and the steric-effect comparison, the optimal immobilized concentration of antibody on the SPR biosensor is determined as 167 nM. These factors must be carefully considered when designing procedures for SPR measurement.

## Acknowledgments

This project is supported in part by National Science and Technology Program in Pharmaceuticals and Biotechnology, National Science Council, Taiwan, R.O.C., NSC 96-2323-B002-001, NSC 96-2323-B002-004, NSC 96-2218-E-002-026, MOEA 96-EC-17-A-05-S1-0017, and COA 95-AS-6.2.1-AD-U1.

## References

- Andrew, P., Barnes, W.L. *Science* **306**, 1002 (2004).
- Babacan, S., Pivarnik, P., Letcher, S., Rand, A.G. *Biosens. Bioelectron.* **15**, 615 (2000).
- Bae, Y.M., Oh, B.K., Lee, W., Lee, W.H., Choi, J.W. *Biosens. Bioelectron.* **21**, 103 (2005).
- Bravo-Abad, J., Degiron, A., Przybilla, F., Genet, C., García-Vidal, F.J., Martín-Moreno, L., Ebbesen, T.W. *Nature Phys.* **2**, 120 (2006).
- Burin, A.L., Ratner, M.A. *J. Phys. Chem. A* **104**, 4704 (2000).
- Chance, R.R., Prock, A., Silbey, R. *J. Chem. Phys.* **62**, 2245 (1975).
- Chiu, N.F., Lee, J.H., Kuan, C.H., Wu, K.C., Lee, C.-K., Lin, C.-W. *Appl. Phys. Lett.* **91**, 83 (2007a).
- Chiu, N.F., Yu, C., Nien, S.Y., Lee, J.H., Kuan, C.H., Wu, K.C., Lee, C.K., Lin, C.W. *Opt. Exp.* **15**, 11608 (2007b).
- Chiu, N.F., Chien, L.C., Chiu, S.J., Lee, J.H., Lin, C.W. *Proc. SPIE*, 63231V (2006a).
- Chiu, N.F., Nien, S.Y., Lee, J.H., Lin, C.W. *Conf. Proc. IEEE Eng. Med. Biol. Soc. Suppl*:6521 (2006b).
- Choong, V.E., Park, Y., Gao, Y., Mason, M.G., Tang, C.W. *J. Vac. Sci. Technol. A* **16**, 1838 (1998).
- Chu, Y.S., Hsu, W.H., Lin, C.W., Lee, K.C. *Micro. Opt. Tech. Lett.* **48**, 955 (2006).
- Enderlein, J., Ruckstuhl, T. *Opt. Expr.* **13**, 8855 (2005).

- Fang, N., Liu, Z., Yen, T.J., Zhang, X. *Opt. Expr.* **11**, 682 (2003).
- Feng, J., Okamoto, T., Simonen, J., Kawata, S. *Appl. Phys. Lett.* **90**, 081106 (2007).
- Gifford, D.K., Hall, D.G. *Appl. Phys. Lett.* **81**, 4315 (2002).
- Gruhlke, R.W., Holland, W.R., Hall, D.G. *Phys. Rev. Lett.* **30**, 2838 (1986).
- Kang, G.Y., Han, G.Y., Kang, J.Y., Cho, I.H., Park, H.H., Paek, S.H., Kim, T.S. *Sens. Actuators B* **117**, 332 (2006).
- Hibbins, A.P., Sambles, J.R., Lawrence, C.R. *J. Mod. Opt.* **45**, 1019 (1998).
- Homola, S.S., Yee, Gauglitz, G. *Sens. Actuators B* **54**, 3 (1999).
- Homola, J. *Anal. Bioanal. Chem.* **377**, 528 (2003).
- Hsieh, H.V., Pfeiffer, Z.A., Amiss, T.J., Sherman, D.B., Pitner, J.B., Biosens. *Bioelectron.* **19**, 653 (2004).
- Hung, Y.J., Smolyaninov, I.I., Davis, C.C., Wu, H.C. *Opt. Exp.* **14**, 10825 (2006).
- Huang, J.G., Lin, H.M., Chuang, T.L., Lee, C.L., Wang, W.S., Juang, R.H., Wang, C.H., Lee, C.K., Lin, S.M., Lin, C.W. *Biosens. Bioelectron.* **22**, 519 (2006).
- Baac, H., Hajos, J.P., Lee, J., Kim, D., Kim, S.J. *Biotechnol. Bioeng.* **94**, 815 (2006).
- Kalinowski, J. *Organic Light-emitting Diodes: Principles, Characteristics, and Processes* (Marcel Dekker, New York, 2005), 1.
- Kalkman, J., Strohhofer, C., Gralak, B., Polman, A. *Appl. Phys. Lett.* **83**, 30 (2003).
- Kim, N., Park, I.S., Kim, D.K. *Sens. Actuators B* **100**, 432 (2004).
- Kitson, S.C., Barnes, W.L., Sambles, J.R. *Opt. Commun.* **122**, 147 (1996).
- Kretschmann, E., Raether, H. *Z. Naturforsch.* **23A**, 2135 (1971).
- Lakowicz, J.R. *Principles of Fluorescence Spectroscopy*, 2nd edn. (Kluwer Academic/Plenum Publishers, New York, 1999), 1.
- Lakowicz, J.R. *Anal. Biochem.* **298**, 1 (2001).
- Lakowicz, J.R., Shen, Y., D'Auria, S., Malicka, J., Fang, J., Gryczynski, Z., Gryzynski, I. *Anal. Biochem.* **301**, 261 (2002).
- Lakowicz, J.R., Malicka, J., Gryczynski, I., Gryczynski, Z., Geddes, C.D. *J. Phys. D; Appl. Phys.* **36**, R240 (2003).
- Lee, W., Oh, B.K., Bae, Y.M., Paek, S.H., Lee, W.H., Choi, J.W. *Biotechnol. Bioeng.* **19**, 185 (2003).
- Li, X., Huang, M., Cao, H., Zhao, J., Yang, M. *Sens. Actuators B* **124**, 227 (2007).
- Lin, C.W., Chen, K.P., Su, M.C., Lee, C.K., Yang, C.C. *Opt. Quantum Electron.* **37**, 1423 (2005).
- Lin, C.W., Chen, K.P., Lin, S.M., Lee, C.K. *Sens. Actuators B* **113**, 169 (2006a).
- Lin, C.W., Chen, K.P., Su, M.C., Hsiao, T.C., Lee, S.S., Lin, S.M., Shi, X.J., Lee, C.K. *Sens. Actuators B* **117**, 219 (2006b).
- Lin, C.W., Chiu, N.F., Huang, J.G., Lee, C.K. 5th International Conference on Nanochannels, Microchannels and Minichannels, Proceedings of ASME ICNMM, Puebla, Mexico, June 18–20 (2007).
- Macleod, H.A. *Tutorials in Optics* (OSA, Washington, DC, 1992), 121.
- Maier, S.A., Kik, P.G., Atwater, H.A., Meltzer, S., Harel, E., Koel, B.E., Requicha, A.A.G. *Nature Mater.* **2**, 229 (2003).
- Joshi, M., Pinto, R., Rao, V.R., Mukherji, S. *Appl. Surf. Sci.* **253**, 3127 (2007).
- Munoz, E., Vidarte, L., Pastor, C. *Eur. J. Immunol.* **28**, 2591 (1998).

- Myszka, D.G. *Curr. Opin. Biotechnol.* **8**, 50 (1997).
- Myszka, D.G., He, X., Dembo, M., Morton, T.A. *Biophys. J.* **75**, 583 (1998).
- Nakano, T., Kobayashi, H., Shinbo, K., Kato, K., Kaneko, F., Kawakami, T., Wakamatsu, T. *Mater. Res. Soc. Symp.* **660**, 1 (2001).
- Natarajan, S., Katsamba, P.S., Miles, A., Eckman, J., Papalia, G.A., Rich, R.L., Gale, B.K., Myszka, D.G. *Anal. Biochem.* **373**, 141 (2008).
- Oh, B.K., Lee, W., Kim, Y.K., Won, H.L., Choi, J.W. *J. Biotechnol.* **111**, 1 (2004).
- Oh, B.K., Lee, W., Bum, S.C., Young, M.B., Won, H.L., Choi, J.W. *Biotechnol. Bioeng.* **20**, 1847 (2005).
- Otto, A. *Z. Phys.* **216**, 398 (1968).
- Patskovsky, S., Kabashin, A.V., Meunier, M., Luong, J.H.T. *Sens. Actuators B* **97**, 409 (2004).
- Piliarik, M., Vaisocherova, H., Homola, J. *Biotechnol. Bioeng.* **20**, 2104 (2005).
- Pockrand, I., Brillante, A. *Chem. Phys. Lett.* **69**, 499 (1980).
- Raether, H. *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* (Springer-Verlag, Berlin, 1988), p. 6.
- Shinbo, K., Toyoshima, S., Ohdaira, Y., Kato, K., Kaneko, F. *J.J. Appl. Phys.* **44**, 599 (2005).
- Sikavitsas, V., Nitsche, J.M., Mountzaris, T.J. *Biotechnol. Prog.* **18**, 885 (2002).
- Teramura, Y., Iwata, H. *Anal. Biochem.* **365**, 201 (2007).
- Sai, V.V.R., Mahajan, S., Contractor, A.Q., Mukherji, S. *Anal. Chem.* **78**, 8368 (2006).
- Waswa, J., Irudayaraj, J., DebRoy, C. *LWT-Food Sci. Technol.* **40**, 187 (2007).
- Wedge, S., Barnes, W.L. *Opt. Exp.* **12**, 3673 (2004).
- Winter, G., Barnes, W.L. *Appl. Phys. Lett.* **88**, 051109 (2006).
- Yager, P., Edwards, T., Fu, E., Helton, K., Nelson, K., Tam, M.R., Weigl, B.H. *Nature* **442**, 412 (2006).

# Carbon-nanotube field emission electron and X-ray technology for medical research and clinical applications

Sigen Wang, Otto Zhou, and Sha Chang

# 19

19.1 Introduction	673
19.2 Electron field emission from carbon nanotubes	674
19.3 Carbon-nanotube field emission electron and X-ray technologies in biomedical applications	677
19.4 Summary and conclusion	694
References	696

## 19.1 Introduction

Ionizing radiation has been widely used for diagnostic imaging (Bushong *et al.* 1997), cancer therapy (Khan *et al.* 1984), industrial inspection (Wong *et al.* 2006), food sterilization (Thayer *et al.* 2004), and security (Liu *et al.* 2008). The basic design of the current X-ray tubes has not been changed much since Roentgen's discovery and the subsequent invention of Coolidge (1956). A conventional X-ray tube comprises a metal filament (cathode), which emits electrons when resistively heated to over 1000 °C, and a metal target (anode) that generates X-radiation when bombarded by the accelerated electrons. Most of the current thermionic cathode X-ray tubes have one hot-filament cathode and one fixed focal spot on the anode. Dual-focus X-ray tubes are also commercially available, usually with two hot filaments of different sizes to give two different focal spot sizes. The thermionic tube based on the design of Coolidge has been a reliable workhorse for X-ray radiology with relatively low cost and high flux. They, however, have several inherent limitations due to the use of thermionic cathodes: (a) the current X-ray tubes cannot be easily miniaturized due to the high temperature required to extract electrons (>1000 °C); (b) they have limited temporal resolution due to the slow response time of thermionic emission; and (c) high operating temperature reduces the lifetime of the X-ray tubes and increases the operating costs. These deficiencies limit the size, the speed, the flux and the overall design flexibility of X-ray imaging and radiotherapy systems. Although other sources of X-ray radiation such as synchrotron (Margaritondo *et al.* 1988) and laser plasma have been developed over the years, their usages are restricted primarily to the research community because of the limited availability.

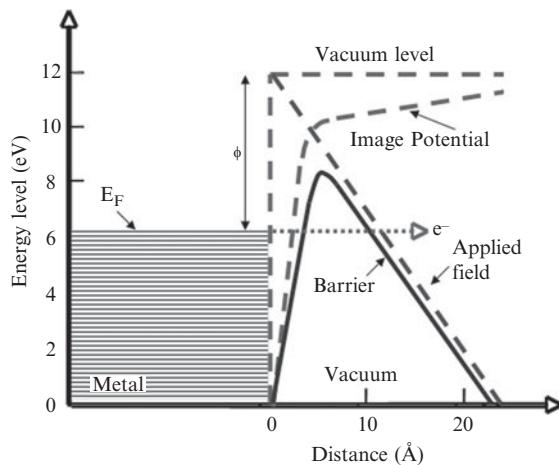
Field emission is a more attractive mechanism to extract electrons compared to thermionic emission because electrons are emitted at room temperature and emission is voltage controllable. The concept of field emission X-ray tube has been discussed in the past and devices with low X-ray flux have been demonstrated (Rangstein *et al.* 2000; Zhou *et al.* 2000; Baptist *et al.* 2001; Sugie *et al.* 2001; Whitlock *et al.* 2001). However, most diagnostic applications require tube currents on the order of 10–100 mA and operating voltage in the range of 30–150 kV, which was difficult to accomplish for the field emission X-ray tubes.

Our group has recently demonstrated a carbon-nanotube (CNT)-based X-ray source (tube), which can generate sufficient X-ray flux for diagnostic imaging applications (Yue *et al.* 2002; Liu *et al.* 2006). The CNT-based X-ray sources have several advantages compared to the thermionic X-ray tubes. One advantage is that by eliminating the thermionic cathode, the life span of the X-ray tubes can potentially be prolonged. Moreover, the size of the X-ray source can be reduced significantly. It also has the ability to produce focused electron beams with very small energy spread and programmable pulse width and repetition rate (Liu *et al.* 2006). This CNT-based cold-cathode X-ray technology can potentially lead to portable and miniature X-ray sources for medical and industrial applications. In this chapter, we briefly review CNT-based X-ray technologies in biomedical applications including an X-ray source (Yue *et al.* 2002), microfocus X-ray tube (Liu *et al.* 2006), microcomputed tomography (micro-CT) scanner (Zhang *et al.* 2005a; Cao *et al.* 2008), stationary digital breast tomosynthesis (Lalush *et al.* 2006; Yang *et al.* 2008), microradiotherapy (micro-RT) system (Wang *et al.* 2007, 2008) and single-cell irradiation system (Chang *et al.* 2006). These are based primarily on our own research at the University of North Carolina (UNC) at Chapel Hill.

## 19.2 Electron field emission from carbon nanotubes

Conventional X-ray tubes use the thermionic electron source. In the thermionic electron source, a material is heated to a high temperature and its electrons gain sufficient energy to overcome the material's work function to be emitted.

In contrast, the field emission X-ray tubes use the cold field emission electron source. This source is based on the phenomenon of field emission, which is associated with a quantum-mechanical tunnelling process whereby electrons tunnel through a (material-dependent) potential barrier under the influence of a high electric field (Gomer *et al.* 1961). In the cold field emission source, a material is subjected to a high electric field (typically a few volts per nanometer) which substantially narrows its electrons' potential barrier to vacuum. The electrons in the material can then quantum-mechanically tunnel through this thinned barrier and be emitted. Figure 19.1 illustrates the process of field emission. Typically, metals such as tungsten and molybdenum are used as cold field emitters.



**Fig. 19.1** Schematic illustration of field emission (quantum-tunnelling effect). When an electric field is applied to the metal, electrons can escape from the metal surface.

The current generated by the field emission process is described by the Fowler–Nordheim (FN) equation (Fowler *et al.* 1928), given by:

$$I = AJ_{\text{FN}},$$

where  $A$  is the emission area and  $J_{\text{FN}}$  is the Fowler–Nordheim current density given by:

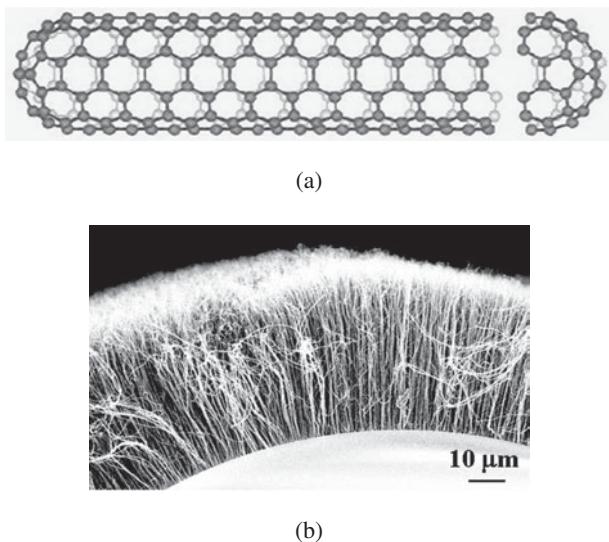
$$J_{\text{FN}} = c_1 F_t^2 \exp(-c_2 \phi^{3/2} / F_t),$$

where  $F_t$  is the electric field at the tip apex,  $\phi$  is the work function, and  $c_1$  and  $c_2$  are constants. The electric field  $F_t$  may also be expressed as  $F_t = \beta (V/d)$ , where  $V$  is the applied potential;  $\beta$  is the field-enhancement factor and it may be described as  $\beta = h/r$ , where  $h$  is the height of the tip and  $r$  is the radius of curvature of the tip apex. Thus,  $J_{\text{FN}}$  may also be given by:

$$J_{\text{FN}} = c_1 (\beta V)^2 / \phi d^2 \exp(-c_2 d \phi^{3/2} / \beta V).$$

The Fowler–Nordheim equation states that the emission current ( $I$ ) increases exponentially with increasing voltage ( $V$ ). According to the above equations, the emission current is strongly dependent on the following three factors: (a) the work function of the emitter surface; (b) the radius of curvature of the emitter apex; and (c) the emission area.

For a metal with a flat surface, however, the threshold field is impractically high, which is typically around  $10^4$  V/ $\mu\text{m}$ . All the electron field emitters rely on the field-enhancement factor  $\beta$  at the sharp tips/protrusions. One way to fabricate sharp field emitters is by using lithography (Brodie *et al.* 1992). It is immediately apparent that a field emitter could be more power efficient than a thermionic emitter, which requires heating. Field emission sources offer several attractive characteristics such as instantaneous response to electric-field variation, resistance to temperature fluctuation, and a high degree of focus ability in electron optics due to their sharp (0.2–0.3 eV) energy spread. Moreover, the field-emitted electrons are confined to a narrow cone angle along the electrical-field direction, where thermal electrons are randomly distributed. However, due to the high electric field experienced at the tips of the materials



**Fig. 19.2** (a) Structure of single-wall carbon nanotube, typical dimension is 1–50 nm in diameter and 1–10  $\mu\text{m}$  in length; and (b) a SEM image of aligned multiwalled carbon nanotubes grown on a curved surface using a chemical vapor deposition technique.

during field emission, the metal atoms often diffuse or electromigrate, causing failure and thermal runaway. Thus, conventional field emitters of Spindt tips have not been used in practical devices since they in general suffer from high turn-on field, low emission current, high cost and poor stability with limited lifetime.

CNT is a relatively new carbon allotrope discovered in 1991 by Iijima *et al.* CNT is a unique form of carbon filament/fiber in which the graphene walls roll up to form tubes, with diameters typically 1–50 nm and lengths of a few micrometers. CNT is comprised of either a single graphene shell (Iijima *et al.* 1991), termed a single-walled carbon nanotube (SWNT) as shown in Fig. 19.2(a), or multiple concentric graphene shells termed multiwalled carbon nanotube (MWNT) (Iijima *et al.* 1993). Figure 19.2(b) shows a scanning electron microscopy (SEM) image of aligned MWNTs grown on a curved surface using a chemical vapor deposition (CVD) technique. Several properties of carbon nanotubes make them favorable for field emission: (a) with graphene walls parallel to the filament axis, nanotubes exhibit high in-axis electrical conductivity at room temperature; (b) nanotubes are high in aspect ratio (length/diameter) which is typically on the order of  $10^3$ , and whisker-like in shape, which is the optimal shape for field emission, with a tip diameter of tens or a few nanometers (i.e. small point source); and (c) nanotubes can be very stable and indeed robust emitters, even at high temperatures due to their strong C–C covalent bonds. Purcell *et al.* (2002) demonstrated that a nanotube emitter, even when heated by its own field-emitted current to 2000 K, remains stable.

In 1995, four years after the discovery of carbon nanotubes by Iijima (1991), three groups reported field emission from CNTs at low-turn-on fields and high current densities (Chernozatonskii *et al.* 1995; de Heer *et al.* 1995; Rinzler *et al.* 1995). From 1998 onwards, the perspective on the use of nanotubes in

field emission devices spurred on efforts worldwide: a first display (Wang *et al.* 1998), as well as a lighting element (Satio *et al.* 2000; Jung *et al.* 2001), were presented. More encouragingly, films with CNTs have also been demonstrated for X-ray sources (Sugie *et al.* 2001; Yue *et al.* 2002). Our group led by Dr. Otto Zhou was among the first to demonstrate the CNT-based field emission X-ray sources (Yue *et al.* 2002).

## 19.3 Carbon-nanotube field emission electron and X-ray technologies in biomedical applications

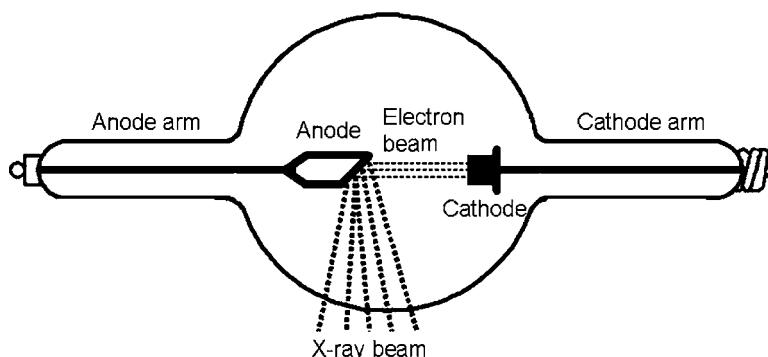
### 19.3.1 X-ray tubes

#### 19.3.1.1 Conventional X-ray tubes

Historically, the first X-ray tube was invented by William Crookes in 1878 and is called a Crookes tube. This tube is a partially evacuated glass bulb containing two electrodes. When an electric current passes through such a tube, the residual gas is ionized and positive ions, striking the cathode, eject electrons from it. These electrons bombard the glass walls of the tube and produce X-rays. Such tubes produce only soft X-rays with low energy.

The Crookes tube was greatly improved by William David Coolidge in 1913 (Coolidge 1930) and such an improved tube was called the Coolidge tube. The Coolidge tube, as shown in Fig. 19.3, consists of a heated filament (cathode) and a target (anode) in a vacuum environment. It is essentially a thermionic vacuum tube in which the cathode emits electrons when an auxiliary current heats the cathode. The electrons emitted from the heated cathode are accelerated by the application of a high voltage across the tube. As the voltage is increased, the energy of the radiation increases. Most of today's X-ray tubes are modified Coolidge tubes.

Nowday, the larger and more powerful tubes have water-cooled anodes to absorb the tremendous amount of heat generated at anodes due to the impact of the electron bombardment. X-ray production is an extremely inefficient process as 99% of the incident electron energy is converted to heat at the anode. Thus, heat dissipation at the anode is a major limitation on the power that can



**Fig. 19.3** The basic shape of the original Coolidge tube.

be applied. A rotating anode tube is also an improvement of the Coolidge tube. By rotating the anode the electron beam hits different parts of the anode and the heat is spread over a larger area. With the exception of dental X-ray tubes, almost all medical X-ray tubes have a rotating-anode design.

The conventional X-ray tubes have several inherent limitations due to the use of thermionic cathodes including limits to temporal resolution and cathode design flexibility.

#### **19.3.1.2 Early studies on field emission X-ray tubes**

X-ray tubes using field emission cathodes have been investigated since the 1970s (Charbonnier *et al.* 1975; Hallenbeck *et al.* 1975). Metal tips were used as the cathodes, and electrons were extracted by applying a pulsed high voltage between the target and cathode using Max Generators.

X-radiation is generated when the field-emitted electrons bombard the target. The advantages of the field emission X-ray tubes have been demonstrated in clinical studies in terms of resolution and exposure time (Hallenbeck *et al.* 1975). However, the metal-tip emitters were inefficient. In order to obtain a high electrical field required for extraction of the electrons, Max Generator, which uses a series of discharging capacitors, was employed (Crooks *et al.* 1972). Unfortunately, such X-ray tubes had a limited lifetime of 200–300 exposures and slow repetition rate. Due to its diode design, the acceleration voltage and the tube current could not be independently controlled.

Field emission X-ray tubes using other types of emitters such as the Spindt tips and diamond have also been investigated (Rangstein *et al.* 2000; Baptist *et al.* 2001; Whitlock *et al.* 2001). However, the highest electron current demonstrated in these X-ray tubes is only of the order of  $\mu\text{A}$  (Rangstein *et al.* 2000). This limits their practical applications.

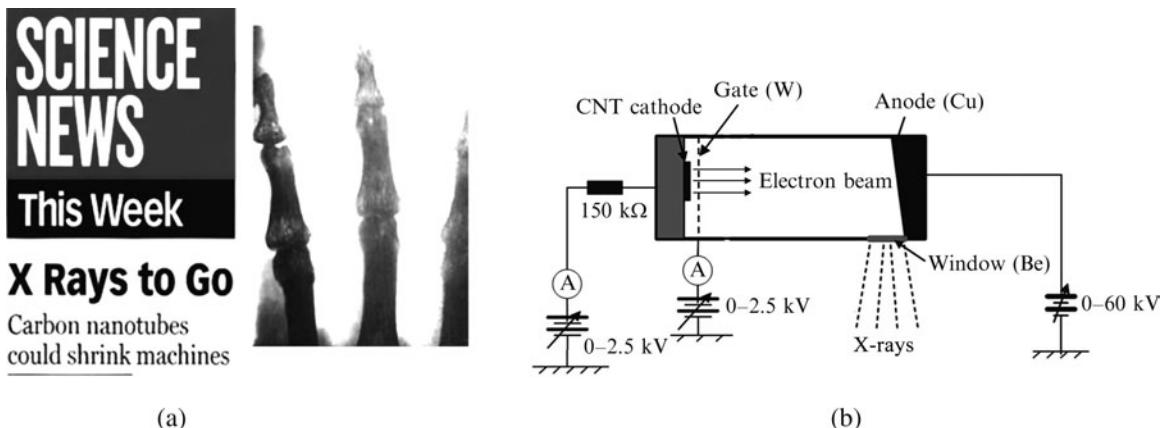
#### **19.3.1.3 Carbon-nanotube field emission X-ray source**

##### **19.3.1.3.1 First carbon-nanotube field emission X-ray tube**

Our group is among the first to demonstrate a CNT-based field emission X-ray source that is capable of producing sufficient flux for imaging human extremities as shown in Fig. 19.4(a) (Gorman *et al.* 2002; Yue *et al.* 2002; Zhou *et al.* 2003). Figure 19.4 (b) presents a schematic of the first CNT field emission X-ray tube developed in Dr. Otto Zhou's lab. It has a triode-type structure with a CNT cathode, a gate electrode and a metal target housed in a vacuum tube with a beryllium (Be) window. The electrons are extracted from the CNT cathode by applying a gate voltage between the cathode and gate. X-rays are produced when the emitted electrons are accelerated and bombard the copper target. The tube current and the acceleration voltage are independently controlled. By replacing the direct current (dc) gate voltage with an amplified signal from a function generator, electron emission and thus X-ray radiation with a programmable waveform is easily produced. Since the initial report, the design of the X-ray tube has undergone several modifications and the performance characterizations have been significantly improved.

##### **19.3.1.3.2 Carbon-nanotube-based field emission microfocus X-ray tube**

A CNT-based field emission microfocus X-ray tube with high spatial resolution, temporal resolution, and stable emission has been demonstrated in our

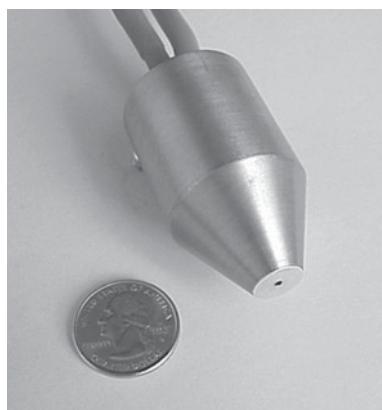


**Fig. 19.4** (a) An X-ray image of a human hand taken using our first CNT-based field emission X-ray tube; and (b) schematic drawing of the CNT-based field emission X-ray source.

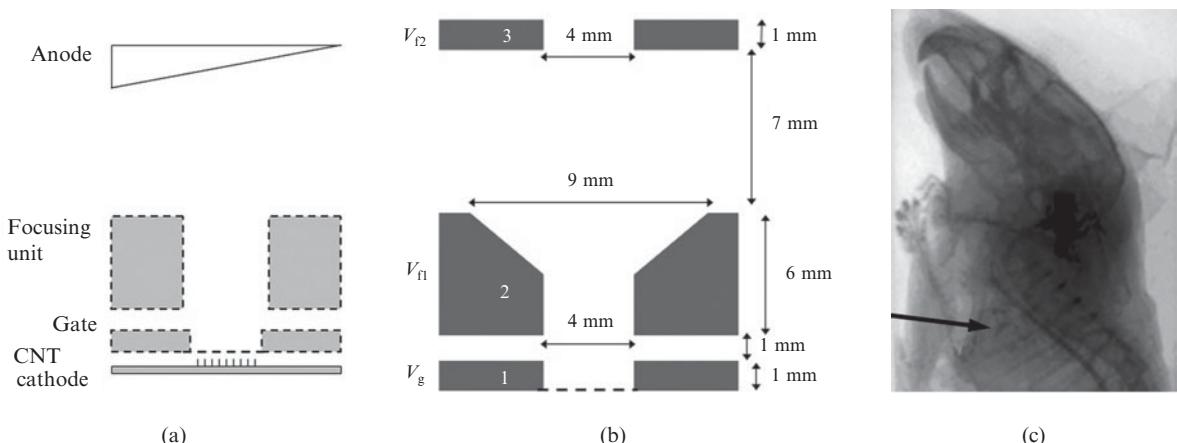
group (Liu *et al.* 2006). This microfocus X-ray tube is schematically illustrated in Fig. 19.5(a). It comprised of a CNT cathode, a gate electrode, an electrostatic focusing unit, and a molybdenum anode with a 9° tilting angle. The entire X-ray source was housed in a vacuum chamber with a base pressure of  $1 \times 10^{-7}$  Torr. Field-emitted electrons were extracted from the CNT cathode by applying a bias voltage on the gate electrode and were subsequently focused by the focusing lenses before reaching the anode. A more detailed drawing with the design and the dimensions of the focusing lenses is shown in Fig. 19.5(b). Electrodes 1 and 3 were made of planar metal diaphragms. The central focusing electrode 2 was in the shape of a truncated cone, which played the role of harnessing the divergence of electrons coming out of the gating grid, and thereby pre-focusing the electrons into a more parallel shape before they reach the final focusing electrode.

The flux generated by this X-ray tube at the 30  $\mu\text{m}$  resolution is higher than those used in the typical micro-CT imaging protocols with a fixed-anode thermionic X-ray source operating at a comparable resolution where the current is less than 0.1 mA at 40 kVp. Electron optics simulation results reveal that the focusing system of this source is very linear in focusing electrons emitted from a cathode with diameter ranging between 0.1 and 2 mm. It is expected that the focal spot size of 10  $\mu\text{m}$  can be obtained using this design but with a smaller CNT cathode. The combined high spatial and temporal resolutions of the CNT-based field emission microfocus X-ray source are highly attractive for dynamic tomography imaging. Figure 19.5(c) shows a projection image of a mouse using the CNT-based field emission microfocus X-ray tube, where the aorta of the heart is clearly revealed as denoted by the arrow.

A miniature CNT-based field emission X-ray tube, as shown in Fig. 19.6, has been jointly developed by Xintek, a UNC start-up company ([www.xintek.com](http://www.xintek.com)), and its partner for X-ray fluorescent applications. The transmission tube operates at 40 kV under the pulse mode and is battery powered. Over 1000 h of stable operation under the dc mode has been demonstrated.



**Fig. 19.6** An image of the prototype 40-kV miniature X-ray tube based on field emission X-ray technology (a dime is on the left).



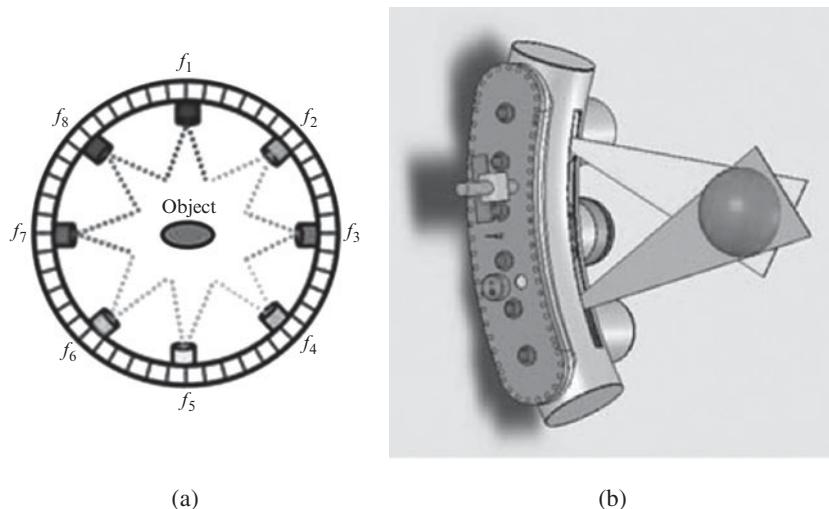
**Fig. 19.5** (a) Schematic of a CNT-based field emission microfocus X-ray tube that consists of a CNT cathode, gate electrode, focusing electrodes, and anode; (b) cross-sectional view of the focusing unit of inhouse developed X-ray tube; and (c) a projection image of a mouse by using the CNT-based field emission microfocus X-ray tube.

#### 19.3.1.3.3 Carbon-nanotube-based field emission multipixel X-ray source

Recently, we have been developing a new multipixel X-ray source that can generate digitized X-radiation with fine control of the spatial distribution and temporal modulation of the X-radiation for applications in gantry-free stationary tomography imaging systems (Yang *et al.* 2008) and microradiotherapy systems (Wang *et al.* 2007, 2008). This new multipixel X-ray source enables the design of rotation-free stationary tomography imaging systems with ultra-fast scanning speed and potentially better imaging quality compared to today's commercial scanners (Yang *et al.* 2008). The multipixel system also enables the design of microradiotherapy systems where an irregular and spatially and temporally modulated small radiation field is formed by electronically "turning on" a user-specified set of X-ray pixel beams (Wang *et al.* 2007, 2008).

Xinray Systems, a joint venture of Siemens and Xintek, has developed and manufactured multipixel X-ray sources for a broad range of applications including diagnostic medical imaging, homeland security and industrial inspection. Figure 19.7(a) shows a schematic illustration of the concept of stationary CT scanner with multiple X-ray pixels, and Fig. 19.7(b) shows a prototype multipixel X-ray tube. Xinray's multipixel X-ray tubes can typically be custom designed to meet the specific requirements of the system vendors.

CNT field emission has several advantages over other tip-material field emission and thermionic emission. Due to their atomically sharp tips and large aspect ratios, CNTs have a much larger field enhancement factor and a much lower emission threshold field that is practical to reach ( $\sim 10^2$  V/ $\mu$ m). CNT field emission X-ray technology offers two key properties that are lacking in the thermionic X-ray sources: temporal modulation and spatial configuration. CNT emitters are stable at high currents (>1 mA) and current density (>10<sup>6</sup> A/cm<sup>2</sup>) (Wei *et al.* 2001; Zhang *et al.* 2004)). These features make CNT field emission X-ray technology ideal for the development of novel imaging and irradiation systems. In our group, several novel imaging and irradiation systems have been developed or are under development. They mainly



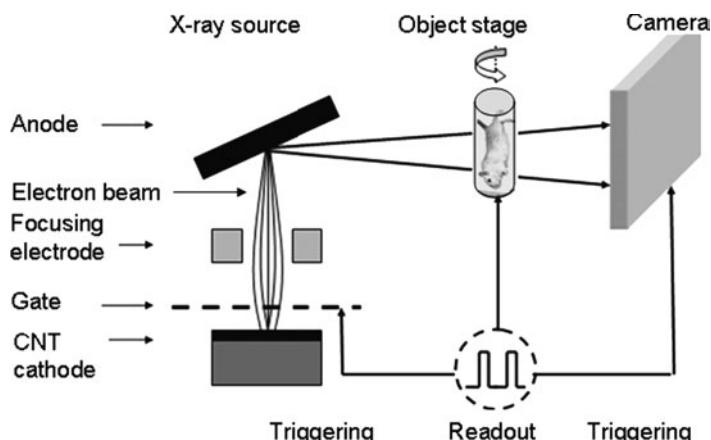
**Fig. 19.7** (a) A schematic illustrating the concept of stationary CT scanner with multiple X-ray pixels; and (b) a prototype multipixel X-ray tube fabricated by XinRay Systems.

include CNT field emission based stationary micro-CT (Zhang *et al.* 2005a; Cao *et al.* 2008), breast tomosynthesis (Lalush *et al.* 2006; Yang *et al.* 2008), multipixel beam micro-RT system for cancer research (Wang *et al.* 2007, 2008), and a multipixel microbeam array for single-cell irradiation (Chang *et al.* 2006).

### 19.3.2 Microcomputed tomography

The development of computed tomography (CT) technologies (Cormack *et al.* 1973; Hounsfield *et al.* 1973) is one of the most important breakthroughs in the field of radiology. CT scanners are now widely used for diagnostic medical imaging and security screening.

Microcomputed tomography (micro-CT) is an important non-invasive tool for pre-clinical cancer study using small animals such as mice and rats (Paulus *et al.* 2000; Holdsworth *et al.* 2002; Ritman *et al.* 2004). Compared to clinical application, a much higher spatial and temporal resolution are needed in order to image the anatomy of small animals. Current commercial micro-CT scanners offer the capability of imaging objects *ex vivo* with high spatial resolution. Studies on lung and colon cancer models have shown that micro-CT is capable of sequentially tracking tumors larger than 2 mm in diameter with high fidelity. Small tumors are still barely distinguishable from the normal tissue. This is partially due to the motion-induced artifacts created by peristalsis, respiration and cardiac motion. The motion-induced artifacts in principle can be minimized by prospective gating (image acquisition is synchronized with the physiological signals), as commonly done in clinical CT scanners. However, performing such measurements on small animals is extremely challenging because their physiological motions can be ten times the rate of those of humans. In anatomical sites that can be greatly affected by respiratory or cardiac motion micro-CT imaging without gating can significantly deteriorate image quality.



**Fig. 19.8** Schematic of the CNT field emission micro-CT scanner.

Several new designs have been considered for high-resolution CT imaging of live animals. One is to use a high-power rotating-anode X-ray source instead of a fixed-anode microfocus X-ray tube (Badea *et al.* 2004). The high flux of the rotating-anode X-ray tube enables a short exposure time at the expense of the focal spot size. To achieve the high resolution, the object is placed as far away from the source and as close to the detector as the photon flux allows. Micro-CT scanners with dual source-detector pairs have also been developed by commercial vendors (SkyScan1178, SkyScan, Belgium and TomoScope, VAMP, Germany). The GE volumetric CT (VCT) uses two area detectors mounted on a rotating gantry, and a high-power rotating-anode X-ray tube (Kiessling *et al.* 2004).

Our group has developed a carbon nanotube (CNT)-based field emission microfocus X-ray source (Zhang *et al.* 2005a; Liu *et al.* 2006). The device utilizes a CNT film as the “cold” field emission cathode. The X-ray waveform can be readily programmed and synchronized with an external triggering signal, thus allowing for high temporal resolution and minimum delay (Cheng *et al.* 2004), both of which are desirable for gated imaging of live animals where the physiological motions are non-periodic. The microfocus X-ray source can be operated in the energy range of 40–60 kVp with an effective focal spot size as low as 30  $\mu\text{m}$  (Liu *et al.* 2006). At a given CNT cathode size, the X-ray focal spot size can be adjusted within a certain range with electronic focusing.

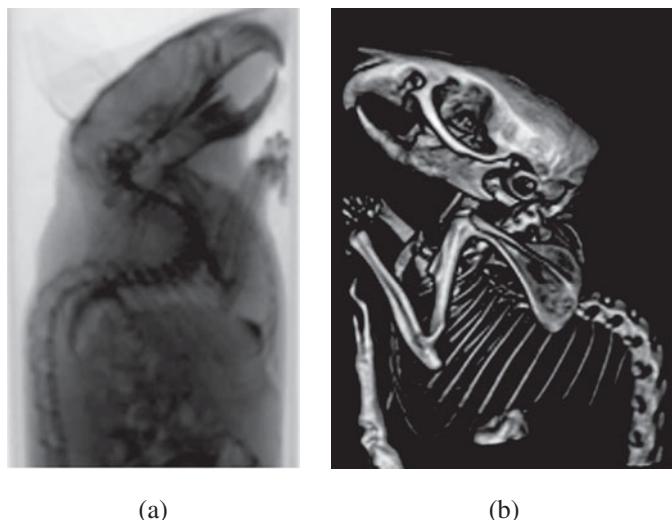
Figure 19.8 illustrates the CNT field emission micro-CT scanner developed in our group. This micro-CT scanner comprises a stationary CNT microfocus X-ray source, a motorized high-precision object rotation stage, an optical sensor for the detection of the object position, and an X-ray camera. The detector was externally triggered by a transistor-transistor logic (TTL) signal such that the frame speed was the same as the triggering signal. The entire operation of the micro-CT scanner, including X-ray exposure, rotation angle, and data collection, was controlled by a PC using a LabView-based program. A modified Feldkamp cone beam backprojection method (Feldkamp *et al.* 1984) written inhouse was used for reconstructing the tomographic images from the acquired projection images. The microfocus field emission X-ray source

has a triode structure, where the X-ray tube current ( $I_a$ ), which determines the X-ray flux, was controlled by the voltage applied on the gate electrode ( $V_g$ ) and the energy of the X-ray photon was set by the voltage applied on the anode ( $V_a$ ). The anode voltage ranges from 0 to 60 kV, which is only limited by the high-voltage power supply (Glassman EW series) used in our lab. An active focusing electrode, made of a metal cylinder, was placed above the gate electrode to focus the field-emitted electrons. The device operating in the reflection mode was housed in a vacuum chamber with a Be window under a  $10^{-7}$  Torr dynamic vacuum. In this prototype system, the emission material was purified SWNTs produced by the laser-ablation method at UNC. The cathode was a 1-mm diameter uniform layer of randomly oriented SWNTs coated on a metal substrate by electrophoresis deposition with fine control of the film thickness and morphology (Gao *et al.* 2001). A metal grid was used as the extraction electrode.

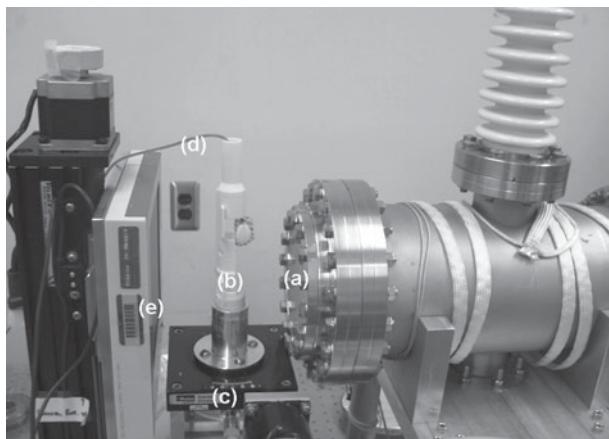
For tomographic imaging, the exact location of the X-ray beam center needs to be determined to reduce artifacts in the reconstructed images. This was accomplished by analyzing the distortion of the X-ray projection images of a phantom that contained several equally spaced identical thin metal disks stacked inside a plastic cylinder (Machin *et al.* 1994). Tomographic images of a normal 8-week-old mouse carcass (25 g) were obtained using this micro-CT scanner. To increase the contrast for the soft tissues, an iodinated contrast agent was injected into the abdomen (0.2 ml) of the mouse. A set of 600 projection images was taken over  $360^\circ$  at 1 s exposure per image. The X-ray source was operated at 40 kV,  $100\ \mu\text{A}$ ,  $150 \times 50\ \mu\text{m}$  effective focal spot size, and cone beam geometry. As shown in Fig. 19.9(a), the projection images demonstrated good bone delineation and soft-tissue contrast.

A 3D skeletal rendering of the carcass is shown in Fig. 19.9(b). Due to a much higher X-ray attenuation coefficient in calcified tissue than in soft tissue, the bony structures are well visualized.

Our group has recently improved and developed a respiratory-gated CNT-based 4D micro-CT system (Cao *et al.* 2008). Figures 19.10 and 19.11 show an



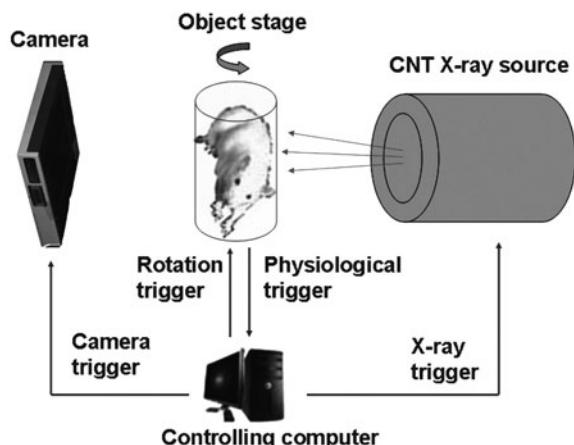
**Fig. 19.9** (a) A projection image demonstrating good bone delineation and soft-tissue contrast; and (b) 3D volume rendering of the whole-body skeletal dataset, the bony structures are well visualized.



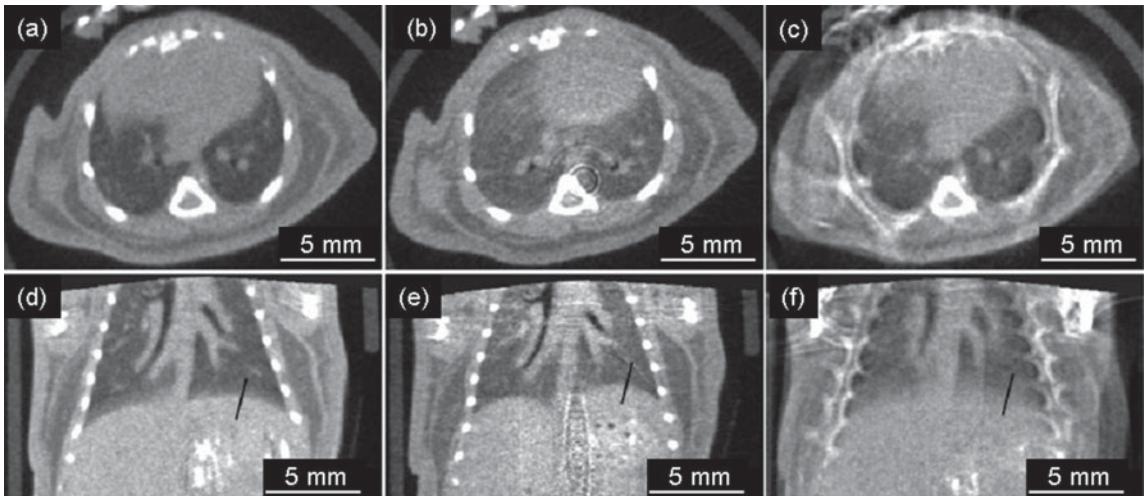
**Fig. 19.10** An image of the CNT-based 4D cone beam micro-CT system. (a) CNT X-ray source; (b) object holder; (c) rotation stage; (d) respiration sensor; and (e) X-ray camera.

image and a schematic of the physiologically gated micro-CT scanner based on the CNT microfocus X-ray source. In this design, the CNT X-ray source and the X-ray camera are stationary while the sample is rotated. The camera is placed as close to the X-ray source as possible to maximize the amount of X-ray flux. The animal is suspended vertically in a custom-built sample holder that is locked to a metal support attached to a computer-controlled rotation stage. The camera is a high-speed CMOS flat-panel sensor with a CsI scintillator plate conventionally coupled to a photodiode array (C7921CA-02, Hamamatsu). The top cover material is 1-mm thick aluminum. The camera has a  $5 \times 5$  cm active field of view with  $1032 \times 1032$  pixels at  $50 \times 50 \mu\text{m}$  pixel resolution. The camera allows flexible read-out and rebinning. Data were acquired without binning.

The source design is similar to that described in our previous publication (Liu *et al.* 2006). It consists of a CNT cathode, a tungsten gate mesh, an Einzel-type electrostatic focusing lens, and a molybdenum anode. The field-emitted



**Fig. 19.11** Schematic illustration of the CNT-based 4D cone beam micro-CT system.



**Fig. 19.12** Multiplanar reformatted images of a single mouse. Images were reconstructed from three consecutive scans of a single mouse using the same imaging protocol at  $67\text{ }\mu\text{m}$  isotropic voxel size. Images shown are from peak inspiration (a) and (d); end expiration (b) and (e); and ungated (c) and (f) in the axial and coronal views, respectively.

electrons were extracted from the CNT cathode by applying a voltage on the gate electrode, focused by the Einzel-type focusing lens, and accelerated to the molybdenum anode to generate X-rays. Electron trajectory simulation from our previous work (Liu *et al.* 2006) has shown that there is a linear relation between the CNT cathode diameter and the source focal diameter. The CNT cathode was capable of reliably delivering  $>1\text{ mA}$  peak cathode current. The gate mesh transmission rate was found to be 82%. The peak anode voltage was set at 40 kVp, and a  $200\text{-}\mu\text{m}$  thick Be window was used.

Micro-CT images of anesthetized free-breathing mice were collected and analyzed using this prototype system. Figure 19.12 shows the multiplanar reformatted images through the volumes produced with three consecutive scans on a single mouse for peak-inspiration-gated, end-expiration-gated, and nongated imaging. As expected, motion blur is visible in the nongated scan (Figs. 19.12(c) and (f)), which obscured the small structures such as vessels and higher-order branching airways. Compared to the nongated imaging, the airways and the diaphragm boundaries in the respiratory-gated imaging are defined substantially better (Cao *et al.* 2008).

The experimental microfocus X-ray source of our current micro-CT system is housed in a large vacuum chamber that prevents it from being rotated. Our group has since designed a smaller version of this source with comparable performance characteristics. This has allowed the construction of a micro-CT scanner with a stationary mouse bed and a rotating source and detector pair. The system is currently under testing at UNC.

The imaging speed can in principle be significantly increased by multiplexing. For the first time, multiplexing has been applied to X-ray radiography in our group (Zhang *et al.* 2006). Using this device, our group has demonstrated the feasibility of multiplexing radiography that enables simultaneous collection of multiple projection images using frequency multiplexing. A drastic

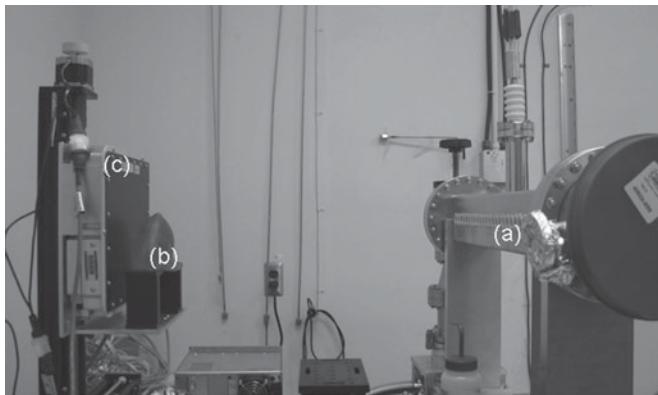
increase of the speed and reduction of the X-ray peak power are achieved without compromising the imaging quality.

### 19.3.3 Breast tomosynthesis

Tomosynthesis is an imaging technique that provides planar information of an object from a set of discrete X-ray projections (Grant *et al.* 1972). In this technique, image slices at arbitrary locations within the object are retrospectively reconstructed from a single scan. Tomosynthesis mammography has been investigated to solve the breast-tissue superimposition problem (Niklason *et al.* 1997; Suryanarayanan *et al.* 2000; Wu *et al.* 2003). Conventional two-dimensional (2D) mammography is currently the most effective screening tool for the early detection of breast cancer. However, there are about 30% of breast cancers missed by mammography, primarily due to overlapped normal breast tissue that obscures the cancer. Meanwhile, the superimposed normal tissues sometimes look like a tumor with ill-defined margins on a mammogram, which generates a large number of unnecessary callbacks. To solve these problems, researchers in the late 1990s developed a novel technique called X-ray digital breast tomosynthesis (DBT), which is a three-dimensional imaging technique that uses a series of projection images acquired at different angles to provide reconstruction planes in the breast. As a three-dimensional (3D) technique, DBT provides more accurate diagnosis by separating overlapping breast tissues.

All the prototype DBT systems that have been manufactured by commercial vendors including GE (Wu *et al.* 2003), Hologic (Ren *et al.* 2005) and Siemens (Bissonnette *et al.* 2005) utilize the standard mammography X-ray tube with a  $\sim 300\text{-}\mu\text{m}$  X-ray focal spot size. Due to the gantry rotation and mechanical instability, the effective focal spot size during image acquisition is larger than the static value, which degrades the image resolution. Two gantry rotation modes have been developed. The GE system (Wu *et al.* 2003) uses a step-and-shoot method. The gantry makes a full stop before taking each projection image. Acceleration/deceleration may cause mechanical instability of the system (Bissonnette *et al.* 2005). The extent of this effect on the system resolution has not been fully evaluated. A continuous rotation mode is used in the Siemens and Hologic systems (Ren *et al.* 2005). The gantry keeps a constant rotation speed during the whole imaging process. In this case, the X-ray focal spot size is enlarged along the motion direction. The value of the enlargement depends on the rotation speed, and the exposure time. Ren *et al.* (2005) reported the X-ray focal spot moves  $\sim 1\text{ mm}$  in a typical scan. The effect of the focal spot blurring on the system, the modulation transfer function (MTF) was found to be insignificant at this speed, especially for thin specimens (Ren *et al.* 2005). However, it does not leave any room for further reduction of the total scanning time, which will require a faster gantry rotation and a larger focal spot blurring.

Our group has investigated the feasibility of constructing a stationary DBT system that can increase the scanning speed, simplify the system design and potentially enhance the image quality (Yang *et al.* 2008) by using the CNT-

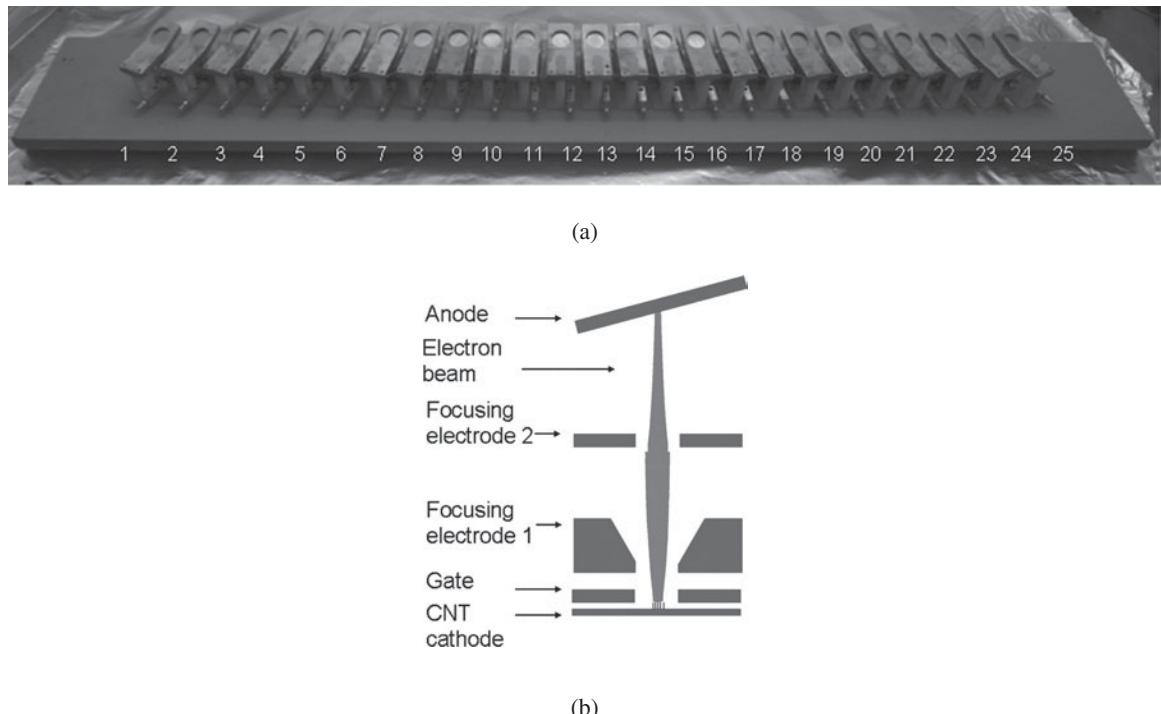


**Fig. 19.13** Image of the Argus system. (a) A 25-pixel X-ray source array; (b) a breast phantom; and (c) a flat-panel detector for full-field mammography (a control unit for X-ray sources and a computer work station are not shown in this image).

based multibeam field emission X-ray (MBFEX) developed in our lab (Zhang *et al.* 2005b, 2007). The stationary DBT system, called Argus, uses a spatially distributed MBFEX source and acquires the projection images without mechanical motion of the sources or the detector. This results in reducing the total imaging time and potentially improving the image quality.

Figure 19.13 shows an image of the assembled Argus system. It is composed of a 25-pixel X-ray source array, a flat-panel detector for full-field mammography, a control unit for X-ray sources, and a computer workstation. The system geometry follows the typical values for the regular mammography devices. The distance between the center of the phantom and the X-ray source is 64.5 cm. The source to detector distance is 69.6 cm, which leaves a 2.5 cm air gap for a normal 5-cm breast phantom. The X-ray pixels are arranged linearly to reduce the system complexity, with even-angular distribution and a 2-degree increment. The total angular coverage of the X-ray source is 48 degrees. In such a design, the distance between the nearest X-ray focal spots varies from 2.5 cm to 2.7 cm, and the total span of the X-ray source array is 57.5 cm.

The key component of the Argus system is the field emission multipixel X-ray source array. Figure 19.14(a) shows an assembled MBFEX source array with 25 individually controllable X-ray pixels, which are tilted towards the isocenter. The central beams of all the 25 X-ray sources intersect at a common isocenter, which is also the center of the phantom. Each X-ray pixel consists of a CNT cathode, a gate electrode, two focusing electrodes, and a molybdenum (Mo) anode, as schematically illustrated in Fig. 19.14(b). The cathode is grounded, and all other electrodes are maintained at constant voltages during imaging acquisition. The gate voltage determines the X-ray tube current. The voltages applied to the two focusing electrodes control the electron trajectory. The X-ray source array is housed in a customized chamber. A 30- $\mu\text{m}$  Mo film serves as the X-ray window as well as the filter. The molybdenum anode and filter combination is commonly used for breast imaging. Up to 30 kV high voltage can be applied to the anode without arcing, which is the energy that would be useful in mammography. The targeted performance for the source design is that each X-ray pixel can provide 10 mA peak current at 200  $\mu\text{m} \times 200 \mu\text{m}$  effective focal spot size.

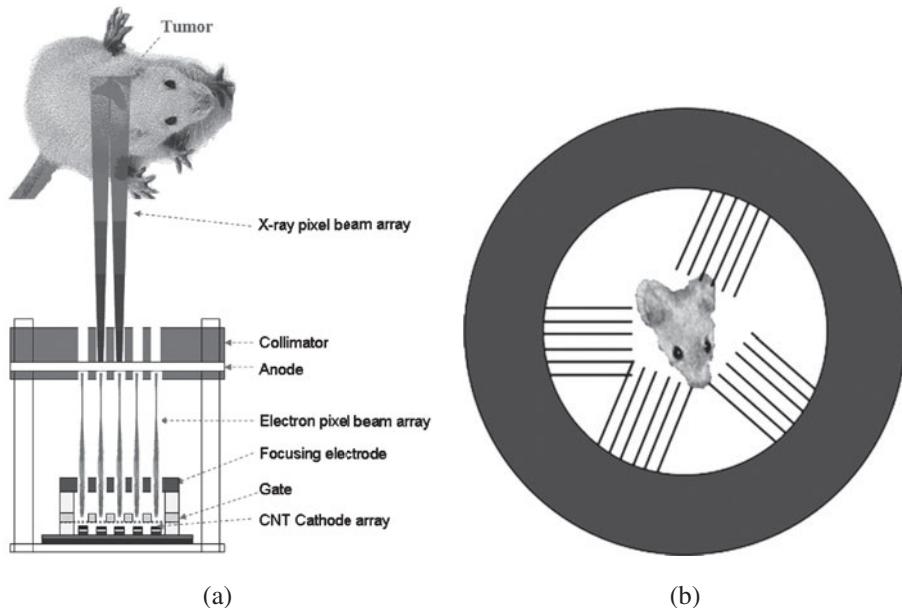


**Fig. 19.14** (a) An assembled MBFEX source array with 25 individually controllable X-ray pixels that are tilted towards the isocenter; and (b) schematic illustration showing that an individual pixel consists of a CNT cathode, a gate, two focusing electrodes, and a Mo anode.

The preliminary test showed that the system functions work as expected. Detailed system calibration and tomosynthesis imaging test are ongoing.

### 19.3.4 Microradiotherapy system

Radiation therapy is one of the basis treatments in cancer management as half of all cancer patients in the US today rely on radiotherapy for cancer local control. In the last decade, significant technological advances have been made in the field of radiotherapy and as a result, more patients receive state-of-the-art radiotherapy in which the radiation dose distribution is highly conformal to the tumor and normal tissue around the tumor is largely spared (Chang *et al.* 2002; Garden *et al.* 2004; Sajja *et al.* 2005). However, available small-animal irradiation technology for cancer research today lacks many of the capabilities found in clinical practice. Most small-animal cancer research today still relies on either irradiation devices designed for humans or simple animal irradiation devices, all of which have very poor spatial resolution and no temporal resolution. Recently, advances in both genetically engineered mouse model and cancer biomarkers promise better pre-clinical studies to understand human cancers and the efficacy of cancer biomarkers (Weiss *et al.* 2002; Becher *et al.* 2006; Whitfield *et al.* 2006). These and other advances in cancer research demand new research tools with high spatial and temporal resolution for pre-



**Fig. 19.15** Schematic illustration of the proposed (a) single array micro-RT system; and (b) multipixel array micro-RT system.

clinical study before proceeding to clinical trials. For small-animal irradiation, there is a need for high-resolution and image-guided microradiotherapy (micro-RT) technology that can deliver radiation treatment similar to human radiotherapy but at mouse scales.

Micro-RT is specifically designed for small-animal irradiation and used for basic and translational cancer research. Currently, there are several prototype micro-RT systems under development in different research labs using the Ir-192 isotope (Stojadinovic *et al.* 2006) or conventional X-ray (Deng *et al.* 2007) as the radiation source. These micro-RT systems have a single radiation beam and rely on mechanical collimation to define the radiation-field shape. These systems have successfully produced small (1 mm or less) single beams. However, the irregular field shaping is not easily achievable in these micro-RT designs that rely on mechanical field-shaping devices. Gated irradiation is also a challenge for the micro-RT systems as mouse respiratory motion is twice as fast as the human heart beat. Taking full advantage of the CNT field emission technology, our group has proposed a novel multipixel X-ray beam array micro-RT system that is based on the CNT field emission technology. The radiation field is formed by electronically “turning on” a user-specified set of the X-ray pixel beams of the micro-RT system. The CNT field emission technology enables the multipixel X-ray beam design and the high spatial and temporal resolution of the system. Unlike imaging devices, radiotherapy device design must be dosimetrically driven and require very different considerations in many aspects of system design and fabrication.

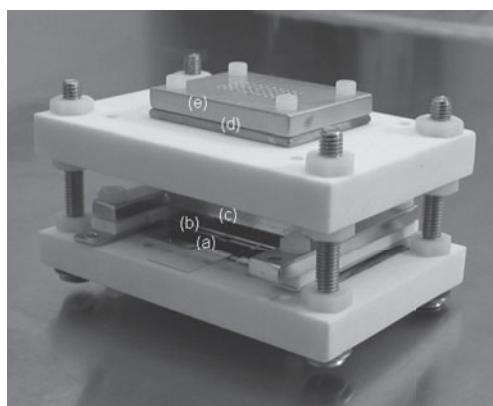
Figure 19.15(a) illustrates a single array of the proposed multipixel X-ray micro-RT system. After the electron pixel beams are generated from the CNT

pixels, they are accelerated towards the target under 80–100 kV voltage and produce X-rays in the target. The X-rays are then collimated into individual X-ray pixel beams. Figure 19.16 shows an image of the prototype of the micro-RT system. Each of the X-ray pixel beams projects a 2 mm × 2 mm radiation field at the center of the treatment target (mouse). The eventual micro-RT system will have a large number of arrays arranged in a donut-shaped structure surrounding the treatment target, as schematically illustrated in Fig. 19.15(b).

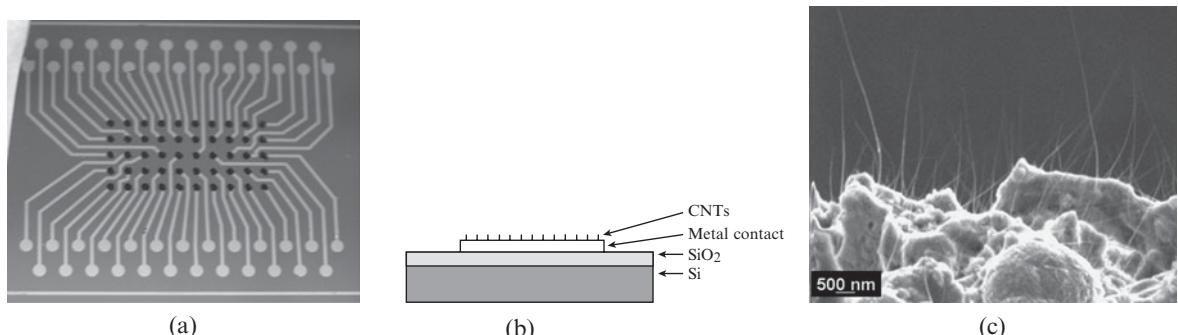
One critical component of the micro-RT system is the multipixel cathode array chip (Wang *et al.* 2008). Figure 19.17(a) shows a completed cathode chip with a 5 × 10 multipixel cathode array. It was fabricated using a combined photolithography and electrophoresis deposition (EPD) method developed in our lab (Gao *et al.* 2001; Oh *et al.* 2004). Figure 19.17(b) presents the schematic diagram for the cathode structure of an individual pixel in the chip. In Fig. 19.17(a), the black dots at the center of the chip are CNT cathodes with a diameter of 1 mm. The center-to-center distance of the CNT cathodes is 2 mm. The bright lines and dots are the electrical connections to the external circuit for individual control. Figure 19.17(c) shows a typical scanning electron microscopy (SEM) image from the fabricated 5 × 10 multipixel electron cathode array chip. It can be seen that the CNTs are roughly vertically deposited on the substrate with a medium CNT density of  $10^7$  emitters/cm<sup>2</sup>. The distribution of CNT height is roughly uniform.

Figure 19.18 shows a typical emission image of a 5 × 10 cathode array chip at an applied voltage of 1400 V. It can be seen that 50 electron beams are roughly uniform. The cathode chip determines the uniformity of the generated X-ray beams, which is critical for delivering predictable and consistent radiation doses to tumors.

The key advantages of the proposed carbon nanotube micro-RT over other micro-RT systems are its ability to electronically, not mechanically, form arbitrary field shapes and intensity modulation pattern for very small radiation fields. Figure 19.19(a) shows that out of 50 pixel beams, only 6 irradiation beams are electronically “turning on”. Different pixel beams can deliver differ-



**Fig. 19.16** An image of the assembled single array micro-RT prototype. (a) Cathode chip; (b) gate electrode; (c) focusing electrode; (d) anode (tungsten foil sitting underneath the collimator); and (e) collimator.

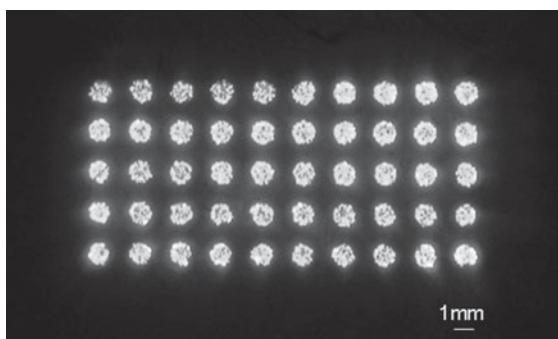


**Fig. 19.17** (a) Image of a completed  $5 \times 10$  pixel cathode array; (b) schematic illustration of the structure (cross-section) of an individual cathode in the cathode chip; and (c) a typical SEM image of CNTs in a cathode array.

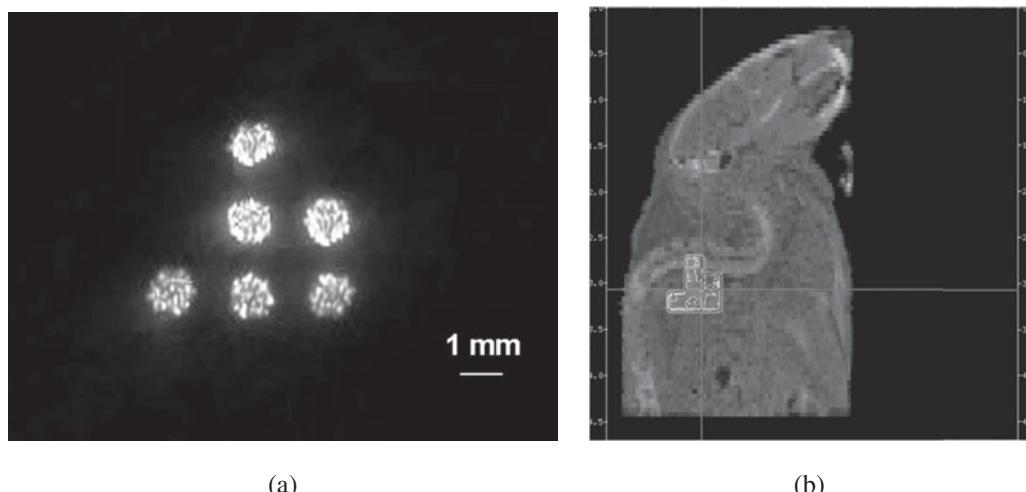
ent doses and dose rates by individual field emission control. Figure 19.19(b) illustrates a mouse image generated by our CNT field emission micro-CT and the Monte Carlo simulation generated dose distribution from an irregular field formed by a 6-pixel beam of the micro-RT system.

The emission current from each pixel is another important issue for micro-RT application, as it is proportional to radiation dose rate. The emission current of a single-pixel CNT cathode in the  $5 \times 10$  cathode array chip was examined. Figure 19.20 shows the typical emission current ( $I$ ) versus applied voltage ( $V$ ) of an individual pixel cathode in the cathode chip. The emission current was 3.7 mA at a gate voltage of 2000 V. The corresponding Fowler–Nordheim (Fowler *et al.* 1928) plot of the emitted current vs. applied voltage ( $\ln (I/V^2)$  vs.  $I/V$ ) from our CNT cathode is roughly linear (see Fig 19.20, inset), indicating the electron current is governed by field emission. The field emission based electron (X-ray) source can turn on/off instantly, resulting in high temporal resolution ( $\sim$ ms level). This ability is another key advantage of our proposed CNT field emission micro-RT device.

Extensive Monte Carlo (MC) simulation of the micro-RT dosimetry in the system design process was used as a guide for design validation and the specification of fabrication parameters has been conducted. MC simulations track the path of individual radiation particles as they travel through each component of the micro-RT system and deposit a radiation dose in the target based on basic physics principles, and are thus ideal for this kind of prototype



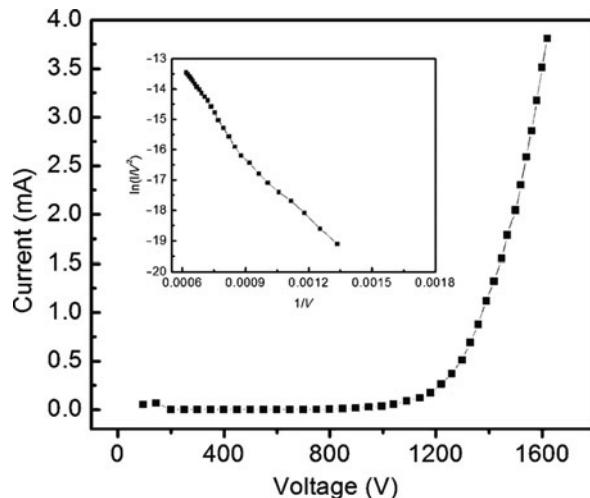
**Fig. 19.18** A  $5 \times 10$  irradiation beam array produced from a carbon nanotube cathode array chip. The irradiation field can be up to  $10 \times 20$  mm.



**Fig. 19.19** (a) Illustration of individual pixel beam field emission control for the proposed micro-RT—six out of fifty pixel cathodes were turned on; and (b) a mouse image from our carbon-nanotube-based field emission micro-CT and the Monte Carlo simulated dosimetry from a 6-pixel beam formed irregular radiation field.

simulation. The MC simulations indicated that for an emission current of 3.7 mA per pixel beam the X-ray pixel beam is expected to generate a dose rate of >1.5 Gy/min at the center of a mouse under the micro-RT irradiation. Although the actual dose rate of the multipixel array micro-RT system will depend on multiple factors yet to be determined at a later stage of the development, the MC results indicate that the micro-RT system will be capable of producing acceptable dose rates.

The micro-RT system is currently still under development. Once developed, it will be integrated with the CNT field emission micro-CT system already



**Fig. 19.20**  $I$ - $V$  curve from one individual cathode in a  $5 \times 10$  pixel CNT cathode array chip. Inset: the corresponding Fowler–Nordheim plot.

developed by our group to form a high-resolution CT image-guided and intensity-modulated irradiation system that a mouse-scale analog to state-of-the-art clinical image-guided radiotherapy systems.

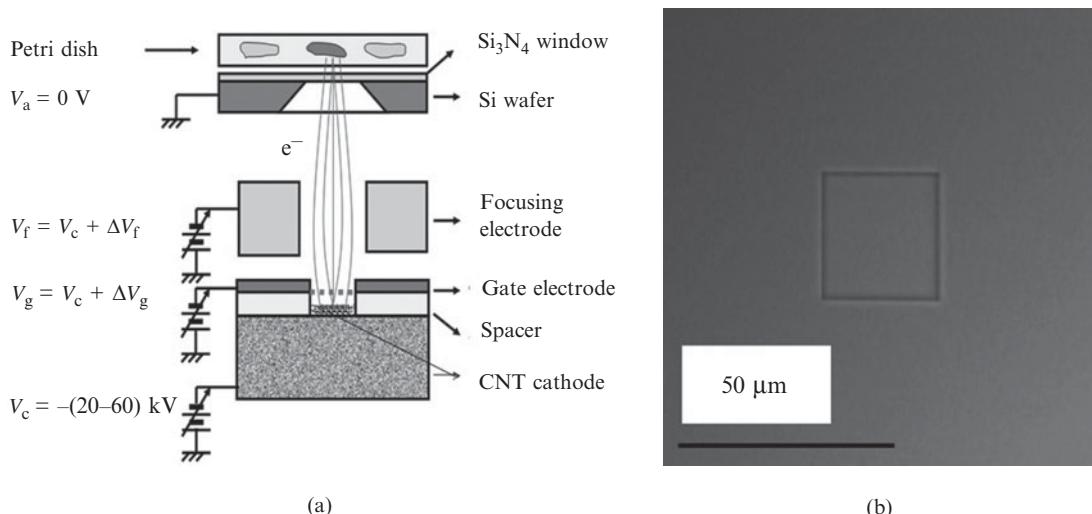
### 19.3.5 Single-cell irradiation system

Recently, there has been increased interest in single-cell irradiation (SCI) devices that use low linear energy transfer (LET) microbeam radiation and are available to research laboratories with average resources (Sowa Resat *et al.* 2004; Sowa *et al.* 2005). SCI systems that can deliver a radiation beam precisely to the targeted cells for “bystander effect” studies, which are vital to improve our understanding of low-radiation dose and dose-rate consequences to our health (Sowa Resat *et al.* 2004). Together with new technological advances in biosensors and cellular imaging, SCI of therapeutic dosage under real-time microscope observation can also play an important role in understanding critical signalling events for both short- and long-term radiation effects occurring immediately after irradiation. At present, technical limitations often prevent researchers from conducting research productively at the very small scales of time and space of these important microscopic events. A suitable research tool would enable better discrimination of the complex processes that govern the ultimate outcome.

We proposed to use CNT field emission technology to develop a novel SCI system that can simultaneously deliver a large number of low-LET electron microbeams to individually selected cells in a petri dish on the timescale of microseconds to minutes (Chang *et al.* 2006). This new technology can be helpful to elucidate the complex molecular events that form the basis of the cellular response to ionizing radiation and other genotoxic stressors. Our group is developing a CNT field emission cellular microbeam system.

Figure 19.21(a) illustrates the single-pixel CNT microbeam system. The electron current is controlled by the gate voltage  $V_g$  and the focusing electrode is used both for isolating the gate structure from the high anode voltage and the first-stage electron-beam focusing. A 400-mm thick Si wafer with a 200-nm thick layer of  $\text{Si}_3\text{N}_4$  is used to fabricate the vacuum window and the electron microbeam beam portal. The thick silicon wafer acts as an electron collimator. The  $\text{Si}_3\text{N}_4$  window is fabricated using a photolithography technique that carefully etches away the silicon in the intended location and forms the  $\text{Si}_3\text{N}_4$  window. The etching chemical selectively etches certain crystal planes and forms the inverted pyramidal opening. Figure 19.21(b) shows one of the  $\text{Si}_3\text{N}_4$  windows fabricated in our lab. A prototype single-pixel SCI is assembled in a large vacuum chamber without the vacuum enclosure as in the finished product. A microbeam with 2 mA peak current was obtained at 30 kV and 1% duty cycle. The dosimetry from a 25-s irradiation is measured by GAFCHROMIC film (HD-810) placed on top of the  $\text{Si}_3\text{N}_4$  window.

Figures 19.22(a) and (b) show the dose profile of the 30-keV electron beam measured using a GAFCHROMIC film. The full width at half-maximum (FWHM) is 28  $\mu\text{m}$ , approximately the size of a large cell. Figure 19.22(c) shows the initial cell irradiation result of the first-generation prototype single-



**Fig. 19.21** (a) Schematic of the single-pixel CNT microbeam system; and (b) an image of the  $\text{Si}_3\text{N}_4$  window.

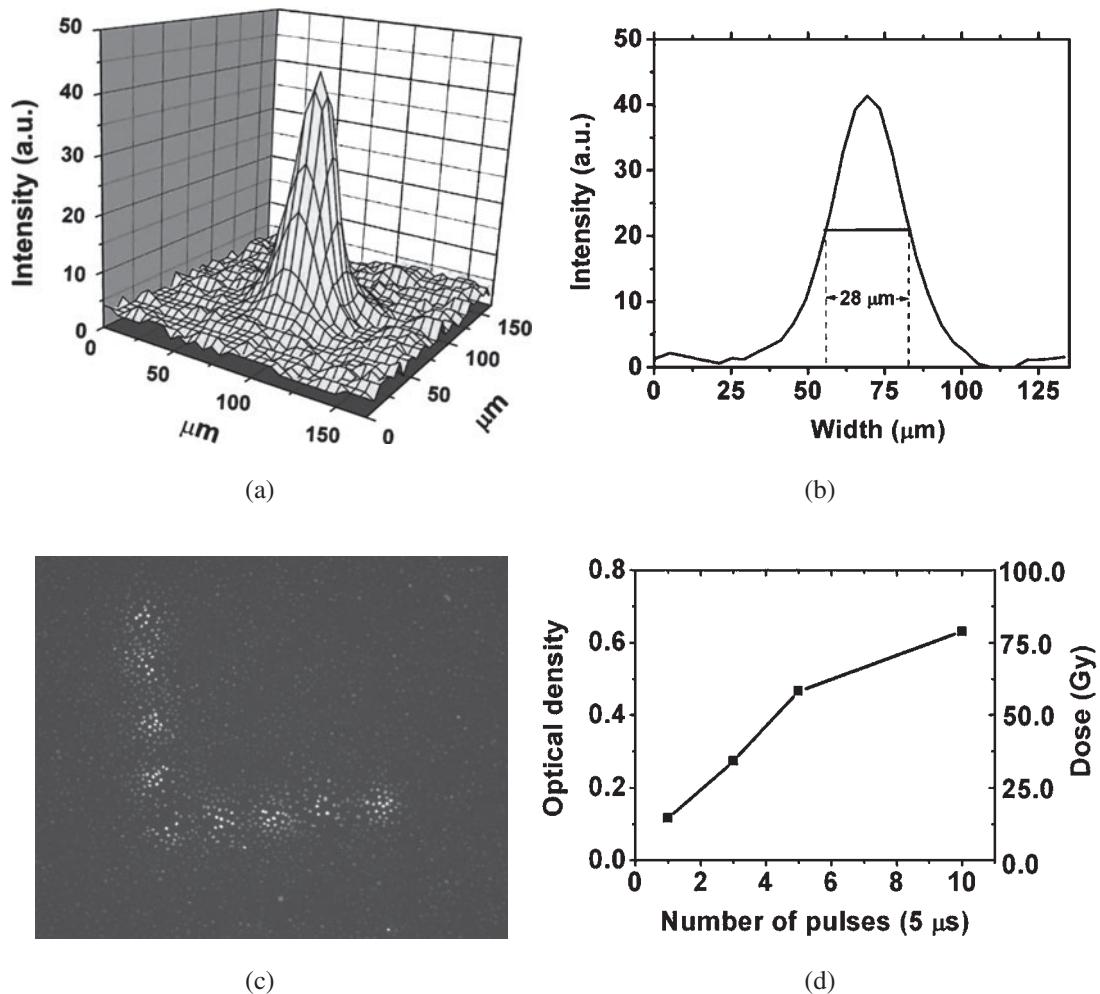
pixel system. Nine radiation sites in the form of an ‘‘L’’ were used for easy identification. Rat-1 cells are used in the experiment and H2AX (bright dots) verified the radiation pattern by regions where double-stranded DNA breaks occurred. The dose per microbeam pulse (5 s) calibration is shown in Fig. 19.22(d). The average dose rate in this case is  $1000 \text{ Gy s}^{-1}$ . By varying the emission current and pulse width, and frequency, the dose rate can be adjusted within a very large dynamic range ( $1 \text{ cGy s}^{-1} - 100 \text{ Gy s}^{-1}$ ).

Several approaches to reduce the microbeam size have been proposed including the use of a different silicon etching technique, adding an additional microbeam collimator, and using a much smaller  $\text{Si}_3\text{N}_4$  window.

## 19.4 Summary and conclusion

In this chapter, we have reviewed the carbon-nanotube field emission principle and current technology development including the development and feasibility demonstration of a CNT field emission microcomputed tomography (micro-CT) scanner, stationary digital breast tomosynthesis system, microradiotherapy (micro-RT) system and a single-cell irradiation system. These systems show great promise in commercial and research application for the future in (a) diagnostic medical imaging; (b) *in-vivo* imaging of small-animal models for pre-clinical cancer studies; (c) security screening; (d) industrial inspection; (e) cancer radiotherapy of small-animal models for pre-clinical cancer studies; and (f) basic cancer research using single-cell irradiation.

Field emission cathode-based radiation devices can be turned on and off at several kHz rates and the cathode can take the form of a point, line, plane, or array of many pixels of high spatial and, more importantly, high temporal resolution. The freedom in the source configuration enables system design with



**Fig. 19.22** (a), (b) Single-pixel SCI electron-beam profiles measured by GAFCHRIC film; (c) the 1st cellular (Rat-1 cell line) irradiation using an “L” pattern irradiation pattern. H2AX (bright) imaging verified the double-strand DNA break sites that are consistent with the irradiation pattern; and (d) the absolute dose per pulse of the prototype CNT irradiation system. The CNT field emission irradiation system is capable of a large range of dose rate.

enhanced performance and novel features. For instance, the technology enables the design of rotation-free stationary tomography imaging systems that can set a new standard for scanning speed in the industry. As carbon-nanotube field emission based technology matures over the next decade we hope to see new research advancement to understand cancer and better cancer detection and treatment.

## References

- Badea, C., Hedlund, L.W., Johnson, G.A. *Med. Phys.* **31**, 3324 (2004).
- Baptist, R. *US Patent* **6**, 259, 765 (2001).
- Becher, O.J., Holland, E.C. *Cancer Res.* **66**, 3355 (2006).
- Bissonnette, M., Hansroul, M., Masson, E., Savard, S., Cadieux, S., Warmoes, P., Gravel, D., Agopyan, J., Polischuk, B., Haerer, W., Mertelmeier, T., Lo, J.Y., Chen, Y., Dobbins III, J.T., Jesneck, J.L., Singh, S. *SPIE* **5745**, 529 (2005).
- Brodie, I., Spindt, C.A. *Adv. Electron. Electron. Phys.* **83**, 1 (1992).
- Bushong, S.C. *Radiologic Science for Technologist* (Mosby, 1997).
- Cao, G., Lee, Y.Z., Liu, Z., Rajaram, R., Peng, R., Calderon-Colon, X., An, L., Wang, P., Phan, P., Lalush, D., Lu, J.P., Zhou, O. *SPIE* **6913**, 691304 (2008).
- Chang, S., Zhang, J., Bordelon, D., Schreiber, E., Cox, A., Zhou, O. *Radiat. Protect. Dos.* **122**, 323 (2006).
- Chang, S.X., Cullip, T.J., Rosenman, J.G., Halvorsen, P.H., Tepper, J.E. *Med. Phys.* **29**, 1130 (2002).
- Charbonnier, F.M., Barbour, J.P., Dyke, W.P. *Radiology* **117**, 165 (1975).
- Cheng, Y., Zhang, J., Lee, Y.Z., Gao, B., Dike, S., Lin, W., Lu, J.P., Zhou, O. *Rev. Sci. Instrum.* **75**, 3264 (2004).
- Chernozatonskii, L.A., Gulyaev, Y.V., Kosakovskaja, Z.J., Sinitsyn, N.I., Torgashov, G.V., Zakharchenko, Y.F., Fedorov, E.A., Valchuk, V.P. *Chem. Phys. Lett.* **233**, 63 (1995).
- Coolidge, W.D. *Am. J. Roentgenol.* **24**, 605 (1930).
- Coolidge, W.D. *Am. J. Roentgenol. Radium Therapy Nucl. Med.* **75**, 166 (1956).
- Cormack, A.M. *Phys. Med. Biol.* **18**, 195 (1973).
- Crooks, H.E., Sangster, J., Ardran, G.M. *Radiology* **456**, 311 (1972).
- de Heer, W.A., Chatelain, A., Ugarte, D. *Science* **270**, 1179 (1995).
- Deng, H., Kennedy, C.W., Armour, E., Tryggestad, E., Ford, E., McNutt, T., Jiang, L., Wong, J. *Phys. Med. Biol.* **52**, 2729 (2007).
- Feldkamp, L.A., Davis, L.C., Kress, J.W. *J. Opt. Soc. Am. A* **1**, 612 (1984).
- Fowler, R.H., Nordheim, L.W. *Proc. Roy. Soc. London Ser. A* **119**, 173 (1928).
- Gao, B., Yue, G.Z., Qiu, Q., Cheng, Y., Shimoda, H., Fleming, L., Zhou, O. *Adv. Mater.* **13**, 1770 (2001).
- Garden, A.S., Morrison, W.H., Rosenthal, D.I., Chao, K.S.C., Ang, K.K. *Semin. Radiat. Oncol.* **14**, 103 (2004).
- Gorman, J. *Science News*, July 6 (2002).
- Gomer, R. *Field Emission and Field Ionization* (Harvard University Press, Cambridge, MA, 1961).
- Grant, D.G. *IEEE Trans. Biomed. Eng.* **19**, 20 (1972).
- Hallenbeck, G.S. *Radiology* **117**, 1 (1975).
- Holdsworth, D.W., Thornton, M.M. *Trends Biotech.* **20**, S34 (2002).
- Hounsfield, G.N. *Br. J. Radiol.* **46**, 1016 (1973).
- Iijima, S. *Nature* **354**, 56 (1991).
- Iijima, S., Ichihashi, T. *Nature* **363**, 603 (1993).

- Jung, I.S., Seonghoon, L., Yoon, H.S., Sung, Y.C., Kyoung, I., Kee, S.N. *Appl. Phys. Lett.* **78**, 901 (2001).
- Khan, F.M. *The Physics of Radiation Therapy* (Williams & Wilkins, Baltimore, MD, USA, 1984).
- Kiessling, F., Greschus, S., Lichy, M.P., Bock, M., Fink, C., Vosseler, S., Moll, J., Mueller, M.M., Fusenig, N.E., Traupe, H., Semmler, W. *Nature Med.* **10**, 1133 (2004).
- Lalush, D.S., Quan, E., Rajaram, R., Zhang, J., Lu, J.P., Zhou, O. *3rd IEEE Inter. Symp. Biomed. Imag.* **6–9**, 1180 (2006).
- Liu, Y., Sowerby, B.D., Tickner, J.R. *Appl. Radiat. Isotopes* **66**, 463 (2008).
- Liu, Z., Yang, G., Lee, Y., Bordelon, D., Lu, J., Zhou, O. *Appl. Phys. Lett.* **89**, 103111 (2006).
- Machin, K., Webb, S. *Phys. Med. Biol.* **39**, 1639 (1994).
- Margaritondo, G. *Introduction to Synchrotron Radiation* (Oxford University Press, 1988).
- Niklason, L.T., Christian, B.T., Niklason, L.E., Kopans, D.B., Castleberry, D.E., Opsahl-Ong, B.H., Landberg, C.E., Slanetz, P.J., Giardino, A.A., Moore, R., Albagli, D., DeJule, M.C., Fitzgerald, P.F., Fobare, D.F., Giambattista, B.W., Kwasnick, R.F., Liu, J.Q., Lubowski, S.J., Possin, G.E., Richotte, J.F., Wei, C.Y., Wirth, R.F. *Radiology* **205**, 399 (1997).
- Oh, S.J., Zhang, J., Cheng, Y., Shimoda, H., Zhou, O. *Appl. Phys. Lett.* **84**, 3738 (2004).
- Paulus, M.J., Gleason, S.S., Kennel, S.J., Hunsicker, P.R., Johnson, D.K. *Neoplasia* **2**, 62 (2000).
- Purcell, S.T., Vincent, P., Journet, C., Binh, V.T. *Phys. Rev. Lett.* **88**, 105502 (2002).
- Rangsten, P., Ribbing, C., Strandman, C., Hok, B., Smith, L. *Sens. Actuators* **82**, 24 (2000).
- Ren, B., Ruth, C., Stein, J., Smith, A., Shaw, I., Jing, Z. *SPIE* **5745**, 550 (2005).
- Rinzler, A.G., Hafner, J.H., Nikolaev, P., Lou, L., Kim, S.G., Tomanek, D., Nordlander, P., Colbert, D.T., Smalley, R.E. *Science* **269**, 1550 (1995).
- Ritman, E.L. *Annu. Rev. Biomed. Eng.* **6**, 185 (2004).
- Saito, Y., Uemura, S. *Carbon* **38**, 169 (2000).
- Sajja, R., Barnett, G.H., Lee, S.Y., Harnisch, G., Stevens, G.H.J., Lee, J., Suh, J.H. *Technol Cancer Res. Treat.* **4**(6), 675 (2005).
- Sowa, M., Murphy, M., Miller, J., McDonald, J., Strom, D., Kimme, G.A. *Radiat. Res.* **164**, 695 (2005).
- Sowa Resat, M., Morgan, W.F. *Cancer Metastasis Rev.* **23**, 323 (2004).
- Stojadinovic, S., Low, D.A., Vicic, M., Mutic, S., Deasy, J.O., Hope, A.J., Parikh, P.J., Grigsby, P.W. *Med. Phys.* **33**, 3834 (2006).
- Sugie, H., Tanemure, M., Filip, V., Iwata, K., Takahashi, K., Okuyama, F. *Appl. Phys. Lett.* **78**, 2578 (2001).
- Suryanarayanan, S., Karellas, A., Vedantham, S., Glick, S.J., D'Orsi, C.J., Baker, S.P., Webber, R.L. *Acad. Radiol.* **7**, 1085 (2000).
- Thayer, D.W. *New England J. Med.* **350**, 1811 (2004).
- Wang, Q.H., Setlur, A.A., Lauerhaas, J.M., Dai, J.Y., Seelig, E.W., Chang, R.-P.H. *Appl. Phys. Lett.* **72**, 2912 (1998).

- Wang, S.G., Liu, Z., Sultana, S., Schreiber, E., Zhou, O., Chang, S. *BioFactors* **30**, 265 (2007).
- Wang, S.G., Liu, Z., An, L., Zhou, O., Chang, S. *Mater. Res. Soc. Symp. Proc.* **1065**, 1065-QQ04-08 (2008).
- Wei, Y., Xie, C.G., Dean, K.A., Coll, B.F. *Appl. Phys. Lett.* **79**, 4527 (2001).
- Weiss, W.A., Chesler, L. *Drug Discovery Today* **7**, 997 (2002).
- Whitfield, M.L., George, L.K., Grant, G.D., Perou, C.M. *Nature Rev. Cancer* **6**, 99 (2006).
- Whitlock, R.R., Bell, M.I., Kerns, D.V., Kerns, S., Davidson, J.L., Kang, W.P. *US Patent* **6**, 333, 968 (2001).
- Whitlock, R.R. *US Patent* **6**, 333, 968 (2001).
- Wong, W.K., Ng, S.H., Xu, K. *Qual. Reliab. Eng. Int.* **22**, 321 (2006).
- Wu, T., Stewart, A., Stanton, M., McCauley, T., Phillips, W., Kopans, D.B., Moore, R.H., Eberhard, J.W., Opsahl-Ong, B., Niklason, L., Williams, M.B. *Med. Phys.* **30**, 365 (2003).
- Yang, G., Rajaram, R., Cao, G., Sultana, S., Lalush, D., Lu, J.P., Zhou, O. *SPIE* **6913**, 69131A (2008).
- Yue, G.Z., Qiu, Q., Gao, B., Cheng, Y., Zhang, J., Shimoda, H., Chang, S., Lu, J.P., Zhou, O. *Appl. Phys. Lett.* **81**, 355 (2002).
- Zhang, J., Tang, J., Yang, G., Qiu, Q., Qin, L.-C., Zhou, O. *Adv. Mater.* **16**, 1219 (2004).
- Zhang, J., Cheng, Y., Lee, Y.Z., Gao, B., Qiu, Q., Lin, W.L., Lalush, D., Lu, J.P., Zhou, O. *Rev. Sci. Instrum.* **76**, 094301 (2005a).
- Zhang, J., Yang, G., Cheng, Y., Gao, B., Qiu, Q., Lee, Y.Z., Lu, J.P., Zhou, O. *Appl. Phys. Lett.* **86**, 184104 (2005b).
- Zhang, J., Yang, G., Lee, Y.Z., Chang, S., Lu, J.P., Zhou, O. *Appl. Phys. Lett.* **89**, 064106 (2006).
- Zhang, J., Yang, Lee, Y.Z., Chang, S., Lu, J.P., Zhou, O. *SPIE* **6510**, 65100W (2007).
- Zhou, O., Lu, J.P. *US Patent* **6**, 553, 096 (2003).
- Zhou, O., Lu, J.P. *US Patent Application*, S/N 09/679, 303 (2000).

# Theory of hydrogen storage in nanoscale materials

*Yufeng Zhao, Yong-Hyun Kim, S.B. Zhang, and Michael J. Heben*

# 20

20.1 Introduction	699
20.2 Basic considerations	701
20.3 Hydrogen–material interaction	705
20.4 Internal interaction in HSMs	714
20.5 Structures of hydrogen sorbents	722
20.6 Required hydrogen-storage properties and design principles (DP)	725
20.7 Summary	731
Acknowledgments	732
References	732

## 20.1 Introduction

Hydrogen is viewed as a clean energy alternative that could one day replace fossil fuels. The so-called hydrogen economy (Turner *et al.* 1999; Schultz *et al.* 2003; Tromp *et al.* 2003; Crabtree *et al.* 2004; Grochala and Edwards 2004; Jacobson *et al.* 2005) is envisioned as a potential technological progress that has the societal impact comparable to the major industrial revolutions. There are three key techniques for the implementation of the hydrogen economy: the first is use of solar energy to split water for hydrogen production, the second is hydrogen storage, and the third is the fuel-cell energy converter. Among them, hydrogen storage is commonly known to be the bottleneck. The U.S. Department of Energy has determined that a system capacity of hydrogen of 6–9 wt% will be required for fuel-cell-powered vehicles to be able to replace petroleum-fuelled vehicles on a large scale (Satyapal *et al.* 2007). However, without a suitable matrix in the hydrogen-storage system, it is extremely difficult to compress this highly flammable gas at moderate pressures required for feasible on-board vehicular storage.

For decades, researchers have been struggling with metal hydrides, chemical hydrides and carbon sorbents. Hydride solids store hydrogen in atomic form and carbon-based high surface area materials adsorb molecular hydrogen. Distinctly, the two types of main stream hydrogen-storage materials (HSM) have advantages and disadvantages for hydrogen storage. Hydride materials can generally have high gravimetric and volumetric capacity but the kinetics of hydrogen charge and/or discharge is too slow because the hydrogen atoms interact strongly with the HSM and form bulk solids with the latter. On the contrary, in hydrogen sorbents both hydrogen charge and discharge are fast because the hydrogen molecules are adsorbed on the surface of the pores open to free space; however, the capacity is too low at near ambient temperature, due to the weak interaction between the sorbent and hydrogen molecules.

Although promise has been demonstrated by researchers (Dillon *et al.* 1997; Chambers *et al.* 1998; Chen *et al.* 1999, 2002; Liu *et al.* 1999,

practical hydrogen storage for vehicular applications is still far from being achieved. Significant technological progresses call for a conceptual breakthrough originated from a systematic theory of hydrogen storage based on fundamental principle of physics, which has not yet been demonstrated. In this chapter, we will focus on a conceptual proposal of unconventional hydrogen sorbents with high internal surface area yet interact with hydrogen (in both atomic and molecular form) in the right binding strength, in order to combine the advantages of solid hydrides and conventional hydrogen sorbents.

The content of this chapter is arranged as follows:

First, the concept of hydrogen sorbent is introduced to describe an optimal type of HSM, in which hydrogen–HSM interaction is much weaker than the internal interactions in the HSMs. Such a sorbent condition in fact ensures that the topological structure and the intactness of the hydrogen sorbent are not affected by hydrogen charge/discharge. In this case, fast kinetics, high efficiency in energy conversion could be realized. Based on the thermodynamics theory, the optimal hydrogen binding energy is discussed for near-ambient operation conditions in hydrogen sorbents.

Second, the mechanism of hydrogen-material interaction is analyzed in terms of bonding between hydrogen and other elements through orbital hybridization. According to the sorbent condition, we show that carbon-based materials could be the best sorbents. For optimal hydrogen binding, boron and/or transition-metal doping of carbon materials are necessary. Physisorption, chemisorptions through weak covalent bonds, non-classical dihydrogen binding, and electrostatic-enhanced binding of  $H_2$  in these materials will be analyzed in detail.

Third, internal interaction in the hydrogen sorbents is discussed. The interaction between components of the HSMs plays a key role in sorbent design and synthesis. Statically, the internal interaction holds the components together for them to form a stable sorbent structure. Dynamically, the internal interaction provides the driving force for structure formation. Of crucial importance for hydrogen storage, the internal interaction affects the interaction between the sorbent and hydrogen. In total, the internal interaction not only determines how to form a stable sorbent, but also allows for the hydrogen-binding properties to be tuned purposely.

Fourth, in terms of structuring and architecture, a hydrogen sorbent can be generally considered as arrays of hydrogen-sorption sites on its surface. Basic quantities such as the density of sorption sites, specific surface area, and their relationship with hydrogen-storage properties will be analyzed. Typical nanoscale building blocks of hydrogen sorbents will be discussed. These nanostructures includes: organometallic buckyballs, B-doped carbon cages and carbon nanotubes, endohedral fullerenes, transition-metal carbide nanoparticles, metal-doped MOFs and porous carbon, etc.

Fifth, design principles (DP) of hydrogen sorbents are derived from the above theory with consideration of practical requirements such as energy efficiency, charging/discharge rate, competitive price, reversibility and durability. These principles are binding-energy optimization, maximum capacity, material stability, etc.

## 20.2 Basic considerations

### 20.2.1 The basis of hydrogen sorbents and hydrogen binding energy

Except for the case of compressed hydrogen tanks, all currently required approaches to solid-state hydrogen storage are based on material–hydrogen interaction. Assume that a series of chemical elements X, Y, ..., and hydrogen H form the hydrogen storage system (X, Y, ..., H). The storage properties associated with the hydrogen charge/discharge reaction (X, Y, ...)<sub>solid</sub> + H<sub>gas</sub> ⇌ (X, Y, ..., H)<sub>solid</sub> can be characterized by the variation of Gibbs free energy:

$$\Delta G = \Delta U + P\Delta V - T\Delta S = \Delta H - T\Delta S, \quad (20.1)$$

where  $U$  is the internal energy,  $H = U + PV$  is the enthalpy,  $V$  is the volume, and  $S$  is the entropy. Quantities in all equations in this chapter are defined in term of charging/discharging a mole of H<sub>2</sub> molecules. The change of quantities in eqn (20.1) and charging/discharging are defined by

$$\begin{cases} \Delta G = G(X, Y, \dots, H) - G(X, Y, \dots) - G(H) \\ \Delta G < 0 : \text{Charging; } \Delta G > 0 : \text{Discharging} \end{cases}. \quad (20.2)$$

Neglecting  $P\Delta V$  for solids, eqn (20.1) can be rewritten as

$$\Delta G = \Delta E - (PV - TS) + \Delta G_0(T), \quad (20.3)$$

with

$$\Delta E = E(X, Y, \dots, H) - E(X, Y, \dots) - E(H), \quad (20.4)$$

and

$$\begin{cases} \Delta G_0(T) = \Delta U_0(T) - T\Delta S_0(T) \\ \Delta U_0 = U_0(X, Y, \dots, H) - U_0(X, Y, \dots) - U_0(H) \\ \Delta U_0 = S_0(X, Y, \dots, H) - S_0(X, Y, \dots) - S_0(H) \end{cases} \quad (20.5)$$

In eqn (20.3),  $P$ ,  $V$ , and  $S$  are specifically associated with the ideal gas without any internal degree of freedom. The temperature-independent  $E$  in eqn (20.4) is the total energy calculated with atomistic methods based on, e.g. density-functional theory. The temperature-dependent correction expressed in eqn (20.5) is contributed from the internal energy ( $\Delta U_0$ ) and entropy ( $\Delta S_0$ ) related with phonons for the solids and vibrational, rotational, translational motion for the hydrogen molecules (Alapati *et al.* 2007). These are universal thermodynamic equations and all the quantities can be evaluated if the crystal structures of the materials are known.

The HSMs characterized by eqns (20.1)–(20.5), however, have serious disadvantage because the energy function with multiple variables entangled together implies a very complicated energy landscape undertaken by the phase transition in hydrogen charge/discharge processes. This difficulty can only be circumvented if all observables  $\mathcal{O}(=G, U, E, \text{etc.})$  satisfy the

approximation of

$$\mathcal{O}(X, Y, \dots, H) = \mathcal{O}[A(X, Y, \dots), H]. \quad (20.6)$$

This simple mathematical condition has a crucially important implication in that the interaction between elements X, Y, ... is much stronger than their interaction with hydrogen. The HSMs with properties satisfying eqn (20.6) are called *hydrogen sorbents*. In this situation, hydrogen charge/discharge can be considered as a perturbation to the sorbent A(X, Y, ...), which always responds as a whole to the effect of hydrogen. The requirement for searching for the right hydrogen sorbents will be summarized in Section 20.6 as the first DP. In this chapter, we are mainly focused on studies of hydrogen sorbents.

An immediate benefit from eqn (20.6) is the greatly simplified kinetics related to the hydrogen charge and discharge processes. In hydrogen sorbents, the sorption/desorption of a hydrogen molecule (or atom) experiences approximately a 1D energy pathway with a single barrier. In special cases when the H-sorption sites are independent and the sorption/desorption barrier is zero, the hydrogen-storage property is completely determined by a *binding energy* defined from eqn (20.4):

$$E_b = -[E(A, H) - E(A) - E(H)]. \quad (20.7)$$

With this definition, positive value of  $E_b$  means hydrogen binding is energetically favorable at  $T = 0\text{ K}$ . Obviously, for non-sorbents, the corresponding quantity is  $\Delta E$  as defined by eqn (20.4). Now, eqn (20.1) can be rewritten as

$$\begin{cases} \Delta G = \Delta G_0(T) - E_b - G'(P, T) \\ G'(P, T) = PV - TS(T) \end{cases}. \quad (20.8)$$

In eqn (20.8), both  $\Delta G_0(T)$  and  $E_b$  can be computed using DFT methods. Usually, the binding energy  $E_b$  is relatively easy to compute, whereas computation of  $\Delta G_0(T)$  is quite tedious. Actually,  $\Delta G_0(T)$  should be quite small and is often neglected in current literature. This is quite reasonable for solid hydrogen sorbents. Because of the topological similarity of the solids A and (A,H), the correction to them should be similar; and the small difference caused by the weak interaction between A and H can be further offset by  $G_0(H, T)$ . It is found that the difference between  $\Delta E$  and the reaction enthalpy is 0–20 kJ/mol-H<sub>2</sub> (Alapati *et al.* 2007) for non-sorbents such as metal hydrides. We expected the difference between binding energy and sorption enthalpy to be smaller in hydrogen sorbent. However, this is overall a state-of-the-art assumption, which should be validated in specific problems through actual calculation.

At a constant temperature, the pure ideal-gas term  $G'(P, T)$  in eqn (20.8) satisfies a derivative form of  $dG'/dp = (1/RT)(1/p)dp$ , with  $PV = RT$  ( $R$  is the gas constant) considered. Integrating the derivative equation over ( $p$ ,  $p_0 = 1\text{ atm}$ ), one can determine  $G'(P, T)$  and obtain (with  $\Delta G_0(T)$  neglected here)

$$\begin{cases} \Delta G = -[E_b + G'(H, T, p^0) + RT \ln(p/p^0)] \\ G'(H, T, p^0) = H'(T, p^0) - TS'(T, p^0) \end{cases}, \quad (20.9)$$

where  $G'(\text{H}, T, p^0)$ ,  $H'(T, p^0)$ , and  $S'(T, p^0)$  are Gibbs free energy, enthalpy, and entropy of hydrogen ideal gas, relative to those at  $T = 0\text{ K}$  (Reuter and Scheffler 2001). Equation (20.9) is a generalized form of the Van't Hoff equation and can be used with eqn (20.2) to evaluate the hydrogen-storage property of a sorbent.

We first examine eqn (20.9) for  $T = 300\text{ K}$ . With the available temperature-dependent entropy and enthalpy for ideal hydrogen gas (Chase Jr 1998), one can easily obtain  $G'(\text{H}, 300\text{ K}, p^0) = -30.74\text{ kJ/mol-H}_2$ . In contrast,  $RT \ln(p/p^0)$  is 0 at 1 atm,  $11.48\text{ kJ/mol-H}_2$  at 100 atm, and  $17.23\text{ kJ/mol-H}_2$  at 1000 atm. Obviously, without sufficient binding energy, the requirement of  $\Delta G(A, \text{H}) < 0$  for hydrogen uptake cannot be realized at near-ambient pressure. Only the binding energy can fight effectively against the entropy to hold hydrogen. Once the hydrogen-binding energy is tuned to  $\sim 30\text{ kJ/mol-H}_2$ , pressure can be used to switch hydrogen charge and discharge. Therefore, hydrogen-binding energy is a key parameter for reversible storage at ambient condition. Tuning the hydrogen-binding energy will be considered as the second DP in Section 6.

### 20.2.2 Langmuir isotherm, optimum hydrogen binding energy, and hydrogen delivery

The above equations provide the basis for discussion of material design and hydrogen-storage properties such as maximum hydrogen capacity and the critical condition for switching the charge and discharge. However, a detailed analysis of the entire charge-discharge cycle is required for a practical operation condition. In such a process, the pressure and/or temperature may vary in a certain range so that maximum hydrogen delivery can be achieved. Also, the above ideal gas model for optimizing binding energy is very rough. Many simulations have showed that the optimal binding energy for near-ambient reversible hydrogen storage is smaller than  $30\text{ kJ/mol}$  (Bhatia and Myers 2006; Blomqvist *et al.* 2007; Mpourmpakis *et al.* 2007). Under the Langmuir isotherm, the hydrogen capacity,  $c$ , of the adsorbent is related to the maximum capacity  $c_m$  through an equation

$$c = \frac{K(T)P c_m}{1 + K(T)P}, \quad (20.10)$$

where  $P$  is the pressure and  $K(T)$  is the equilibrium constant related to the reaction enthalpy  $\Delta H'$  and entropy  $\Delta S'$  relative to those at standard condition through

$$K = \frac{1}{P_0} \exp\left(\frac{\Delta S' T - \Delta H'}{RT}\right). \quad (20.11)$$

Based on eqn (20.10), the hydrogen storage condition ( $T_1$ ,  $P_1$ ) and exhaustion condition ( $T_2$ ,  $P_2$ ) give a delivery

$$D(T_1, P_1; T_2, P_2) = c_m \left[ \frac{K(T_1)P_1}{1 + K(T_1)P_1} - \frac{K(T_2)P_2}{1 + K(T_2)P_2} \right]. \quad (20.12)$$

Maximization of  $D$  at a constant temperature  $T$ , i.e.  $K(T_1) = K(T_2) = K$ , requires  $K = 1/\sqrt{P_1 P_2}$ . The optimum enthalpy is therefore derived from eqn (20.11) as

$$\Delta H'_{\text{opt}} = T \Delta S' + \frac{RT}{2} \ln \left( \frac{P_1 P_2}{P_0} \right). \quad (20.13)$$

In this way, Bhatia and Myers obtained an optimum enthalpy  $-15 \text{ kJ/mol}$  and the maximum delivery 63.5% of the total hydrogen capacity and a storage pressure  $P_1 = 30$  and exhaustion pressure  $P_2 = 1.5 \text{ bar}$  (Bhatia and Myers 2006) for homogeneous sorbents at  $T = 298 \text{ K}$ . We will see in Section 6, the percentage of deliverable can be increased if temperature change is allowed.

### 20.2.3 The “interaction–structure–property” paradigm

Compared to the reaction enthalpy currently used in the study of nonsorbents, binding energy is related more straightforwardly to the microscopic nature of hydrogen–material interaction. Through the binding energy, the interaction between the HSM and hydrogen are reduced to localized bonding, and the complicated chemical environment in the HSM is considered to tune the binding energy in a single variable  $A(X, Y, \dots)$ .

In principle, the way that hydrogen interacts with other materials is inherently determined by the nature of the bonds between hydrogen and the constituent elements in the materials. Therefore, one can only choose the right material for a particular type and strength of interaction, because the chemical nature of the elements cannot be manipulated. We call this inherent part of the “interaction” *intrinsic bonding*. The elemental materials in ground states are usually formed through unalterable intrinsic bonds, and offer little freedom of design. However, the complexed material structure is not a part completely subjected to the intrinsic interaction. The art of structuring allows us to go beyond the ground state and create great variety of metastable, complex sorbents  $A(X, Y, \dots)$  and enrich the chemical environment. In correspondence to a particular chemical environment, a structure modulates the internal interaction, which in turn fine tunes the external H–sorbent interaction within a pre-determined range. The interaction mechanism changing with the totality of a chemical environment can be called *extrinsic interaction* because it can be externally fabricated through material structuring. It is the extrinsic interaction (upon the intrinsic bonding) and metastable complexity (beyond the ground-state structures) that grants a spacious freedom for material design and offers great opportunities for basic research as well.

Traditionally, the basic “structure–property” paradigm is widely used to guide materials science. The “property” part, representing function, is basically the interface between materials science and engineering. The “structure” part is believed to be the core of materials science, because structures represent the materials. However, we show here the “interaction” part has to be added to the basic paradigm. As is discussed above, the term “interaction” used here refers not only to localized chemical bonds, long-range electrostatic attraction or repulsion, delocalized electronic effect, etc.; but also to any underlying driving force, by which structuring and functionalization could be realized. Therefore, we construct the framework of theory based on an

“interaction–structure–property” paradigm and emphasize especially the “interaction” part, which includes both hydrogen–sorbent interaction and the internal interaction inside the HSMs. In the following sections, we will try to show a picture of systematic relationship between these three parts, out of which a few general principles will be extracted to guide the HSM design.

## 20.3 Hydrogen–material interaction

Besides the thermodynamic considerations, we describe here a quantum-mechanical basis of orbital hybridization for analysis of the hydrogen–material interaction. Following Harrison (1980), we use here the simplest picture, which, however, provides sufficiently clear insight for the relevant issues to be considered here.

Without losing generality, the problem of chemical interaction or chemical bonding can be approximated by hybridization of two atomic or molecular orbitals,  $\phi_1$  and  $\phi_2$ . The hybridized wavefunction can be expressed as a linear combination of  $\phi_1$  and  $\phi_2$ . Through solving the Schrödinger equation within the variational approach, the energy levels of the hybridized states are roughly

$$E_{\pm} = \frac{H_{11} + H_{22}}{2} \pm \sqrt{\left(\frac{H_{11} - H_{22}}{2}\right)^2 + H_{12}H_{21}}, \quad (20.14)$$

with  $H_{ij} = \langle \phi_i | H | \phi_j \rangle$ . The + and – signs correspond, respectively, to the antibonding and bonding states. When the energy separation between the lower state  $\phi_1$  and the higher one  $\phi_2$ ,  $H_{22} - H_{11}$ , is small, a non-polar bond is formed after the hybridization. On the contrary, when  $H_{22} - H_{11}$  is large, the bond is polar and eqn (20.14) can be simplified to

$$E_1 = H_{11} + \frac{H_{12}H_{21}}{H_{11} - H_{22}} \text{ and } E_2 = H_{22} + \frac{H_{12}H_{21}}{H_{22} - H_{11}}, \quad (20.15)$$

where the bonding-state level  $E_1$  and the antibonding state level  $E_2$  can be considered as a result of perturbation to the original energy levels for  $\phi_1$  and  $\phi_2$ . Ionic bonds are the extreme case to the polar ones.

Because eqns (20.14) and (20.15) are basic quantum-mechanical relations, they do not depend on the occupation of the original states  $\phi_1$  and  $\phi_2$ . Therefore, the concept of hybridization is applicable to both normal chemical bonding between two half-occupied orbitals, and the so-called non-classical bonding between a fully occupied orbital and an empty orbital (hole). For a non-classical binding of dihydrogen as we will discuss later, the lower state  $\phi_1$  is the  $\sigma$ -bond state of a free  $H_2$  molecule, and the higher state  $\phi_2$  is the hole state in the sorbent. In this case, the dihydrogen binding energy can be qualitatively analyzed with eqn (20.15) and approximated by  $H_{11} - E_1 = \frac{H_{12}H_{21}}{H_{22} - H_{11}}$ . Obviously, the larger the separation between the  $H_2$   $\sigma$ -bond level and the sorbent hole level, the smaller is the  $H_2$  binding energy. This is one of the two reasons for the increase in  $H_2$  binding energy to 3d TM atoms from Sc to Ni with gradually dropping of the TM  $d$ -orbitals (Zhao *et al.* 2005). The other reason is backdonation from the more and more occupied TM  $d$ -orbitals to the antibonding  $H_2$  orbital. Also, the more localized the hole state (the larger

$H_{12}H_{21}$ ), the larger is the binding energy. This explains the larger  $H_2$  binding energy in B-doped smaller carbon cages than that in B-doped large carbon cages (Kim *et al.* 2006).

### 20.3.1 Intrinsic bonding between hydrogen and other elements

According to the above quantum-mechanical concept, the intrinsic bonding between chemical elements can be qualitatively illustrated in a simple picture of state hybridization as shown in Fig. 20.1. Conventionally, the formation of a chemical bond is the pairing up of two single electrons occupying the energy levels localized, respectively, at two atomic centers. In the cases of pure covalent or non-polar bonding, the two initial states (lighter lines) occupied by single electrons (single arrows) are approximately in the same level. Hybridization gives rise to two split levels (heavier lines). The two electrons pair up in the lower level (bond state), leaving the upper level (antibond state) empty. The gravity center of bonding electron density (bond charge) is localized right in the middle of the two atoms, therefore, the system shows no polarity. When the initial states of the two atoms are in different energy levels, the hybridized bond state contains more of the lower initial state and less of the higher one. Therefore, the gravity center of bond charge is closer to the atom with higher electronegativity, and the bond is a polar bond. An extremely polarized bond is an ionic bond, in which case the electron of one atom is simply transferred to the energy level of another atom and the shift of energy levels can be neglected. In contrast to the pairing up of two single electrons in the above three types of bonds, a non-classical bond is the bonding between an empty orbital or localized hole state and an electron pair. The electron pair could be a lone-electron pair localized at a single atom, e.g., N in  $NH_3$ , or the two electrons in a covalent bond between two atoms, e.g. H–H in a hydrogen molecule. The non-classical bonding that hybridizes an empty orbital of one atom and

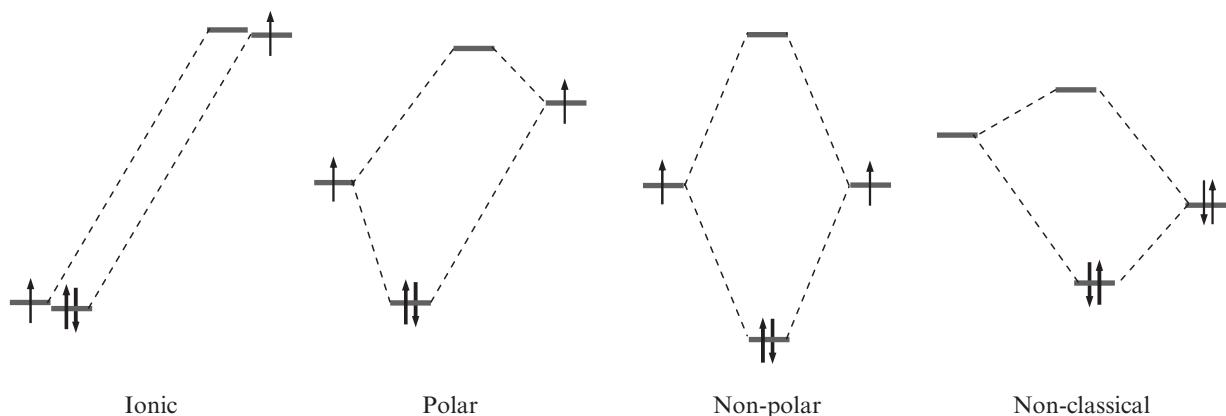


Fig. 20.1 Basic bonding mechanism.

a bond orbital of the other two atoms is called three-center two-electron (3c–2e) coordination.

#### 20.3.1.1 *Ionic bonds in metal hydrides*

With a relatively high electronegativity, hydrogen interacts with most metals through ionic bonds, forming metal hydrides as a type of ionic solid. Two observations about the ionic crystals are important for hydrogen storage. (1) The cations (metal, M) and anions (H) are always alternatively distributed in the lattice, due to the rule that “opposites attract, and likes repel”. An ionic solid favors most a cation–anion stoichiometric ratio of 1:1, because a higher ratio unavoidably leads to a crowd of like charges. (2) The formation of metal hydrides is always energetically favorable and the higher the ionicity of the metal–hydrogen bonds, the larger is the formation energy (Phillips 1970). These imply that M–H interaction is stronger than M–M interaction and, metal hydrides do not satisfy eqn (20.6), therefore, are not hydrogen sorbents.

Hydrogen storage in metal hydrides has been intensively studied for half a century and is still an active research area. According to the above properties of ionic solids, metal-hydride solids with higher H ratio, e.g. AlH<sub>3</sub>, are superior to those with an H–M stoichiometric ratio of 1:1, e.g. LiH, which are too stable to release hydrogen. However, for most metal-hydride solids, easier H discharge usually means harder charge, or vice versa. In order to optimize the reaction enthalpy for reversible charge/discharge, more ingredients including catalysts, are added to metal hydrides to form complex hydrides (Meisner *et al.* 2006; Alapati *et al.* 2007), which, however, often risks fulling into another paradox—the more ingredients for optimizing enthalpy to improve reversibility, the more complicated is the kinetics against reversibility. This is because more variables introduced to the energy functions lead to a more complicated energy landscape.

Although pure metals are not good as hydrogen sorbents, metal can be used as an ingredient for metastable hydrogen sorbents along with other materials, such as carbon. In these kinds of sorbents, metal atoms should be dispersed on the surface of carbon substrate. Here, a crucial consideration is to avoid the formation of ground-state metal-hydrides solid. For example, alkaline-metal hydrides, with the highest ionicity and the most favorable H–M stoichiometric ratio (1:1), can be easily formed. To avoid this, a high degree of electron depletion of the alkaline-metal atoms must be realized through charge transfer from the metal to the substrate.

#### 20.3.1.2 *Non-polar covalent bonds*

A non-metal atom with electronegativity similar to hydrogen can form a non-polar molecule with hydrogen through very strong covalent bonds. For example, the formation energy of CH<sub>4</sub> molecules from C atoms and H<sub>2</sub> molecules is as high as 400 kJ/mol-H<sub>2</sub>, highlighting the strongest interaction between hydrogen and other materials. Because of the single coordination number of hydrogen and multiple coordination number of carbon, hydrocarbon structures always terminate with hydrogen. The interaction inside a hydrocarbon structure is strong carbon–carbon bonding, and between two hydrocarbon structures is weak van der Waals (vdW) force. Consequently, carbon materials could be

good sorbents for atomic hydrogen if the H–C bond strength can be weakened via tuning. Hydrogen storage in hydrocarbons is one of the most studied methods in traditional chemical complexes, and small hydrocarbon molecules ( $\text{CH}_4$ ,  $\text{C}_2\text{H}_6$ , etc.) in natural gas are currently the major fuel in fuel cells. But “reforming” these small hydrocarbon molecules to hydrogen generates carbon dioxides. In recent years, carbon-based nanoscale sorbent has become a hot area, stimulated by the fact that carbon nanostructures may be recyclable in the process of hydrogenation and dehydrogenation.

#### 20.3.1.3 Polar covalent bonds

If a non-metal atom has a higher electronegativity than that of hydrogen, e.g. nitrogen, oxygen, it will form molecules with polar bond(s) with hydrogen. Compared to  $\text{CH}_4$ , the  $\text{H}_2\text{O}$  and  $\text{NH}_3$  molecules interact more strongly with each other through hydrogen bonds and condense into liquids at ambient conditions, therefore, are much easier to store. Water has been occasionally proposed as a fuel, but so far most scientists seriously doubt it. Ammonia ( $\text{NH}_3$ ) is a hydrogen-rich fuel, but it is not as good as  $\text{CH}_4$ , because nitrogen oxide can be more harmful to the environment than carbon dioxide.

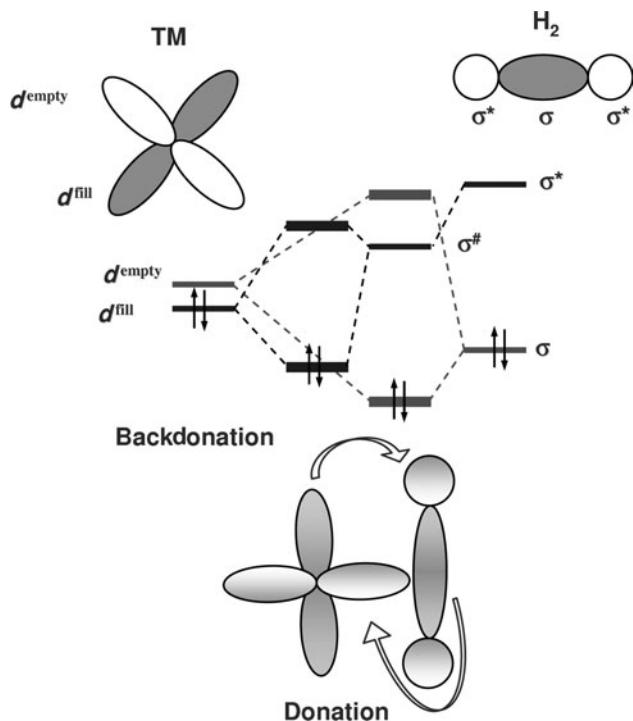
Although not good as direct fuels, the small polar molecules ( $\text{H}_2\text{O}$ ,  $\text{NH}_3$ ) are not abandoned for energy applications. An interesting approach to hydrogen storage is the use of the so-called molecule compounds of these small polar molecules (Mao and Mao 2004). Scientists learn this from a natural material called methane hydrate (Suess *et al.* 1999), in which the  $\text{CH}_4$  molecules stabilize a clathrate structure of water, which is another phase of ice with microcages formed of several tens of water molecules but unstable for pure water. The guest  $\text{CH}_4$  molecules are trapped in the cages and stabilize them through vdW interaction added to the hydrogen bonds between the  $\text{H}_2\text{O}$  molecules. The methane hydrate may have a melting point close to room temperature. Hydrogen molecules themselves are not good guest molecules for stabilizing the clathrate phases of ice. The trick is to use some guest molecules as stabilizer occupying a fraction of the number of cages, while  $\text{H}_2$  molecules can be trapped in the majority of cages (Florusse *et al.* 2004; Lee *et al.* 2005). In this method, the light hydrogen gas is compressed and restricted in numerous microcontainers, which can be closed and opened through phase transition. This phenomenon poses a few questions for fundamental thermodynamics: (1) How to deal with the phase-transition process in a thermodynamics system with several different particles and different types of interaction? (2) Can the guest molecules and hydrogen be treated as a perturbation in the clathrate framework, and how to deal with the role of the guest molecules and hydrogen gas? (3) What new behaviors can be predicted in such a phase transition? (4) Is the phase transition reversible and easy to control at normal conditions? (5) Is the hydrogen gas released gradually or abruptly during the phase transition? Currently, these issues may not have been fully exposed because the quantity of the samples in experiments is usually too small.

#### 20.3.1.4 Kubas coordination

About twenty years ago, Kubas discovered an unconventional coordination between a hydrogen molecule and transition-metal (TM) atoms (Kubas *et al.* 1984). Interestingly, it is less acknowledged that Saillard and Hoffmann (1984)

published, at almost the same time, a theoretical study on stable chemisorbed non-dissociated hydrogen molecules (namely, dihydrogen). This is an historical discovery not only in coordination chemistry, but also for hydrogen storage. It was soon demonstrated that the dihydrogen molecules can be reversibly adsorbed/desorbed (Mediati *et al.* 1992). The intermediate binding strength bridges the weak physisorption and conventional chemisorption and gives new hope for practical hydrogen storage at near ambient conditions. Very recent simulations of multiple dihydrogen binding to a single TM atom (Gagliardi and Pyykko 2004) initiated a promising new direction of hydrogen storage based on the organometallic approaches (Yildirim and Ciraci 2005; Zhao *et al.* 2005). This further stimulates a topic study of molecular hydrogen binding (Lochan and Head-Gordon 2006).

Kubas coordination originates from the unique property of the TM  $d$ -orbitals. The basic fact is that there are five  $d$ -orbitals, which are partly occupied in most free TM atoms; and the energy difference between the occupied and empty  $d$ -orbitals is not large. This gives rise to a lot of unique properties (e.g. magnetic property) of TM materials. Specifically in Kubas coordination, the empty  $d$ -orbital(s) play a key role. These empty orbitals have strong ability to accept external electrons because they are below the vacuum level. On the other hand, the occupied  $d$ -orbital(s) are ready to donate an electron. Consequently, the mechanism of Kubas coordination (Kubas 2001) can be illustrated as two components of charge donation or non-classical orbital hybridization (Fig. 20.2). The hydrogen molecule donates a fraction of electrons in its bonding state ( $\sigma$ ) to the empty  $d$ -orbital of the TM atom and the total energy



**Fig. 20.2** Mechanism of Kubas coordination.

is lowered because the two electrons in the hydrogen molecule now occupy a lower orbital hybridized from  $\sigma$  and  $d^{\text{empty}}$ . This in turn pulls the antibonding  $\text{H}_2$  state  $\sigma^*$  down to  $\sigma^\#$ , which helps the backdonation of electron from  $d^{\text{fill}}$  to  $\sigma^*$  and further lowers the total energy. These two components of hybridization stretch the H–H bond, but the hydrogen molecule remains undissociated (dihydrogen). Quantitative decomposition of dihydrogen binding energy into the two components is possible within the charge-decomposition analysis (Dapprich and Frenking 1995) or extended transition-state analysis (Li and Ziegler 1996).

The key to dihydrogen binding is a localized empty orbital, which is below the vacuum level and has affinity for the bonding electrons of a  $\text{H}_2$  molecule. The lower the empty orbital, the more favorable is the formation of a dihydrogen complex. Generally, the empty orbital does not have to be a TM  $d$ -orbital. For example, positive simple-metal ions can, too, bind dihydrogen (Lochan and Head-Gordon 2006) because the lower bonding orbital(s) is now empty. Without backdonation, the 3c–2e coordination is already completed. Actually, the 3c–2e coordination is only rigorous when there is no backdonation, because backdonation involves more (than 2) electrons occupied in another orbital. Backdonation takes the advantage of lowering of  $\sigma^*$  due to donation of  $\sigma$  charge, adding up to the dihydrogen binding. But too strong backdonation eventually splits the  $\text{H}_2$  molecule to form a hydride. The dihydrogen–dihydride transition depends on the energy levels of the occupied and empty  $d$ -orbitals and their separation. The higher the  $d$ -orbitals and the smaller the separation between the empty and occupied  $d$ -orbitals, the more easily the transition happens (Tomas *et al.* 1998).

The above mechanism explains why simple metals only form hydrides: (1) simple metals do not have low-lying empty orbitals, therefore do not satisfy the condition for 3c–2e coordination; (2) simple metals have a very strong tendency to “backdonate” all their valence electrons to hydrogen. For the TM elements close to the simple metals in the periodic table, e.g. Sc, hydride formation is also highly favorable. However, after using up all valence electrons to form hydrides, it still binds dihydrogen with its empty  $d$ -orbital(s) (Zhao *et al.* 2005). The TM atoms to the right of the Periodic Table bind dihydrogen more strongly, because they have sufficient valence electrons for backdonation, yet their valence electrons are not radical enough to split the  $\text{H}_2$ . In this case, the binding energy of dihydrogen could be larger than 100 kJ/mol, which is good for static observation of the dihydrogen identity but too strong for reversible hydrogen storage at near room temperature.

#### 20.3.1.5 Electrostatic interaction between hydrogen molecule and charges

Although poorly polarizable, a hydrogen molecule does interact weakly with charges through charge quadrupoles and charge-induced dipoles. Quantum-mechanical calculations have showed an electrostatic potential map of the hydrogen molecules (Vitillo *et al.* 2005), in which the potential field is positive at the two ends and negative at the side plane perpendicular to the H–H bond axis. Therefore, a hydrogen molecule interacts with negative charges in the head-on geometry but interacts with positive charges in the side-on geometry (Lochan and Head-Gordon 2006). Because of the poor polarizability,

electrostatic interaction of H<sub>2</sub> with most ions is relatively weak. However, when the radius of the ion is small, e.g. Li<sup>+</sup>, the binding can be strong enough for hydrogen storage (Lochan and Head-Gordon 2006).

### 20.3.2 Hydrogen interaction with carbon-based sorbents

In most cases, the hydrogen–sorbent interactions are much more complicated than the intrinsic interactions, due to the fact that the interaction between elements, X, Y, . . . , in the sorbent A(X, Y, . . .) can alter the sorbent–hydrogen interaction. Here, we chose the carbon-based materials as examples to discuss hydrogen-binding mechanisms. For clarity, we distinguish molecule and atom sorbents, which bind, respectively H<sub>2</sub> molecules and atomic H.

#### 20.3.2.1 Physisorption

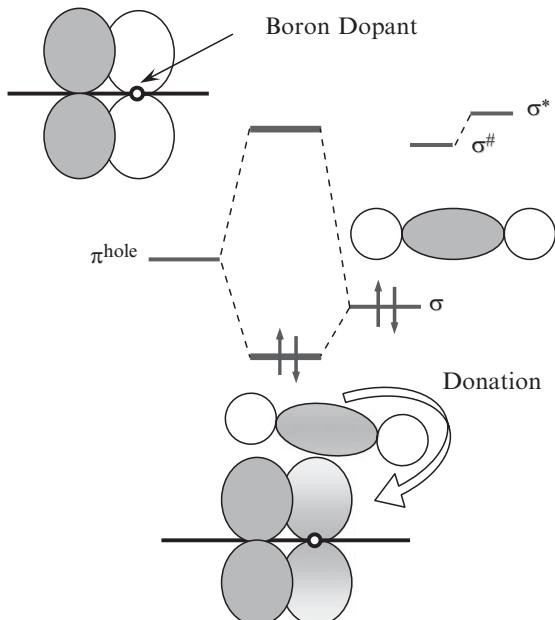
The simplest H<sub>2</sub> sorbent is perhaps a *sp*<sup>2</sup>-bonding network of pure carbon, in which the closed electronic shell and strong C–C bonds leave no reactive sites for hydrogen chemisorption, thus only physisorption of hydrogen molecules is allowed in a weak dispersive force field (vdW interaction). Strict computation of the vdW force quantum mechanically at the atomic level is still a challenge. Even if the continuum dispersive force potential is known, a H<sub>2</sub> molecule often has to be treated as a quantum-mechanical particle, due to its small mass (Wang and Johnson 1998; Hathorn *et al.* 2001; Lu *et al.* 2003). Here, we employ a simple continuum model, in which the binding energy of physisorbed H<sub>2</sub> is proportional to the effective contact area *s* per hydrogen molecule on the surface of the sorbent. That is,

$$E_b = s B_0, \quad (20.16)$$

where *B*<sub>0</sub> can be calibrated with the binding energy of a hydrogen molecule encapsulated in an ideally spherical shell with a radius of *r*<sub>Ω</sub>, which is the equilibrium distance of vdW interaction between the H<sub>2</sub> molecule and the shell. If we presumably use H<sub>2</sub>@C<sub>60</sub> as a gauging system with a binding energy of 0.3 eV/H<sub>2</sub> or 7 kcal/mol (Slanina *et al.* 2006) and an effective contact area of *s*<sub>Ω</sub> = 4π*r*<sub>Ω</sub><sup>2</sup> = 1.453 nm<sup>2</sup>, then *B*<sub>0</sub> = 0.207 eV/nm<sup>2</sup> is obtained. It is known that the binding energy of a H<sub>2</sub> physisorbed to a flat surface is ~0.04 eV/H<sub>2</sub>, therefore, the effective contact surface area of a H<sub>2</sub> molecule to such a flat surface is *s*<sub>Π</sub> = 0.19 nm<sup>2</sup>/H<sub>2</sub>. Interestingly, this effective contact area is comparable to the geometrical occupancy area of 0.14 nm<sup>2</sup>/H<sub>2</sub>, estimated within different methods (Nijkamp *et al.* 2001; Panella *et al.* 2005), and that of 0.17 nm<sup>2</sup>/H<sub>2</sub>, measured at 77 K and 1 bar (Zuttel *et al.* 2004). Equation (20.16) and the parameters *B*<sub>0</sub>, *s*<sub>Π</sub> measure the binding strength of H<sub>2</sub> molecules in the dispersive force field, independent of the material.

#### 20.3.2.2 Non-classical chemisorption

When the above sorbent is doped with acceptors, for example, substitutional boron atoms, hole states or empty orbitals will be generated, which interact with the bonding orbital of hydrogen molecules through non-classical bonding, as shown in Fig. 20.3. The binding energy ranges from 10 kJ/mol-H<sub>2</sub> to 20 kJ/mol-H<sub>2</sub>, depending on the degree of localization of the hole state



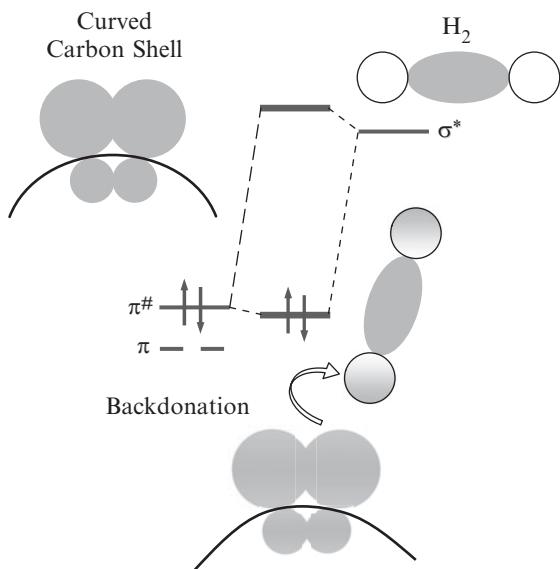
**Fig. 20.3** Dihydrogen bonding to a localized hole state.

(Kim *et al.* 2006), which is understandable from eqn (20.15) in terms of wavefunction overlap—the more localized the hole, the more effectively its wavefunction overlaps with the  $\sigma$ -bond orbital of the  $\text{H}_2$ . In the  $\text{BH}_5$  complex, the  $\text{BH}_3$  molecule binds a dihydrogen through exactly such a mechanism (Schreiner *et al.* 1994; Tague and Andrews 1994). For  $\text{BH}_5$ , unlike Kubas coordination between a  $\text{H}_2$  and a TM atom, there is no backdonation from the B atom to the antibond  $\sigma^*$  state of hydrogen molecules. For boron-doped carbon cages or sheets, backdonation from the occupied  $\pi$ -orbital to the  $\sigma^*$  may not be substantial, due to the delocalization and relatively low energy of the  $\pi$ -electrons.

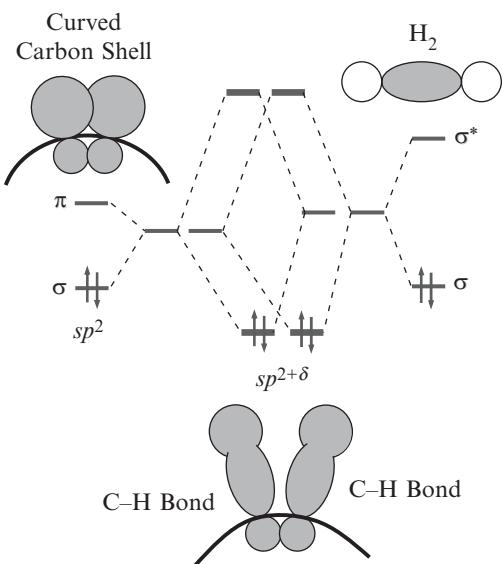
However, Cheng *et al.* suggested that the highly curved carbon shell, e.g. very small diameter carbon nanotubes (CNTs), could induce charge transfer from carbon to  $\sigma^*$  state of  $\text{H}_2$ , therefore slightly enhancing the binding of the latter (Cheng *et al.* 2005). The mechanism is that the curvature or strain pushes the original  $\pi$  state up to  $\pi^\#$  (Fig. 20.4), which could hybridize with  $\sigma^*$ . First-principle calculations indicate that the binding enhancement is quite small (Cheng *et al.* 2005) because the  $\pi - \sigma^*$  gap is large. In fact, the curvature-enhanced  $\text{H}_2$  binding to the carbon shell is just metastable. When the curvature becomes larger,  $sp^{2+\delta}$  ( $0 < \delta < 1$ ) rehybridization (Park *et al.* 2003) would be energetically much more favorable so that C–H bonds will be formed if the  $\text{H}_2$  dissociated barrier can be overcome.

### 20.3.2.3 Weak C–H bonds

Figure 20.5 shows the mechanism of weak chemisorption of H atom to carbon sorbents through the so-called  $sp^{2+\delta}$  rehybridization (Park *et al.* 2003). The advantage of this type of sorbent is the tunable H binding energy. Basically, the degree of rehybridization, measured by  $\delta$ , can be tuned by the curvature of the



**Fig. 20.4** Dihydrogen bonding to carbon sites activated by large curvature.



**Fig. 20.5** Weak C–H bonds formed through  $sp^{2+\delta}$  rehybridization.

carbon shell. The larger the curvature, the bigger the value of  $\delta$ , and the larger is the average H binding energy. This sorbent also has serious disadvantages. First, strain accumulation with H loading usually makes H binding less and less favorable; therefore the hydrogen capacity is not sufficiently high. Second, the kinetics of hydrogen charge/discharge is slow, even if a catalyst is used. This mechanism is currently applied to hydrogen storage using a “spillover” method (Lueking and Yang 2002, 2004; Li and Yang 2006).

To summarize this section, we see that a continuum of energies exists for hydrogen bound in/on solids and molecules. On the weak side of the continuum

is non-dissociative physisorption due purely to the vdW force on, e.g. planar graphite (4 kJ/mol). On the strong side is the full C–H chemical bond in methane with an energy of  $\sim$ 400 kJ/mol. Between these two limits are, broadly speaking, chemical hydrides, metal hydrides, and sorbents.

## 20.4 Internal interaction in HSMs

The interaction between components of the HSMs plays a key role in both sorbent design and synthesis. Statically, the internal interaction holds the components together forming a stable sorbent structure. Dynamically, the internal interaction provides the driving force for structure formation. Specifically important for hydrogen storage, the internal interaction affects the interaction between the sorbent and hydrogen. Overall, the internal interaction not only determines how to form a stable sorbent, but also allows for the hydrogen-binding properties to be tuned purposely.

Consider a HSM containing atoms  $X_1, X_2, \dots, X_N$ . The internal interaction can be measured by the following cohesive energy ( $E_c$ ) or formation energy with respect to the free atoms:

$$E_c = \left[ \sum_i e(X_i) - E(X_1, X_2, \dots, X_N) \right] / N, \quad (20.17)$$

where  $e(X_i)$  is the energy of an isolated  $X_i$  atom. Assume that the HSM form a hydride solid with  $2m$  hydrogen atoms, then the average formation energy per H atom is

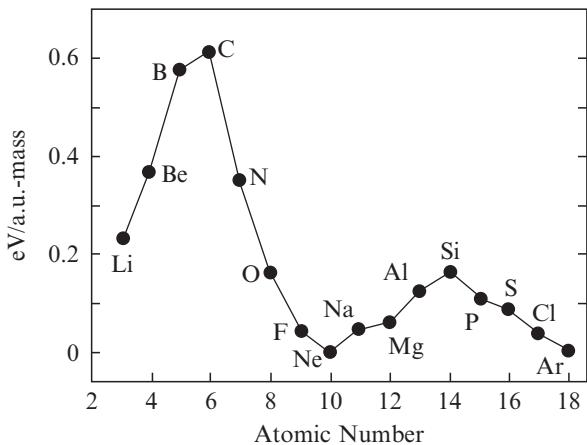
$$F_H = [E(X_1, X_2, \dots, X_N) + mE(H_2) - E(X_1, X_2, \dots, X_N, 2mH)] / 2m. \quad (20.18)$$

When  $N$  is similar to  $2m$ , the condition for the HSM to be a good sorbent would be

$$E_c \gg F_H. \quad (20.19)$$

If  $E_c < F_H$ , the hydrogen atoms could easily penetrate into the HSM, which is therefore burned out by hydrogen to form solid hydride. This is the situation of hydrogen reaction with metal elements, especially alkaline metals.

To find which element is the most suitable to form a hydrogen sorbent, the ratio of cohesive energy over atomic mass is plotted in Fig. 20.6 from element 3 (Li) through 18 (Ar). The cohesive energy is taken from Kittel (1960). Carbon stands out in the plot, followed by B, Be, and N, among all those elements. This simple illustration seems to be reminiscent of several historical movements in hydrogen-energy research since the mid-twentieth century. First, hydrogen storage with active carbon was initiated in 1967 and today blossoms into a huge research area of carbon-based hydrogen sorbents (Dillon and Heben 2001), which holds perhaps the greatest promise for integration of all hydrogen-storage methods into one practical system. Second, solid or liquid boranes as potential rocket fuels were intensively studied fifty years ago (Hermanek 1992) and now have stimulated a series of subfields of complex metal hydrides



**Fig. 20.6** Ratio of cohesive energy over atomic mass of elemental bulk materials.

or chemical hydrides by introducing other elements such as metals (Zuttel *et al.* 2003) and nitrogen (Chandra and Xu 2006).

### 20.4.1 Internal interaction in pure carbon and hydrogenated carbon structures

Going from the cohesive energy down to the detailed bonding mechanism, one finds that the uniqueness of carbon materials lies in the multiple configurations of hybridization, i.e.  $sp$ ,  $sp^2$ , and  $sp^3$  hybridization. The strong, topologically flexible  $sp^2$  bonding configuration imparts this light element a structure of single-atom-layered film, which can be wrapped up on the nanoscale to form tubular and spherical shells with extraordinary mechanical properties (Yakobson *et al.* 1996). All these materials have high specific surface area (SSA, surface area per unit mass) for gas sorption.

More interestingly, the  $sp^2-sp^3$  transition is quite straightforward because (1) the promotion energy for rehybridization  $sp^{2+\delta}$  is almost a continuous function of  $0 < \delta < 1$ , and (2) no topological rearrangement is needed in such a transition. The coexistence of  $sp^2$  and  $sp^3$  domains in a partially hydrogenated carbon shell or sheet introduces a few key interactions as follows.

#### 20.4.1.1 Attraction between dangling $\pi$ bonds

The first-order effect is the attraction of dangling  $\pi$  bonds (single  $\pi$  electrons) created in the boundaries between the  $sp^2$  and  $sp^3$  domains. In any conditions, the dangling  $\pi$  bonds have to be paired up, which instantly excludes the odd number of unhydrogenated (naked) carbon atoms from the stable patterns. For example, in the hydrogenation of carbon cages, a naked five-membered ring is obviously forbidden. Concerning the reaction dynamics, in the early stage of hydrogenation, the energetically favorable sorption sites and the H hopping pathway are not arbitrary (Zecho *et al.* 2002; Ferro *et al.* 2003; Hornekaer *et al.* 2006). When a small domain of hydrogenated carbon is formed, the structure of domain boundaries and its propagation should largely be determined by the attraction of the dangling  $\pi$  bonds.

**Table 20.1** Reaction heat (kJ/mol-H<sub>2</sub>) of hydrogenation of small hydrocarbon molecules and graphene sheet.

$\text{C}_2\text{H}_4 + \text{H}_2 \rightarrow \text{C}_2\text{H}_6$	200	$\text{C}_6\text{H}_6 + 3\text{H}_2 \rightarrow \text{C}_6\text{H}_{12}$	90	$\text{Graphene} + n\text{H}_2 \rightarrow \text{Hydrocarbon sheet}$	40
--	-----	--	----	--	----

#### 20.4.1.2 Resonant effect

Even the isolated  $\pi$  bonds (paired  $\pi$  electrons) are not sufficiently stable. In the second order, these  $\pi$  bonds attract each other through the resonant effect (Chen and King 2005; Kertesz *et al.* 2005) such as aromaticity. In fact, this interaction could be strong enough to conform the domain shape. Table 20.1 shows the reaction heat associated with hydrogenation of ethane molecules, benzene molecules, and infinite graphene. The electronic resonant interaction is so strong that the energy is lowered by more than 1.0 eV per  $\pi$  bond in benzene molecules (110 kJ/mol), and nearly 1.7 eV in graphene!

Another famous example is the hydrogenated fullerene, C<sub>60</sub>H<sub>36</sub>. Haufler *et al.* proposed a structure with T<sub>h</sub> symmetry, in which the 12 naked carbon pairs form isolated  $\pi$  bonds and are evenly distributed across the cage (Haufler *et al.* 1990). Later it was found that a tetrahedral isomer (Taylor 1992), with four aromatic rings formed of the 12  $\pi$  bonds, is 3 eV lower in energy (Buhl *et al.* 1995). In this case, aromaticity only lowers the energy by 0.25 eV per  $\pi$  bond. The possible reasons are: (1) the 12  $\pi$  bonds in the T<sub>h</sub> structure may not be isolated as they look alike geometrically; (2) the tetrahedral structure may cause more strain.

#### 20.4.1.3 Strain effect

The lattice mismatch between sp<sup>2</sup> and sp<sup>3</sup> domains causes substantial strain in a carbon sheet. If one defines a positive sign for one side and negative for the other, the interaction of sorption sites follows the rule that “opposites attract, and likes repel”. According to this rule, two-side hydrogenation favors a full coverage of all carbon atoms with H atoms. For example, if a hydrogen atom is simply added on top of each carbon atom alternately on both sides of a graphene layer, one obtains a single-layered, diamond-like hydrocarbon. In this hydrocarbon sheet, the hydrogen capacity is ~7 wt.% with an almost ideal binding energy of ~40 kJ/mol-H<sub>2</sub> (Table 20.1). However, to apply such a simple concept for hydrogen storage is still a challenge. In reality, hydrogen atoms or molecules have to be added one by one and the reaction has to experience complicated kinetics (Stojkovic *et al.* 2003) due to all the above interactions.

A carbon cage is different from a carbon sheet in two aspects. First, it has inherent strain due to the curvature. Second, it only opens the outer side to hydrogenation. In such a system, low coverage of H chemisorption actually releases the inherent strain if the hydrogenation follows a particular pattern. An example is linear hydrogenation of a CNT along the axial direction, in which hydrogen (Gulseren *et al.* 2002) or other chemisorbed species (Kudin *et al.* 2001) form axial lines dividing the CNT circumferentially into flat sp<sup>2</sup> facets. It is also found (Yoshida *et al.* 1993) in C<sub>60</sub>H<sub>x</sub> that strain relief

is maximized at  $x = 36$ . After the critical coverage is exceeded, hydrogen becomes exceptionally less favorable.

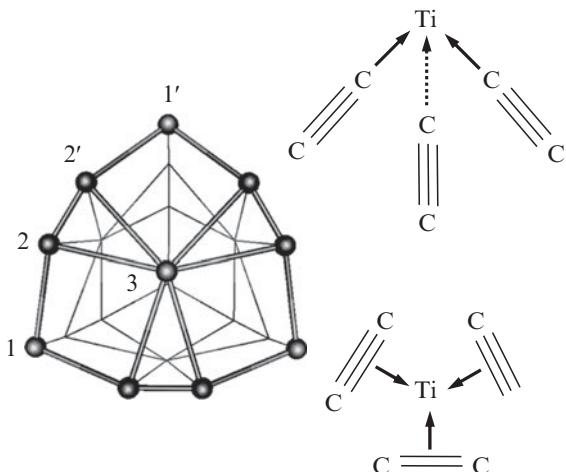
So far, we have not found sufficient researches from which a general rule can be extracted to clarify all these types of interaction and their effect on hydrogenation of carbon materials. But this should be an interesting direction and important for hydrogen storage. Carbon materials offer an excellent scaffold for development of HSMs. If pure carbon materials cannot fulfill the purpose, at least they can serve as good backbone frameworks for further functionalization with, for example, metals.

### 20.4.2 Internal interactions in carbon–metal systems

Metals dominate the majority of the Periodic Table and offer great variety of ingredients for carbon-based HSMs. Compared to non-metal elements (e.g. Si, N, O, F), metals are chemically more complementary to carbon in terms of electronegativity and bonding behavior. Therefore, combination of metal and carbon could give novel properties, as is demonstrated in the rapidly growing organometallic chemistry in recent years (Frenking 2001). Because both metal and carbon have been extensively studied as hydrogen-storage materials, combining the two may be a good strategy for picking up the advantages and circumventing the disadvantages of the HSMs formed of either pure metal or pure carbon. The role of metals in carbon-based hydrogen sorbents lies in two aspects: first, dispersed TM atoms are dihydrogen sorption centers; and second, metals could be used to improve the reaction kinetics of hydrogen chemisorption and tune the H binding energy to carbon atoms (Zhao *et al.* 2007). We will see that both these purposes rely crucially on understanding carbon–metal interaction and its effect on carbon–carbon, metal–metal, carbon–hydrogen, and metal–hydrogen interaction.

#### 20.4.2.1 Carbon–metal interactions

Most TMs form metal carbides (Oyama and Keiffer 1992) through polar covalent bonds, as a consequence of the moderate difference in electronegativity of TM and carbon. The covalent-bonding nature in metal carbides is not as obvious as in the typical covalent crystals such as diamond. Both carbon and metal have large coordination number in TM carbides. For example, TiC has a rocksalt structure, in which each carbon or Ti atoms has six nearest neighbors. In this sense, the TiC solid is similar to an ionic crystal and Ti is indeed less electronegative than C. On the other hand, if we do not distinguish the two types of atoms, the TiC structure is a simple cubic structure like a pure metal, which favors higher coordination numbers. In fact, carbides formed with the TM elements from the left side of the Periodic Table are more ionic because these TM atoms are more electropositive. Carbides formed with TM elements from the right side are more metallic, due to more valence electrons present in these metals. This explains the decreasing melting point of TM carbide with the TM moving from left to right in the Periodic Table (Oyama and Keiffer 1992). But the covalent-bonding nature becomes clearer in small metal–carbide clusters, such as metallocarbohedrene (MetCar,  $Ti_8C_{12}$ ) (Guo *et al.* 1992), in which carbon atoms appear as dicarbon units. The two



**Fig. 20.7** Bonding framework in a  $\text{Ti}_8\text{C}_{12}$  MetCar molecule.

carbon atoms in each dicarbon are connected with a triple bond in between and bond covalently with Ti atoms from both sides (Fig. 20.7) (Pilgrim and Duncan 1993). The  $\text{C}_2\text{Ti}_2$  motif in MetCar resembles the  $\text{C}_2\text{H}_2$  molecule.

In MetCars, non-classical coordination is also important. Dewar coordination between carbon  $\pi$ -orbitals and TM  $d$ -orbitals (Frenking 2001) is similar to Kubas coordination between the  $\sigma$ -orbital of a  $\text{H}_2$  molecule and TM  $d$ -orbitals. Dewar coordination allows for the TM atoms to be decorated on carbon molecules such as fullerenes or CNTs (Zhang and Dai 2000; Yildirim and Ciraci 2005; Zhao *et al.* 2005), but the binding is not as strong as C-TM covalent bonds. Boron-doped carbon cages can enhance the cage–TM binding significantly (Zhao *et al.* 2005).

In principle, carbon can form pure ionic bonds with extremely electropositive alkaline-metal atoms. However, alkaline-metal carbides are practically unstable because of the high M–C stoichiometric ratio (4:1) required by their relative valence. If carbon atoms form a low-energy  $sp^2$  shell, whose lowest-unoccupied energy levels are still low enough to accept electrons from the alkaline-metal atoms, this metal-intercalated carbon compound is energetically more favorable. Therefore, in alkaline-metal–carbon compounds, the normal ionic bonds (localized Coulomb interaction) do not exist. In other words, they transform into delocalized Coulomb interaction arising from the charge transfer from metal to carbon sheets.

#### 20.4.2.2 The effect of metal on carbon–carbon interaction

In mild reaction conditions, the extremely strong C–C interaction can largely maintain its identity even when strong metal–carbon interaction is coupled into the carbon network. In the simple-metal intercalated carbon systems, only the anti- $\pi$  orbitals hybridize with the valence orbitals of the simple-metal atoms, with little disturbance of the carbon  $\pi$  bonds. The transition metals have a stronger influence and their empty  $d$ -orbital(s) hybridize with both the anti- $\pi$  and  $\pi$  bonds in carbon. But the C–C  $\sigma$ -bond states are too low to be affected by the TM atoms. Generally, a TM atom can interact with any covalent

bonds through 3c–2e non-classical coordination, which is the origin of TM-based catalysis. The strength of the interaction depends on the energy level or stability of the covalent bond (Kubas 2001). The order of energy levels from high to low is  $\pi(\text{C–C})$ ,  $\sigma(\text{H–H})$ ,  $\sigma(\text{C–H})$ , and  $\sigma(\text{C–C})$ , therefore the strength of the 3c–2e interaction drops from TM– $\pi(\text{C–C})$ , to TM– $\sigma(\text{H–H})$ ,  $\sigma(\text{C–H})$ , and TM– $\sigma(\text{C–C})$ . This understanding is important for choosing space-opening blockers in designing organometallic hydrogen sorbents: the best blockers should have no  $\pi(\text{C–C})$  bonds to avoid blocker–TM interaction so that the TM centers can be kept highly unsaturated.

The presence of metal in carbon not only affects the intrinsic carbon–carbon bonds, but also alters the complex interaction such as the resonant effect and strain. A famous example is  $\text{C}_{60}\text{Li}_{12}$  (Kohanoff *et al.* 1992), in which aromatic five-membered rings are created after 1-e is transferred from each Li to each of the pentagons in  $\text{C}_{60}$ . Such an unusual aromaticity in fullerenes was termed an hexagon isolation rule (Jemmis *et al.* 2000). Another example is the violation of the isolated pentagon rule (IPR) in endohedral metallofullerenes (Beavers *et al.* 2006) by an interaction mechanism unknown so far.

#### 20.4.2.3 *The effect of carbon on metal–metal interaction*

Carbon is more electronegative than metals, thus it positively charges the metal atoms. This effect dominates in metal-intercalated carbon materials. Metal-intercalated carbon compounds cannot be directly used for hydrogen storage, because there is no open space for hydrogen gas. However, the principle of metal dispersion in these compounds is applicable to hydrogen sorbents with dispersed metal atoms. Transfer of the metal valence electrons to carbon on the one hand weakens the metal–metal bonds, and on the other hand induces Coulomb repulsion between the metal atoms. The more electropositive and the smaller the cohesive energy, the easier are the metal atoms separated. Currently, only the alkaline metals are routinely intercalated in carbon, and the least electropositive metal intercalated in carbon is calcium (Calandra and Mauri 2005). It would be a big challenge to disperse the initial 3d TM atoms of the Periodic Table (Sc, Ti). To achieve this, the only way is to increase the electronegativity of the substrate by incorporation of the electron-deficient boron atoms.

#### 20.4.2.4 *The effect of metal on carbon–hydrogen binding*

In the spillover approach to hydrogen storage, TM particles are used as catalyst for  $\text{H}_2$  dissociation (Lueking and Yang 2002). It is proposed that the catalytic metal particles actually provide a source for hydrogen atoms, which diffuse through the deliberately designed “bridges” to the surface of carbon materials such as activated carbon or CNTs (Yang *et al.* 2006). In this situation, the metal has no direct influence on carbon–hydrogen interaction. If the metal particles are placed on the surface of the carbon materials, charge transfer and locally induced strain may help the initial hydrogenation and even hydrogen diffusion. Overall, the metal influence on carbon–hydrogen interaction is not significant.

However, if metal centers are evenly embedded in carbon, the C–H interaction will be changed substantially, as is demonstrated in hydrogenation of MetCars (Zhao *et al.* 2006). Here, the TM atoms not only catalyze the hydrogenation of carbon atoms, but also weaken the H binding to the car-

bon atoms significantly. The interaction mechanism can be explained with Fig. 20.7, the Ti atoms 1 and 1' are more reactive and can catalyze dissociation of H<sub>2</sub> molecules, which spill H atoms over to the dicarbon (2 – 2') in-between. But the C–H bonds so formed are much weaker than that in a free C<sub>2</sub>H<sub>4</sub> molecule, because the triple carbon bond in MetCar is stabilized by the Ti atom 3 through Dewar coordination. Importantly, the hydrogen dissociation and recombination is highly reversible through the catalytic effect (Zhao *et al.* 2006).

An unexpectedly interesting case of carbon hydrogenation affected by metal has been found in the endohedral metallofullerenes. With the increase of hydrogen coverage, hydrogenation of empty fullerenes usually is divided into two stages by a critical coverage (e.g.  $x = 36$  for C<sub>60</sub>H<sub>x</sub>), at which the strain is optimized. The hydrogen binding is strong ( $\sim 1.0\text{ eV/H}_2$  for C<sub>60</sub>) before reaching this critical coverage, and drops sharply after that. This is not good for hydrogen storage, which requires a flat linear binding behavior with intermediate value ( $\sim 0.3\text{ eV/H}_2$ ) in a wide coverage range (see Section 20.6.2). Surprisingly, metal atoms encapsulated inside the fullerene cages generally weaken the H binding in the earlier stage and strengthen it in the later one, independent of the reaction pathways (Zhao *et al.* 2007). The charge transfer from the inside metal atoms stabilizes the outside carbon cage and therefore H binding in the earlier stage is weakened. However, hydrogenation in the later stage becomes more favorable because it induces negative curvature at the naked carbon sites, which form chemical bonds with the encapsulated metal atoms. It is quite surprising that a few hidden metal atoms can systematically affect such a complicated reaction by altering the patterns of hydrogen coverage on the surface of the fullerenes. For example, the inside metal atoms may also stabilize some special structures, e.g. the naked five-membered rings, which are forbidden in hydrogenation of empty fullerenes.

### 20.4.3 Internal interaction in boron-based materials

Boron is placed right before carbon in the Periodic Table of elements and has an atom size and orbital configuration very similar to those of the latter. Therefore, one expects for boron a chemical richness and important role in organic chemistry similar to carbon. However, boron shows its uniqueness in dramatic difference from carbon, due simply to the “shortage” of one valence electron compared with carbon. The most straightforward consequence of such an electron deficiency is that pure boron-based materials favor a coordination number higher than that of carbon. In the famous 12-atom icosahedral structure, each boron atom has a coordination number of 5, in sharp contrast to the layered structure of graphite. This explains why boron does not favor high surface area structures like active carbon. In this sense, boron-based materials may not be good gas sorbents. However, with compensation of one electron per boron atom, boron-based materials spontaneously transform into bonding networks similar to these of carbon materials. The most common examples are boron nitrides and boron–metal compounds.

**Table 20.2** Reaction heat (kJ/mol-H<sub>2</sub>) of hydrogenation of small hydrocarbon and ammonia-borane molecules.

$\text{C}_2\text{H}_2 + \text{H}_2 \rightarrow \text{C}_2\text{H}_4$	$\text{C}_2\text{H}_4 + \text{H}_2 \rightarrow \text{C}_2\text{H}_6$	$\text{C}_6\text{H}_6 + 3\text{H}_2 \rightarrow \text{C}_6\text{H}_{12}$
180	200	90
$\text{BNH}_2 + \text{H}_2 \rightarrow \text{BNH}_4$	$\text{BNH}_4 + \text{H}_2 \rightarrow \text{BNH}_6$	$\text{B}_3\text{N}_3\text{H}_6 + 3\text{H}_2 \rightarrow \text{B}_3\text{N}_3\text{H}_{12}$
168	12	-3.5

#### 20.4.3.1 BN compounds

The BN system is analogous to the C material in that they both have 1D atomic chains, 2D sheets or shell, and 3D diamond-like structures via different types of hybridizations. The  $sp^3$  hybridization in BN crystal is similar to that in diamond, except for that the B–N (C–C) bonds are polar (non-polar) ones. But the hybridizations in 1D chains or 2D sheets of BN systems differ more from the standard  $sp$  and  $sp^2$  for carbon. To describe the origin of this difference, we first examine the electronic structures of  $\text{BH}_3$  and  $\text{NH}_3$  molecules, both of which are  $sp^2$  configuration. However, there is a hole state in  $\text{BH}_3$ , while the  $\text{NH}_3$  has a lone electron pair. Therefore, a double bond in BN sheets should contain a  $\sigma$  bond and a non-classical bond, rather than a  $\sigma$  bond and a classical  $\pi$  bond. For the same reason, each triple bond in a B–N chain has a  $\sigma$  component, a  $\pi$  component, and a non-classical component. The different nature of hybridization in BN and C systems should be reflected in their  $sp$ – $sp^2$ – $sp^3$  transitions, as shown in Table 20.2 for the reaction heats of sequential hydrogenation of their small molecules. The  $sp$ – $sp^2$  transitions (shown by the first column in Table 20.2) in carbon and BN systems are very similar, but  $sp^2$ – $sp^3$  transitions (the second column) are completely different. The difference originates from the fact that the rehybridization consumes, respectively, a classical  $\pi$  bond in the carbon system and a non-classical bond in the BN system. Obviously, the non-classical B–N bond is more stable than the C–C or B–N  $\pi$  bonds. This is also evidenced by the fact that TM atoms bind less strongly to  $\text{B}_3\text{N}_3\text{H}_6$  than to the benzene ring (Shevlin and Guo 2006). The other distinguishing result is the reaction-heat differences of the  $sp^2$ – $sp^3$  transitions in diatomic molecules and the aromatic rings. Due to the aromaticity, hydrogenation of benzene ring gives reaction heat 110 kJ/mol-H<sub>2</sub> less than that of  $\text{C}_2\text{H}_4$ . Such a difference is only 15.5 kJ/mol-H<sub>2</sub> for the BN systems, indicating a much less pronounced resonant effect of the non-classical bonds (at least in this particular case) than that of the delocalized  $\pi$  bonds. Obviously, hydrogen storage through hydrogenation of BN nanotube is not feasible.

The intermolecular interaction in BN systems is stronger than that in hydrocarbon, due to the polarized B–N, B–H, and N–H bonds. For example,  $\text{C}_2\text{H}_6$  is a liquid with low vaporization temperature, but  $\text{BH}_3\text{NH}_3$  molecules form a solid at ambient condition. This renders the ammonia-borane a potential material for hydrogen storage. But it suffers from slow kinetics and irreversibility once larger BN molecules such as borazine ( $\text{B}_3\text{N}_3\text{H}_6$ ) are formed.

#### 20.4.3.2 Boron–metal compounds

Boron can form layered structures in a honeycomb lattice with metal atoms sandwiched by the B-sheets (Nagamatsu *et al.* 2001; Ivanovskaya *et al.*

2003), in analogy to metal-intercalated graphene (Calandra and Mauri 2005). Recently, metal-boride nanotubes were proposed theoretically for potential application in superconductivity (Buzea and Yamashita 2001; Ivanovskaya *et al.* 2003) and hydrogen storage (Meng *et al.* 2007). Due to the electron deficiency, boron substrate (sheets, tubes, cages) are much more electronegative than their carbon analogies, which enhances significantly the charge transfer from metal to boron substrate and eventually can separate TM atoms with large cohesive energy like Ti. TM-coated boron nanotubes or cages repel each other because of the radial polarity with the same charge in the outer sphere (Zhao *et al.* 2005). This is good for opening space for hydrogen sorption, however, also a serious challenge for synthesis, because the stronger the interlayer Coulomb interaction, the higher the energy of the nanostructure wrapped up from a single layer than the bulk stack of flat layers. Even if the nanostructures can be formed, when hydrogen is charged, possible hydride formation could firmly seal off the interstitial space and again form a solid. In order to prevent this, one should use the right blockers or cross-linkers, and obviously, more spherical nanoparticles are easier to open space than nanotubes. In fact, TM-embedded metal-boride nanostructures are more stable than the TM-coated ones (Zhao *et al.* 2008).

## 20.5 Structures of hydrogen sorbents

### 20.5.1 Sorption sites and basic quantities for measurement

A hydrogen-storage structure can be considered as arrays of hydrogen-sorption sites. For physisorption, a sorption site is a particular unit dispersive surface area defined by the effective contact area  $s$  in eqn (20.16). For chemisorption, a sorption site is an active atom, e.g. a carbon atom or a metal atom. The occupation number, or number of hydrogen atoms adsorbed on a sorption site is:

$$n_H = \begin{cases} 2, & \text{for physisorption} \\ 1; & \text{for hydrocarbon} \\ 18 - n_s^e - n_v^e; & \text{for TM atoms} \end{cases}, \quad (20.20)$$

where  $n_s^e$  is the number of electrons contributed by the substrate to the TM orbitals, and  $n_v^e$  is the number of valence electrons in the TM atom.

We define here a quantity measuring the total number of sorption sites ( $N_s$ ) divided by the total mass ( $M$ ) of the HSMs, or the density of sorption sites, which is related to the SSA and effective contact surface area per sorption site ( $s_0$ ) by:

$$D_s = N_s/M = S_{SSA}/s_0. \quad (20.21)$$

For simplicity, here the mass is measured in atomic units and that of hydrogen atom is approximated by a unit mass. If the material density is  $D_{HSM}$ , the gravimetric capacity ( $C_W$ ) and volumetric capacity ( $C_V$ ) are:

$$C_W = N_s n_H / M = D_s n_H, \quad (20.22)$$

and

$$C_V = D_s n_H D_{HSM}. \quad (20.23)$$

The condition (19) satisfied in a sorbent defined by eqn (20.6) does not guarantee a high hydrogen-storage capacity, unless it is reinforced by a second condition:

$$\text{Max}(D_s = S_{SSA}/s_0). \quad (20.24)$$

These structure-related quantities are useful for evaluation of the basic properties of a hydrogen sorbent. For an immediate application of the above theory, below we briefly analyze the case of physisorption.

The gravimetric capacity of a physisorption sorbent is

$$C_W = 2S_{SSA}/s_0. \quad (20.25)$$

Under the condition of constant  $s_0$ , the gravimetric capacity of a physisorption sorbent is proportional to the SSA. When there is no micropore in the sorbent (local curvature radius  $r \gg r_\Omega$ ), using  $s_0 = s_\Pi = 0.19 \text{ nm}^2/\text{H}_2$  (see Section 20.3.2), one obtains the scaling constant of 1 wt.%  $\sim 570 \text{ m}^2/\text{g}$ , which is close to the Chahine rule of 1 wt.%  $\sim 500 \text{ m}^2/\text{g}$  (Chahine and Bose 1994). Notice that for graphene sheet, we have  $S_{SSA} \sim 3000 \text{ m}^2/\text{g}$ . So the ideal gravimetric capacity would be 6 wt.%. But this has to be achieved at cryogenic temperature because the binding energy is too small. One could increase the binding energy or  $s_0$  by introducing micropores, but this will lead to a lowering of capacity according to eqn (20.21). For example, to achieve a minimum H binding energy of 20 kJ/mol, the effective contact area per  $\text{H}_2$  molecules must exceed  $0.8 \text{ nm}^2$ . A hydrogen sorbent with  $S_{SSA} \sim 3000 \text{ m}^2/\text{g}$  can only store 0.8 wt.%. Therefore, pure physisorption sorbents are not suitable for practical hydrogen storage, agreeing with large-scale simulation (Wang and Johnson 1999). However, a recent theory suggested an extraordinarily high capacity with nanographene (Patchkovskii *et al.* 2005).

## 20.5.2 Nanoscale building blocks for hydrogen sorbents

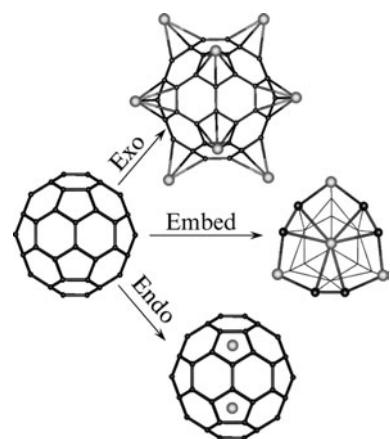
In a pioneering work of fullerene synthesis and hydrogenation, Haufler *et al.* concluded their paper with the following vision:

*“These substances may require a new and divergent meaning for the terms ‘outer sphere’ and ‘inner sphere’ in describing organometallic complexes. For the fullerenes there is no doubt that there will be a rich chemistry associated with the ‘outside’ of the molecule. We have given one such example here in the formation of  $C_{60}\text{H}_{36}$ . True outside complexes will certainly also exist for these fullerenes with a wide range of metals and ligands. But the fullerenes possess a unique sort of ‘inside’ space as well. Here a wide range of atoms from throughout the periodic table may reside, perhaps with little direct influence on the overall ‘outside’ chemistry. Although chemically rather hidden, their ability to fine-tune the optical and redox properties of the fullerenes may be highly useful.”* (Haufler *et al.* 1990)

On the basis of observation of the unique structure of the fullerene molecule, the authors envisioned an exciting research direction in chemistry, which has vigorously become true in terms of fullerene functionalization from both “outside” (Balch and Olmstead 1998; Reed and Bolksar 2000; Sawamura *et al.* 2002) and “inside” (Bethune *et al.* 1993; Shinohara 2000; Akasaka and Nagase 2002). Here, we emphasize that these stimulating statements also initiate a philosophy of hydrogen-sorbent structuring in nanoscale. That is, functionalized carbon nanostructures, from both outside and the inner side with ligands or metal, play a central role in hydrogen sorbents. These nanoscale building blocks are functional cores, which may go beyond carbon, though the carbon-based nanostructures are extraordinarily important.

In addition to the core structure units, building blocks such as cross-linkers and blockers are necessary in order to assemble them into a macroscopic material or to open space for H<sub>2</sub> gas.

#### 20.5.2.1 Nanosorbents



**Fig. 20.8** Three types of organometallic nanosorbents with metal atoms supported, embedded, and encapsulated by carbon.

Other organic, inorganic nanosorbents are often studied, for example, ZnO cluster in MOF, SiO<sub>2</sub> in zeolites, and TiO<sub>2</sub> in porous framework (Hu *et al.* 2006).

#### 20.5.2.2 Cross-linkers and blockers

The role of cross-linkers is to link the nanosorbents into a macroscopic framework, just like the organic hydrocarbon units linking the ZnO clusters in MOF. The hydrocarbon linkers have been proposed to be able to cross-link carbon nanotubes to form carbon nanoframeworks (Ding *et al.* 2007; Weck *et al.* 2007). In this case, the cross-linkers open space in CNT bundles, overcoming the vdW interaction.

In contrast to the linker motifs that are chemically parts of the framework, independent molecules can serve as blockers to open space in a compound, which is stable without the blockers. For example, NH<sub>3</sub> and CH<sub>3</sub>NH<sub>2</sub> molecules can further intercalate the K<sub>3</sub>C<sub>60</sub> fullerenes to form new compounds with larger lattice constant (Rosseinsky *et al.* 1993; Margadonna *et al.* 2004; Ganin *et al.* 2006). Other small molecules such as CO, Cp, benzene, and tetrahydrofuran (THF) can be used as blockers to open space. Recently, we proposed THF-Li cointercalation of graphite through first-principles calculation. This novel material may be readily synthesizable for high-capacity hydrogen storage (Zhao 2008).

### 20.5.3 Architecture of hydrogen sorbents

The first type of hydrogen sorbents are assemblies of building blocks, including the functional core and the cross-linkers, into either periodic structures, e.g. metalorganic frameworks (MOF) (Rosi *et al.* 2003; Chae *et al.* 2004) or potentially hierarchical structure, e.g. dendrimers (Tomalia 2004). These ordered solid sorbents are usually formed in a “bottom-up” procedure. The second type of solid sorbents are less ordered, such as activated carbon and carbide-derived-carbon (Gogotsi *et al.* 2003), which are formed with a “top-down” procedure. These structures are highly porous solids with all the sorption sites accessible from free space. If the barrier of hydrogen sorption/desorption is very small, the solid sorbents would have high performance for fast charge and discharge.

However, when the barrier of hydrogen sorption/desorption is large, solid sorbents will suffer from slow kinetics, no matter how good is the binding energy. This happens in carbon sorbents without a metal, where the H atoms are chemisorbed through C–H bonds. To overcome this difficulty, the nanoscale building blocks, e.g. fullerenes, have to condense into liquid or gas rather than solids, so that the nanoscale sorbents can carry hydrogen directly to the charge/discharge devices (e.g. fuel-cell membrane plates) with catalyst. Here, the hydrogenated carbon nanosorbent is similar to CH<sub>4</sub> as a fuel for fuel cells, but the carbon nanoparticles are not burnt into CO<sub>2</sub> gas and can be repeatedly used. This architecture works only when the interaction between the carbon nanosorbents is weak, which is fortunately true, because the carbon or hydrocarbon particles interact with each other through vdW forces. If the carbon nanosorbents are small (like C<sub>60</sub>, C<sub>70</sub>, and C<sub>80</sub>), they will exist as liquid or gas, at least at a slightly elevated temperature.

## 20.6 Required hydrogen-storage properties and design principles (DP)

At least four basic requirements should be satisfied *simultaneously* by any practical hydrogen-storage systems for vehicular application. These are system capacity and net hydrogen delivery, efficiency of energy output, charge/discharge rate, material durability. From a viewpoint of economics and the environment, low cost and low toxicity are also required. Although a full optimization of the whole hydrogen-storage system ultimately relies on engineering design, these practical requirements add fundamental restraints to material research from the very beginning.

The material gravimetric capacity of hydrogen in a HSM is,

$$C_w^m = M_H / (M_H + M_{HSM}) \sim M_H / M_{HSM}, \quad (20.26)$$

where  $M_H$  and  $M_{HSM}$  are the mass of hydrogen and the storage medium, respectively. Ultimately, the weight of the container plus serving systems ( $M_{\text{container}}$ ) must also be accounted. That is, the system capacity is defined as

$$C_w^s = M_H / (M_H + M_{HSM} + M_{\text{container}}). \quad (20.27)$$

**Table 20.3** Delivery percentage of maximum capacity at different operating temperature ranges for hydrogen binding energy of 15.1 kJ/mol, storage pressure 30 bar, and exhaustion pressure 1.5 bar.

$T_2 = 350\text{ K}$		$T_2 = 400\text{ K}$	
$T_1 = 250\text{ K}$	$D = 85.2\%$	$T_1 = 250\text{ K}$	$D = 89.0\%$
260 K	83.3%	260 K	87.1%
270 K	81.1%	270 K	84.9%
280 K	78.6%	280 K	82.3%
290 K	75.8%	290 K	79.6%
300 K	72.8%	300 K	76.6%

The goal for system capacity set by the United States Department of Energy standard is 6–9 wt.% (Satyapal *et al.* 2007). A parallel requirement is the volumetric capacity, which is set at 45 kg/m<sup>3</sup>. As far as residual hydrogen in the HSM cannot be avoided at any particular exhaustion condition, these capacities should be net hydrogen delivery rather than the maximum hydrogen amount contained in the storage system. The maximum material capacity can be determined from eqns (20.7)–(20.9), but the 63.5% deliverable percentage of hydrogen determined from eqns (20.10)–(20.13) for the fixed operation temperature of 298 K is obviously not very satisfying. The delivery can be increased if temperature can be manipulated in an acceptable range, e.g. 250–350 K. Substitute  $T_1 = 250\text{ K}$ ,  $P_1 = 30\text{ bar}$ ,  $T_2 = 350\text{ K}$ ,  $P_2 = 1.5\text{ bar}$ ,  $\Delta S' = -8R$ , and  $\Delta H' = 15.1\text{ kJ/mol}$  into eqns (20.11) and (20.12), the delivery percentage is obtained of 85.2%. Table 20.3 lists the delivery percentage at different operating temperature ranges.

Energy efficiency requires that the storage system is not supposed to consume or waste substantial energy in operation of itself. Reaction control is a major part that subtracts energy from power output. First, binding energy or reaction enthalpy is a key factor that influences the energy efficiency. If the reaction enthalpy is too small, cooling and/or high pressure is needed to enhance hydrogen charging, and in either case extra energy has to be consumed. On the other hand, big binding means a large amount of reaction heat has to be removed (added) in the charging (discharging) process. This may greatly reduce the energy efficiency.

Charging and discharging rates can be different. A charging process must be fast and should last for less than 10 min. The minimum discharge rate  $R$  should be no less than the value determined by:

$$R (\Delta H)_{\text{H}_2} \eta_{\text{HSM}} \eta_{\text{FC}} = P_{\text{vehicle}}, \quad (20.28)$$

where  $(\Delta H)_{\text{H}_2}$  is the combustion heat of hydrogen,  $\eta_{\text{HSM}}$  and  $\eta_{\text{FC}}$  are, respectively, the energy efficiency of HSM system and the fuel-cell system,  $P_{\text{vehicle}}$  is the output power of the vehicle. In fact, any discharge rate larger than  $R$  determined by eqn (20.28) can meet practical use, because the actual discharge rate can always be controlled by pressure. The charge/discharge rate is mostly determined by the reaction pathway of hydrogen and HSM. Generally, the more complex the reaction process, the slower is the charge/discharge kinetics.

A hydrogen-storage system should have the same lifespan as a vehicle, with a typical mileage of 200 000. The minimum mileage for each charge should be greater than 200, therefore, a hydrogen sorbent should last  $\sim$ 1000 charge/discharge cycles. This is quite a tough requirement for the material stability.

Based on these practical requirements, the design principles (DPs) can be derived for hydrogen-storage materials.

### 20.6.1 Sorbent DP

*DP1. The best hydrogen-storage materials should be hydrogen sorbent.*

A good hydrogen sorbent should simultaneously fulfill eqns (20.6), (20.13), (20.19), and (20.24). The sorbent DP is the most important design principle that relates to many aspects of hydrogen-storage properties, as summarized in Table 20.4.

The superior properties of sorbents is associated with their structure, which is featured as surface adsorption of hydrogen, due to the weaker H–HSM interaction than the HSM internal interaction, as is shown in eqn (20.19). This simplifies the charge/discharge process from a phase transition over a complicated energy landscape in non-sorbents to the one of straightforward add-on/jump-off in sorbents. Taking the dihydrogen as an example, at the binding strength of 15 kJ/mol, each individual  $H_2$  molecule keeps jumping on and off the sorption site (TM center) at a frequency as high as  $(10^8 \sim 10^9)/s$  at room temperature. In fact, in this situation, one can not even observe a stable dihydrogen experimentally, but statistically sufficient amounts of dihydrogen are held to the TM centers in every moment at near-ambient pressure. All the dihydrogen molecules are immediately available if the pressure is lowered. For atomic hydrogen sorbents, comparably fast kinetics can be achieved when the catalytic TM atoms are available nearby all sorption sites (Zhao *et al.* 2006).

In Section 2, the direct treatment of non-sorbent is circumvented, because quantitative solution of the phase-transition problem in non-sorbents is currently impossible, though it can be characterized experimentally in some way. Here, we show that a simple picture based on classical statistics may shed some light on this issue.

Theoretically, there are multiple phases of a complex material system and all these phases are linked by intercrossed reaction pathways. The whole collection of the minimum-energy paths (MEPs) sketches the energy landscape quantified by the energy function in eqns (20.1) and (20.2). Reversible

**Table 20.4** Comparison of the properties of the two types of storage media.

Sorbents	Non-sorbents
Structure	Porous solid
H–HSM interaction	Weak to intermediate
Charge/discharge rate	Fast
Energy efficiency	High
Reversibility	Good
	Close-packed solid
	Strong
	Slow
	Low
	Poor

hydrogen storage requires that the two particular phases, corresponding to charged and discharged material systems, are energetically most favorable and they are linked by two exceptionally favorable MEPs. In this situation, we can roughly model the charge/discharge processes as hydrogen diffusion in and out of a sphere through the surface. The discharge rate depends mainly on three parameters. The first is the average diffusion barrier  $E_a$  in H hopping. The second is the average radial length  $\lambda$  per H-hopping step, a quantity similar to the mean-free path projected to the radial direction. The third is the ratio of number of surface atoms over those in the whole volume, which sensitively depends on the size ( $R$ ) of the sphere. In particular, when the ratio approaches 1, the non-sorbent transforms into a sorbent.

Both  $E_a$  and  $\lambda$  are determined by the considered MEPs. As we have discussed in Section 3.1A,  $E_a$  could be tuned to a smaller value with additives. However, more complicated complexes often lead to two risky situations. First, new undesired phases with lower energy could be introduced, which could trap the complexes in such a structure phase (no reversible reaction can happen) or even destroy the HSMs, e.g. by emission of other gases (Meisner *et al.* 2006). Second, introduction of more undesired reaction pathways reduces  $\lambda$ ; therefore, the charging/discharging rate is lowered. Even if these two issues can be solved, the size of the sphere will substantially hinder the charge/discharge. Currently, ball milling of the non-sorbent into powder is widely used to accelerate the kinetics, but ball milling is energy consuming and can hardly be used in a practical system. Recently, it was observed that the hydrogen-discharge rate significantly increases when the HSM are synthesized as nanoparticles dispersed in a porous framework (Gutowska *et al.* 2005). All these features point to an unavoidable trend that the development of non-sorbents eventually has to be directed towards that of sorbents.

## 20.6.2 Binding energy DP

*DP2. The binding energy principle means: (1) The binding energy should fall in an ideal range of 10–30 kJ/mol-H<sub>2</sub>, corresponding to a near ambient operation condition; (2) The ideal binding energy as a function of sorption/desorption process should be a flat line (i.e. no gradient); and (3) The barrier of H sorption / desorption should be low.*

The concept of binding energy, eqn (20.7), can be well defined only for sorbents. Therefore, the binding energy DP puts further constraints on HSM after DP1, for optimization the design of hydrogen sorbents. The three requirements by DP2 are discussed below.

### 20.6.2.1 Binding-energy range

First, an ideal binding-energy range can be determined with eqn (20.8), in association with the near-ambient conditions. The binding energy obviously affects energy efficiency and the system capacity. According to eqn (20.8), the binding energy corresponding to ambient operation conditions is  $\sim 15$  kJ/mol, which is the right value to hold the hydrogen (cancel the entropy term) leading to  $\Delta G(A, H) = 0$ , meaning there is little accompanying heat in the charge/discharge process. However, if the binding energy is too big, the extra

part has to be dissipated out during charging, and same amount of heat has to be provided for discharge. This not only consumes more energy, but also reduces the system capacity because of heat control.

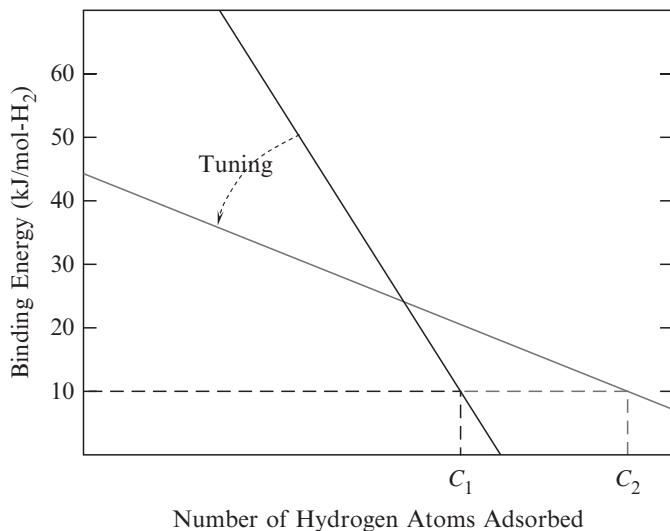
It should be pointed out that if the system works at ambient pressure, a light container with  $M_{\text{container}} \sim 0.1 M_{\text{HSM}}$  should be feasible; the system weight percentage is almost the same as the material weight percentage according to eqn (20.27). However, high-pressure operation may require heavy containers, which significantly reduces the system capacity.

#### 20.6.2.2 Flat linearity of binding energy

Usually, it is expected that there is no long-range correlation of hydrogen binding in sorbents. In the same functional core, the binding energy of hydrogen, especially hydrides, often changes dramatically with H loading, because hydrogenation gradually modifies the property of the sorbent building block. A general decreasing trend of hydrogen binding energy would be expected as shown ideally by the straight lines in Fig. 20.9. The cutoff number of H atoms up-taken ( $C_1$ ,  $C_2$  in Fig. 20.9) determines the hydrogen capacity. Through structural modification of the sorbents, convert the dropping line of binding energy to a more flat line so that the binding energy of all the sequentially added hydrogen can be confined into the ideal window. As a result, the hydrogen capacity is simultaneously increased from  $C_1$  to  $C_2$ .

#### 20.6.2.3 Sorption/desorption barrier

The barrier of hydrogen sorption and desorption is an important part of the hydrogen-binding process. For physisorption and Kubas coordination, the barrier can be neglected, but atomic hydrogen sorption and desorption in carbon cages experiences a substantial barrier. Under the adiabatic approximation, the barrier may not affect the energy efficiency, but will slow down the charge/discharge rate. A catalyst must be used in order to accelerate the reaction.



**Fig. 20.9** Schematic illustration of binding energy dependence of H coverage in nanosorbents and requirement for binding-energy tuning.

### 20.6.3 Capacity DP

*DP3. The gravimetric and volumetric capacity may be simultaneously maximized through integration of all binding mechanisms in one HSM with optimized pore size and structure of building units.*

According to eqns (20.20)–(20.25), there are three ways to increase the gravimetric capacity. Maximization of  $S_{SSA}$  is the first and most effective method. The smaller the building blocks and the lighter is the constituent elements of a hydrogen sorbent, the larger is its specific surface. In designing the hydrogen sorbent, elements such as B, C, N, should be used as the main ingredients to form nanostructures with all the atoms exposed on the surface. The second way is to minimize the effective area per sites  $s_0$ , in order to increase the density of sorption sites. Therefore, it is important to activate every surface atom as sorption sites. Third, increasing the inherent capacity,  $n_H$ , is also a key to capacity maximization. Because  $n_H$  is constant for physisorption and hydrides, the only possibility left is using transition metals as sorption sites. This is critical for organometallic approaches to hydrogen storage (Zhao *et al.* 2008).

Once the gravimetric capacity is determined, the volumetric capacity is proportional to the mass density of the sorbent, according to eqn (20.23). That means too large a pore volume will waste space. Therefore, optimized pore size is important for simultaneous maximization of the volumetric capacity, but the pores size should not be devised only for the two types of weak chemisorption, i.e. strained C–H bonds and Kubas binding. In addition, physisorption space should also be reserved, so that an integration of all binding mechanisms could be realized in one system for maximum capacity. This may be applicable to the nanoparticle fluid, in which interstitial pores may exist, due to the spherical shape and relatively large size. A recent study showed some evidence suggesting that additional physisorption can be enhanced by non-classical dihydrogen binding (Hu *et al.* 2007).

### 20.6.4 Stability DP

*DP4. Stable structure is essential for a hydrogen sorbent with required durability. However, generally there is a contradiction between the stability of the sorbent and its capacity. The optimized design is a proper balance of the two through seeking the metastable structures that are kinetically stable in the operation condition.*

According to eqns (20.24) and (20.25), high capacity requires high SSA, but high SSA often implies poor stability. Calculation showed that graphene segments of carbon hexagonal rings have an extremely high SSA, above  $7000\text{ m}^2/\text{g}$  (Chae *et al.* 2004), but it is not a stable system. An ideal crystal of MOF5 is reported to have a SSA as high as  $4500\text{ m}^2/\text{g}$ , however, even water can destroy MOF5 at room temperature (Greathouse and Allendorf 2006). Generally, there is a contradiction between capacity and stability, which calls for a subtle compromise between these two during its design.

For organometallic sorbents, the main challenge is to fix the highly under-coordinated TM centers firmly on the substrate. According to eqn (20.20), the

less the coordination number, the higher is the inherent capacity of a transition metal. Three tricks can be used for balancing stability and capacity: first, using the strongest but least number of bonds to hold the TM metal; second, taking advantage of the ionic bonds, which hold the TM atoms by sucking its valence electrons and simultaneously separate the TM atoms from each other; third, using light blockers that do not coordinate with the TM atoms, e.g. C–C  $\sigma$ -bonds.

For hydrocarbon nanostructures, the capacity and stability can be considered separately. The main concern for stability is to prevent the carbon nanostructure from coalescing, especially in the presence of external catalysts. Pure carbon cages coalesce easily under catalysis, but encapsulated metal atoms may effectively stabilize the carbon cages against coalescence because the active orbitals (near the Fermi level) of the cages are already coordinated with the endohedral metal atoms.

## 20.7 Summary

In this chapter, we outlined an integrated theory for hydrogen storage starting from the fundamental law and concepts of thermodynamics and quantum mechanics. The concept of hydrogen sorbent is introduced to describe an optimal type of HSMs, in which hydrogen–HSM interaction is much weaker than the internal interactions in the HSMs, to realize fast kinetics, and high efficiency in energy conversion. Optimal hydrogen binding energy and the isotherm of the complete cyclic process for hydrogen charge/discharge are discussed for near-ambient operation conditions in hydrogen sorbents. The mechanism of hydrogen–material interaction is analyzed in terms of bonding between hydrogen and other elements through orbital hybridization. According to the sorbent condition, we show that carbon-based materials could be the best sorbents. For optimal hydrogen binding, boron and/or transition-metal doping of carbon materials are necessary. Physisorption, chemisorptions through weak covalent bonds, non-classical dihydrogen binding, and electrostatic enhanced binding of H<sub>2</sub> in these materials were analyzed in detail. As guidance for HSM design and evaluation, internal interaction in the hydrogen sorbents is discussed. The interaction between components of the HSMs plays a key role in sorbent design and synthesis. Statically, the internal interaction holds the components together for them to form a stable sorbent structure. Dynamically, the internal interaction provides the driving force for structure formation. Of crucial importance for hydrogen storage, the internal interaction affects the interaction between the sorbent and hydrogen. In total, the internal interaction not only determines how to form a stable sorbent, but also allows for the hydrogen-binding properties to be tuned purposely. Based on this theory, structuring and architecture of a hydrogen sorbent is generally considered as arrays of hydrogen-sorption sites on its surface. Basic quantities such as the density of sorption sites, specific surface area, and their relationship with hydrogen-storage properties were analyzed. Typical nanoscale building blocks including, organometallic buckyballs, B-doped carbon cages and carbon nanotubes, endohedral fullerenes, transition-metal carbide nanoparticles, metal-doped MOFs and porous carbon, etc. have been listed and discussed. Finally,

the design principles (DP) of hydrogen sorbents are derived from the above theory with consideration of practical requirements such as energy efficiency, charging/discharge rate, competitive price, reversibility and durability. These principles are binding energy optimization, maximum capacity, material stability, etc.

Hydrogen storage in solid-state materials has been studied for the last half a century and is becoming an extremely active area. The research activities eventually have to be directed to the creation of a reliable manual book for design and synthesis, which contains complete databases of categorized motifs of material structures, interactions, property requirements and their quantified relationship sorted by detailed design rules and general principles. With such a guiding manual, efficient and motivated R&D activities could lead to a practical HSM.

## Acknowledgments

This work is supported by the Office of Science, Basic Energy Sciences, Division of Materials Science, the Office of Energy Efficiency and Renewable Energy Hydrogen, Fuel Cell, and Infrastructure Technologies Program of the US Department of Energy through the Hydrogen Sorption Center of Excellence under Grant No. DE-AC36-99GO10337.

## References

- Akasaka, T., Nagase, S. *Endofullerenes: A New Family of Carbon Cluster* (Kluwer Academic Publishers, Dordrecht, 2002).
- Alapati, S.V., Johnson, J.K., Sholl, D.S. *Phys. Chem. Chem. Phys.* **9**, 1438 (2007).
- Balch, A.L., Olmstead, M.M. *Chem. Rev.* **98**, 2123 (1998).
- Beavers, C.M., Zuo, T.M., Duchamp, J.C., Harich, K., Dorn, H.C., Olmstead, M.M., Balch, A.L. *J. Am. Chem. Soc.* **128**, 11352 (2006).
- Bethune, D.S., Johnson, R.D., Salem, J.R., Devries, M.S., Yannoni, C.S. *Nature* **366**, 123 (1993).
- Bhatia, S.K., Myers, A.L. *Langmuir* **22**, 1688 (2006).
- Blomqvist, A., Araujo, C.M., Srepusharawoot, P., Ahuja, R. *Proc. Natl. Acad. Sci.* **104**, 20173 (2007).
- Buhl, M., Thiel, W., Schneider, U. *J. Am. Chem. Soc.* **117**, 4623 (1995).
- Buzea, C., Yamashita, T. *Supercond. Sci. Technol.* **14**, R115 (2001).
- Calandra, M., Mauri, F. *Phys. Rev. Lett.* **95**, 237002 (2005).
- Chae, H.K., Siberio-Perez, D.Y., Kim, J., Go, Y., Eddaoudi, M., Matzger, A.J., O'Keeffe, M., Yaghi, O.M. *Nature* **427**, 523 (2004).
- Chahine, R., Bose, T.K. *Int. J. Hydrogen Energy* **19**, 161 (1994).
- Chambers, A., Park, C., Baker, R.T.K., Rodriguez, N.M. *J. Phys. Chem. B* **102**, 4253 (1998).
- Chandra, M., Xu, Q. *J. Power Sources* **156**, 190 (2006).
- Chase, Jr., M.W. (ed.) *NIST-JANAF Thermochemical Tables*, 4th edn, vol. **9**. *J. Phys. Chem. Ref. Data Monograph* (1998).

- Chen, P., Wu, X., Lin, J., Tan, K.L. *Science* **285**, 91 (1999).
- Chen, P., Xiong, Z.T., Luo, J.Z., Lin, J.Y., Tan, K.L. *Nature* **420**, 302 (2002).
- Chen, Z.F., King, R.B. *Chem. Rev.* **105**, 3613 (2005).
- Cheng, H.S., Cooper, A.C., Pez, G.P., Kostov, M.K., Piotrowski, P., Stuart, S.J. *J. Phys. Chem. B* **109**, 3780 (2005).
- Crabtree, G.W., Dresselhaus, M.S., Buchanan, M.V. *Physics Today* **57**, 39 (2004).
- Dapprich, S., Frenking, G. *Angew. Chem. Int. Ed. Engl.* **34**, 354 (1995).
- Dillon, A.C., Heben, M.J. *Appl. Phys. A: Mater. Sci. Process.* **72**, 33 (2001).
- Dillon, A.C., Jones, K.M., Bekkedahl, T.A., Kiang, C.H., Bethune, D.S., Heben, M.J. *Nature* **386**, 377 (1997).
- Ding, F., Lin, Y., Krasnov, P.O., Yakobson, B.I. *J. Chem. Phys.* **127**, 164703 (2007).
- Ferro, Y., Marinelli, F., Allouche, A. *Chem. Phys. Lett.* **368**, 609 (2003).
- Frenking, G. *J. Organomet. Chem.* **635**, 9 (2001).
- Florusse, L.J., Peters, C.J., Schoonman, J., Hester, K.C., Koh, C.A., Dec, S.F., Marsh, K.N., Sloan, E.D. *Science* **306**, 469 (2004).
- Gagliardi, L., Pyykkö, P. *J. Am. Chem. Soc.* **126**, 15014 (2004).
- Ganin, A.Y., Takabayashi, Y., Bridges, C.A., Khimyak, Y.Z., Margadonna, S., Prassides, K., Rosseinsky, M.J. *J. Am. Chem. Soc.* **128**, 14784 (2006).
- Gogotsi, Y., Nikitin, A., Ye, H.H., Zhou, W., Fischer, J.E., Bo, Y., Foley, H.C., Barsoum, M.W. *Nature Mater.* **2**, 591 (2003).
- Greathouse, J.A., Allendorf, M.D. *J. Am. Chem. Soc.* **128**, 10678 (2006).
- Grimes, R.N. *Chem. Rev.* **92**, 251 (1992).
- Gulseren, O., Yildirim, T., Ciraci, S. *Phys. Rev. B* **66**, 045409 (2002).
- Guo, B.C., Kerns, K.P., Castleman, A.W. *Science* **255**, 1411 (1992).
- Gutowska, A. et al. *Angew. Chem., Int. Ed. Engl.* **44**(23), 3578 (2005).
- Harrison, W.A. *Electronic Structure and the Properties of Solids* (San Francisco, W.H. Freeman, 1980).
- Hathorn, B.C., Sumpter, B.G., Noid, D.W. *Phys. Rev. A* **64**, 022903 (2001).
- Haufler, R.E. et al. *J. Phys. Chem.* **94**, 8634.
- Hermanek, S. *Chem. Rev.* **92**, U175 (1992).
- Hornekaer, L., Sljivancanin, Z., Xu, W., Otero, R., Rauls, E., Stensgaard, I., Laegsgaard, E., Hammer, B., Besenbacher, F. *Phys. Rev. Lett.* **96**, 156104 (2006).
- Hu, X., Skadtchenko, B.O., Trudeau, M., Antonelli, D.M. *J. Am. Chem. Soc.* **128**, 11740 (2006).
- Hu, X., Trudeau, M., Antonelli, D.M. *Chem. Mater.* **19**, 1388 (2007).
- Ivanovskaya, V., Enjashin, A.N., Sofronov, A.A., Makurin, Y.N., Medvedeva, N.I., Ivanovskii, A.L. *J. Molec. Struct.-Theochem.* **625**, 9 (2003).
- Jacobson, M.Z., Colella, W.G., Golden, D.M. *Science* **308**, 1901 (2005).
- Jemmis, E.D., Manoharan, M., Sharma, P.K. *Organometallics* **19**, 1879 (2000).
- Kertesz, M., Choi, C.H., Yang, S.J. *Chem. Rev.* **105**, 3448 (2005).
- Kim, Y.-H., Zhao, Y.F., Williamson, A., Heben, M.J., Zhang, S.B. *Phys. Rev. Lett.* **96**, 016102 (2006).
- Kittel, C. *Introduction to Solid State Physics*, 7th edn (New York, Wiley, 1960).
- Kohanoff, J., Andreoni, W., Parrinello, M. *Chem. Phys. Lett.* **198**, 472 (1992).
- Kubas, G.J. *J. Organomet. Chem.* **635**, 37 (2001).

- Kubas, G.J., Ryan, R.R., Swanson, B.I., Vergamini, P.J., Wasserman, H.J. *J. Am. Chem. Soc.* **106**, 451 (1984).
- Kudin, K.N., Scuseria, G.E., Yakobson, B.I. *Phys. Rev. B* **64**, 235406 (2001).
- Lee, H. *et al. Nature* **434**, 743 (2005).
- Li, J., Ziegler, T. *Organometallics* **15**, 3844 (1996).
- Li, Y.W., Yang, R.T. *J. Am. Chem. Soc.* **128**, 8136 2006.
- Liu, C., Fan, Y.Y., Liu, M., Cong, H.T., Cheng, H.M., Dresselhaus, M.S. *Science* **286**, 1127 (1999).
- Lochan, R.C., Head-Gordon, M. *Phys. Chem. Chem. Phys.* **8**, 1357 (2006).
- Lu, T., Goldfield, E.M., Gray, S.K. *J. Phys. Chem. B* **107**, 12989 (2003).
- Lueking, A., Yang, R.T. *J. Catal.* **206**, 165 (2002).
- Lueking, A., Yang, R.T. *Appl. Catal. A: General* **265**, 259 (2004).
- Mao, W.L., Mao, H.-K. *Proc. Natl. Acad. Sci. USA* **101**, 708 (2004).
- Margadonna, S., Iwasa, Y., Takenobu, T., Prassides, K. Structural and Electronic Properties of selected fulleride salts. In *Fullerene-based Materials: Structures and Properties* (Springer-Verlag, Berlin, 2004), 127–164.
- Mediati, M., Tachibana, G.N., Jensen, C.M. *Inorg. Chem.* **31**, 1827 (1992).
- Meisner, G.P., Scullin, M.L., Balogh, M.P., Pinkerton, F.E., Meyer, M.S. *J. Phys. Chem. B* **110**, 4186 (2006).
- Meng, S., Kaxiras, E., Zhang, Z. *Nano Lett.* **7**, 663 (2007).
- Mpourmpakis, G., Froudakis, G.E. *Catal. Today* **120**, 341 (2007).
- Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y., Akimitsu, J. *Nature* **410**, 63 (2001).
- Nijkamp, M.G., Raaymakers, J., van Dillen, A.J., de Jong, K.P. *Appl. Phys. A: Mater. Sci. Process* **72**, 619 (2001).
- Oyama, S.T., Keiffer, R. *Kirk-Othmer Encyclopedia of Chemical Technology*, (ed.) Howe-Grant, M.E. (New York, John Wiley & Sons, 1992).
- Panella, B., Hirscher, M., Roth, S. *Carbon* **43**, 2209 (2005).
- Park, S., Srivastava, D., Cho, K. *Nano Lett.* **3**, 1273 (2003).
- Patchkovskii, S., Tse, J.S., Yurchenko, S.N., Zhechkov, L., Heine, T., Seifert, G. *Proc. Natl. Acad. Sci. USA* **102**, 10439 (2005).
- Phillips, J.C. *Rev. Mod. Phys.* **42**, 317 (1970).
- Pilgrim, J.S., Duncan, M.A. *J. Am. Chem. Soc.* **115**, 9724 (1993).
- Reed, C.A., Bolskar, R.D. *Chem. Rev.* **100**, 1075 (2000).
- Reuter, K., Scheffler, M. *Phys. Rev. B* **65**, 035406 (2001).
- Rosi, N.L., Eckert, J., Eddaudi, M., Vodak, D.T., Kim, J., O’Keeffe, M., Yaghi, O.M. *Science* **300**, 1127 (2003).
- Rosseinsky, M.J., Murphy, D.W., Fleming, R.M., Zhou, O. *Nature* **364**, 425 (1993).
- Saillard, J.Y., Hoffmann, R. *J. Am. Chem. Soc.* **106**, 2006 (1984).
- Satyapal, S., Petrovic, J., Read, C., Thomas, G., Ordaz, G. *Catal. Today* **120**, 246 (2007).
- Sawamura, M., Kuninobu, Y., Togano, M., Matsuo, Y., Yamanaka, M., Nakamura, E. *J. Am. Chem. Soc.* **124**, 9354 (2002).
- Schreiner, P.R., Schaefer, H.F., Schleyer, P.V. *J. Chem. Phys.* **101**, 7625 (1994).

- Schulz, M.G., Diehl, T., Brasseur, G.P., Zittel, W. *Science* **302**, 624 (2003).
- Shevlin, S.A., Guo, Z.X. *Appl. Phys. Lett.* **89**, 153104 (2006).
- Shinohara, H. *Rep. Progr. Phys.* **63**, 843 (2000).
- Slanina, Z., Pulay, P., Nagase, S. *J. Chem. Theory Comput.* **2**, 782 (2006).
- Stojkovic, D., Zhang, P., Lammert, P.E., Crespi, V.H. *Phys. Rev. B* **68**, 195406 (2003).
- Suess, E., Bohrmann, G., Greinert, J., Lausch, E. *Sci. Am.* **281**, 76 (1999).
- Tague, T.J., Andrews, L. *J. Am. Chem. Soc.* **116**, 4970 (1994).
- Taylor, R. *J. Chem. Soc.-Perkin Trans. 2*, 1667 (1992).
- Tomalia, D.A. *Aldrichimica Acta* **37**, 39 (2004).
- Tomas, J., Lledos, A., Jean, Y. *Organometallics* **17**, 4932 (1998).
- Tromp, T.K., Shia, R.L., Allen, M., Eiler, J.M., Yung, Y.L. *Science* **300**, 1740 (2003).
- Turner, J.A. *Science* **285**, 687 (1999).
- Vitillo, G.J., Damin, A., Zecchina, A., Ricchiardi, G. *J. Chem. Phys.* **122**, 114311 (2005).
- Wang, Q.Y., Johnson, J.K. *Molecular physics* **95**, 299 (1998).
- Wang, Q.Y., Johnson, J.K. *J. Chem. Phys.* **110**, 577 (1999).
- Weck, P.F., Kim, E., Balakrishnan, N., Cheng, H.S., Yakobson, B.I. *Chem. Phys. Lett.* **439**, 354 (2007).
- Yakobson, B.I., Brabec, C.J., Bernholc, J. *Phys. Rev. Lett.* **76**, 2511 (1996).
- Yang, F.H., Lachawiec, A.J., Yang, R.T. *J. Phys. Chem. B* **110**, 6236 (2006).
- Yildirim, T., Ciraci, S. *Phys. Rev. Lett.* **94**, 175501 (2005).
- Yoshida, Z.I., Dogane, I., Ikehira, H., Endo, T. *Chem. Phys. Lett.* **201**, 481 (1993).
- Zecho, T., Guttler, A., Sha, X.W., Lemoine, D., Jackson, B., Kuppers, J. *Chem. Phys. Lett.* **366**, 188 (2002).
- Zhang, Y., Dai, H.J. *Appl. Phys. Lett.* **77**, 3015 (2000).
- Zhao, Y., Kim, Y.-H., Dillon, A.C., Heben, M.J., Zhang, S.B. *Phys. Rev. Lett.* **94**, 155504 (2005).
- Zhao, Y., Kim, Y.-H., Simpson, L.J., Dillon, A.C., Wei, S.-H., Heben, M.J. *Phys. Rev. B* **78**, 144102 (2008).
- Zhao, Y., Lusk, M.T., Dillon, A.C., Heben, M.J., Zhang, S.B. *Nano Lett.* **8**, 157 (2008).
- Zhao, Y., Dillon, A.C., Kim, Y.-H., Heben, M.J., Zhang, S.B. *Chem. Phys. Lett.* **425**, 273 (2006).
- Zhao, Y., Heben, M.J., Dillon, A.C., Simpson, L.J., Blackburn, J.L., Dorn, H.C., Zhang, S.B. *J. Phys. Chem. C* **111**, 13275 (2007).
- Zuttel, A., Rentsch, S., Fischer, P., Wenger, P., Sudan, P., Mauron, P., Emmenegger, C. *J. Alloys Compd.* **356**, 515 (2003).
- Zuttel, A., Sudan, P., Mauron, P., Wenger, P. *Appl. Phys. A: Mater. Sci. Process.* **78**, 941 (2004).

# 21

# Electron cold sources: Nanotechnology contribution to field emitters

21.1 Introduction	736
21.2 Driving forces for the evolution of cold cathodes	737
21.3 Single-atom emitters	739
21.4 Use of single-atom nanotip: The Fresnel projection microscope	742
21.5 Use of single-atom nanotip: The microgun	746
21.6 Material issues for field emitters: Carbon nanocompounds	748
21.7 Carbon-nanotube field emitters	749
21.8 Carbon-nanopearl field emitters	761
21.9 Applications and uses of carbon nanocompounds, CNTs and CNPs, as cold cathodes	765
21.10 Conclusions	769
Appendix 1: Electron emission from solids	770
Appendix 2: Tip-profile evolutions by mass-transport surface self-diffusion	774
Appendix 3: Field factor $\beta$ and field-enhancement factor $\gamma$	782
References	785

*Vu Thien Binh*

## 21.1 Introduction

The evolution of the production and the uses of electron beams are intimately linked with the advancement of physics and simultaneous growth of technology occurring over different periods. From the first discovery of electrons ejected from solid carbon material at high temperature (Thompson 1897), the subsequent industrial applications of electron beams in various fields<sup>1</sup> may be traced to the success achieved in realizing vacuum inside a closed chamber, a technology initially based on the so-called Geissler glass tubes of 1854.

A second breakthrough for electron emission occurred with the introduction of quantum mechanics during the 1930s. The concept of electron emission by tunnelling through the surface barrier led to the field emission (FE) or cold emission process. The main instigator of this field was Erwin Muller when he started his thesis in FE (1935) under the direction of Gustav Hertz from Siemens (Drechsler 1978). Muller's main initial contribution was the introduction of well-outgassed field emitter tips within a radial projection microscope (Muller 1938). The tipped geometry cathodes allowed him to reach easily electric fields in the range of  $10^8$  V/cm, required to initiate the tunnelling process. It further allowed him to gain experience in physics of controlled ultraclean surfaces at a time when the concept of ultrahigh vacuum had remained to be established. The strict control of the field emitter surface state with comparison of emission data using quantum mechanics of Fowler and Nordheim (1928) have led to the development of field (electron) emission microscopy (FEM) and field ion (emission) microscopy (FIM) techniques. There was a marked growth of related basic physics of solid surfaces and the advent of industrial applications such as, for example, the FE flat-panel displays (FED).

<sup>1</sup> As for example, CRTs (Braun's tube in 1897), vacuum electronic diodes (Fleming's "valve" in 1904), triodes (De Forest's triode in 1906) and X-ray tubes (Coolidge tube in 1917).

This chapter presents a comprehensive account of the recent evolution of FE cathodes and their applications taking stock of exciting possibilities emerging from nanotechnology. Surely, it is not meant to be an extensive review of cold cathodes for which the reader may refer to several review articles and books already available in the literature (Herring and Nichols 1949; Muller 1953; Dyke and Dolan 1956; Gomer 1961; Muller and Tsong 1969; Swanson and Bell 1973; Modinos 1984; Miller *et al.* 1996; Xu and Huq 2005). Broadly, cold cathodes can be classified into two categories: the planar cathodes and the tipped-geometry cathodes called field emitters. The planar cathodes have already been the subject of a review (Binh and Semet 2007); therefore this chapter will mostly address the field emitters with their specifications. The structure of this contribution is divided into a core text and three appendices. In the core part, the known physical specifications related to a tipped geometry will be reviewed. This will be followed by different examples where the advent of nanotechnology has contributed towards improving new cold cathodes leading to a global miniaturization of instrumentations and devices. Most of the physics details concerning the different basic processes or mechanisms are presented in the three appendices: (1) Electron emission from solid, (2) Tip-profile evolutions by mass-transport surface self-diffusion and (3) Field factor  $\beta$  and field-enhancement factor  $\gamma$ .

## 21.2 Driving forces for the evolution of cold cathodes

### 21.2.1 Fundamental aspects and industrial applications

The use of field emitters was an important step for electron microscopy when the field emission gun (FEG) replaced the conventional thermionic electron gun from the mid-1960s (Reimer 1993), in particular for scanning electron microscopy (SEM) and scanning transmission electron microscopy (STEM). The brightness of the electron probes increased drastically, providing subsequently high-resolution observations that cannot be achieved with hot cathodes. Such a breakthrough is a consequence of decreasing the blurring of the contrast-transfer function caused by a better spatial and temporal coherence of field emitters as compared to thermionic cathodes. Therefore, the continuing efforts to improve resolution have been directed to increase the emitter coherence by reduction of energy spread and the emitter size. This holds relevance to improve holography in the hope of overcoming the resolution limit imposed mainly by spherical aberration. Single-atom emitters (Binh 1988; Binh and Marien 1988; Binh *et al.* 1992), called nanotips are the outcome of these efforts; thanks to its reduced energy spread between 0.10 to 0.25 eV and stability of FE from the single topmost atom located at the tip apex.

From the industrial and commercial points of view, FE devices offer unique advantages over vacuum tubes. Among the pros are the modest power consumption and the possibility to integrate with solid-state electronics for extreme miniaturization. However, industrial development of vacuum micro-or

nanoelectronic devices is more demanding than the uses of FE for surface studies or for standalone FEGs in electron microscopes. Considering different applications using electron beams, an ideal cathode from the industrial point of view can be outlined and characterized by the following key factors, with their relative importance depending on the specific applications targeted:

- room-temperature operating condition;
- possibility of a fast modulation of the emission;
- low operating voltage;
- high brightness and efficiency;
- insensitivity to medium or poor vacuum conditions, i.e. FE stability;
- long lifetime and reproducible operating conditions;
- high throughput and ease of fabrication for large-area cathodes.

Most current research is now focused on the fabrication techniques for the preparation of emitter arrays for well-aligned nanoemitters and, on the other hand, in new nanomaterials with physics features relevant for FE with a view to developing a robust technology matching with the upper dictates of industrial devices. The present attention is focused on carbon nanotubes (CNTs) mainly because they can be grown in array form and further can be vertically aligned to the substrate surface (Teo *et al.* 2003). Associated to the chemically inert nature of the graphene surface, the high aspect ratio of the vertically aligned CNTs (va-CNTs) has made them potential and privileged candidates as alternatives to metal tips for uniform emission and high toughness FE arrays (FEAs) (Semet *et al.* 2002). Recently, studies of cold cathodes based on carbon nanopearls (CNPs) were carried out to explore the possibility of using them in applications like FED, compact mass spectrometers and X-ray sources. Both of them allow some of the bottlenecks associated with the use of metallic field emitters to be overcome, as described below.

### **21.2.2 Bottlenecks of metallic field emitters**

Until recently, most field emitters are metallic tips obtained either by electropolishing/chemical etching for standalone tips (Muller and Tsong 1969), or by microfabrication techniques for Spindt-type integrated micro-cathodes in field emission arrays (FEAs) (Spindt 1968). The fabrication of standalone tips is now routine for applications using FEG. However, the microfabrication of Spindt FEAs requires relatively expensive and complicated processes and the FE uniformity over large-area arrays is still uncontrollable. This is the first drawback of metallic tips for applications requiring high throughput and ease of fabrication for large-area cathodes.

From the physics point of view, the metallic tips are mainly subject to the following peculiar characteristics that constitute other bottlenecks for their use in most industrial applications. They originate from two intrinsic features of metallic surfaces: a sticking coefficient near one and a relatively small activation energy for mass-transport surface diffusion, both of which induce the following instabilities.

### 21.2.2.1 FE current instabilities

As detailed in Appendix 1, the FE current is exponentially dependent on the tip work function  $\Phi$  that is itself subject to adsorption. With a sticking coefficient near one at room temperature, for metallic tips it is therefore very difficult to keep  $\Phi$  constant over the FE area, in particular in medium or poor vacuum condition. Subsequently, not only is the total FE current unstable in time but the current distribution over the emitting area is fluctuating due to the differential spatial adsorption. This means a rapid destruction of the tip by a thermal runaway process, as described below.

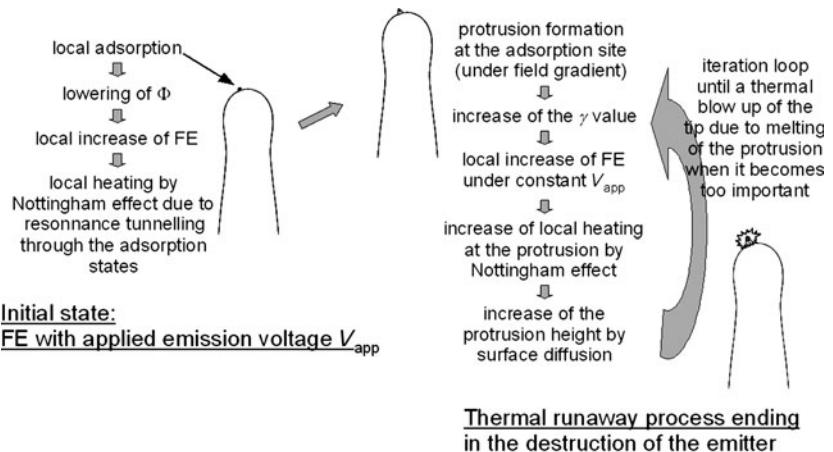
### 21.2.2.2 FE emitter-morphological instabilities

The morphological instabilities of metallic emitters are related mainly to the relatively small values of the activation energy for surface diffusion, which is in the range of 15 to 60 kcal/mole. From a practical point of view, mass transport by surface diffusion of atoms becomes effective for temperatures over one third of the metal melting temperature  $T_M$ . The consequence is the following morphological instabilities, ending in the destruction of the metallic emitters:

- The first one is tip blunting and grain-boundary grooving; they are morphological evolutions due to the conical geometry of the tip and to the presence of grain boundaries along the shank. These evolutions by surface diffusion are mainly consequences of heat treatments of the tips to temperatures over  $0.7T_M$ , in order to thermally clean the tip surface. The observed effects are not only a continuous increase of the tip radius, but also a sudden blow out of the tip due to a detachment of part of it following an ovulation process or grain-boundary grooving. Characteristic times are in the range of minutes for field emitter geometry (see Appendix 2).
- The second morphological instability is the uncontrolled and very rapid formation of local nanometric protrusions, due both to a local heating by the Nottingham effect (Fleming and Henderson 1940; Nottingham 1941) and a surface diffusion under high field gradient, as described in Appendix 2. If the FE applied voltage is kept constant during this protrusion growth, the FE currents from the nanoprotuberances may be considerably increased due to the augmentation of the field-enhancement factor  $\gamma$  over them (see Appendix 3), ending in the destruction of the tip by a local melting. This is the thermal runaway process. A schema of this process is presented in Fig. 21.1.

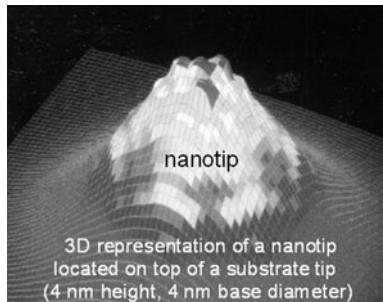
## 21.3 Single-atom emitters

In 1965, (Crewe 1965; Crewe *et al.* 1968) made the first experimental demonstration that a dramatic improvement in resolution could be made by using FEG because it allowed the electron beam to be focused on a probe area of only a few angstroms in diameter. Moving from thermionic to FE cathodes principally allows reduction of the emitting area and the energy dispersion of



**Fig. 21.1** Schema of the thermal-runaway iterative process ending in the destruction of the tip. The protrusion formation is detailed in Appendix 2, it is all the faster for higher protrusions because of more intense FE currents due to the increase in the field-enhancement factor  $\gamma$ .

the e-beam. Further developments can be expected if new improvements in the FE emitters can be realized by: (i) decreasing the size of the emission area, (ii) decreasing the angular dispersion of emitted beam, (iii) decreasing the width of the energy distribution of the emitted electrons and (iv) increasing the stability of the emission. Improvements in some or all of these aspects can be realized by using single-atom tips, as described hereafter. The growth mechanisms of these single-atom tips are detailed in Appendix 2.



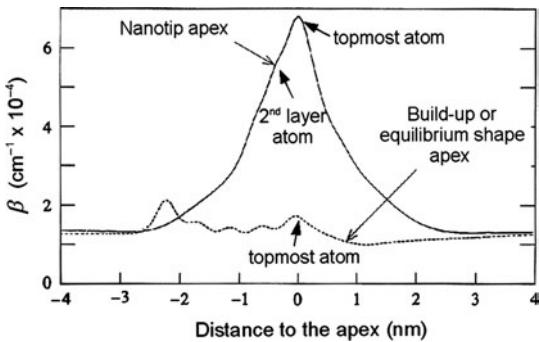
**Fig. 21.2** A pyramidal (111) nanotip on top of a 50-nm radius substrate tip. The dark areas at the apex atoms represent the enhancement of F over these atoms.

### 21.3.1 Confinement of the field emitting area

The first issue to be faced in improving tip performance is the reduction of the FE area. As the FE area is governed principally by the tip geometry and in particular the apex structure and composition, confining the FE area to the topmost atom can be done by an increase of the field-enhancement factor  $\gamma$  over this atom (see Appendix 3). This region of high curvature can be obtained, either by a build-up process (Sokolovskaja 1956; Fu *et al.* 2001), or by a field-induced growth of a pyramidal nanotip (Fig. 21.2) (Binh *et al.* 1993).

During the build-up process—made either by the anisotropic equilibrium shape with a monolayer deposition of Pd on W<111> tip (Fu *et al.* 2001), or by a field-induced build-up technique (Sokolovskaja 1956; Bettler and Charbonnier 1960)—taking the {112} facet enlargement to its limits ends in the intersection of the three facets, with the formation of a corner at the <111> apex. The angle between (112) and (111) facets is about 20° and, if the (111) corner ends with one atom, the  $\gamma$  factor is in the range of 1.4 compared to the surrounding field (Fig. 21.3, Atlan *et al.* 1992). It is enough to allow preferential FE over the protruding apex by a ratio of  $\sim 15$  for the FE currents, but without being exclusive in particular for FE at high fields.

To obtain exclusive FE from a protrusion, calculations indicated that a minimum protrusion height of about 2 nm is necessary. Such a geometry can only be obtained by the field-induced pyramidal nanotip (Fig. 21.2) as detailed in Appendix 2. For such a geometry, calculation of  $\gamma$  at the atomic scale yields a value between 6 and 7 at the apex of the nanotip compared to the surrounding



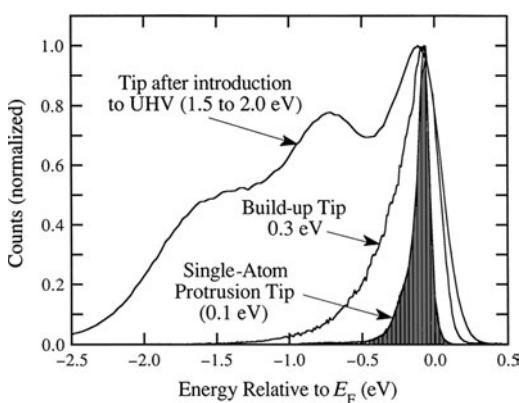
**Fig. 21.3** Field distribution ( $\beta$ ) over the (111) apex of a build-up tip and of a nanotip of 2 nm height and 4 nm cap diameter.

field (Atlan *et al.* 1992), as illustrated in Fig. 21.3. For such a value of  $\gamma$ , the FE patterns are exclusive from the topmost atom at the apex of the nanotip. This means that a nanotip delivers a self-collimated e-beam with opening angles of 4 to  $6^\circ$  and keeping the same size for the emission spot within a large interval of FE voltages/currents, for example from 1 pA (500 V) to 10 nA (700 V).

### 21.3.2 Characteristics of the e-beam

The first characteristic of FE from nanotips is its long-term stability for currents less than 1 nA. The FE stability of the single-atom emitters is explained if one considers the very small probability of having adsorbed atoms on the atom size apex coming from the surrounding gas phase. The atom size of the emitting area also explains the observed discrete jumps of the currents when the stability is broken either by an increase of the FE current over 1 nA or after a long period in vacuum.

The second characteristic is the total energy distribution (TED) of the emitted electrons. In Fig. 21.4, the TEDs of a non-clean W (111) microtip, a build-up W(111) single-atom tip and a W nanotip are plotted for comparison. These plots show that the build-up tip has the same energy distribution as a clean microtip (with a full width half-maximum (FWHM)  $\sim 0.3$  eV), only the pyramidal nanotip presents an energy distribution in the range of 0.1 eV due to



**Fig. 21.4** Measured TEDs from non-clean W microtip, build-up single-atom tip and pyramidal nanotip.

its height, which is in the range of a few nm (Purcell *et al.* 1995; Gohda and Watanabe 2001).

In a nutshell, single-atom emitters can be obtained either by a build-up technique or by a high-field-induced pyramidal nanoprolusion. However, due to its geometry, only the pyramidal nanotips present the most interesting features for two applications, which are described hereafter.

## 21.4 Use of single-atom nanotip: The Fresnel projection microscope

### 21.4.1 The Fresnel projection microscope (FPM)

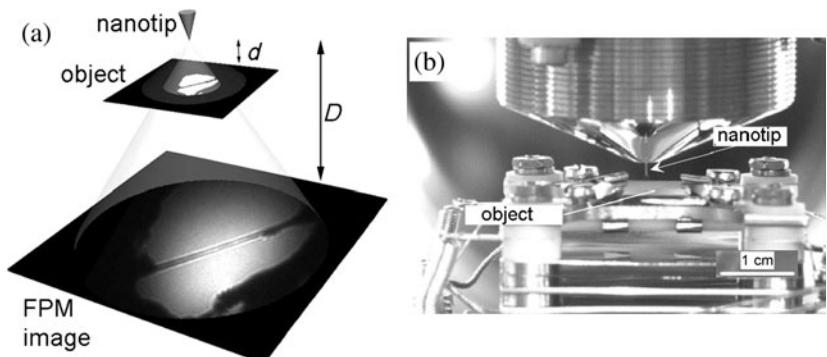
The projection microscope was introduced by Morton and Ramberg (1939) and was called point projector electron microscope. It uses a the quasi-radial propagation of electrons coming from a FE tip when the object is inside the beam path. This projection or shadow microscope (Fig. 21.5) is essentially a lensless microscope; with a magnification factor  $M$  given by:

$$M \approx D/d, \quad (21.1)$$

where  $D$  and  $d$  are the distances between the projection point and the screen or the object, respectively. Equation (21.1) shows that the magnification increases by approaching the object to the projection tip and could reach values in the range of  $10^6$  to  $10^7$  for  $d$  between 100 to 10 nm with  $D$  in the range of 10 cm. With the recent technological development in piezo-drives, tip–sample distances of less than 1 nm can now be routinely handled.

The above approach of defining the magnification only considered the projection microscope within the geometric point of view. If the electron source is a single-atom tip, the coherence of the e-beam has to be considered and the projection image may also reflect the diffraction process by the object, i.e. the formation of interference patterns. Within the last condition, the projection microscope is called a Fresnel projection microscope (FPM) in order to point out the characteristic of coherence of the electron source.

As the FE voltage needed to extract electrons from a nanotip is in the range of hundreds of volts, the wavelength of the electrons  $\lambda$  is about 0.1 nm. For object dimensions in the range of 1 nm or greater, the interaction between the



**Fig. 21.5** (a) Schematic description of the Fresnel projection microscope. (b) Photo of the FPM head with the piezo-drive system for controlling the tip–sample distance with nanometric resolution.

e-beam and the object can be depicted by the classical electron optics wave theory. Considering the atomic dimension of the source, the wave front can be considered as spherical within the object dimension. So, in the near-field diffraction process, the wavefunction  $\Phi(I)$  at each point  $(x, y)$  of the image projected onto a screen is given by the Fresnel–Kirchhoff integral:

$$\Phi(I) = (iA/\lambda\rho)(\exp(ikD)/D) \iint_{\text{object}} f(x, y) \exp(-ik/2\rho((x - x_0)^2 + (y - y_0)^2)) dx dy, \quad (21.2)$$

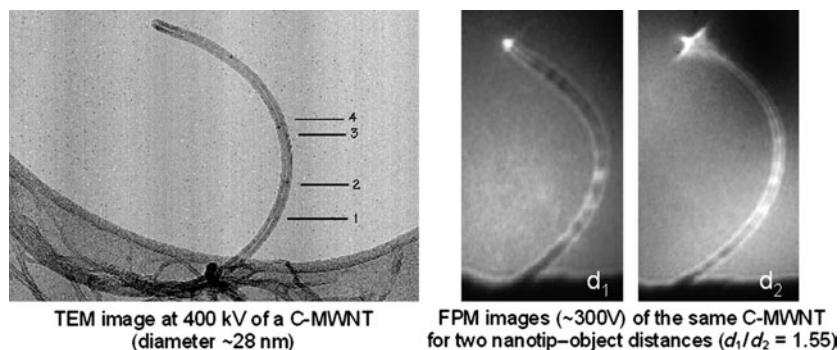
with  $|\Phi(I)|^2$  the intensity at each point on the screen. The term  $\exp(ikD)/D$  represents the cast shadow of the object on the screen and the integral term stands for the surrounding diffraction fringes. The function  $f(x, y) = \exp(-a(x, y) + i\zeta(x, y))$  is the complex transmission function of the object, with  $a(x, y)$  the attenuation factor and  $\zeta(x, y)$  the phase shift factor. For a pure amplitude object  $\zeta = 0$ , with  $a = \infty$  for a complete opaque object and  $a = 0$  for a free opening. In the case of pure phase objects, such as magnetic or electric fields,  $a = 0$  and  $\zeta \neq 0$ . For a phase object, it is necessary to consider the phase shift  $\Delta\zeta$  between two electron beams starting from the point source, passing through electric and magnetic fields along two different paths and terminating together at one point in the image plane (Aharonov and Bohm 1959).  $\Delta\zeta$  is given by:

$$\Delta\zeta = (e/\hbar) \left( \int V dt - \iint B dS \right),$$

where the first integral corresponds to the phase difference induced by an electrostatic potential  $V$  and the second by a magnetic field  $B$ .

Therefore, in experimental situations where the electron source is atom sized, such as the nanotip, the resulting projection diffractograms are not only highly correlated with the direct space representation of the object, but also give information about the eventual presence of localized magnetic or electric fields.

For illustration, Fig. 21.6 is a comparative example of the same object, a multiwall carbon nanotube (CNT), observed with a high-resolution transmission electron microscope (HRTEM) at 400 kV and imaged with FPM. Defects along the CNT, indicated by arrows 1 to 4, are detected with both techniques. They have different contrast in FPM images. However, only the FPM images



**Fig. 21.6** HRTEM and FPM observations of the same CNT. (For easy comparison, the  $d_2$  FPM image is reduced to the same size as  $d_1$  and HRTEM images.)

show clearly the presence of a localized electric field (diffraction fringes of the phase object) at the end of the CNT, due to the FE applied voltage ( $\sim 300$  V) between the nanotip and the CNT. The diffraction fringes are also more important for higher values of the electric field, as  $d_2 < d_1$ ,  $d_1$  and  $d_2$  are, respectively, the point source–CNT distances of the two FPM images.

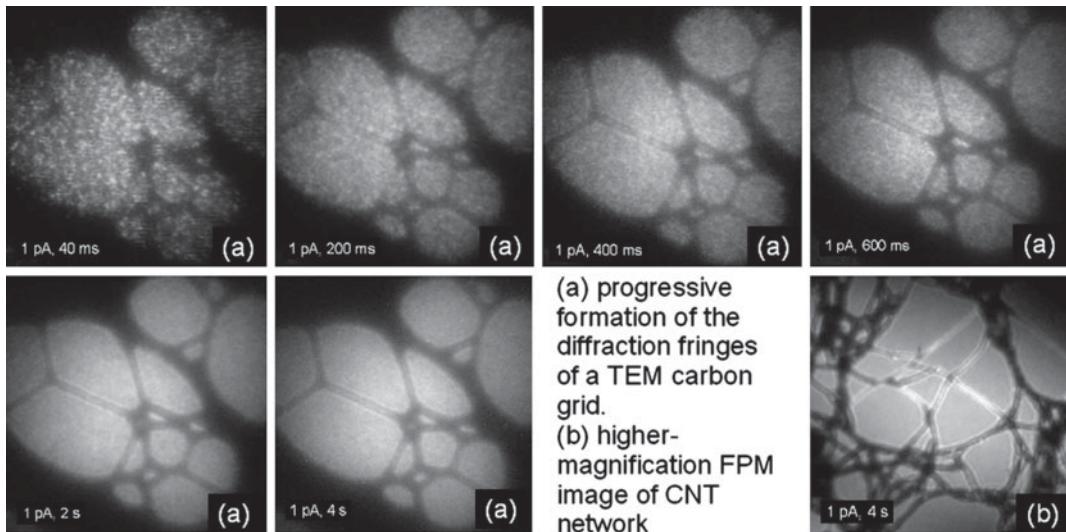
### 21.4.2 Object reconstruction from FPM images

The best resolution obtained by a direct approach of the diffractograms is limited to about 0.5 nm. To improve this resolution, (Mayer and Vigneron 2000; Mayer 2001) have proposed a protocol using simultaneously the projection images obtained for different positions of the source relative to the object, in particular for various orientations. This results in diffraction tomography. An inverse scattering technique was therefore developed by using the information present in all diffraction directions, allowing a 3D reconstruction of a sample at the atomic level from the FPM diffraction images. In its present form, this technique is limited to real-valued potential energies, i.e. opaque and transparent objects, and an exact knowledge of the incident wave is assumed; but it indicates that a computerized tomography approach of the FPM could result in 3D observations of opaque and phase objects with atomic resolution.

### 21.4.3 Brightness and coherence of the e-beam from a nanotip

The experimental brightness  $B$  of the nanotip is given by  $B = J/\Delta\Omega$ , where  $J$  is the FE current density and  $\Delta\Omega$  the solid-angle beam opening. For a nanotip emitting 1 nA under a FE voltage of 200 V,  $B \approx 2 \times 10^8$  A cm $^{-2}$  sr $^{-1}$ . Its degeneracy factor  $\delta$  (Appendix 1) is about  $10^{-3}$ , this value has to be compared with conventional tips in electron microscope FEG ( $\delta \approx 10^{-6}$  for 1  $\mu$ A at 150 kV) and with thermionic sources that have  $\delta$  values between  $10^{-12}$  to  $10^{-10}$ . The brightness of the nanotip can be determined experimentally by considering the gradual formation of the diffraction fringes of FPM images, as presented in Fig. 21.7. It shows that with a total e-beam current of 1 pA, an exposure time between 2 and 4 s gives enough electron flux for a complete formation of the FPM images. The brightness information can also come indirectly from coherence parameters inferred from interference pattern (Spence *et al.* 1994).

The coherence of the e-beam from a nanotip can be estimated by the following parameters: (i) the size of the FE area and the angular opening that defines the transverse coherence length, and (ii) the energy distribution  $\Delta E$  of the e-beam, which defines the coherence time and the longitudinal coherence length. For a single-atom nanotip— $\Delta E = 0.1$  eV at a FE voltage of 200 V—values of 7 nm,  $6.6 \times 10^{-15}$  s and 55 nm are obtained for transverse coherence length, coherence time and longitudinal coherence length, respectively. From the experimental point of view, the estimation of the coherence may be appreciated by looking on the sharpness of the fringes at high magnification,

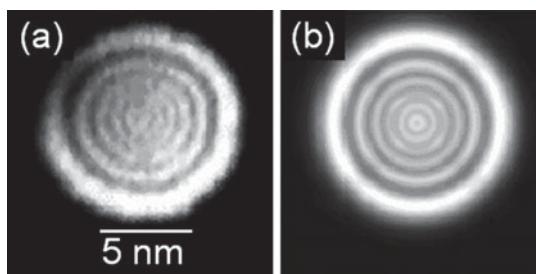


**Fig. 21.7** Building-up of electron interference pattern with time. (a) FPM images of fibers from a TEM carbon grid at different exposure times (40 ms to 4 s). With a total current of 1 pA, the exposure time for a complete formation of the diffraction fringes is between 2 to 4 s. (b) FPM image of CNTs with diameter about 30 nm. The sharpness of the fringes can be easily perceived for higher magnification, i.e. for closer distance  $d$  between tip-sample. Please note that the ratio between the diameters of C fibers in (a) and CNTs in (b) is about 10.

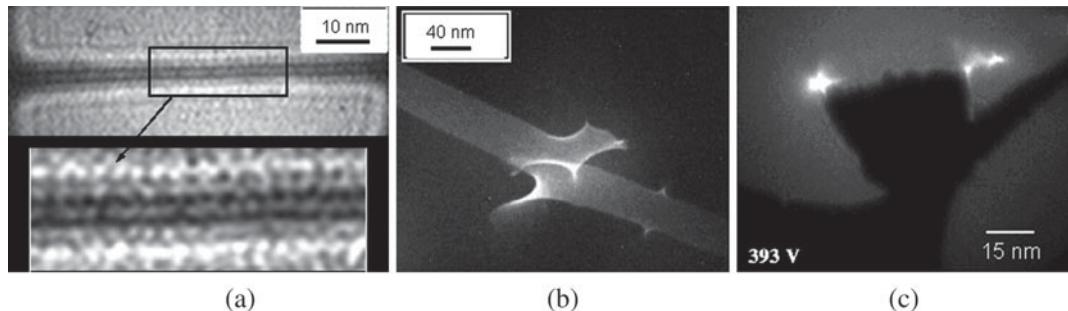
i.e. for small values of  $d$ , in order to have enough spatial resolution (Fig. 21.7). Figure 21.8 shows the diffraction patterns of a 10-nm carbon hole, Fig. 21.8(a) is a FPM image and Fig. 21.8(b) is a numerical simulation of the Fresnel diffraction pattern from eqn (21.2) when the size of the source is one atom. When this size increases to a three-atom facet, most of the fringes present in Fig. 21.8(b) vanished. This confirms the actuality and stability of a one-atom size apex of the nanotips for FE currents up to a few nA used in FPM (see Appendix 2).

#### 21.4.4 Some observations with the FPM

Of all the known types of particle sources, nanotip FE electron sources are the brightest and the most coherent currently available. Used as a projection source in FPM, it converts this projection microscope into a low-voltage, high-resolution and lensless microscope given nanometric resolution in the hundred-volt energy range. It is a perfect tool for observations of organic



**Fig. 21.8** Fresnel diffraction images of a 10-nm carbon hole. (a) Experimental FPM image. (b) Numerical calculation from the Huygens–Fresnel formalism eqn (21.2) with the size of the source equal to 0.3 nm.



**Fig. 21.9** FPM images. (a) Amplitude object: ARN fiber, the periodicity of the diffraction pattern is related to the periodicity of the organic fiber profile. (b) Phase object: at the connection zone of a PMMA fiber, in the center of the photo, the large diffraction pattern indicates the presence of a localized electric field. (c) Phase object: The two zones of intense bright fringes at the corners of an isolated magnetic (iron-oxide) nanoparticule show important localization of the magnetic field at the corners of a cube.

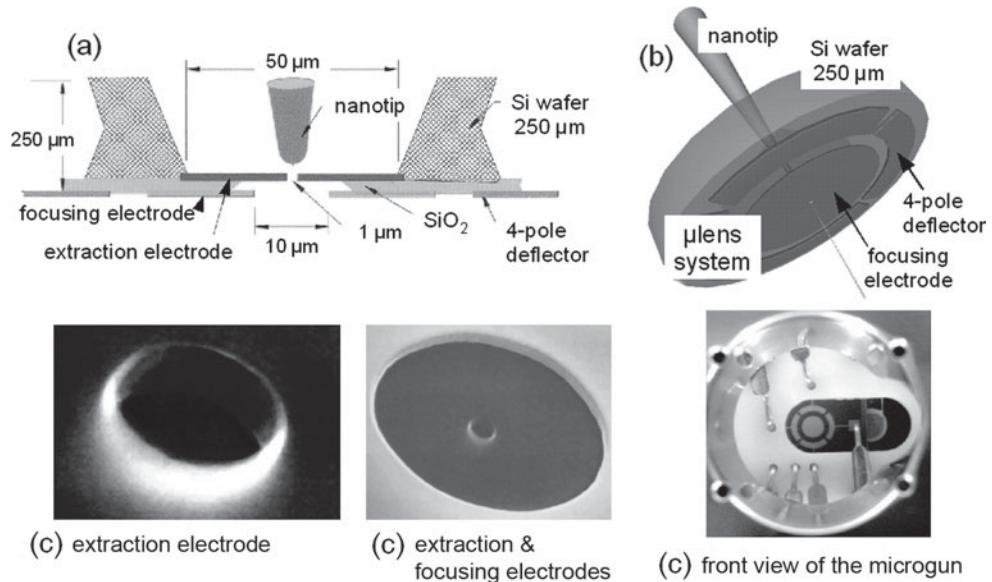
materials—because of the low-energy of the beam—and a unique microscope for observations of phase objects. (Binh *et al.* 1994, 1998; Binh and Semet 1998). Figure 21.9 presents some FPM images to illustrate the great potentiality of the present projection microscope.

## 21.5 Use of single-atom nanotip: The microgun

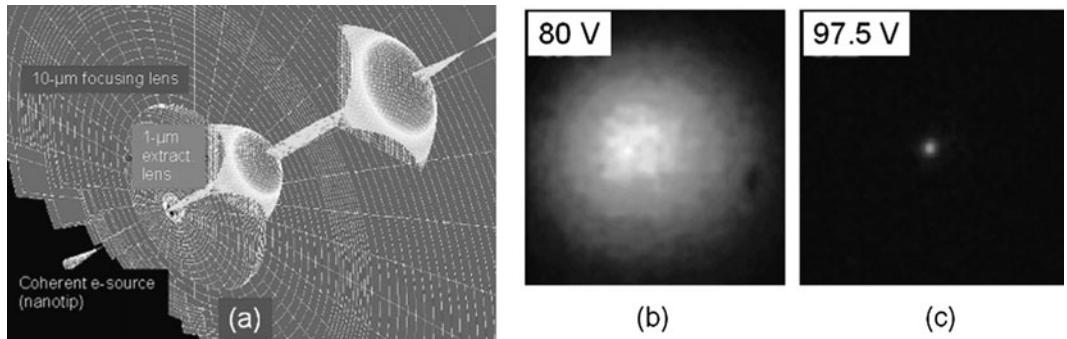
Arrays of independent electron guns for massively parallel electron lithography, microscopy and related applications are the main driving forces responsible for the miniaturization of electron-beam columns (Chang *et al.* 1990). The microcolumn proposed by Chang is based on individually microfabricated electron optical components on silicon membranes, using Schottky emitters as FE sources. Another route to the microcolumn approach is the microgun ( $\mu$ gun), which is associated with a single-atom field emitter, the nanotip, placed in front of a  $\mu$ lens system. These are two-electrode lenses having micrometer-size bores and incorporating a coplanar four-pole deflector (Binh *et al.* 1999) (Fig. 21.10).

The  $\mu$ gun system presents three main advantages: (i) the use of a nanotip, which is an atom-size dimension cathode, delivering a stable self-collimated coherent e-beam. (ii) The microfabrication on the same chip of the combined system of microelectrodes and deflectors contributes to the compactness of the system, reduces drastically the cost of the column and permits a simple and very precise alignment of the different electrodes. (iii) The combination of atom-size electron source, coherent beam and micrometer-dimension electrodes minimizes the electron-optics aberrations.

This is shown by some extensive 3D numerical simulations, Fig. 21.11(a), with the size of a cross-over between 1 to 4 nm. Experimentally, the quality of the focusing features of the  $\mu$ lens system can be appreciated by projecting the e-beam on a fluorescent screen as shown in Figs. 21.11(b) and (c). In this experiment the total current is around 1 pA with a brightness value of about  $10^5 \text{ A cm}^{-2} \text{ sr}^{-1}$  and it reveals the following characteristics: (i) the circular symmetry of the spot is not affected during the whole focusing sequence, i.e. distortions and astigmatism due to field aberrations are not discernible, even



**Fig. 21.10** Structure of a  $\mu$ gun. (a): Side view. (b) 3D representation. (c) SEM images of the experimental setup.



**Fig. 21.11** (a) 3D numerical simulation of the focusing possibility of the  $\mu$ gun. (b) Non-focus beam projected on a screen showing a circular symmetry all along the focusing sequence. (c) Focus spot projected on a screen. This corresponds to a size of the cross-over between 1 to 4 nm given by the numerical simulations. (Experimental conditions: extracting voltage 100 V, the insets of (b) and (c) are the focusing voltages.)

when the beam is completely focused; (ii) due to the absence of a diaphragm, the focal spot keeps the same total FE current of 1 pA delivered by the nanotip; (iii) an appearance of a spot distortion is only observed when the misalignment of the nanotip with the extraction bore is over 5 nm; (iv) average deflection values of about 2.5 mrad/V may be obtained with the 4-pole coplanar deflection system with a distortion of the beam by about 10% for deflection in the range of  $\pm 8^\circ$ .

The unique focusing and deflection characteristics at 100 V of the  $\mu$ gun, compared to a conventional FEG, comes from the quality of the extracting and focusing lenses with edge defect size less than 10 nm and their alignment within 0.1  $\mu\text{m}$ , and on the other hand, from the minimalism of the lens system

because the initial size of the emitting area is already atom sized. In other words, FPM does not need diaphragms and lenses to get a virtual source in the range of an atom, which are the main causes of aberrations in conventional FEG.

## **21.6 Material issues for field emitters: Carbon nanocompounds**

Most of the weaknesses of field emitters are inherent to the features of metallic surfaces. Metals can provide a large electron supply for FE, but suffer from high reactivity with residual gases in the vacuum. Moreover, its malleable nature shows a poor resistance to backspattered damage and deformation by surface diffusion. Therefore, metal field emitters are subject to rapid blow-up by a thermal-runaway process, with lifetimes too short for industrial applications. The ideal material for a field emitter, to meet the industrial use of cold cathodes, should have high electron concentration, high thermal conductivity and, in particular, should be hard and non-reactive in medium and poor vacuum environments.

Paths to improving emitter performance have followed modifications in fabrication methods and materials. A large variety of non-metals have been suggested as candidates, covering materials such as carbon—graphite, amorphous carbon, diamond, nanotubes and nanopearls—and metallic carbides for their low reactivity and high hardness, as well as semiconductors, which possess fewer conduction electrons than metals but have the advantages of their possible integration and availability of established routes for microfabrication of micrometer-size emitter structures.

### **21.6.1 FE planar cathodes**

Semiconductor materials have low reactivity and high surface hardness. Moreover, the electron affinity can be controlled either intrinsically by the ability of doping them, or with external excitations on multilayer nanostructures made available with the improvement in the epitaxial growth of ultrathin films. These emitters, which can have negative electron affinity, can emit electrons at very low electric field, i.e. without the need of a large field enhancement. This means that most of these cathodes could have a planar emitting surface and use a variety of principles for electron emission. Semiconductor planar cathodes have the advantage of producing a uniform FE current density and the methods of manufacturing them are very similar to well-known microelectronics processing. Details about multilayer planar cathodes are presented in greater detail in a review paper (Binh and Semet 2007) and are beyond the scope of the present chapter.

A lot of research over the past decade has targeted diamond or diamond-like carbon as a material for FE planar cathodes. This is because certain surfaces of diamond exhibit negative electron affinity combined to other advantages like low reactivity, high hardness and high thermal conductivity. However,

the primary drawback of diamond is that it is an insulator, thus, electron concentration in the conduction band, which sees a smaller barrier to emission, is not compatible for large emission currents. The mechanisms for FE from diamond surfaces are not fully understood yet. Therefore, attempts to dope diamond or to provide a better contact to supply electrons to the conduction band have had rather limited success.

### 21.6.2 Field emitters

Planar cathodes and tipped-geometry field emitters have both strengths and weaknesses as cold cathodes. The advantages of field emitters are that they operate at high efficiency with high current density without the need for a low surface barrier. Another advantage of field emitters is that many of them can be put in parallel, up to  $10^8$  to  $10^9$  emitters per  $\text{cm}^2$ , in order to increase the total current available. Since the introduction of Spindt-type integrated microtips, the potential for high current and high efficiency have made field emitter arrays (FEAs) the most widely studied cold cathodes. Even with the significant drawbacks of metallic field emitters, namely their short lifetime and low reliability in poor vacuum, their main setback comes from their non-standard fabrication steps, in particular for very large area FEAs, making them difficult to use for retail industrial applications.

Recently, most of these issues have been resolved with advances in materials, fabrication processes and proper design of FE cathodes. Attention has been drawn recently to carbon nanotubes (CNTs) because of the success in deterministic growing CNTs in array form and the capability of making them vertically aligned to the substrate surface. The success of carbon graphene nanocompounds, such as CNTs and carbon nanopearls (CNPs), also comes from the intrinsic features of the graphene surface that are: a low reactivity, a high electron concentration and a high surface hardness, in particular for surface diffusion. The electron concentration is similar to metals, as well as for the work function. The main advantage is the possibility for cost-effective growth, with spatial deterministic location of nanosize compounds with sharp, tip-like features providing large field-enhancement factor. Of all the carbon nanocompounds, two types of products have shown encouraging results by taking advantage of the emission features of field emitters: the first one is CNTs for array structures and the second one is CNP strings for large emission areas. Both of them will be detailed hereafter.

## 21.7 Carbon-nanotube field emitters

There are three important parameters describing a field emitter: (i) the work function or surface barrier height, (ii) the field-enhancement factor and (iii) the emission area. For most field emitters, the surface barrier is intrinsic to the material, and for a graphene surface the work function is between 4 to 5 eV. Such a value for the work function implies a FE field larger than  $5 \times 10^7 \text{ V/cm}$ , thus needing a large field-enhancement factor  $\gamma$ . For isolated vertically aligned field emitters on a plane,  $\gamma$  can be increased by reducing the radius of curvature

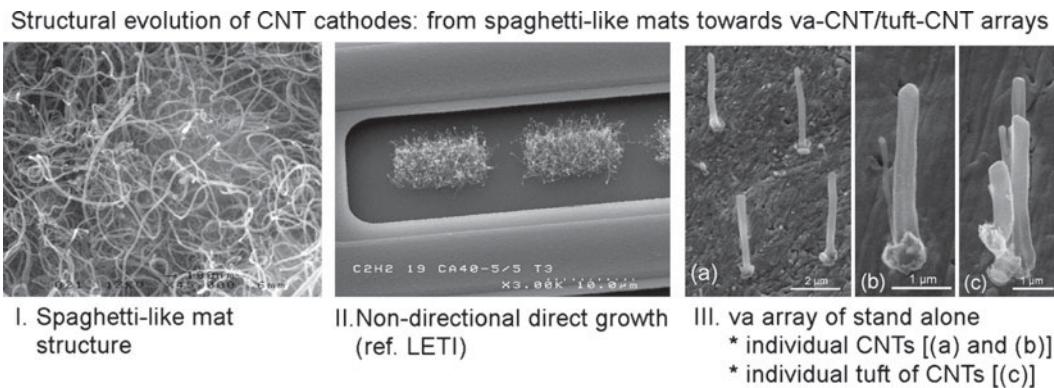
and increasing the aspect ratio of the emitter (see Appendix 3). However, the emission area is determined by the geometry, in particular the apex radius, and in general the emission area decreases as the radius decreases. This means that an optimum radius of curvature for the field emitter may not be the smallest possible, but involves a tradeoff of emission area and field enhancement. Moreover, the mechanical resistance of the field emitter must be considered. This is why multiwall carbon nanotubes (MWNTs), with diameters in the range of a hundred nm, are preferred to single-wall carbon nanotubes (SWNTs) with diameters in the range of a few nm.

For arrays of vertically aligned field emitters, it is highly recommended that the emitters are spaced between 1 to 2 times their heights apart in order to avoid electrostatic field screening (Nilsson *et al.* 2000). Moreover, the structural variation between emitters in the array must be as small as possible in order to achieve good FE uniformity over the entire array. To meet these two requirements, there has to be an evolution of the CNT cathode structure, presented in Fig. 21.12.

At the start of the use of CNTs as field emitters (de Heer *et al.* 1995) non-directional CNTs were deposited by screen printing on a plane substrate followed by a seasoning process to activate FE sites, this is called a “spaghetti-like” structure cathode (Fig. 21.12(I)). The following improvement is the direct growth on the cathode substrate of non-directional CNTs (Fig. 21.12(II)); it provides a greater density of FE sites compared to spaghetti-like cathodes. However, the optimum structure for CNT cathodes is an array of vertically aligned CNTs (va-CNTs) spaced at about one height ensuring that each individual CNT, or tuft of CNTs, is a FE site (Fig. 21.12(III)). Good FE uniformity will be reached when the variation of  $\gamma$ , i.e. geometry, between emitters is negligible. Fabrication processes of va-CNTs or CNT tufts, meeting these requirements, are presented hereafter.

### 21.7.1 The catalytic process of the CNT growth

The catalytic mechanism of the chemical vapor deposition (CVD) growth of carbon filaments, as nanotubes and nanofibers, was described in the work of

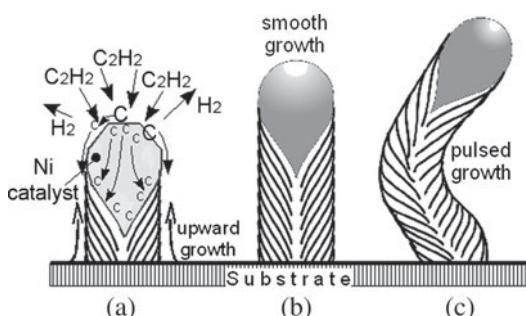


**Fig. 21.12** Evolution of CNT cathodes to reach higher current density and better FE uniformity.

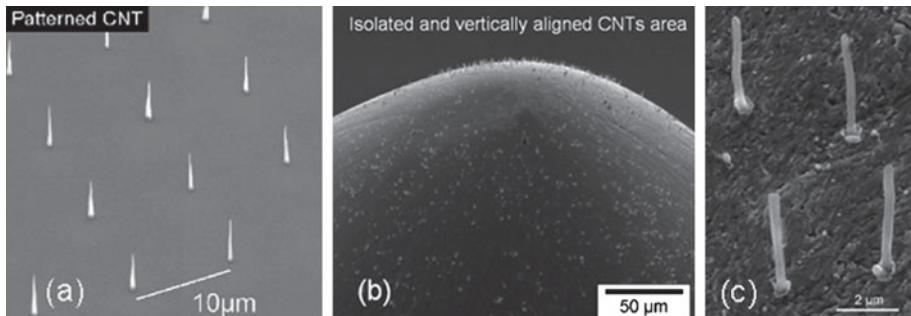
Robertson (1969) although the first observations began in 1889 (Hughes and Chambers 1889). It was reported that carbon filaments were grown from the interaction of carbon-containing gases and metal surfaces that led to graphitic carbon at relatively low temperatures. Since then considerable progress has been made with different CVD methods, in particular with the recent fabrication of vertically aligned multiwall CNTs by plasma-enhanced CVD (PECVD) (Ren *et al.* 1999; Bower *et al.* 2000; Merkulov *et al.* 2000; Chhowalla *et al.* 2001; Chen *et al.* 2004). Although these processes differ in growth conditions—such as substrate, catalyst, feed gas (such as acetylene, methane, carbon monoxide, ethane ...), temperature (in the range of 700–1200 K), root or tip growth and the presence or absence of an electric field—the growth process, suggested by most of the groups, essentially adopts a common established concept of CVD carbon fiber growth developed in the 1970s (Baker and Harris 1978; Snoeck *et al.* 1997; de Jong and Geus 2000), which is schematically depicted in Fig. 21.13.

The catalytic metals (such as pure or alloys of Fe, Co, Ni, Cr, V or Mo) are metals that must decompose, dissolve carbon and/or form metal carbides. The hydrocarbon molecules decompose at the contact with the free-metal surface of the catalyst and form carbon atoms with concomitant desorption of molecular hydrogen. The carbon atoms dissolve in and diffuse through the bulk of the metal, in concomitance with surface diffusion of metal and C atoms (Fig. 21.13(a)). The C atoms precipitate in the form of graphene layers at the other side of the metal particle and push it out from the substrate.

As long as the gradient of concentration of C exists across the catalyst particle, a steady-state process occurs with a continuous and smooth growth of the CNTs, resulting in a straight fiber (Kock *et al.* 1985) (Fig. 21.13(b)). Next to that, if the presence of fresh catalytic surface is erratic/not regular, either in area or location, then a pulsed growth mode is observed and the CNTs are no longer straight (Fig. 21.13(c)). The CNT growth stops when the catalyst particle is completely encapsulated, by a graphene layer for example, because no reaction between vapor and metal can occur. From this model of individual CNT growth, called the tip-growth mechanism, which was first suggested by Hoogenraad (1995), the precipitation of graphene layers from the Ni-graphite surface is crucial for the nanofilament growth. It also explains two things: why the metal particles are found at the tip of the carbon fiber as the graphite fiber pushes the metal particle from the support and continues to grow at the back



**Fig. 21.13** Tip-growth mechanism during the catalytic process of an individual CNT growth.

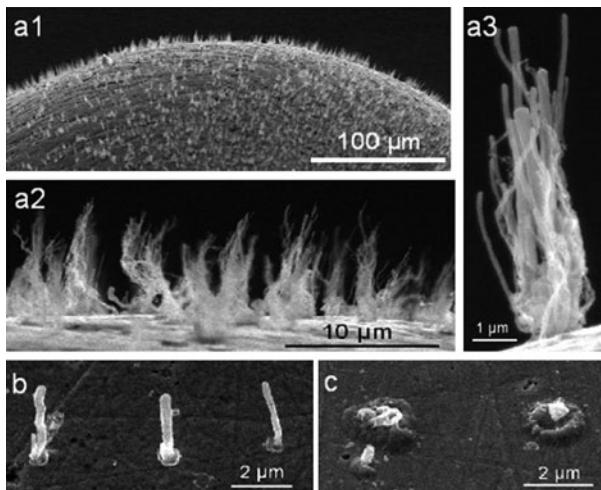


**Fig. 21.14** (a) Well-aligned array of CNTs controlled by electron lithography patterning of the Ni catalytic dots ( $C_2H_2 : NH_3$  flow of 40:200 sccm at 3.5 mbar, 700 °C substrate temperature and -600 V substrate bias (Teo *et al.* 2003). (b) and (c) Direct growth of free-standing, isolated and vertically aligned CNTs on the apex of a metallic substrate tip from a manual dispersion of Ni clusters confined on the apex of the Ta tip. The feed gas is  $C_2H_2$  with a ratio  $C_2H_2 : NH_3$  kept constant at 2:3 sccm at 0.1 to 0.2 Torr, 700 °C and -100 V substrate bias (Mauger *et al.* 2007).

of the particle; and also the close relation between the diameters of the CNT to each catalyst particle.

The main characteristics of the growth and the related features (Teo *et al.* 2003; Mauger *et al.* 2007) are the following:

- (1) The spacing between the different CNTs across the arrayed distribution can be strictly controlled by the localization of the catalyst particles. This spatial distribution can be done either by an electron lithography patterning, or by a manual dispersion of individual Ni nanoclusters for a pitch in the range of few micrometers that is about the height of the grown CNTs (Fig. 21.14).
- (2) In most cases, with a size of catalysts of less than ~100 nm, there is only the nucleation of a single CNT per dot.
- (3) At some catalyst sites, two or more CNTs can nucleate. These situations happen when the size of the catalyst is greater than 100 nm, or when an aggregation of several Ni nanoclusters happens. A former systematic study of va-CNTs growth with PE-CVD (Teo *et al.* 2003), established that at ~700 °C the size of the catalytic Ni dots, obtained by electron lithography, determines the diameter and number of CNTs per dot. For a diameter of the Ni catalyst dots of less than 100 nm, the yield of one va-CNT per dot is very high (88–100%). For larger catalyst dots, up to 800 nm, tufts of va-CNTs are obtained, which means the presence of several va-CNTs per catalytic dot. When the catalysts are Ni nanoparticles, having a size dispersion centered at 100 nm, the control for obtaining single CNTs or tufts of CNTs per Ni particle is now achieved by modulating the growth temperature (Mauger *et al.* 2007). For a temperature of ~650 °C, the PE-CVD growth usually gives one tuft of CNTs per particle (Fig. 21.15(a1) to (a3)), whereas only a single CNT per particle is obtained when the growth temperature is increased to 700 °C (Fig. 21.15(b)). At higher temperature, 800 °C, the dissolution of Ni into the metallic substrate impeded the formation of the CNTs (Fig. 21.15(c)). Please note, within the tuft structure only one of the CNTs had a dominant height and it will then be the preferential field emitting CNT.



**Fig. 21.15** PE-CVD growth of carbon nanotubes (Mauger *et al.* 2007). (a1) to (a3): Growth temperature 650 °C, formation of isolated CNT-tufts at each catalyst dot. (a1): Overview of the tuft array at the apex of a Ta tip; (a2): a closer and side view of the array showing the small discrepancy of the height of the CNTs tufts; (a3): a close view of one tuft located at one Ni catalyst dot. (b): At 700 °C; formation of one CNT per dot. (c): At 800 °C, dissolution of the Ni catalyst into the Ta substrate.

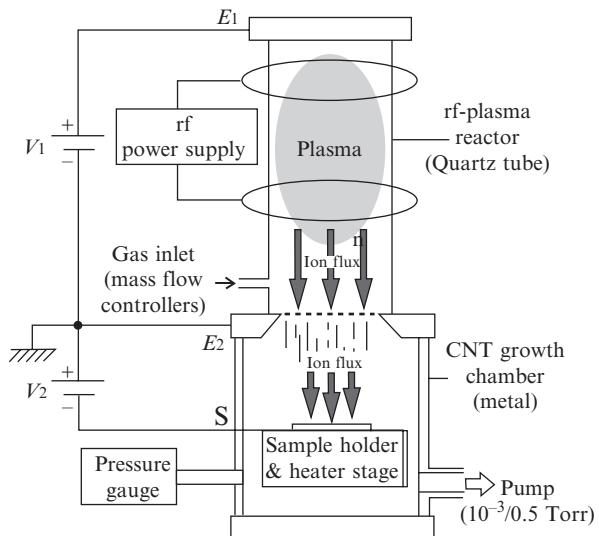
- (4) The histogram of the heights of the CNTs within an array showed a tight dispersion around a mean value, this last value is dependent on the growth conditions. It has been shown that, with 100-nm Ni dots, the average height of the va-CNTs is 5.9 micrometers, with a standard deviation of the distribution in CNT height equal to 0.4 micrometer.

### 21.7.2 Electric-field-guided growth of va-CNTs during PE-CVD

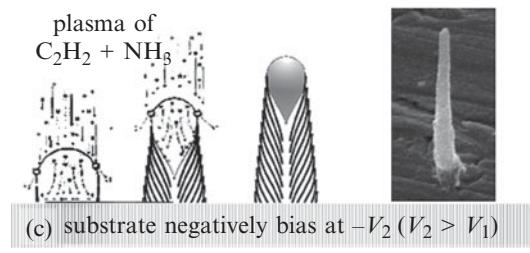
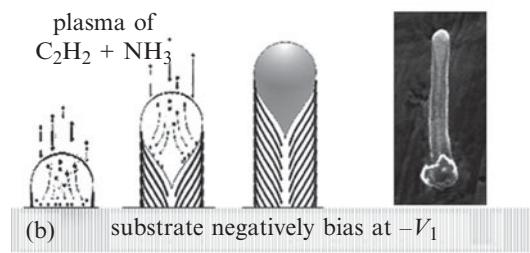
CNTs grown in PE-CVD present the particularity of being standalone, straight and perpendicular to the support surface and the role of the electric field has been established by PE-CVD growth in a dual-chamber reactor (Fig. 21.16) (Mauger and Binh 2006). In the first chamber, the plasma is formed under radio-frequency (rf) excitation. The ionized species are then transferred to an adjacent growth chamber by an electrostatic system. In this case, the applied field during the growth is now disconnected from the strength of the electric field inside the plasma sheath in a conventional single chamber PE-CVD reactor.

Without any electric field, we can notice the presence of an undefined mass at the roots of the CNTs on the substrate, which is attributed to amorphous carbon (*a*-C); and we can also emphasize that the CNTs obtained are very tortuous, with frequent changes in the growth direction and with uncontrolled lengths from one catalytic site to another. An example is given in Fig. 21.17(a).

According to the expectation, when an electric field is applied, the growth direction of the CNTs is perpendicular to the surface. This directional growth is observed for values of macroscopic fields (see Appendix 3) starting from  $F_{\text{macro}} \sim 50 \text{ V/cm}$ . No change in the growth behavior is observed for  $F_{\text{macro}}$  up to 100 V/cm (Fig. 21.17(b)), except for the length. For higher values of



**Fig. 21.16** Schematic drawing of the two-chamber PE-CVD reactor.



**Fig. 21.17** (a) to (c) Growth processes of CNTs under different applied fields, with  $|V_2| > |V_1|$ . On the left are schematic representations of the different steps during the growth showing the geometry of the CNTs and the size of the catalyst nanoparticles. The scanning electron microscopy photos at the right side are as-grown CNTs; they indicate the presence of amorphous carbon and non-directional growth only when the CVD is done without applied field. The macroscopic fields, for  $V_1$  and  $V_2$ , are respectively  $\sim 75$  and  $150 \text{ V/cm}$ .

**Table 21.1** Dependence of the va-CNT geometry versus the macroscopic applied field. The ion current densities are measured at the substrate tip.

$F_{\text{macro}}$ (V/cm)	0	50	75	100	150
Ion current (ions/nm <sup>2</sup> /s)	15	75	190	300	370
Geometry of the CNTs	non-directional / tubular / $\alpha$ -C	Directional / tubular / clean	Directional / tubular / clean	Directional / tubular / clean	Directional / conical / clean

the fields, the CNTs present a conical geometry, as Fig. 21.17(c) shows for  $F_{\text{macro}} = 150$  V/cm. One common aspect during the growth with an applied electric field is the absence of  $\alpha$ -C surrounding the CNTs, i.e. all the as-grown CNTs are “clean” without any post-treatment. The results are also summarized in Table 21.1.

To assess the roles of the electric field during the CNT growth, the three different main effects associated with its presence are reviewed:

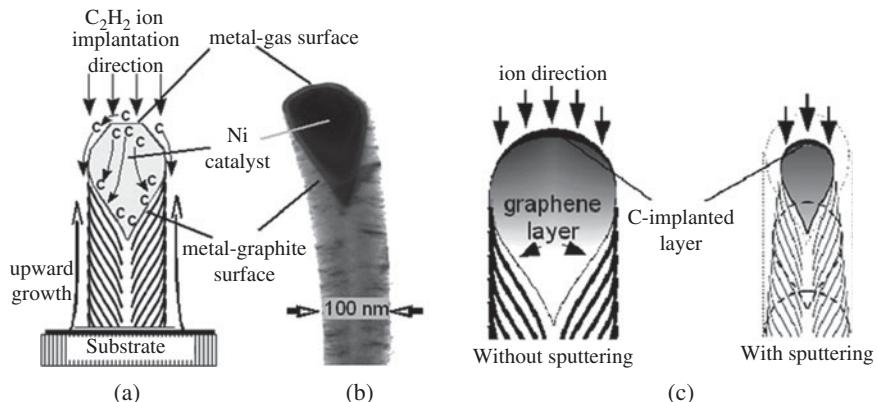
### (1) Electrostatic force stretching

Charge redistribution takes place along the carbon nanotubes, thus creating a giant dipole. This dipole is oriented in the direction of the applied electric field and can then fix the growth direction. However, a first calculation shows that this electrostatic force (Muller and Tsong 1969; Miller *et al.* 1996) is negligible compared, for example, to the gravity force on the metallic nanoparticles. It also contradicts the ability to grow CNTs aligned at angles not perpendicular to a substrate (AuBuchon *et al.* 2006).

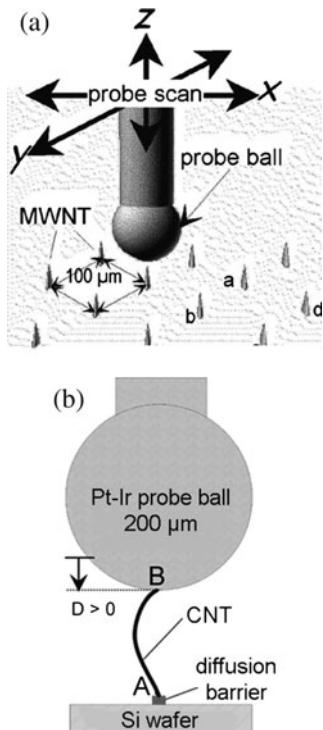
### (2) Carbon implantation in Ni catalyst

The plasma created by the rf-excitation released  $5 \times 10^8$  positive ions/cm<sup>3</sup> in the growth chamber. When the sample holder is polarized negatively these ions bombard the surface, and in particular the Ni nanocatalyst, with flux values between 70 to 400 ions/nm<sup>2</sup>/s function of the applied field (Table 21.1). These values correspond to a high fluence implantation (Ziegler *et al.* 1985), with a C-implanted layer having a mean projected range  $R_p$  and a range straggling  $\Delta R_p$ , both of them are estimated to be of the order of 1 to 2 nm (Fig. 21.18(a)).

Two effects can be anticipated from the C implantation into the Ni nanoparticles: (i) The maintenance of a continuous and uniform feed of C, because the ion-implantation process is less susceptible to be influenced by the surface conditions and in particular the need for a free-metal surface as for CVD without field. (ii) The localization of the implantation zone will always be at the apex of the metallic nanoparticles. In the model where the driving force for CNT growth is the concentration gradient of C atoms from the metal-gas surface to the metal-graphite surface, these two effects then induce a continuous growth of the CNTs in the vertical direction and a same length of the CNTs if the nanocatalysts have the same dimension. This is confirmed experimentally by Teo *et al.* (2003). By a strict definition of the size of the Ni catalyst, va-CNTs produced in arrays of thousand of individuals exhibit remarkable uniformity in terms of diameter and height.



**Fig. 21.18** (a) Schematic representation of the va-growth model for individual CNTs. The role of the electric field is to maintain a fixed direction for ion implantation of the ionized carbon species. (b) TEM image of individual CNT obtained. (c) The continuous decrease of the catalyst size by sputtering induces a conical growth geometry for the CNTs.



**Fig. 21.19** Schematic drawing of the active part of SAFEM showing (a) the scan over a CNT array, (b) a direct contact with a mechanical stress on a single CNT of the array.

Conversely, without ion implantation, i.e. in the absence of an electric field, the feeding of C necessary for the CNT growth depends drastically on the existence of a free-metal surface for the decomposition and dissolution of C. The erratic location of this area across the Ni nanoparticles surface, due to impurity surface adsorption, leads then to tortuous CNTs because of a pulsed growth. Moreover, a closing of the free-metal surface by uncontrolled surface adsorption causes a considerable dispersion in the CNT length. This is the well-known spaghetti-like CVD growth of CNTs without field.

### (3) Catalyst particle sputtering

For higher values of the field and in concomitance with the carbon-atom implantation, the sputtering of the Ni particle occurs. Within the smooth growth model at the back of the particle, this leads to a conical CNT in relation with the continuous decrease of the Ni particle diameter due to the sputtering (Fig. 21.18(c)).

### 21.7.3 Characteristics of CNTs in relation to FE features

Parameters that may influence the FE characteristics of CNTs are the field-enhancement factor  $\gamma$ , the emission area and the surface reactivity in medium and poor vacuum. From an analytical point of view, and in particular for FE characterizations, the essential point to take into consideration is that field emitters inherently produce a FE conical e-beam because of the radial direction of the electric field surrounding the emitter. Therefore, the conventional FEM analyses of the projection image on a screen located a few cm away cannot be used to probe the features of individual CNTs within an array. To probe individually each CNT of an array, a scanning-anode field emission microscope (SAFEM) must be used (Fig. 21.19). With a SAFEM, a spherical anode is brought in front of each field emitter, by using piezo-driven displacements with a resolution of one nm, and if the diameter of the anode is less than the spacing between two adjacent FE sites, the FE characteristics investigated can

only be related to the individual emitter in front of the microanode (Semet *et al.* 2005). Nanometer-controlled displacement also allows a soft contact between the probe ball and the CNT for direct conduction measurements.

#### 21.7.3.1 Field-enhancement factor $\gamma$

The cylindrical CNTs can be considered as a hemisphere on a post, so the theoretical value for  $\gamma$  is around 200 for a length  $\sim 6\text{ }\mu\text{m}$  and a diameter  $\sim 50\text{ nm}$  (see Appendix 3); a value that has been confirmed experimentally by SAFEM measurements (Milne *et al.* 2003). This  $\gamma$  value can be increased by a conical geometry of the CNTs.

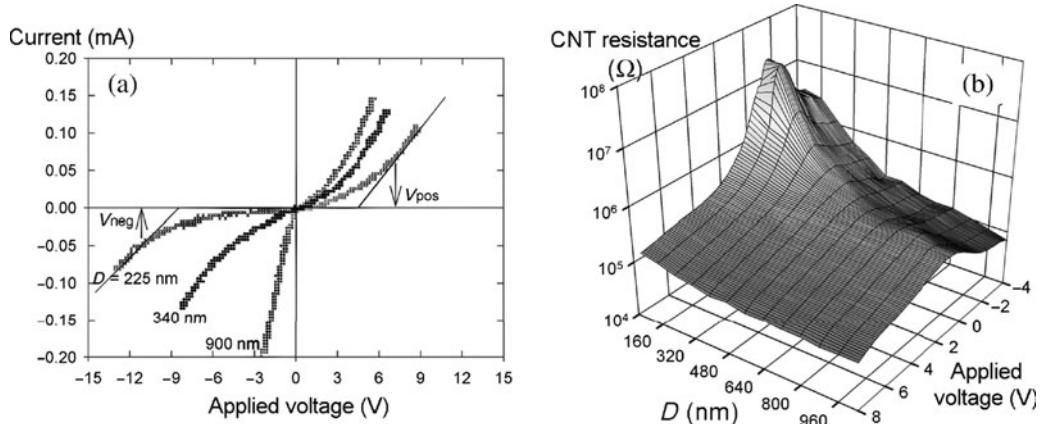
#### 21.7.3.2 Electrical-conductivity measurements

Compared to a cone, cylindrical geometry has the best aspect ratio for  $\gamma$  (Utsumi 1991) but, conversely, it gives the lowest electrical conductivity, which results in the heating of the CNT for high FE currents (Purcell *et al.* 2002). Three complementary experimental measurements have been done to characterize the CNT conductivity: with suspended nanotubes between two contacts (Lee *et al.* 2002), from the shifting of the TEDs of the FE electrons (Purcell *et al.* 2002) and directly on each CNT of an array with SAFEM (Semet *et al.* 2005). Following are the main results:

- (1) At room temperature, resistances of the nanotubes are mostly in the range of  $1\text{--}10\text{ k}\Omega/\mu\text{m}$  length. If conduction is assumed to be through the entire cross-sectional area of the cylindrical nanotube, the resistivity of the nanotube is between  $10^{-6}$  and  $10^{-5}\Omega\text{ m}$ .
- (2) Temperature-dependent resistance measurements show a small increase in resistance as the temperature is lowered from 300 to 4.2 K and also indicated a drop of about 40% in the resistance of the CNT as the temperature increases from 300 to 2000 K. This decrease in resistance at higher temperatures indicates that CNTs have thermally activated defects that do contribute to the conduction at higher temperature, with an activation energy estimated to be about 30 meV.
- (3) Direct SAFEM measurements on individual CNTs within an array show non-linear and non-symmetric  $I\text{-}V$  characteristics and a conductance that is dependent on the axial mechanical compression applied to the CNT. There is a reduction in the resistance with bending (Fig. 21.20). Simulation-derived data supports the hypothesis that, as the nanotube is mechanically compressed, its overall conduction increases as a result of the increasing number of parallel conduction channels that can participate in the overall conduction. This is in agreement with the upper notion of thermally activated defects. By modifying the mechanical strain, one can tune back and forth the number of active conduction channels, and thus modulate the resistance of the nanotubes just by bending.

#### 21.7.3.3 Emission area and surface reactivity

For as-grown CNTs, the FE presents poor current stability above 10 nA, with highly curved Fowler–Nordheim (FN) plots, distinct and irreproducible peaks in the TEDs. These effects are known to be specific to FE from surfaces with adsorbed molecules or structures (Swanson and Crouser 1970). Each

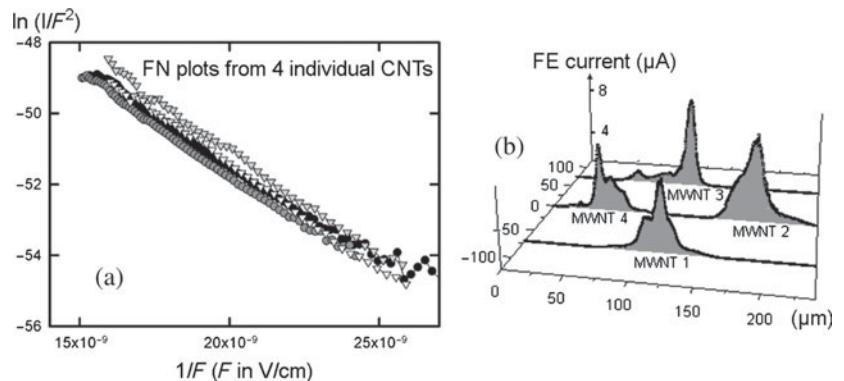


**Fig. 21.20** (a) Non-symmetric  $I$ - $V$  characteristics of CNTs under different mechanical stresses. (b) Correlated CNT resistances versus applied voltages and stresses (Semet *et al.* 2005).

MWNT therefore requires a cleaning/conditioning procedure in order to obtain reproducible FE characteristics and in particular to present  $I$ - $V$  measurements that strictly follow the conventional FN formula, i.e. plotting  $\ln(I/V)$  versus  $1/V$  resulted in a straight line, as in Fig. 21.21(a) (Semet *et al.* 2002).

The room-temperature cleaning/conditioning process is a three-step procedure:

- **Step 1:** Just after the start of emission ( $\sim 1$  pA), the emission current is steadily increased until a sudden decrease in the current is observed. This sharp drop in the FE current generally occurs when it reaches a value of  $\sim 0.1 \mu\text{A}$ .
- **Step 2:** After this sharp decrease, the FN plots, which show high curvature and saturation, evolve towards straight lines as the maximum conditioning currents increase gradually from  $0.1 \mu\text{A}$  to a few  $\mu\text{A}$ . Concurrently, there is a noticeable decrease in both the occurrence and the amplitude of the observed fluctuations in the emission current.
- **Step 3:** Thereafter, by increasing the conditioning current to  $5 \mu\text{A}$  or more, reproducible straight lines for the F-N with very stable emission current are obtained.

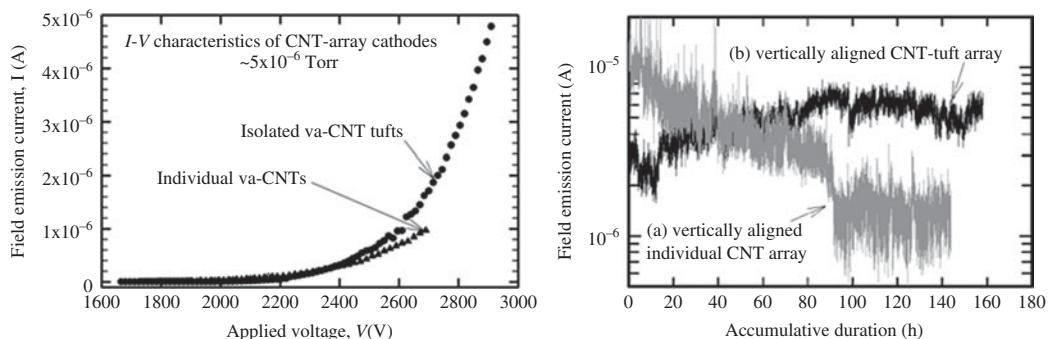


**Fig. 21.21** FE characteristics from four individual CNTs of an array after the conditioning process. (a) Similar FN plots/slopes indicate a same work function for the four CNTs. (b) The four CNTs have also the same FE current, which is a consequence of the same geometry.

This process is interpreted as a cleaning of the surface at the apex of the CNT by FE. Just after the introduction of the sample into vacuum, FE with an adsorbate-covered apex begins at localized areas that have the smallest work function. This causes a formation of nanoprotusions due to electric-field-driven surface diffusion of the physisorbed adsorbates. From the measurement point of view, this corresponds to spatial instabilities and a large fluctuation rate in the FE current. At larger currents, up to a few  $\mu\text{A}$ , the local temperature becomes high enough to field evaporate the physisorbed particles. Consequently, an increase of the global work function at the apex occurs, with a stabilization of the FE current. For FE currents  $> 5 \mu\text{A}$ , most of the adsorbates are field evaporated and those that stay are strongly bonded to the surface. From a measurement point of view, this corresponds to a stable emission with reversible straight line F–N characteristics, with a work function value in the range of 4 to 5 eV. The combination of high temperature and field leads to the well-known mechanism of field-driven thermal runaway, which is an extremely unstable situation for metal field emitters that generally break down without warning. In contrast, as the resistance of CNTs decreases substantially with temperature, giving a negative feedback to heating, and as surface diffusion is much slower for covalent carbon, both of these peculiar qualities inhibit the field-driven nanoprotusion formation, allowing the possibility of cleaning the surface by field desorption with no fear of the CNT destruction by thermal runaway.

After a cleaning/conditioning process, the electron emission features of individual isolated CNTs, spaced  $100 \mu\text{m}$  apart, are investigated using a SAFEM system, as seen in Fig. 21.21. Each of the four adjacent emitters is subject to a constant applied bias of 260 V from the anode, and exhibits the same FE current in the  $\mu\text{A}$  range (Fig. 21.21(b)). They also reveal the same work function, represented by the same slope for the four FN plots (Fig. 21.21(a)). These two FE characteristics suggest that the apex of the different CNTs of the array is mostly identical after the conditioning process. This is possible as the geometry of the nanotubes obtained by PECVD is almost identical, and in particular if they have the same apex radius. Furthermore, the graphitic nature of the surface at the apex favours physical adsorption, so that the field evaporation of the adsorbates during the conditioning process leaves the apex surface in the same relatively clean state, having the same work function. This result indicates that, with an adequate conditioning process, reproducible and uniform FE from CNT arrays can be obtained.

The maximum emission currents observed from each CNT, after conditioning, are 10 to  $20 \mu\text{A}$  before failure. Until this limit current, FE from CNTs is very stable due not only to the small sticking coefficient intrinsic to a graphene surface, but also to the possible heating of the CNTs by a FE current-induced Joule effect, which prevents readsorption. For FE currents  $> 20 \mu\text{A}$ , the electrostatic forces exert on the CNTs—which is estimated to be greater than  $10^6 \text{ N}$ —in concomitance with higher temperature—estimated to be more than 2000 K—leads to a rapid shortening of the CNTs by field evaporation, therefore to a sharp drop in the FE current or CNT failure.



**Fig. 21.22** On the left-hand side, plots of  $I$ - $V$  FE characteristics within a diode structure for individual va-CNT array and isolated va-CNT tuft array. On the right-hand side is the corresponding FE current evolutions with a constant FE voltage at 3 kV. For va-CNT tuft array the FE current of  $\sim 5 \mu\text{A}$  can be kept constant for a total duration of 600 h.

#### 21.7.3.4 Arrayed CNT cathode lifetime, individual versus tuft structures (Mauger *et al.* 2006)

A va-CNT array with a controlled spacing is a recommended structure for cathodes using CNTs as field emitters. However, to improve the FE lifetime and stability of the emitted current, isolated-tuft-CNT arrays should be favored to standalone CNT arrays (Fig. 21.22).

There are two main characteristics that differentiate FE of arrays of individual CNTs from the tuft-CNT structure. The first one concerns the cathode lifetime. In the case of arrays of tufts of CNTs, the collective behavior induces a self-replacing process that increases the lifetime. The second characteristic relates to the fluctuations of the FE currents. For tuft-CNT arrays the FE currents are less noisy compared to individual CNT array emission. This may come from a better yield in the FE from the tuft arrays or, in other words, the currents are obtained from integration of a larger number of FE sites due to the tuft structure. These FE characteristics can be explained as follows:

- (1) The loss in field enhancement, by using tufts instead of isolated CNTs, is between 5% and 10% in most of the observed tuft structures. This is why, for the same applied FE voltage, the initial current in the tuft structure is smaller than for isolated CNTs.
- (2) For individual va-CNT arrays, after  $\sim 150$  h of continuous FE nearly 60% of the CNTs are missing. This behavior has been analyzed in detail by Mauger *et al.* (2004). Therefore, under large FE currents and in a vacuum of  $5 \times 10^{-6}$  Torr, the ion backscattering means a progressive destruction of the CNTs in the arrays, leading then to the failure of the cathode when most of them are destroyed. The backscattering process appears to be the main cause that limits the lifetime of individual va-CNT array cathodes.
- (3) For va-CNT tuft-array cathodes, after a 600-h run, the cathode shows a decrease, not only in the number of CNTs per tuft, but also in the density of the tufts. Such results indicate that the first consequence of the e-beam-induced ion sputtering is to prune out the tuft. The ion backscattering process still appears to be the main cause that limits the lifetime of the cathode, but due to the collective behavior of the CNTs

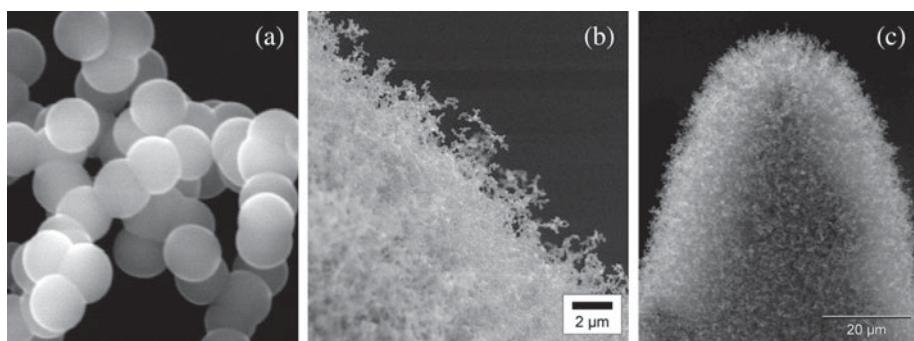
under the ion backspattering phenomena, the progressive pruning out of the tuft leads to an increase of the FE cathode lifetime, by a factor that is experimentally measured to be between 5 to 10 times more compared to individual CNTs. The reason is that each tuft stays as a FE site as long as one CNT exists.

- (4) For tuft cathodes, the better stability (less noise) of the FE currents indicates that the density of emitting sites for the tuft cathodes is larger than for individual va-CNT array cathodes.

## 21.8 Carbon-nanopearl field emitters

The quest to develop improved and cheaper cold emitters led to the emission characterization of CNTs, which represent the current benchmark for cost-effective field emitters. However, CNTs show problems of their own: the tubular structure of the CNT confines the FE to the apex region. For layered films of CNTs, the mats that are obtained by a low-cost technology, such as paste deposition, have an overall field-enhancement factor drastically reduced due to the mutual screening effect within the spaghetti-like structure. The consequence is a sparsely distributed emission sites resulting in non-uniformity of the emission area. The solution to this problem is the much more expensive and time-consuming growth of parallel arrays of individual, “standalone” CNTs/tuft-CNTs perpendicular to the substrate surface, significantly adding to the cost of fabrication of large electron sources. Moreover, for the va-CNTs arrays, the drawback is the fixed number of emitting CNTs. In poor vacuum, the gradual destruction of the CNTs by retro-ion sputtering, leads to an irreversible decrease of the emission sites. The lifetime of these cathodes is then limited, it is dependent on the density of the emission sites and they have an upper limit to avoid the mutual electrostatic screening.

There is a recent cost-effective carbon nanomaterial, “carbon nanopearls” (CNP) (Levesque *et al.* 2004) that are monodisperse solid 150-nm diameter nanospheres of nanocrystalline carbon, and form continuous 3D chains analogous to strings of pearls (Fig. 21.23).



**Fig. 21.23** (a) High-resolution SEM of carbon nanopearls ( $\sim 150$  nm diameter). (b) and (c) show the surface structure of the CNP layer deposited on a metallic tip.

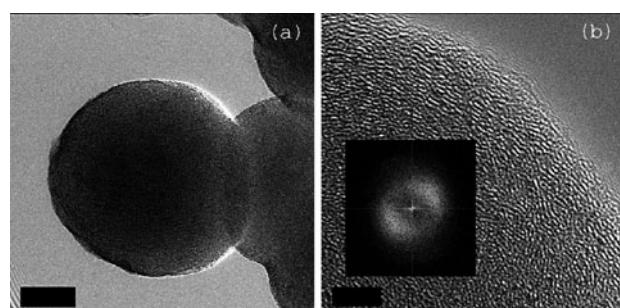
CNPs have the principal advantage that they are readily produced and do not require structural orientation for stable and uniform FE. Compared to the tubular geometry of CNTs, the nanopearls have the advantage of statistically presenting a high density of apex areas with a small radius of curvature ( $\sim 75$  nm) when deposited on a planar surface. Moreover, these spheres are composed of graphitic flakes that are unclosed at the surface and therefore believed to exhibit many dangling bonds with the potential to enhance the FE current (Mayer *et al.* 2002; Oshima *et al.* 2002). These two features give the nanopearls excellent prospects as a cathode material.

### 21.8.1 CVD growth process of carbon nanopearls

The CVD synthesis uses a mixture of acetylene (as the source of carbon) and nitrogen (20%  $\text{C}_2\text{H}_2$ , 80%  $\text{N}_2$ ). This gas mixture flows through a quartz tube (80 sccm ( $\text{C}_2\text{H}_2$ ) / 400 sccm ( $\text{N}_2$ )), in which Ni clusters of around 100 nm diameter have been placed as catalyst for the CVD reaction. The CNPs are formed at 700 °C and with higher temperatures producing larger diameters. The product, a volumetric foam-like structure (Fig. 21.23), is formed in the reaction vessel where the Ni catalysts are deposited.

The CNPs show the following characteristics:

- (1) High-resolution SEM observations shows strings of nanopearls organized into a 3D foam-like configuration with the sphere diameter of about 150 nm and about 85% monodispersity. The length of the strings can be up to hundreds of micrometers, with frequent changes in the direction. Figure 21.23(a) shows the as-grown product, illustrating the very high purity nanosphere composition without any by-products. The foam-like macroscopic appearance of the product can have a volume of tens of  $\text{cm}^3$ .
- (2) High-resolution transmission electron microscopy (HRTEM) observations confirm the spherical shape of the nanopearls and a uniform diameter  $\approx 150$  nm (Fig. 21.24(a)). High-magnification observations indicate that they are solid spheres composed of flakes having 2D dimension of a few nanometers (2 to 6 nm) and organized in concentric layers (Fig. 21.24(b)). The interfringe distances are about 0.47 nm. Due to the flake structure, the surface of these spheres exhibits atomic corrugations



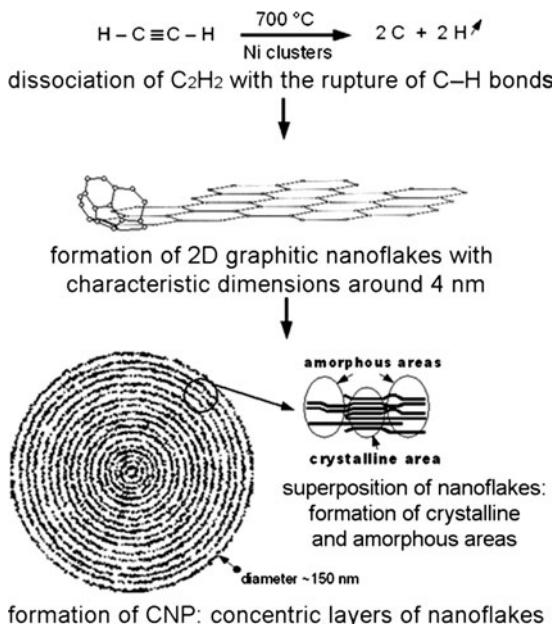
**Fig. 21.24** HRTEM of a CNP. The bar sizes in (a) and (b) are 40 nm and 5 nm, respectively.

corresponding to the unclosed graphitic flakes on the surface. The diffraction patterns are diffuse rings with a directional intensity enhancement (insert in Fig. 21.24(b)) indicating a mixture of amorphous and oriented crystalline carbon species. The presence of nanocrystalline and disordered graphite is confirmed by Raman spectroscopy with spectra showing two peaks, a G-peak at  $1590\text{ cm}^{-1}$  (graphite), and a second D-peak around  $1350\text{ cm}^{-1}$  (amorphous).

- (3) The sizes of the graphitic crystallites are estimated from X-ray diffraction and Raman spectroscopy measurements. Raman analyses indicate a plane coherence length of about 4 nm, whereas the X-ray analyses show an average size of crystallites about 2 nm with a basic cell characteristic near the conventional graphitic structure, with a distortion less than 10%.
- (4) From EDX measurements the composition of the product is mainly carbon with traces of oxygen, silicon and sulfur due to contamination from the chamber and during the transfer of the samples. No traces of Ni are observed inside the CNP foam, indicating that if present its percentage concentration in the nanopearls is less than 0.1%. This is the main difference with CNT growth. Contrary to the CNTs for which one Ni cluster is always associated to each CNT, there is no Ni associated with each of the carbon nanopearls or strings of nanopearls.

The following steps (Fig. 21.25) are proposed for the growth of these nanopearls, based on mechanisms previously suggested for C nanocompound formation (Wang and Kang 1996):

- (1) Decomposition of the acetylene into atomic carbon;
- (2) Formation of nanosize wavy flakes of graphene. The wavy 2D structure of these flakes is obtained by the combination of pentagonal and heptagonal



**Fig. 21.25** Schematic representation of the mechanism for the formation of the CNPs in three steps. Wavy nanoflakes can be obtained by an insertion of pentagonal and heptagonal carbon rings within the planar hexagonal carbon rings.

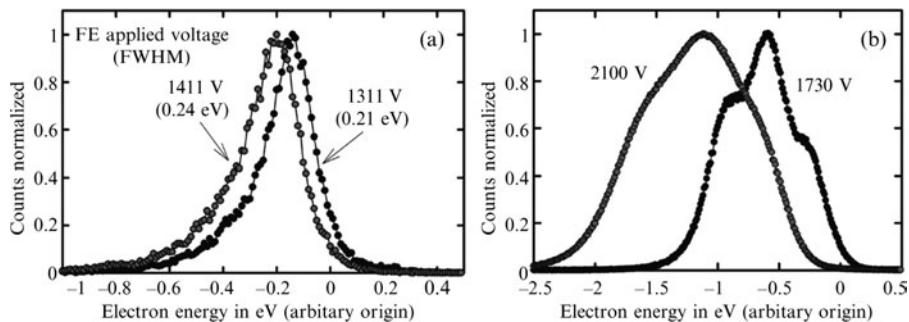
onal structures with the planar hexagonal structure of graphene in order to accommodate the curvature of the carbon sphere.

- (3) Aggregation in a concentric way to form the solid carbon nanospheres. As the flakes have 2D dimension in the range of 4 nm, the superposition of these flakes, concentric layer by layer, should statistically create 3D nanocrystallites in the range of 2 nm, in concomitance with amorphous areas. Note that there are two possibilities for this last step, one is the formation of aggregates of the graphitic flakes before their inclusion in the concentric layers to form the carbon nanopearls and the second is the direct aggregation of individual flakes on the surface of the carbon nanopearls to form the concentric layers of nanocrystallites. The volume of product formed is directly dependent on the reaction time and the Ni nanoclusters are the key for the formation of the 3D foam-like macroscopic structure of the product.

### 21.8.2 Nanopearls as graphitic nanostructures for field emission

Cold cathodes using CNPs may be obtained either by a direct growth on a metallic substrate during CVD reaction (Fig. 21.23(c)), or by incorporating them into a solvent and screen printing to deposit a layer on a substrate. The substrate could be the apex of a tip in order to optimize the global field-enhancement factor. Initially, the newly grown cathodes exhibited unstable FE currents. However, an *in-situ* conditioning process consisting of a stepped increase of the FE current, which is similar to the one used for CNTs, eliminated these instabilities. The typical current–voltage characteristics after the conditioning process follow the expected metallic behavior, i.e. a linear variation of  $\ln(I/V^2)$  vs.  $(1/V)$ , even if the total current  $I$  results from the contribution of multiple FE sites. The CNPs surface is made of graphene, these FE sites exhibit the same quality for current stability as observed with the CNTs, and currents in the range of tens  $\mu\text{A}$  are maintained stable for hours in a poor vacuum environment. At high currents, over a few tens of  $\mu\text{A}$ , bright spots appear at the surface of the cathode indicating that Joule heating of some CNP strings is occurring due to the high currents through the strings, a phenomenon that has also been observed with CNTs. However, unlike CNTs, when an arc occurs at high currents with CNP cathodes, any structural breakdown is self-repairing and the same cathode returns to stable emission after a reconditioning process. This self-repairing feature allows good current stabilities, up to hours, to be obtained even in poor vacuum of  $10^{-4}$  Torr of nitrogen or under a partial oxygen pressure of  $10^{-5}$  Torr (Levesque *et al.* 2005). Actually, due to the foam-like structure of the material, the degradation process of the outermost CNP layer caused by ion sputtering does not bring a rapid and total destruction of the cathode, providing its immediate replacement by the underneath layer until the total deletion of the deposited CNP layer.

FE electron spectroscopy from a single string of CNPs shows a metallic-like behavior for the TEDs of the emitted electrons (Mouton *et al.* 2008). This means a conventional single-peak shape TED with a FWHM in the range of



**Fig. 21.26** (a) Characteristic TED from a single nanoparticle string indicating an energy distribution of  $\sim 0.2$  eV and a shift towards the lower-energy side for increasing applied voltage. (b) TED spectra from a carbon-nanoparticle-tipped cathode with a mean density of FE sites in the range of 1 to 2 sites/ $\mu\text{m}^2$ .

0.2 eV (Fig. 21.26(a)). However, as the resistance of the string is not negligible, the position of the peak is then dependent on the potential drop along the string, which is a function of the FE current, as shown by the shift of the TED peak towards the lower-energy side for increasing FE voltage. After a conditioning process, more than one string contributes to the FE pattern. The resulting TED is then the convolution of different TED peaks related to the independent FE sites. The TED spectrum is then a multiple peak spectrum with values of the FWHM in the range of 0.5 to 1 eV, behavior specific to the foam-like structure of the carbon nanoparticle film (Fig. 21.26(b)).

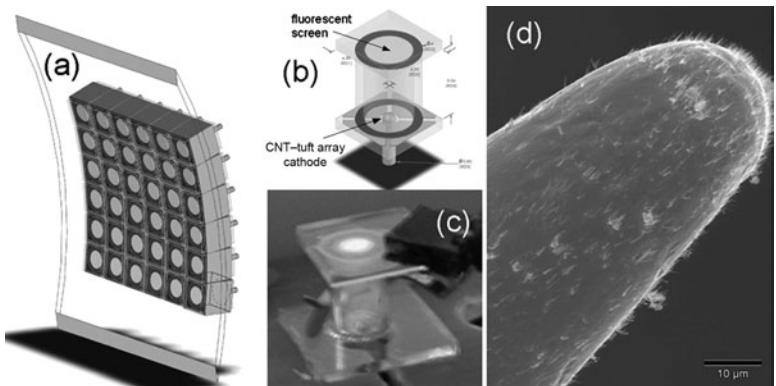
## 21.9 Applications and uses of carbon nanocompounds, CNTs and CNPs, as cold cathodes

Nowadays, there are three main applications of the carbon nanocompound cold cathodes: FEDs, mass spectrometry and X-rays.

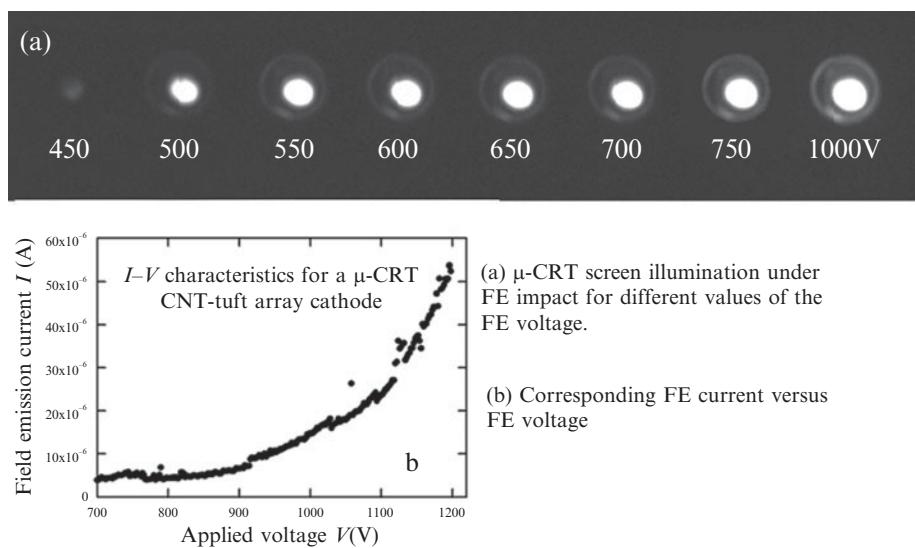
### 21.9.1 FED applications

In most FED applications CNTs are used as field emitters either under non-directional mats (spaghetti-like CNTs) or as va-CNT arrays. Among them, one application is the giant field emission displays obtained by an array of millimetric CRTs ( $\mu$ -CRTs) disposed into a two-dimensional matrix (NANOPAGE European project 2004) (Fig. 21.27). The cathode of the  $\mu$ -CRT is composed of a substrate metallic tip with an apex radius in the range of 20 to 400  $\mu\text{m}$  on which individual vertically aligned CNTs/tuft-CNTs are grown directly within an arrayed structure (Fig. 21.27(d)).

The va-CNT tuft field emitters are inserted into a  $\mu$ -CRT, as schematically represented in Fig. 21.27(b), having a diameter of the fluorescent screen equal to 3 mm. The direct growth on metallic tips allows a direct use of these arrays



**Fig. 21.27** (a) Schematic representation of a  $6 \times 6$  matrix of  $\mu$ -CRTs of a giant flexible FED. (b) Schematic representation of one micro-CRT. (c) FE illumination of the fluorescent screen of a  $\mu$ -CRT. (d) SEM of the apex of the substrate metallic tip showing vanadium CNT tuft array.



**Fig. 21.28** Screen illumination under FE impact for different values of the applied voltages from 450 to 1000 V. The diameter of the  $\mu$ -CRT fluorescent screen is 3 mm (see Fig. 21.27).

as cold cathodes after a conditioning process to ensure an optimum setting for a uniform and stable overall FE of the arrays. Figure 21.28(a) shows the illumination of the screen for different applied FE voltages.

Note that a uniform illumination of the screen is obtained and only a driving voltage of  $\sim 50$  V is enough to flip from a dim state (450 V) to a fully lightened state (500 V). A saturation of the fluorescent screen is reached for  $V \sim 550$  V. Figure 21.28(b) is a plot of the  $I - V$  characteristics showing the possibility to have FE currents over  $50 \mu\text{A}$  for each  $\mu$ -CRT.

### 21.9.2 Mass spectrometry (ToF-MS) and X-ray applications

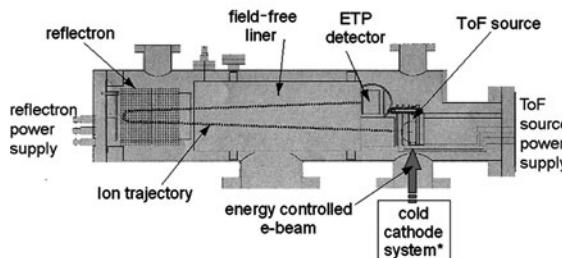
A nanoparticle-coated emission surface has a certain redundancy and resilience to damage, because if the top layer of nanoparticles is destroyed, the next layer can

begin to emit. Free CNP layer cathodes have proven to be good emitters under poor vacuum conditions, offering superior performance to conventional field emitter designs and even to CNT array cathodes. In order to further improve CNP cathode reliability and reduce the cost of fabrication, CNPs are mixed with polymer solutions of ethyl-cellulose or polymethyl-methylmethacrylate (PMMA), and this composite is deposited on conductive substrates to be the cathode surface. The viscosity of the composite solution is controlled by adjusting the relative proportions of the solvent, polymer and CNP components. This allows the features of the CNP polymer paste to be tuned to suit a variety of deposition techniques. The CNP composite layer may be deposited on sharp-tipped substrates, in order to take advantage of the field-enhancement factor of the substrate, or on planar surfaces generating 2D area emitters.

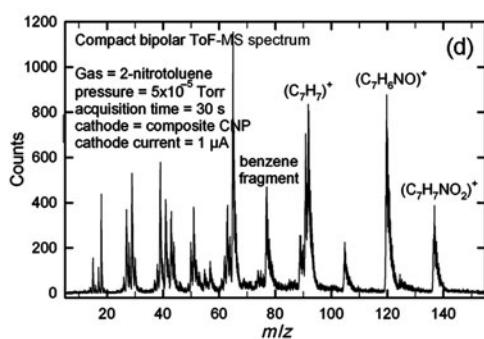
The role of the embedding polymer is to mechanically stabilize the CNP strings during FE, transportation and general handling. Therefore, only those nanopearls that protrude above the polymer surface would emit. The use of the composite CNP in polymer surface results in noticeably increased emission lifetime and stability over the free nanopearl surface, in particular in the working pressure range of the ToF-MS and X-ray chamber ( $\sim 10^{-5}$  to  $10^{-6}$  Torr). Embedding the CNPs in an insulating polymer matrix, can extend emission lifetimes to in excess of 200 h at mass spectrometer vacuums—i.e. in the presence of organic polymer gases, such as acetophenone ( $C_6H_5COCH_3$ ), 2-nitrotoluene ( $C_6H_4(CH_3)(NO_2)$ ), acetonitrile ( $CH_3CN$ ) and toluene ( $C_6H_5CH_3$ ), and in the range of  $10^{-5}$  to  $10^{-6}$  Torr—without developing filamentous adsorption compounds that are the main cause for FE cathode breakdown under this environment.

These composite CNP cold cathodes have been used in a compact time-of-flight mass spectrometer (ToF-MS) (Mouton *et al.* 2008) (Fig. 21.29). Due to the directional emission imposed by the applied field, the ratio between total emitting current and effective current for ionization is greater than 50%, which has to be compared with a value of about 10% for the thermionic cathode. Figure 21.29(c) indicates that an effective current in the ionization chamber of order 5 nA is ample for the formation of ion species in a total vacuum pressure of  $10^{-7}$  Torr of residual gas for an acquisition time of order 600 s. However, the usual cathode current to have acquisition time of the order of a few tens of seconds is in the range of 1 to  $10\ \mu A$ , as shown in Fig. 21.29(d) for a pressure of  $2 \times 10^{-5}$  Torr.

The directional and high current density extracted by FE from CNP cold cathodes are also used in compact X-ray units (Binh *et al.* 2007) (Fig. 21.30(a)). The vacuum inside the X-ray chamber is  $5 \times 10^{-5}$  Torr and the CNP cathodes have had an effective duration of more than 200 h at  $15\ \mu A$ . The X-ray spectra obtained for a FE current of  $15\ \mu A$  are presented in Fig. 21.30(b) and are obtained without the need to cool the anode. This opens the possibility for an evolution towards an extreme miniaturization and fast modulation of the X-ray emission.



Schema of the compact ToF-MS using the CNP cold cathode



(a), (b), (c) and (d): ToF-MS spectra

For (a), (b) and (c) all the settings of the Tof were kept identical.

(a): W filament cathode

(b), (c) and (d): polymer-embedded CNP cold cathode

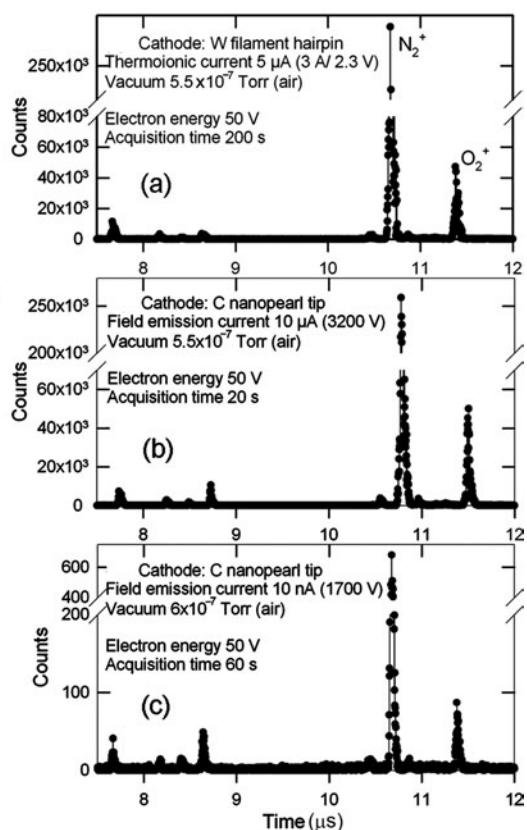


Fig. 21.29 Schema of the compact bipolar ToF-MS, the cold cathode system includes CNP cathode, extraction and retarding lenses and diaphragms (top left). (a) to (d) are ToF spectra. The ratio between cathode current and effective current for ionization is 10% for the thermionic cathode for spectrum (a) and 50% for the FE cathode for spectra (b), (c) and (d).

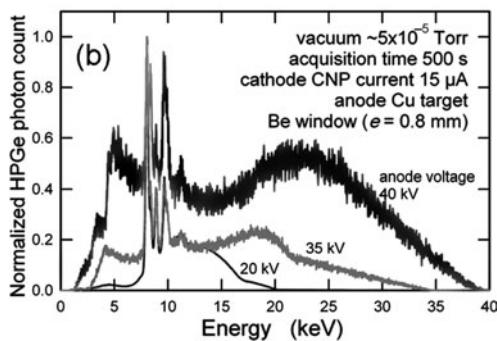
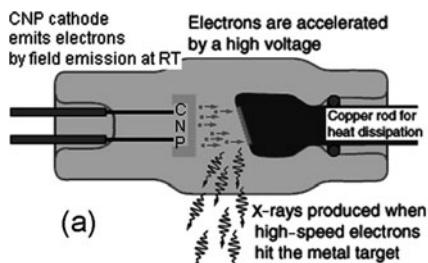


Fig. 21.30 (a) Schema of the compact X-ray tube using a composite CNP cathode. (b) Spectrum evolution with the impact energy of the electrons from 20 to 40 kV, with the characteristic peaks and brehmsstrahlung continuum.

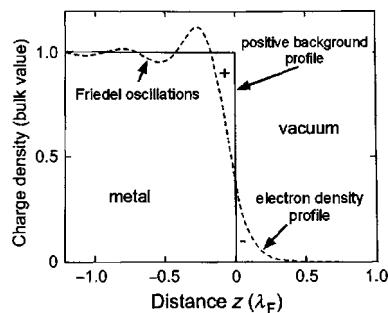
## 21.10 Conclusions

Cold cathodes can be classified, based on their geometry, into field emitters and planar cathodes. Field emitters operate by tunnelling electrons from a material into vacuum as the result of an applied electric field. As the FE process is dependent on a high electric field at the surface, a sharp, tip-like geometry is necessary. Planar cold cathodes have a flat emitting surface with low surface barrier and use a variety of principles for electron emission at very low electric fields and, therefore, there is no need for severe field enhancement. First and foremost, the very low or negative electron affinity cathodes can emit electrons because the surface vacuum level is near or below the bulk conduction-band edge. Another type of planar cathode is the biased junction emitters (Schottky and p-n junctions) in which electrons are injected into a surface region that, consequently, presents a very low surface barrier.

Planar cathodes and field emitters both have strengths and weaknesses as cold cathodes. Some of the weaknesses are inherent to the structures, while others may be resolved with improved technology or advanced materials. Lowering the voltages of operation has been achieved by increasing the sharpness of emitters and shrinking the distances between the extraction electrodes and emitters. But the operating voltage can also be reduced by using materials with lower surface barriers. Nanotechnology, which has essentially emerged from physics, chemistry, materials science and biology, is perhaps the best technical approach to reach the ideal material for a cold cathode with low surface barriers, high electron concentration, high thermal conductivity, mechanically hard and non-reactive in a vacuum environment. Moreover, with its paradigm of “bottom-up” fabrication paths, nanotechnology can take up the challenge of a low-cost, precise fabrication of nanostructures for cold cathodes that bring together the strong points of planar cathodes and field emitters. This last condition is essential to consider industrial applications.

Cold cathodes do not rely on the heating of a material to emit electrons over the vacuum barrier. The lack of heating decreases the demands on the cathode power supply and obviates the need for complex separation of the cathode from other devices, lending cold cathodes to smaller dimensions and denser integration than is achievable with thermionic cathodes. Above all the applications for field emitters currently being investigated, two types of devices have generated most of the research and show a solution taking advantage of the emission features of field emitters. The first type is flat-panel displays that can benefit from low voltage and small size of field emitter arrays and the second is high-power vacuum tubes and e-beam-based instrumentations that can take advantage of the high current density, directional emission and fast modulation of the emission current.

Above all the well-known types of particle sources, FE electron sources are the brightest currently available. With the recent development of single-atom field emitters, the nanotips produce beams brighter by two or more orders of magnitude than the conventional FE tips, with a degeneracy factor, i.e.



**Fig. 21.31** Charge redistributions at a step-like surface ( $\lambda_F$  is the Fermi wavelength).

coherence, comparable to arc-light sources. This may augur well for interferometry of correlated electrons, as well as for the development of related instrumentations.

## Appendix 1: Electron emission from solids

### A.1.1 Surface barrier

The escape of electrons from a solid without external excitations, such as high temperatures or irradiations, or external perturbations, such as electric or magnetic fields, is prevented by the presence of a surface barrier resulting mainly from charge redistribution at the solid-vacuum interface (Fig. 21.31).

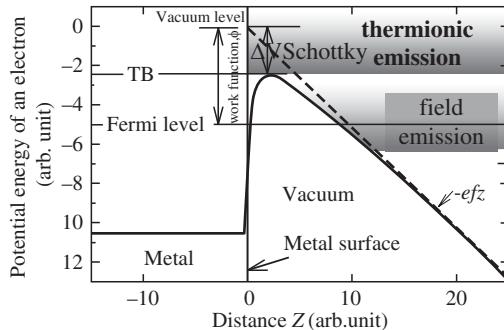
The surface barrier can be characterized by the work function  $\Phi$ , which is defined as the difference between the chemical potential  $\mu$  of the electrons inside the conductor and the electrostatic potential difference  $\Delta\varphi_{i-o}$  of an electron between the inside and the outside of the conductor (Herring and Nichols 1949):

$$\Phi = \Delta\varphi_{i-o} - \mu. \quad (\text{A.1.1})$$

The chemical potential  $\mu$  is a volume property that is independent of the structure of the surface. On the contrary,  $\Delta\varphi_{i-o}$  depends on the condition of the surface as well as upon the structure of the interior just underneath the surface; it is therefore a tuning parameter for  $\Phi$ . This is why the work function varies from one crystallographic facet orientation to another and why electronegative adsorbates such as oxygen usually increase the work function, whereas electropositive adsorbates such as Ba or Cs decrease the work function.

### A.1.2 Electron emission

In this appendix the two basic mechanisms to extract electrons from a solid surface will be reviewed. Within a metal, an electron current density  $j_m$  of roughly  $10^{12} \text{ A cm}^{-2}$  impinges on the inner surface. Only a small fraction of this current escapes from the metal either jumping over the surface barrier, a



**Fig. 21.32** Potential energy  $V(z)$  for an electron in the vicinity of a metal surface with an applied field  $F$  ( $V(z) = \Phi - eFz - e^2/16\pi\epsilon_0 z$ ,  $\epsilon_0$  is the vacuum dielectric constant). The decrease in the effective surface barrier  $\Delta V_{\text{Schottky}}$  due to the image effect is  $\sim 3.8 F^{1/2}$  (for  $\Delta V_{\text{Schottky}}$  in eV and  $F$  in  $\text{V}/\text{\AA}$ ).

process called thermionic emission, or tunnelling through it when the width of this surface barrier is lowered and rounded by the presence of an applied field  $F$  (Fig. 21.32), a process called field emission (FE). Thermionic emission is also called “hot emission” because it needs a supply of energy  $kT$  to raise the electrons from the Fermi level to the top of the barrier by heating the cathode to a temperature  $T$ . In FE, there is a conservation of the electron energy during its tunnelling through the deformed surface barrier and detectable currents come out when the width is in the range of the Fermi wavelength, which is of the order of 1 nm. As no supply of energy to the electrons is needed, it is called “cold emission”.

The total current density extracted from a cathode is obtained by integrating the product of the charge of an electron  $e$ , the number of electrons per second per unit area incident on the barrier and the penetration probability  $D(F, W)$ , over all accessible energies defined by the Fermi–Dirac distribution inside the solid.

The current density is given by one of the following relations depending on the emitted electron energy (Murphy and Good 1956), in Hartree units:

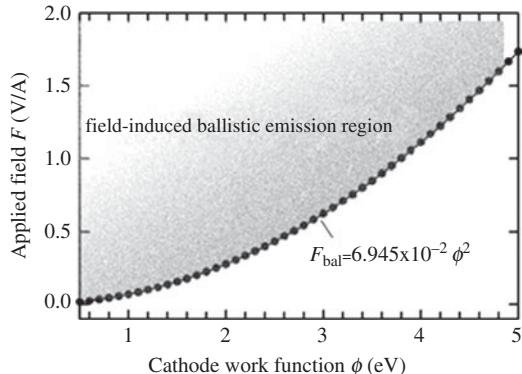
$$J_{\text{Schottky}} = \frac{(kT)^2}{2\pi^2} \left( \frac{\pi d}{\sin \pi d} \right) \exp \left( -\frac{\Phi - F^{1/2}}{kT} \right), \quad (\text{A.1.2})$$

$$J_{\text{FE}} = \frac{F^2}{16\pi^2 \Phi t^2(y)} \left( \frac{\pi c k T}{\sin \pi c k T} \right) \exp \left( -\frac{4\sqrt{2} \Phi^{3/2} v(y)}{3F} \right), \quad (\text{A.1.3})$$

$$J_{\text{TF}} = \frac{F}{2\pi} \left( \frac{kT t(y)}{2\pi} \right)^{1/2} \exp \left( -\frac{\Phi}{kT} + \frac{F^2 \Theta}{24 (kT)^3} \right). \quad (\text{A.1.4})$$

$J_{\text{Schottky}}$ , is for electrons over the Fermi level with energy in excess of the actual surface barrier height  $TB$ , with  $d = F^{3/4}/\pi kT$ . Equation (A.1.2) is called the thermionic Richardson–Schottky relation.  $J_{\text{FE}}$  is related to the electrons having energy less than  $TB$  that tunnel through the field-deformed surface barrier when its thickness is in the range of 1 nm or less. Equation (A.1.3) is called the field emission (FE) relation, with  $c = c(F, \Phi, t(y))$  an approximate evaluation function of  $D(F, W)$  about the Fermi energy, the variable  $y = F^{1/2}/\Phi$ , and  $v(y)$  and  $t(y)$  are the Nordheim elliptic functions. When both thermionic and FE are concomitant, a regime that is called T–F emission, the emission current density is given by  $J_{\text{TF}}$ , with  $\Theta = \{3/t^2(y)\} - \{2v(y)/t^3(y)\}$ .

The upper analytical resolution has not considered the direct electron emission from the Fermi sea and over the surface barrier. This last process, called field-induced ballistic emission (Forbes 1999), corresponds to a situation where the top of the barrier is pulled down under the Fermi level. This is a consequence of a drastic decrease of the surface barrier height by the Schottky effect,  $\Delta V_{\text{Schottky}}$  (Fig. 21.32). This means that the top of the barrier  $TB$  is pulled down below the Fermi level for applied field values  $F > F_{\text{bal}} = 6.945 \times 10^{-2} \Phi^2$ , where  $F_{\text{bal}}$  is the threshold field when  $TB$  is at the Fermi level. Under such a condition, comparable to a situation of negative electron affinity, electron emission no longer takes place solely by tunnelling



**Fig. 21.33** The field-induced ballistic region lies over the curve  $F_{\text{bal}}$ , it corresponds to a situation where the top of the barrier is pulled down under the Fermi level by the Schottky effect.

through the barrier, but occurs dominantly by flowing over the top of the barrier directly from the Fermi sea. This mechanism is not relevant for metal cathodes having  $\Phi > 4.5$  eV because it only happens for  $F \geq 1.5$  V/Å. However, for cathodes having  $\Phi < 2$  eV it rapidly becomes the dominant mechanism when  $F > 0.2$  V/Å, as shown in Fig. 21.33.

### A.1.3 Thermionic emission versus field emission

Strictly speaking, the energy-level intervals of thermionic electrons are limited for the lower energy side by TB (without or with deformation due to applied field) and the upper energy side is defined by the Fermi–Dirac distribution function in temperature (see Fig. 21.32). On the other side, the interval of energy levels of electrons emitted by FE is defined for the lower-energy side by the tunnelling barrier width, which must be less than  $\sim 1$  nm, and for the upper energy side also by the Fermi–Dirac distribution. Therefore, two main differences result from these two processes.

The first one resides in their brightness, a consequence of the important difference regarding the electron densities at each energy level between the Fermi sea and the Fermi–Dirac distribution over the Fermi level. The brightness  $B$  is experimentally defined as the current density  $J$  (current per unit area normal to the beam) emitted into a solid angle  $\Omega$  (Silverman 1994). Thus, the number of electrons received in a time interval  $\Delta t$  within a solid angle  $\Delta\Omega$  through a detecting surface  $\Delta A$  can be written as follows:

$$\Delta n = \frac{B}{e} \Delta A \Delta t \Delta\Omega. \quad (\text{A.1.5})$$

By introducing a beam degeneracy factor  $\delta$  as the mean number of particles per cell of phase space,  $\Delta n = \delta \times$  (number of occupied cells), i.e.

$$\begin{aligned} \Delta n &= \delta \times \left[ \frac{2\Delta p_x \Delta p_y \Delta p_z \Delta x \Delta y \Delta z}{h^3} \right] \\ &= \delta \times \left[ \frac{2(p^2 \Delta p \Delta \Omega)(v \Delta t \Delta A)}{h^3} \right]. \end{aligned} \quad (\text{A.1.6})$$

The experimental brightness  $B$  is related to a maximum brightness  $B_{\max}$  by the expression

$$B = \delta B_{\max}, \quad (\text{A.1.7})$$

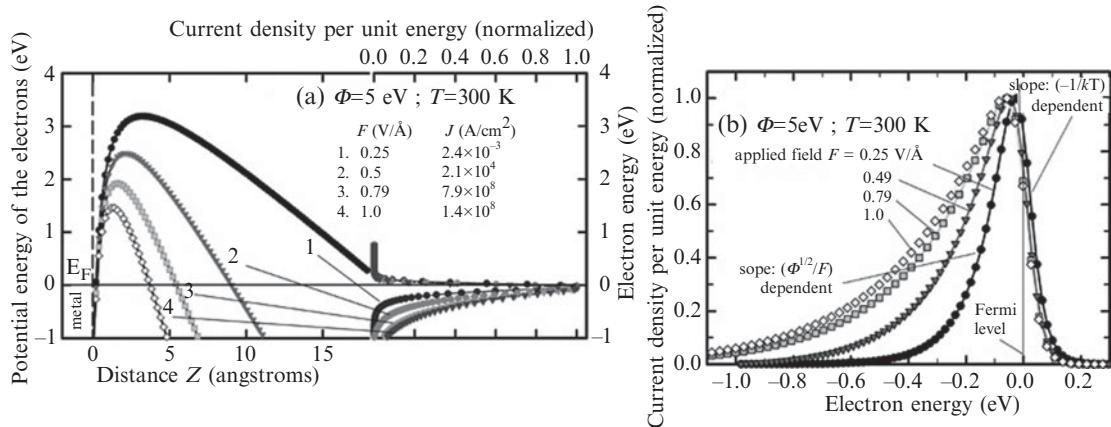
$$B_{\max} = \frac{4meE \Delta E}{h^3}, \quad (\text{A.1.8})$$

where  $E$  and  $\Delta E$  are, respectively the energy and the energy dispersion of the electrons and  $h$  is Planck's constant.  $B_{\max}$  is then directly related to the theoretical maximum number of electrons that can be extracted from a solid.

Taking the example of conventional cathodes with work function in the range of 4 to 5 eV, thermionic emission has  $\delta$  between  $10^{-12}$  and  $10^{-10}$ , whereas for FE the values of  $\delta$  are within the range of  $10^{-6}$  to  $10^{-3}$ . This means from the experimental point of view that current densities from hot cathodes have an upper limit of the order of  $100 \text{ A cm}^{-2}$ , while cold cathodes can deliver currents from  $10^5$  to  $10^9 \text{ A cm}^{-2}$ .

The second main difference between the two electron emission processes is the total energy distribution (TED) of the emitted electrons whose upper and lower limits are set by the penetration probability  $D(F, W)$ . For thermionic emission, as the emitted electrons jump over the barrier, this width is defined essentially by the Fermi–Dirac distribution over the top of the surface barrier, which could be in the range of a few eV for high-enough temperature in order to have a detectable number of emitted electrons. In the case of FE, at room temperature for example, the TED of the field-emitted electrons is governed by the tunnelling probability through the field-rounded triangular barrier (Fig. 21.34).

The FE TED spectrum therefore presents a high-energy side slope that is mostly temperature dependent (i.e.  $\propto -1/kT$ ) and a low-energy side slope that is field dependent (i.e.  $\propto \Phi^{1/2}/F$ ), with its maximum pinned near the Fermi level (Fig. 21.34(b)). The half-width (FWHM) of this distribution is less than  $\sim 0.3$  eV and, for a given temperature, the FWHM increases with the applied



**Fig. 21.34** (a) Evolution of the surface barrier with the applied field (left-hand plots) with the corresponding TED of the emitted electrons (right-hand plots). (b) Detailed plots of the TEDs.

field due to a larger deformation of the surface barrier by the field (i.e.  $-eFz$ ) (Fig. 21.34(b)).

## Appendix 2: Tip-profile evolutions by mass-transport surface self-diffusion

The solid vacuum interface is not a static system, and several kinetic processes occur depending on the substrate temperature  $T$ ; among them and of interest to this review are the surface diffusion processes. This surface diffusion over large distances is associated with possible shape changes of the solid objects, and in this case, one is strictly speaking of mass-transport surface diffusion (Herring 1953; Mullins 1959; Bonzel 1981; Seebauer 1995).

In general, surface diffusion occurs due to a gradient in chemical potential  $\mu$  and is observed either as a change in the concentration of a diffusing species or as a modification in the topography of the sample. The diffusion flux  $J$  with a concentration  $n$  of diffusing species along the direction  $x$  with the drift velocity given by the Nernst–Einstein relation, is:

$$J = -n(D/kT)(d\mu/dx), \quad (\text{A.2.1})$$

where  $D$  is the surface diffusion coefficient and  $\mu(x)$  is the chemical potential in one dimension. The mass-transport surface diffusion coefficient  $D_S$  is given by:

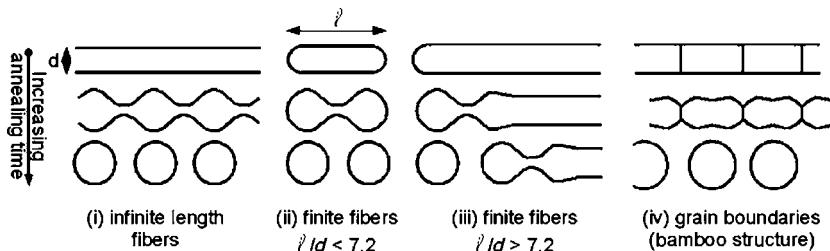
$$D_S = (n/N_0)D = D_{S0} \exp \{ -(E_f + E_m)/kT \} \quad (\text{A.2.2})$$

where  $N_0$  is the total number of adsorption sites per unit area =  $\Omega^{-2/3}$ , with  $\Omega$  the atomic volume. The activation energy of mass-transport surface diffusion is the sum of the energies of formation  $E_f$  and migration  $E_m$  of the diffusing species.

The gradient of the chemical potential along the surface, which is the driving force, has multiple origins: surface-curvature gradient (capillarity forces), free-energy gradient (equilibrium shape), and electric-field-strength gradient are among the main ones that will be detailed in the following sections. It is convenient to distinguish between intrinsic instabilities that derive from the reduction of the stored energy in the object—as with capillarity—and extrinsic instabilities that are due to interaction with external potential field—as surface free-energy gradient or electric-field gradient.

### A.2.1 Intrinsic instabilities due to capillarity forces

Intrinsic instabilities, as presented in this section, derive from reduction in the stored energy of the object, e.g. capillarity (Herring 1950; McLean 1978). Figure 21.35 indicates the four manners in which this can be achieved: (i) Raleigh instabilities of infinitely long fibers, (ii) relaxation of short fibers, (iii) retraction and progressive pinching off (ovulation process) from the end of semi-infinite fibers, (iv) development of grain-boundary grooves (bamboo structure) which grow to penetrate to fibers. All these phenomena have been



**Fig. 21.35** Schematic illustration of different modes of intrinsic microstructural instabilities for tip-geometry emitters.

observed experimentally with tip-geometry emitters, and with annealing times in the range of minutes for characteristic dimensions in the range of micrometers.

To analyze the geometrical changes due to mass-transport surface diffusion, we shall consider the free energy of a given system:

$$dF = -S \, dT + V \, dP + \gamma \, dA + \sum_i \mu_i^0 \, dn_i, \quad (\text{A.2.3})$$

where  $S$  is the entropy,  $T$  the temperature,  $V$  the volume,  $p$  the pressure,  $\gamma$  the free surface energy,  $A$  the area of the surface,  $\mu_i^0$  and  $n_i$  the chemical potential and concentration of species  $i$ . For curved surfaces and under mass-transport surface diffusion, the area  $A$  is dependent on  $n_i^2$ , therefore for solids with isotropic  $\gamma$ , the chemical potential  $\mu$  on a curved surface, with  $r_1$  and  $r_2$  the principal radii of curvature at any point of the surface, is:

$$\mu = \mu_0 + \gamma \Omega (1/r_1 + 1/r_2), \quad (\text{A.2.4})$$

where  $\mu_0$  is the chemical potential on a flat surface and  $\Omega$  is the atomic volume.

For solids with anisotropic surface energy  $\gamma(\theta)$ ,  $\theta(x)$  is the angle of orientation, the expression of chemical potential as a function of surface curvature is (Herring 1951):

$$\mu(x) = \mu_0 + (\gamma(\theta) + (\delta^2 \gamma / \delta \theta^2) \Omega K(x)), \quad (\text{A.2.5})$$

where  $K(x)$  is the curvature of the surface described by the shape function  $y(x)$ :

$$K(x) = -d^2 y / dx^2 \{1 + (dy/dx)^2\}^{-3/2}. \quad (\text{A.2.6})$$

For surface diffusion over a small range of orientations, average values  $\gamma$  and  $\gamma''$  can replace the chemical potential and its second derivative and the flux of atoms  $J$ , given by eqn (A.2.1) and modified with eqns (A.2.5) or (A.2.6), becomes:

$$J = -(\Omega N_0 D_S / kT)(\gamma + \gamma'') dK / dx. \quad (\text{A.2.7})$$

The diffusion flux  $J$  induces a change of the surface profile. The rate of motion of a surface element normal to the surface,  $dy/dt$ , is obtained by taking the divergence of  $(-J)$  and multiplying it by the atomic volume  $\Omega$ :

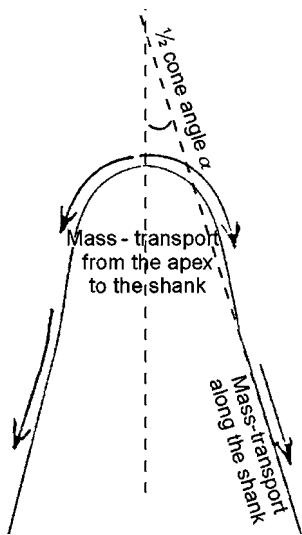
$$dy / dt = -B d^4 y / dt^4, \quad (\text{A.2.8})$$

<sup>2</sup>Demonstration for a spherical surface of radius  $r$  :  $dA = (2/r) dV$ ;  $V = \sum_i \Omega_i n_i$  ( $\Omega_i$  is the atomic volume) therefore  $dA = \sum_i (2\Omega_i / r) dn_i$ . It follows that  $\mu_i = (dF/dn_i)_{T,P,nj} = \mu_i^0 + (2\gamma\Omega_i / r)$ .

for an approximation of small slope ( $dy/dx \ll 1$ ) and with

$$B = -(\Omega^2 N_0 D_S / kT)(\gamma + \gamma''). \quad (\text{A.2.9})$$

Equation (A.2.8) can now be solved for a variety of typical profiles with different boundary conditions in order to obtain the evolution of these profiles due to mass-transport surface diffusion. It also means that, during this mass transport, the flux of atoms (eqn (A.2.7)) migrates from the surface area with high curvature towards the area having lower curvature in order to minimize the total surface energy of the system. In the following section, we will discuss the typical geometries for which the problem has been solved for conical tip geometry: field emitter tip blunting and grain-boundary grooving.



**Fig. 21.36** Due to capillarity forces the surface atoms migrate from the apex of the tip towards and along the conical shank as well.

#### A.2.1.1 Tip blunting (Nichols and Mullins 1965; Binh et al. 1974)

The blunting of a conical field emitter is a phenomenon resulting directly from capillarity forces and at quite appreciable rates for temperatures greater than about half the melting temperature. The conical geometry implies a decrease of the local surface curvature  $K$  from the apex of the tip towards the shank, therefore, a migration of surface atoms from the apex towards and along the tip shank (Fig. 21.36), as illustrated by eqn (A.2.7).

This mass transport results in an increase of the apex radius and a decrease of the tip length, a phenomenon called tip blunting. This tip-profile evolution depends strongly on the “competition” between the mass transport from the apex and along the shank, therefore, on the cone angle of the tip.

Resolution of eqn (A.2.8) gives the following results:

*For half-cone angle  $\alpha$  from  $0^\circ$  to  $3^\circ$ :*

As the mass-transport surface diffusion along the shank is slow due to the small value of the cone angle, because of a small surface-curvature gradient, surface diffusion from the apex will therefore lead to the formation of a bulbous shape near the apex and a neck immediately behind it (Fig. 21.37(a) from 0 to 13). The further decrease of the neck diameter (Fig. 21.37(a) time 39 to 94) will end in the detachment of a solid drop (Fig. 21.37(a) time 100), a process termed an “ovulation” by Nichols and Mullins.

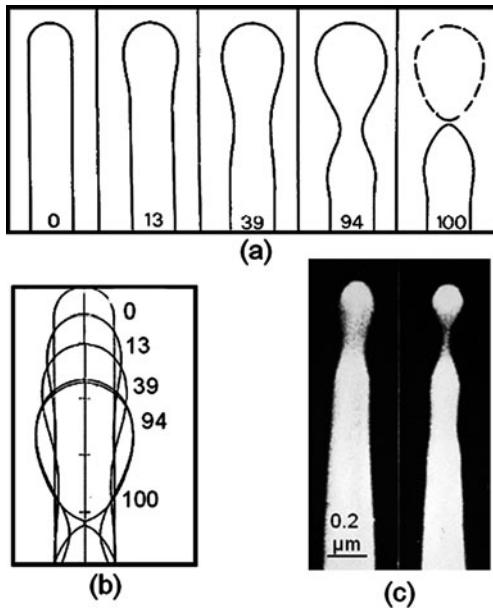
The time for ovulation is cone-angle dependent and is given in Table 21.2 for a value of  $B = 10^{-20}$ , which is typical for metal tips at temperatures  $> 0.7 T_M$ ,  $T_M$  is the melting temperature.

*For half-cone angle  $\alpha \geq 3^\circ$ :*

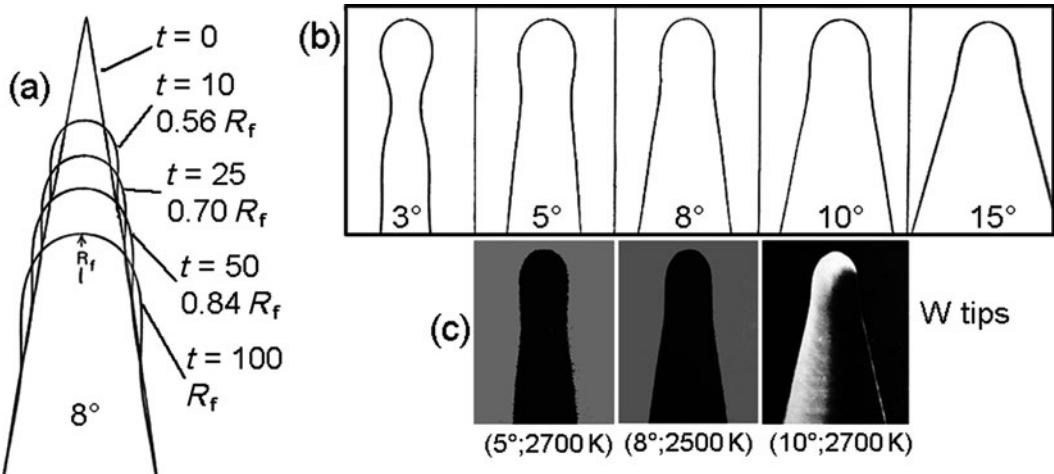
The mass transport from the apex is now balanced with the migration along the shank; therefore no spheroidization or solid-drop formation occurs and the tip blunts monotonically by keeping a steady-state shape, as shown in Fig. 21.38(a). The steady-state shapes are cone-angle dependent (Fig. 21.38(b)). The variations of the tip radius  $R$  and position  $z$  with heating

**Table 21.2** Ovulation times.

	$\alpha = 0^\circ$	$1^\circ$	$2^\circ$
$R_0 = 0.1 \mu\text{m}$	6.6 s	13 s	31 s
$R_0 = 1 \mu\text{m}$	18.3 h	36 h	86 h



**Fig. 21.37** (a) Formation of a solid drop and its ovulation (detachment time of the drop is normalized to 100). (b) Formation of a solid drop and its ovulation with the recession of the tip. (c) Scanning electron microscopy observations of the solid-drop formation on a W tip at 2700 K.



**Fig. 21.38** (a) Numerical simulation of a steady-state profile evolution with time ( $\alpha = 8^\circ$ ). (b) Numerical simulation of the steady-state profiles for different tip half-cone angles  $\alpha$ . (c) Scanning electron microscopy observations of steady-state tip profiles for three different values of half-cone angle  $\alpha$ .

time  $t$  are given by Herring's scaling law (Herring 1950):

$$R_2^4 - R_1^4 = A_\alpha B(t_2 - t_1) \quad (\text{A.2.10})$$

$$z_2^4 - z_1^4 = B_\alpha B(t_2 - t_1). \quad (\text{A.2.11})$$

Table 21.3 presents the times needed to double the tip radius for two initial values of 0.1 and 1 μm and for  $B = 10^{-20}$ .

**Table 21.3** Heating times to double the initial radius.

$\alpha$	3°	4°	5°	8°	10°	15°	20°	30°	45°
0.1 => 0.2 $\mu\text{m}$	50.7 s	31.5 s	27.0 s	13.0 s	11.4 s	6.36 s	3.69 s	1.50 s	0.41 s
1 => 2 $\mu\text{m}$	140 h	87.4 h	75.1 h	35.9 h	31.6 h	17.7 h	10.2 h	4.17 h	1.13 h

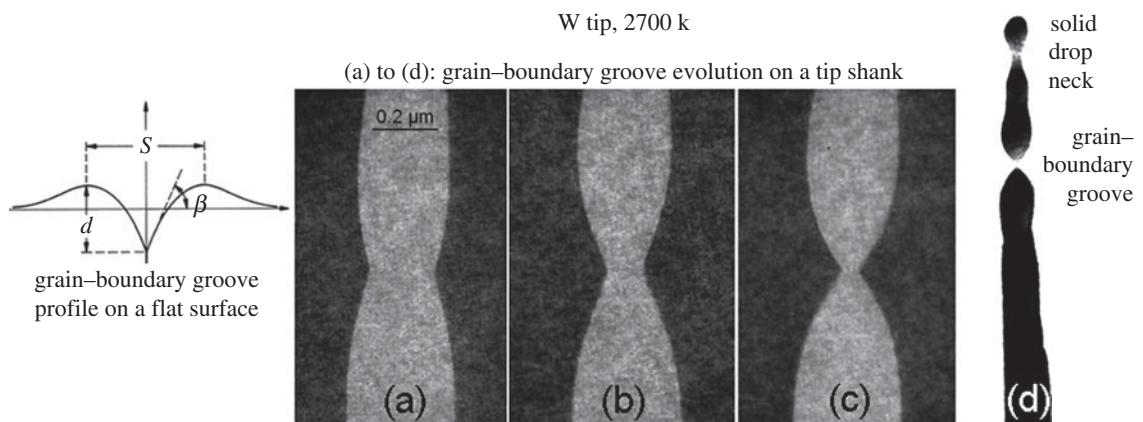
The time values indicate that for tip radii in the range of 0.1  $\mu\text{m}$  or less—which are conventional values for field emitters—tip blunting is a phenomenon that must be taken into account for any heating time more than 1 min.

#### A.2.1.2 Grain-boundary grooving (Bailey and Watkins 1950; Mullins 1957; Binh et al. 1976)

An initially flat surface of polycrystalline material, such as a metal, develops a groove at a grain boundary by mass-transport surface diffusion in order to balance surface energy  $\gamma$  and grain-boundary energy  $\gamma_B$ , as shown in Fig. 21.39, with a groove angle  $\beta$  given by:

$$2\gamma \sin \beta = \gamma_B. \quad (\text{A.2.12})$$

For a conical-geometry object as a tip, the grain-boundary grooving induces a bamboo-like structure with a decrease of the diameter of the groove neck ending in a potential separation into two parts (Fig. 21.39). As the number and positions of grooves are erratic, they cannot be controlled for a polycrystalline tip. Therefore, such grain-boundary grooving introduces an unpredictable source of tip failure, with a detachment into two parts when the groove neck diameter becomes zero. Moreover, the time for the grooving process is very comparable with the ovulation time. Figure 21.39(d) shows an example of a solid-drop process in concomitance with a grain-boundary groove formation, and for which the neck diameter of the grain-boundary groove is smaller than the solid-drop neck diameter.



**Fig. 21.39** SEM observations of the evolution of a grain-boundary groove on a tip shank (a) to (c); with the possible detachment of the two parts when the diameter of the groove neck is zero as in (d).

## A.2.2 Crystal equilibrium shape from anisotropic surface energy

Beside the intrinsic instabilities, the extrinsic instabilities are due to interaction with an external potential field—as surface free-energy gradient or electric-field gradient—and lead to specific relaxing equilibrium shapes detailed hereafter.

The equilibrium shape of a crystal corresponds to the shape of minimum surface energy (Stranski 1947, 1949). The geometry of the equilibrium shape is given by the Wulff construction:

$$\gamma_1/h_1 = \gamma_y/h_2 = \dots \dots \dots = \gamma_i/h_i = \gamma_{hkl}/h_{hkl} = \text{constant}, \quad (\text{A.2.13})$$

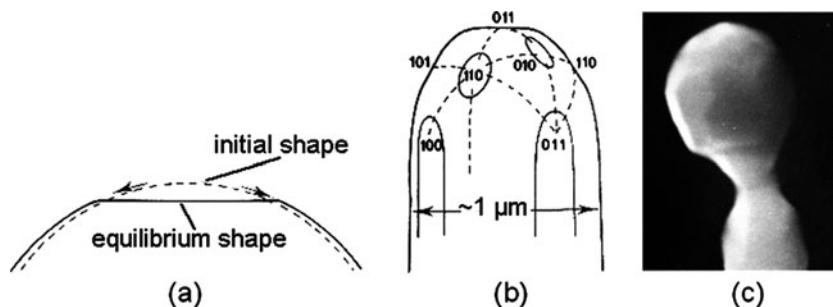
where  $h_i$  is the distance of the  $i$ th facet, with crystallographic indexes ( $hkl$ ), to the center of the crystal and  $\gamma_i$  is the  $i$ th facet surface energy.

For a clean crystal, the  $\gamma_i$  are nearly the same, so the equilibrium shape is nearly a sphere.  $\Delta\gamma_i$ , the change of  $\gamma_i$  for a Langmuir adsorption (Lacmann 1968) is given by:

$$\Delta\gamma_i = kT N_0 \ln(1 - \theta_i), \quad (\text{A.2.14})$$

where  $\theta_i$  the coverage degree of the facet  $i$ , which is dependent on  $E_{\text{ad}}^{(i)}$  the adsorption energy, and is given by  $\theta_i \sim \exp(E_{\text{ad}}^{(i)}/kT)$ .

Consequently, from the equilibrium shape point of view, when the mean coverage of an adsorbate increases from  $\theta_i = 0$  to 1, the initially nearly spherical shape evolves towards a polyhedron equilibrium shape. In the course of such a coverage increase, the total number of facets decreases, while the size of some facets increases. The equilibrium-shape changes are important for strong chemisorptions, such as carbon, oxygen, sulfur or metal on metals for example. For weak adsorption—such as physisorption—the shape changes are negligible because  $\Delta\gamma_i$  is very small and also the physisorbed particles are normally desorbed at surface self-diffusion temperatures. Figures 21.40(a) and (b) are schematic representations of the formation of an equilibrium shape, and Fig. 21.40(c) is a scanning electron microscopy (SEM) observation of a polyhedral Ni solid drop obtained after a heat treatment in a vacuum of  $10^{-5}$  Torr.



**Fig. 21.40** (a) Formation of a plane facet on an initially spherical crystal. (b) Equilibrium shape of the end of an emitter tip calculated from Wulff's law (bcc crystal) (Drechsler 1981). (c) SEM micrograph of the end of a Ni tip (1250 K, 4 min,  $10^{-5}$  Torr).

Another example is, with a monolayer of Pd on W(111) tip, the facet enlargement ends in the formation of an atomic corner at the <111> apex (Fu *et al.* 2001).

The growth time of the equilibrium shape  $\tau$  is proportional to (diameter of the crystal)<sup>3</sup> and can be estimated by the following practical equation (Drechsler 1981):

$$\tau(s) = 150 \times (10^{-6}/D_S) \times (2R)^3, \quad (\text{A.2.15})$$

where  $D_S$  is in  $\text{cm}^2 \text{s}^{-1}$  and  $R$  is the radius of the crystal in  $\mu\text{m}$ . From a practical point of view, the equilibrium shape should be formed at a temperature greater than  $0.6T_M$  and only for crystals of diameters up to  $\sim 10 \mu\text{m}$ . As illustration, for a copper crystal at an annealing temperature of 1200 K the characteristic growth times of the equilibrium shape are  $10^{-6}$ ,  $10^{-3}$ , 1 and  $10^3$  s for, respectively, crystal diameters of 0.01, 0.1, 1 and  $10 \mu\text{m}$ .

### A.2.3 Equilibrium shapes under the presence of a gradient of electric field

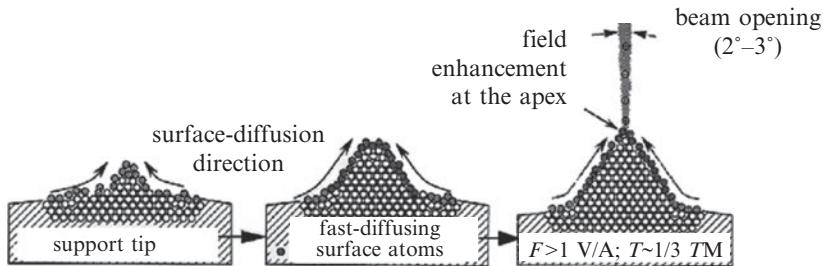
If the tip is heated in the presence of a sufficiently strong electric field, polarization forces will draw migrating atoms towards the highest-field regions (Dyke and Dolan 1956). The normal surface stress  $p_{xx}$  due to a field  $F$  at a conductor surface is:

$$p_{xx} = 1/2\epsilon_0 F^2, \quad (\text{A.2.16})$$

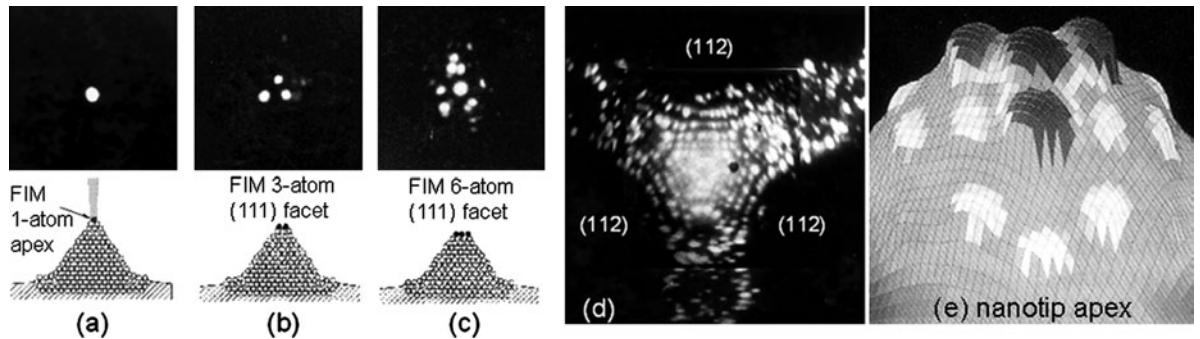
where  $\epsilon_0$  is the permittivity of the vacuum. In the presence of an externally applied stress to the surface  $p_{xx}$ , eqns (A.2.5) and (A.2.6) have an additional term that is  $(-1/2\Omega\epsilon_0 F^2)$ . Subsequently, eqn (A.2.1) also has an additional term proportional to the field gradient,  $(\Omega \epsilon_0 F dF/dx)$ , indicating atom surface migration towards the highest-field regions (in absolute value), with eventually the formation of nanoprotuberances and asperities. There are two main consequences for tip-profile evolutions (Binh *et al.* 1996), which are described hereafter.

#### A.2.3.1 Field-induced build-up

Due to the anisotropy of  $\gamma_{hkl}$  an annealed tip apex will always present some flat facets, in particular for low index value orientations. The presence of an applied voltage induces a field gradient between the center and the edge of the facets with subsequently a migration of surface atoms from the center towards the periphery edge. For  $F$  in the range of  $0.5 \text{ V}/\text{\AA}$ , the principal crystal facets are enlarged and the edges between them are accentuated, a process known experimentally as field-induced “build-up” (Sokolovskaja 1956; Bettler and Charbonnier 1960). If the heating process is carefully controlled, the edges between two neighboring facets and the protrusion between three neighboring facets may be of single-atom sharpness, as for a build-up W<111> tip for example (Binh and Marien 1988).



**Fig. 21.41** Field-induced nanotip formation process.



**Fig. 21.42** FIM images showing the atomic structure of a single-atom W nanotip. (a) The first layer consists of only one atom; (b) the second layer has three atoms; (c) the third layer has seven atoms; (d) after evaporating many layers, we can observe the triangular base of the nanotip that is located on the apex of a build-up structured W  $<111>$  substrate tip. (e) Numerical reconstruction, from the FIM images, of the nanotip apex showing the field distribution over the topmost atoms (the darker the color the higher the field).

#### A.2.3.2 Field-induced pyramidal nanotips

At higher field values in the range of 1 to 3 V/ $\text{\AA}^3$ , the activation energy for surface diffusion is now considerably lowered by the contribution of atomic polarizability and permanent dipole moments (Muller and Tsong 1974). The surface diffusion coefficient  $D_S$  may then reach values in the range of  $10^{-5} \text{ cm}^2/\text{s}$  at temperatures  $\sim T_M/3$ .

The high diffusivity facilitates an increase of the height of some existing thermally induced atomic-size corrugations due to the field-gradient driving force, leading to the formation of nanoprotrusions. The geometry of the formed nanoprotrusions is determined by the equilibrium between the pulling-up by the electric-field-gradient force and the blunting due to the capillarity forces. When the field enhancement over the apex of these protrusions is high enough the last atom is field ionized and a quasi-steady-state profile may be obtained, in which the rate of supply of material to the protrusion apex by surface diffusion is exactly matched by the rate of loss by field evaporation. A schematic drawing of this mechanism is given in Fig. 21.41.

For W, the high protrusion geometry remains intact upon quenching and the resulting protrusions are generally triangular  $<111>$  pyramids of 2 to 3 nm dimensions ending in one atom, as shown by the field ion microscopy (FIM) analysis presented in Fig. 21.42.

<sup>3</sup>For such high fields and in order not to destroy the tip by an too large electron emission, a positive field is applied to the emitter.

## Appendix 3: Field factor $\beta$ and field-enhancement factor $\gamma$

Field emission of electrons from the surface of a solid can be obtained by a strong electric field (Guth and Mullin 1942; Kleint 1993). This phenomenon was first reported by Wood (1897) and occurs at fields in the order of 5000 V/ $\mu\text{m}$  for surfaces with work functions in the range of 4 to 5 eV (Earhart 1901; Hobbs 1905; Kinsley 1905). Such large fields are extremely difficult to realize on flat surfaces, but can be generated by the field-enhancing features of tip-like structures. Muller introduced the notion of a geometrical field-enhancement factor at the apex of a microtip, making available fields higher than 5000 V/ $\mu\text{m}$  with voltages only in the range of thousands of volts (Muller 1938). For a standalone tip emitter, the relation between the applied voltage  $V_{\text{appl}}$  and the electric field at the apex  $F$ :

$$F = \beta V_{\text{appl}}, \quad (\text{A.3.1})$$

introduces the notion of the field factor  $\beta$ .

With the introduction of Spindt-type integrated microtips obtained by micro-fabrication techniques (Spindt 1968) the cathode–anode configuration is more similar to a parallel-plate capacitor with a geometrical perturbation due to the tip emitter. Under such a configuration, one introduces a parameter called the field-enhancement factor  $\gamma$ , which is the ratio between the local field  $F$  over the emitter and a macroscopic field,  $F_M$ , only defined by two parallel-plate capacitors without the emitter. The value of  $\gamma$  depends on the geometry of the emitter that is now considered as a local protrusion on the plane surface. For the geometry of protrusions with high aspect ratio,  $\gamma$  could reach values of a thousand and confusion of the local field over the emitter with the macroscopic field could lead to severe misinterpretation of experimental results. Confounding the field-enhancement factor  $\gamma$  and the field factor  $\beta$  is a mistake that has to be avoided as well.

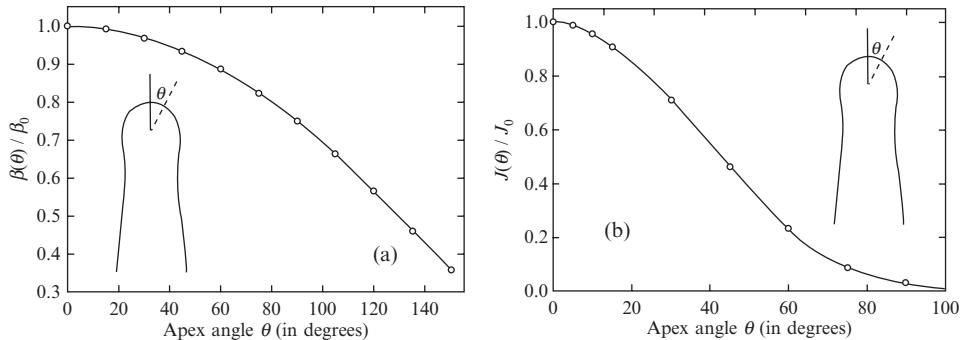
### A.2.4 Field factor $\beta$

Calculation of field factor  $\beta$  for a standalone needle-shaped emitter having a millimeter-sized ending with an apex in the range of 0.1  $\mu\text{m}$ , typical of a field emitter used in FEM, was established by Dyke and Dolan (1956) using typical geometries of experimental tungsten cathodes and the results are plotted in Fig. 21.43.

The typical cathode geometries (see Fig. 21.38 of Appendix 2) are considered to be the equipotential surfaces surrounding a core, which is a charged isolated sphere-on-orthogonal cone, so the value  $\beta_0$  at the emitter apex is:

$$\beta_0 = (R_0^{n-1}/d^n)\{n + (n+1)(a/R_0)^{2n+1}\}(V_R/V_{\text{appl}}), \quad (\text{A.3.2})$$

where  $V_R$  is the potential difference between the core and the anode,  $R_0$  is the apex radius,  $d$  is the cathode–anode spacing,  $a$  is the radius of the core sphere and  $n$  is a parameter depending on the emitter cone angle. For example, for an emitter with a half-cone angle  $\alpha$  of about 5° (see Appendix 2) and an apex



**Fig. 21.43** (a) Field factor distribution, relative to field factor at the apex  $\beta_0$ , of an emitter with a half-cone angle of  $5^\circ$  as a function of the apex angle. (b) FE current density, relative to current density at the apex  $J_0$ , of an emitter with a half-cone angle of  $5^\circ$  as a function of the apex angle.

radius of  $R_0 = 0.4 \mu\text{m}$ , the values  $a = 0.1235 \mu\text{m}$  and  $n = 0.10$ . However, eqn (A.3.2) is not easy to use experimentally to estimate  $\beta_0$ , calculation needs to take into account the exact geometry of the blunt tip after each thermal treatment (see Appendix 2), so the field factor at an emitter apex is usually estimated with a rather good approximation either by one of the following empirical relations (A.3.3) to (A.3.5):

$$\beta_0 \sim 1/5R_0; \quad (\text{A.3.3})$$

or within an accuracy of a factor of 2, by using either the hyperboloidal approximation (Eyring *et al.* 1928):

$$\beta_0 = 2/R_0 \ln(4d/R_0), \quad (\text{A.3.4})$$

or the paraboloidal approximation (Becker 1951):

$$\beta_0 = 2/R_0 \ln(2d/R_0). \quad (\text{A.3.5})$$

These equations are valid for  $d >> R_0$  and it means a value of the field factor between  $10^4$  to  $10^5 \text{ cm}^{-1}$  for tip radii in between 1000 and 100 Å.

The merit of the Dyke and Dolan analysis was to show that the field factor  $\beta$  (in other words the local field distribution at an emitter) decreases rapidly from the apex to the cone shank (Fig. 21.43(a)) which means that, for a field emitter, most of the FE current is only extracted from the apex region limited by an apex angle  $\theta$  of about  $20^\circ$  (Fig. 21.43(b)).

## A.2.5 Field-enhancement factor $\gamma$

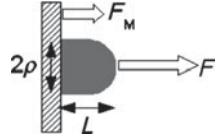
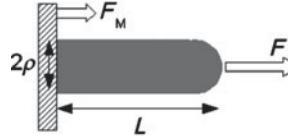
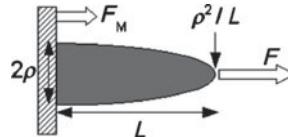
The above analysis is no longer viable when the tip dimensions and the distances between cathode–anode are in the range of micrometers.

Therefore, two categories of electric fields are defined:

On the one hand, a macroscopic field  $F_M$ , which is the field between the two parallel electrodes considered as flat surfaces:

$$F_M = \frac{V_{\text{appl}}}{d}, \quad (\text{A.3.6})$$

**Table 21.4** Relations giving the field-enhancement factors for different geometries of protrusions on a plane surface.

Hemisphere on a post (for $v < 4$ )	$\gamma = 2 + v$	
Hemisphere on a post (for $4 < v < 3000$ )	$\gamma = 1.2 \times (2.15 + v)^{0.90}$	
Hemiellipsoid on a plane (for $v > 1$ ) (apex radius $r_a = \rho/v = \rho^2/L$ )	$\gamma = \frac{\zeta^3}{(\nu \ln(\nu + \zeta)) - \zeta}$	
Field-enhancement factor $= \gamma = F / F_M$ ; $v = L/\rho$ ; $\zeta = (\nu^2 - 1)^{1/2}$ .		

where  $V_{\text{appl}}$  is the voltage applied between the cathode and the anode, being a distance  $d$  apart;

On the other hand, a local field  $F$ , close to the emitting surface, which is the field actually acting at an emitting surface that determines the surface barrier as described in the above Appendix 1.  $F$  is strongly dependent on the local conditions at the surface and in particular the local geometrical corrugation. It is related to  $F_M$  by a field-enhancement factor  $\gamma$ :

$$\gamma = \frac{F}{F_M}. \quad (\text{A.3.7})$$

The value of  $\gamma$  depends on the geometry of the local protrusion on the plane surface and in particular on the aspect ratio  $v = L/\rho$  between the protrusion length (height)  $L$  and its base radius  $\rho$ . For a perfectly flat cathode surface  $\gamma = 1$ ,  $F_M$  and  $F$  are identical.

Estimates of  $\gamma$  for different protruding geometries have been calculated (Forbes *et al.* 2003), in particular for two of them: a hemisphere on a post model and a hemiellipsoid on a plane. They are presented in Table 21.4.

The results in Table 21.4 are given for assumptions of a gap  $d \gg L$  and for protrusions on a flat-surface substrate. Otherwise, there are two main corrections to be highlighted:

- (1) If the anode is relatively close to the protrusion, i.e.  $d$  is in the range of  $L$ , then the influence of  $d$  may need to be taken into account by using the relation (Miller 1967):

$$\gamma(d) = \gamma \times (1 - L/d). \quad (\text{A.3.8})$$

- (2) If the underlying surface is not flat, the field enhancement of the substrate,  $\gamma_{\text{substrate}}$ , has to be taken into account, resulting in a total field-enhancement factor  $\gamma_{\text{total}} = \gamma_{\text{substrate}} \times \gamma$ . Otherwise, care may be needed over the definition of the macroscopic field (Miller *et al.* 1996). For example, if the protrusion is located on the apex of a substrate tip having a radius  $R_0$ , the local field at the apex of the protrusion is given by eqn (A.3.7), but with  $F_M = (V/5R_0)$ .

## References

- Aharonov, Y., Bohm, D. *Phys. Rev.* **115**, 485 (1959).
- Atlan, D., Gardet, G., Vu Thien Binh, Garcia, N., Saenz, J.J. *Ultramicroscopy* **42**, 154 (1992).
- AuBuchon, J.F., Chen, L.H., Gapin, A.I., Jin, S. *Chem. Vap. Depos.* **12**, 370 (2006).
- Bailey, G.L.J., Watkins, H.C. *Proc. Phys. Soc. B* **63**, 350 (1950).
- Baker, R.T.K., Harris, P.S. in *Chemistry and Physics of Carbon*, (eds) P.L. Walker and P.A. Thrower (Marcel Dekker, New York, 1978) Vol. 14.
- Becker, J.A. *Bell Sys. Tech. J.* **30**, 907 (1951).
- Bettler, P., Charbonnier, C. *Phys. Rev.* **119**, 85 (1960).
- Binh, V.T., Piquet, A., Roux, H., Uzan, R., Drechsler, M. *Surf. Sci.* **44**, 598 (1974).
- Binh, V.T., Chaudier, M., Couturier, J.C., Uzan, R., Drechsler, M. *Surf. Sci.* **57**, 184 (1976).
- Binh, V.T. *J. Microsc.* **151**, 355 (1988).
- Binh, V.T., Marien, J. *Surf. Sci.* **202**, L539 (1988).
- Binh, V.T., Purcell, S.T., Garcia, N., Doglioni, J. *Phys. Rev. Lett.* **69**, 2527 (1992).
- Binh, V.T., Garcia, N., Purcell, S., Semet, V. in *Nanosources and Manipulation of Atoms under High Fields and Temperatures: Applications*, (eds) Vu Thien Binh, N. Garcia and K. Dransfeld, NATO ASI Series E Applied Sciences, Vol. **235**, 59 (Kluwer, NL, 1993).
- Binh, V.T., Semet, V., Garcia, N. *Appl. Phys. Lett.* **65**, 2493 (1994).
- Binh, V.T., Garcia, N., Purcell, S. *Adv. Imag. Electron. Phys.* **95**, 63 (1996).
- Binh, V.T., Purcell, S.T., Semet, V., Feschet, F. *Appl. Phys. Lett.* **72**, 975 (1998).
- Binh, V.T., Semet V. *Ultramicroscopy* **73**, 107 (1998).
- Binh, V.T., Semet, V., Guillot, D. *Appl. Phys. Lett.* **73**, 2048 (1999).
- Binh, V.T., Semet, V. *Planar Cold Cathodes in Advances in Imaging and Electron Physics*, (ed.) P. Hawkes, vol. **148**, (Elsevier Academic Press Inc., San Diego, USA, 2007) 1–73.
- Binh, V.T., Brookes, M., Semet, V., Mouton, R., Kilgour, D., Jupp, I. *Gordon Res. Conf. Detecting Illicit Substances: Explosives & Drugs*, 16–21 Sept. (Montana, USA, 2007).
- Bonzel, H.P. *Surface Mobilities on Solid Materials: Fundamental Concepts and Applications*, Vu Thien Binh, NATO ASI Series B: Physics, Vol. 86 (Plenum Press, New York, 1981).
- Bower, C., Zhu, W., Jin, S., Zhou, O. *Appl. Phys. Lett.* **77**, 830 (2000).
- Chang, T.H.P., Kern, D.P., Murray, L.P. *J. Vac. Sci. Technol. B* **6**, 1698 (1990).

- Chen, L.H., Aubichon, J.F., Gapin, A., Daraio, C., Bandaru, P., Jin, S., Kim, D.W., Yoo, I.K., Wang, C.M. *Appl. Phys. Lett.* **85**, 5373 (2004).
- Chhowalla, M., Teo, K.B.K., Ducati, C., Rupasinghe, N.L., Amaratunga, G.A.J., Ferrari, A.C., Roy, D., Robertson, J., Milne, W.I. *J. Appl. Phys.* **90**, 5308 (2001).
- Crewe, A.V. *Conference on Non-conventional Electron Microscopy* (Cambridge, England, 1965).
- Crewe, A.V., Walls, J., Welter, L.M. *J. Appl. Phys.* **39**, 5861 (1968).
- de Heer, W.A., Chatelin, A., Ugarte, D. *Science* **270**, 1179 (1995).
- de Jong, K.P., Geus, J.W. *Catal. Rev. Sci. Eng.* **42**(4), 481 (2000).
- Drechsler, M. *Surf. Sci.* **70**, 1 (1978).
- Drechsler, M. in *Surface Mobilities on Solid Materials: Fundamental Concepts and Applications*, (ed.) Vu Thien Binh, NATO ASI Series B: Physics, vol. 86 (Plenum Press, New York, 1981).
- Dyke, W.P., Dolan, W.W. *Adv. Electroni. Electron. Phys.* **8**, 89 (1956).
- Earhart, R.F. *Philos. Mag.* **1**, 147 (1901).
- Eyring, C.F., Mackeown, S., Millikan, R.A. *Phys. Rev.* **31**, 900 (1928).
- Fleming, G.M., Henderson, J.E. *Phys. Rev.* **58**, 887 (1940).
- Forbes, R.G. *Ultramicroscopy* **79**, 11 (1999).
- Forbes, R.G., Edgcombe, C.J., Vadrè, U. *Ultramicroscopy* **95**, 57 (a review of the different estimations of  $\gamma$  are given in this reference) (2003).
- Fowler, R.H., Nordheim, L.W. *Proc. Roy. Soc. (London) A* **119**, 173 (1928).
- Fu, T.Y., Cheng, L.C., Nien, C.H., Tsong, T.T. *Phys. Rev. B* **64**, 113401 (2001).
- Gohda, Y., Watanabe, S. *Phys. Rev. Lett.* **87**, 177601 (2001).
- Gomer, R. *Field Emission and Field Ionization* (Harvard University Press, Cambridge, MA, USA, 1961).
- Guth, E., Mullin, C.J. *Phys. Rev.* **61**, 339 (1942).
- Herring, C. *J. Appl. Phys.* **21**, 301 (1950).
- Herring, C. in *Physics of Powder Metallurgy*, (ed.) W.E. Kingston (McGraw Hill, New York, 1951).
- Herring, C. in *Structures and Properties of Solid Surfaces*, (eds) R. Gomer and C.S. Smith (University of Chicago Press, 1953).
- Herring, C., Nichols, M.H. *Rev. Mod. Phys.* **21**, 185 (1949).
- Hobbs, G.M. *Philos. Mag.* **10**, 617 (1905).
- Hoogenraad, M.S. PhD thesis (Utrecht University, 1995).
- Hughes, T.V., Chambers, C.R. *US Patent* **405**, 480 (1889).
- Kinsley, C. *Philos. Mag.* **9**, 692 (1905).
- Kleint, Ch. *Prog. Surf. Sci.* **42**, 101 (2004); *Surf. Interf. Anal.* **36**, 387 (The author reviewed the early history of field emission) (1993).
- Kock, A.J.H.M., de Bokx, P.K., Boellard, E., Klopp, W., Geus, J.W. *J. Catal.* **96**, 468 (1985).
- Lacmann, R. *Springer Tracts in Modern Physics* **44**, 1 (1968).
- Lee, S.B., Teo, K.B.K., Chhowalla, M., Hasko, D.G., Amaratunga, G.A.J., Milne, W.I., Ahmed, H. *Microelectron. Eng.* **61/62**, 475 (2002).

- Levesque, A., Vu Thien Binh, Semet, V., Guillot, D., Fillit, R.Y., Brookes, M.D., Nguyen, T.P. *Thin Solid Films* **464**, 308 (2004).
- Levesque, A., Vincent, P., Vu Thien Binh, Guillot, D., Brookes, M.D. *J. Vac. Sci. Technol. B* **23**(2), 665 (2005).
- Mauger, M., Vu Thien Binh, Levesque, A., Guillot, D. *Appl. Phys. Lett.* **85**, 305 (2004).
- Mauger, M., Vu Thien Binh *J. Vac. Sci. Technol. B* **24**, 997 (2006).
- Mauger, M., Mouton, R., Hamzaoui, O., Vu Thien Binh *J. Vac. Sci. Technol. B* **25**, 575 (2007).
- Mayer, A. *Phys. Rev. B* **63**, 35408 (2001).
- Mayer, A., Vigneron, J.P. *Phys. Rev. B* **62**, 16023 (2000).
- Mayer, A., Miskovsky, N.M., Cutler, P.H. *Phys. Rev. B* **65**(15), 155420 (2002).
- McLean, M. *Met. Sci.* **12**, 113 (1978).
- Merkulov, V.I., Lowndes, D.H., Wei, Y.Y., Eres, G., Voelkl, E. *Appl. Phys. Lett.* **76**, 3555 (2000).
- Miller, H.C. *J. Appl. Phys.* **38**, 4501 (1967).
- Miller, M.K., Cerezo, A., Heatherington, M.G., Smith, G.D.W. *Atom Probe Field Ion Microscopy* (Clarendon, Oxford, 1996).
- Milne, W.I., Teo, K.B.K., Chhowalla, M., Amaratunga, G.A.J., Lee, S.B., Hasko, D.G., Ahmed, H., Groening, O., Legagneux, P., Gangloff, L., Schnell, J.P., Pirio, G., Pribat, D., Castignolles, M., Loiseau, A., Semet, V., Vu Thien Binh, *Diam. Rela. Mater.* **12**, 422 (2003).
- Modinos, A. *Field, Thermoionic and Secondary Electron Emission Spectroscopy* (Plenum Press, New York, 1984).
- Morton, G.A., Ramberg, E.G. *Phys. Rev.* **56**, 705 (1939).
- Mouton, R., Semet, V., Kilgour, D., Brookes, M.D., Vu Thien Binh *J. Vac. Sci. Technol. B* (2008).
- Muller, E.W. *Z. Phys.* **108**, 668 (1938).
- Muller, E.W. *Ergeb. Exakt. Naturwiss.* **27**, 290 (1953).
- Muller, E.W., Tsong, T.T. *Field Ion Microscopy, Principles and Applications* (Elsevier, Amsterdam, 1969).
- Muller, E.W., Tsong, T.T. *Prog. Surf. Sci.* **1**, 1 (1974).
- Mullins, W.W. *J. Appl. Phys.* **28**, 333 (1957).
- Mullins, W.W. *J. Appl. Phys.* **30**, 77 (1959).
- Murphy, E.L., Good, R.H. *Phys. Rev.* **102**, 1464 (1956).
- NANOPAGE European project IST-FP6 #004251 (2004–2006).
- Nichols, F.A., Mullins, W.W. *J. Appl. Phys.* **36**, 1826 (1965).
- Nilsson, L., Groening, O., Emmenegger, Ch., Kuettel, O., Schaller, E., Schlapbach, L., Kind, H., Bonard, J.M., Kern, K. *Appl. Phys. Lett.* **76**, 2071 (2000).
- Nottingham, W.B. *Phys. Rev.* **59**, 906 (1941). The Nottingham effect stated that the difference in the average energy of the electrons emitted by the cathode and those supplied to the cathode—which are at the Fermi level—results in an exchange of power and in a corresponding variation of temperature. Electrons that are emitted from energy levels above the average energy of replacement electrons serve to cool the material; conversely, electrons emitted from levels below the average energy of the replacement electrons heat the material. If the FE electrons are coming from resonant

tunnelling at levels under the Fermi level, the net energy exchange is concentrated at the emitting surface, resulting in an increase in temperature directly as a function of the FE current density. Such a situation corresponds generally to strong chemisorptions, such as carbon, oxygen, sulfur or metal on metals, for example.

- Oshima, C., Matsuda, K., Kona, T., Mogami, Y., Komaki, M., Murata, Y., Yamashita, T., Kuzumaki, T., Horike, Y. *Phys. Rev. Lett.* **88**, 038301 (2002).
- Purcell, S., Vu Thien Binh, Garcia, N. *Appl. Phys. Lett.* **67**, 436 (1995).
- Purcell, S., Vincent, P., Journet, C., Vu Thien Binh, *Phys. Rev. Lett.* **88**, 105502 (2002).
- Reimer, L. *Transmission Electron Microscopy, Physics of Image Formation and Micro-analysis*, Springer Series in Optical Sciences (Springer-Verlag, Berlin, Germany, 1993).
- Ren, Z.F., Huang, Z.P., Wang, D.Z., Wen, J.G., Xu, J.W., Wang, J.H., Calvet, L.E., Chen, J., Klemic, J.F., Reed, M.A. *Appl. Phys. Lett.* **75**, 1086 (1999).
- Robertson, S.D. *Nature* **221**, 1044 (1969).
- Seebauer, E.G., Allen, C.E. *Progr. Surf. Sci.* **49**, 265 (1995).
- Semet, V., Vu Thien Binh, Vincent, P., Guillot, D., Teo, K.B.K., Chhowalla, M., Amarasingha, G.A.J., Milne, W.I., Legagneux, P., Pribat, D. *Appl. Phys. Lett.* **81**, 343 (2002).
- Semet, V., Mouton, R., Vu Thien Binh *J. Vac. Sci. Technol. B* **23**, 671 (2005).
- Semet, V., Vu Thien Binh, Guillot, D., Teo, K.B.K., Chhowalla, M., Amarasingha, G.A.J., Milne, W.I., Legagneux, P., Pribat, D. *Appl. Phys. Lett.* **87**, 223103 (2005).
- Silverman Markk, P. *More than One Mystery, Exploration in Quantum Interference* (Springer-Verlag, New York, 1994).
- Snoeck, J.W., Froment, G.F., Fowles, M. *J. Cataly.* **169**, 240 (1997).
- Sokolovskaya, I.L. *J. Tech. Phys. (URSS)* **26**, 1177 (1956).
- Spence, J.C.H., Qian, W., Silverman, M.P. *J. Vac. Sci. Technol. A* **12**, 542 (1994).
- Spindt, C.A. *J. Appl. Phys.* **39**, 3504 (1968).
- Stranski, I.N. *Disc. Faraday Soc.* **5**, 13 (1949).
- Stranski, I.N., Suhrmann, R. *Ann. Phys. 6F* **1**, 753 (1947).
- Swanson, L.W., Bell, A.E. *Adv. Electron. Electron. Phys.* **23**, 193 (1973).
- Swanson, L.W., Crouser, L.C. *Surf. Sci.* **23**, 1 (1970).
- Teo, K.B.K., Lee, S.B., Chhowalla, M., Semet, V., Vu Thien Binh, Groening, O., Castignolles, M., Loiseau, A., Pirio, G., Legagneux, P., Pribat, D., Hasko, D.G., Ahmed, H., Amarasingha, G.A.J., Milne, W.I. *Nanotechnology* **14**(2), 204 (2003).
- Thompson, J.J. *Philos. Mag.* **44**, 293 (1897).
- Utsumi, T. *IEEE Trans. Electron. Devices* **38**, 2276 (1991).
- Wang, Z.L., Kang, Z.C. *J. Phys. Chem.* **100**, 17725 (1996).
- Woods, R.W. *Phys. Rev.* **5**, 1 (1897).
- Xu, N.S., Huq, S.E. *Mater. Sci. Engi.* **R48**, 47 (2005).
- Ziegler, J.F., Biersack, J.P., Littmark, U. *The Stopping and Range of Ions in Solids* (Pergamon, New York, 1985) Vol. 1. (Values of 1 to 2 nm are calculated with SRIM code for C into Ni with 1 keV energy.)

# Free-standing grid-like nanostructures assembled into 3D open architectures for photovoltaic devices

X.Y. Kong, Y.C. Wang, X.F. Fan, G.F. Guo, and L.M. Tong

## 22

22.1 Introduction	789
22.2 Fabrication of photoelectrodes with 2D grid-like nanostructures by the biotemplating approach	791
22.3 Assembly and photophysics of grid-like nanostructures into 3D open architectures for the photocatalytic electrodes	794
22.4 Performance of DSSCs working with dye-sensitized $TiO_2$ stacked-grid array photoelectrodes	799
22.5 Characteristics and performance of DSSCs working with $TiO_2/NiO$ composite photoactive electrodes	802
22.6 Summary	805
Acknowledgments	806
References	806

## 22.1 Introduction

Dye-sensitized solar cells (DSSCs) are currently undergoing rapid development in an effort to obtain robust, efficient, and cheap devices that are suitable for practical use (Hagfeldt *et al.* 1994; Hagfeldt and Grätzel 1995). Dye-sensitized solar cells (DSSCs) were first reported by Grätzel and O'Regan in 1991. A typical DSSC comprises a mesoporous nanocrystalline  $TiO_2$  film sensitized by dye molecules and a liquid electrolyte containing an iodide/triiodide redox couple (Kay and Grätzel 1993, 2002). The nanocrystalline  $TiO_2$  electrode provides a high specific surface area for the adsorption of a large number of dye sensitizers. Since 1993, the improvement of the efficiency of DSSCs has seemed very slow and came to a bottleneck stage (Gratzel 2001). Nevertheless, the theoretical upper limit of overall solar to electrical conversion efficiency for a cell with only one photoanode dye-sensitized electrode is around 30%, whereas the corresponding value for a composite device with two photoactive semiconductors is around 43% (Gratzel 2000). In order to break through the bottleneck, considerable attention was devoted in the past to developing photoelectrodes with larger surface areas that could adsorb a large amount of dye, and synthesizing dyes with broader absorption ranges. Significant improvements in the performance of a dye-sensitized solar cell have been mainly due to the development of high-performance nanoporous  $TiO_2$  thin-film electrodes that have a large surface area capable of adsorbing a large amount of photosensitizer, and due to the synthesis of new Ru complex photosensitizers capable of absorbing in the wide visible range (Nazeeruddin *et al.* 1993; Fisher *et al.* 2000).

The diverse strategies have also been devoted to understanding the electrode architecture for efficient electron diffusion and transport for targeting the increase in the output voltage or the photogenerated current, alternatively. However, the electron diffusion in the nanocrystalline particles is more than two orders of magnitude smaller than that in bulk anatase crystals (Bisquert *et al.* 1999). In this sense, it has been reported that the use of a network structure of TiO<sub>2</sub> single-crystal-like nanorods instead of TiO<sub>2</sub> nanoparticles results in significant improvements, achieving rapid electron transfer and high efficiency (Montanari *et al.* 2002). The TiO<sub>2</sub> nanofibers and nanorods recently gained attention for fabrication of DSSC due to the channelled electron transfer in them. Conversion efficiencies of ~6.2% and ~9.3% are reported in polycrystalline TiO<sub>2</sub> fibers and single-crystalline nanorods, respectively (Adachi *et al.* 2004; Song *et al.* 2005). Recently, a photoanode for solar cells was also made from ZnO nanowires, but its efficiency was limited by the small surface area of the nanowires (Baxter and Aydil 2005; Law *et al.* 2005). Replacing a nanoparticle network with an array of nanowires perpendicular to the substrate and further increasing their length, where every possible point of electron injection is directly connected to the substrate with a minimal number of interfaces and grain boundaries, could further improve charge transport in these devices (Mor and Grimes 2006).

It has also been recognized that the optimized optical design of electrode architecture for enhancing light-harvesting efficiency (LHE) and making a more efficient use of the solar spectrum would improve the photoconversion efficiency. To date, three-dimensionally (3D) ordered architectures have attracted extensive interest for high photocatalytic and photovoltaic efficiencies. It depends on strengthening of the interaction between a 3D photonic crystal structure and light. Mallouk and colleagues achieved a 26% enhancement of the photogenerated current across the visible spectrum (400–750 nm) by coupling a photonic crystal of TiO<sub>2</sub> nanocolloids, with respect to a conventional photoelectrode with random nanoparticle networks (Nishimura *et al.* 2003). This effect arises from the slow group velocity of light in the vicinity of the stop band, and the consequent localization of light intensity in the voids or in the dye-sensitized TiO<sub>2</sub> portions of the photonic crystal (Halaoui *et al.* 2005; Mihi and Miguez 2005).

In this chapter, we demonstrate 3D open architectures with grid-like nanostructure arrays as photocatalytic electrodes for a new type of dye-sensitized solar cell. A novel technique to fabricate a series of semiconducting oxides with grid-like nanostructures replicated from the biotemplates is introduced systematically. These semiconducting oxides, including n-type TiO<sub>2</sub> or p-type NiO nanogrids, were sensitized with the dye molecules, then assembled into 3D stacked-grid arrays on a flexible substrate by means of the Langmuir–Blodgett method or the ink-jet printing technique for the photocatalytic electrodes. This kind of DSSC could achieve conversion efficiencies around 7%. We will provide the understanding of the light localization in the stacked-grid arrays (*slow down photon*) for the increase the efficiency of harvesting solar energy and enhancement of photocatalytic activity. Meanwhile, it is found that the stacked-grid array provides a direct path to improve the *efficient electron diffusion and rapid transport* through the network to the conductive

substrate, resulting in an increase of the output voltage or the photogenerated current.

We also report a new type of DSSC with composite structure photoelectrode. The new type of cells employed two photoactive semiconductors electrodes, one using n-type TiO<sub>2</sub> as a photoanode, and another using p-type NiO as a photocathode. Both are grid-like nanostructures and stacked into 3D open architectures and provide more surface area sensitized by more dye molecules. The solar cells with composite semiconducting electrodes we report here perform great conversion efficiencies around 8.5%, which opens up an alternative approach to break through the bottleneck and solve some of the problems encountered in fabrication of traditional DSSCs. It is still a challenge to understand the dependence of photoconversion efficiency of the composite photochemical cells based on the stacked-grid electrodes combined with n-type and p-type oxide semiconductors.

## 22.2 Fabrication of photoelectrodes with 2D grid-like nanostructures by the biotemplating approach

Nature routinely produces unique nanostructures in biological systems with useful properties. The synthetic replication of evidently useful biological structures by a simple casting process is expected to introduce biological structures into man-made materials (Caruso 2004). A lot of reliable and sophisticated biotemplating techniques to fabricate nanostructures with unique functionalities for practical applications (Cook *et al.* 2003; Huang and Kunitake 2003; Chik and Xu 2004) have been reported. The 2D grid-like nanostructure is a potential nanoscale building block for nanodevice assembly. However, it is difficult to be routed directly from the chemical synthesis approach. The biotemplating approach is a cost- and time-effective way to fabricate the functional materials with 2D grid-like structure replicated from the biological system. It is well known that the beautifully iridescent wings of a butterfly comprise the textured leaf-like scales in order, and the individual scale exhibits the 2D regular grid configuration (Anderson and Richards 1942). This naturally ordered and periodic structure can perform photonic-bandgap-like behaviors in sunlight, not only for the resplendent physical colors, but also for the absorption of solar energy to regulate the temperature of butterfly wings (Biro *et al.* 2003; Vukusic and Hooper 2005). Thus, the butterfly-wing scale is an ideal biotemplate for fabricating functional materials with 2D grid-like structure. It is also expected that 3D open architectures stacked from 2D grid-like morphology could enhance the light-harvesting efficiency (LHE) due to the lattice spacings of the order of wavelengths of light.

In recent decades, semiconducting titania (TiO<sub>2</sub>) has received special attention because of its unique properties, such as high refractive index, high photocatalytic and photovoltaic efficiencies. Herein, we demonstrate the biotemplating approach to efficiently fabricate free-standing 2D TiO<sub>2</sub> grid-like nanostructures. The processing of free-standing 2D TiO<sub>2</sub> grids is described as coating

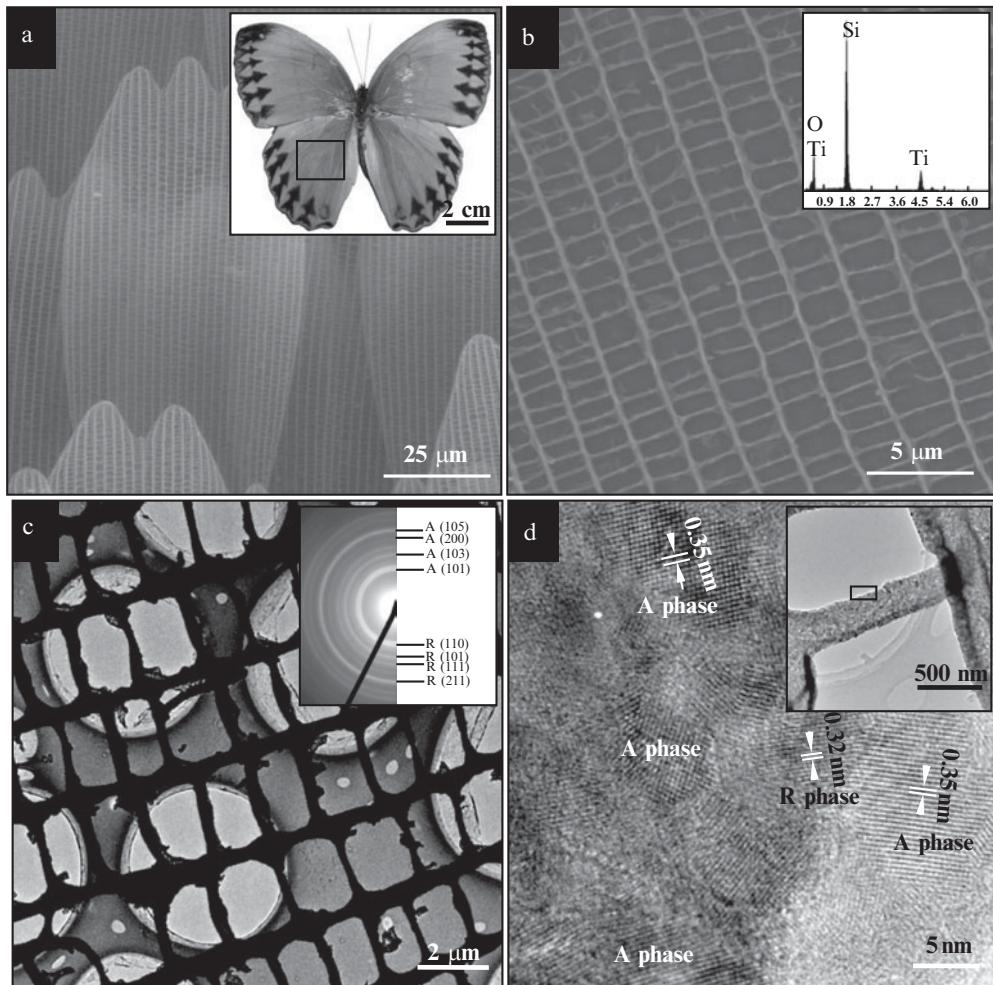
titanium thin film on the individual butterfly-wing scales, and calcinating for the removal of biotemplates, while preserving the grid-like morphology.

We produced the nanocrystalline  $\text{TiO}_2$  grids replicated from the butterfly-wing scale templates. The leaf-like wing scales are taken from butterfly wings and dispersed individually on fused silica substrates, and uniformly coated with pure titanium (99.99%) by thermal vaporization under the pressure at  $2 \times 10^{-4}$  Pa. The thickness of coated film on each wing scale is about 50–100 nm. Consequently, the samples were calcinated in a furnace at 650 °C for 30 min in air. It is very sophisticated for oxidizing titanium film on wing scales into nanocrystalline  $\text{TiO}_2$ , while removing the underlying biotemplates and preserving the uniform grid-like configuration. The replicas are easily transferred to silicon substrate or TEM grids. They are examined by FEI XL30 environmental FE-SEM for the observation of morphologies, EDX for the composition analysis, and FEI Tecnai F30 TEM for microstructure analysis.

In our experiments, we used a kind of butterfly specimen named *Mimathyma chevana* from Sichuan, China, shown in the inset of Fig. 22.1(a). The SEM image shows that the wing scales exhibit a leaf-like configuration and naturally assemble in order. The leaf-like wing scale is a fine grid structure periodically consisting of longitudinal ridges and cross-ribs with a flat surface. The unit window of the grid is about 2.5  $\mu\text{m}$  in length and 1  $\mu\text{m}$  in width. The diameter of the cross-ribs is about 200 nm.

The wing scales are made of organic material called chitin, which can be removed by calcination. This kind of butterfly-wing scale is an ideal template candidate for replication in our experiments. Figure 22.1(b) shows a free-standing  $\text{TiO}_2$  grid replicated from individual wing scales and transferred on the silicon substrate. The typical spectra of EDX recorded from a grid on silicon substrate show the elements of Ti, O and Si, corresponding to the  $\text{TiO}_2$  replica and silicon substrate, without any other impurity. We have examined a dozen replica samples by EDX, and found that the elements of carbon, nitrogen and other elements corresponding to biotemplates were not detectable in the replica. This indicates that the organic materials of biotemplates are removed completely by high-temperature calcination, while the titanium films coated on wing-scale surfaces are oxidized and converted into nanocrystalline  $\text{TiO}_2$ . During this replication process, the oxidation of titanium film and removal of organic materials will produce thermal stress, which plays a significant role in the good preservation of the grid-like configuration. It can also be found that there are little breaks between ridges and cross-ribs in Fig. 22.1(b), corresponding to the release of local thermal stress. It is very important to optimize the appropriate processing conditions, such as the coating thickness of the film, oxidation temperature and annealing time, etc., for minimizing the thermal stress in the grid replica. In our case, the coating thickness of titanium film is about 50–100 nm, and the samples were calcinated in a furnace at 650 °C for 30 min in air.

The grid-like replica of  $\text{TiO}_2$  can be easily transferred onto TEM-supporting grids. A TEM image as shown in Fig. 22.1(c), clearly shows that the grid-like replica from the wing-scale template is the result of mesoporous nanocrystalline  $\text{TiO}_2$  constituting the longitudinal ridges and cross-ribs. The surfaces of the longitudinal ridge and cross-rib are uniform. From the selected-area



**Fig. 22.1** (a) SEM image of typical leaf-like morphology of butterfly-wing scales. Inset: a photo of a butterfly specimen named *Mimathyma chevana* from Sichuan, China. (b) SEM image of free-standing  $\text{TiO}_2$  grid replicated from an individual butterfly-wing scale on the silicon substrate. The corresponding EDX spectrum indicates the replica is pure  $\text{TiO}_2$ . (c) Bright-field TEM image of  $\text{TiO}_2$  grid replica, and corresponding SAED pattern; (d) HRTEM image taken from a cross-rib of a grid. Inset: a typical TEM bright-field image of a grid rib.

electron diffraction (SAED) pattern, the diffraction rings represent the lattice planes (101), (103), (200), (105) of anatase phase of  $\text{TiO}_2$  and (110), (101), (111), (211) of rutile phase of  $\text{TiO}_2$ , respectively. This result suggests that the grids are composed of two phases of anatase and rutile  $\text{TiO}_2$ . The high-resolution TEM image (Fig. 22.1(d)) taken from a typical cross-rib edge, shows that two kinds of dominated fringe spaces of 0.35 nm and 0.32 nm, in agreement with the lattice spacing of the (101) plane of the anatase phase and the (110) plane of rutile phase, respectively. It can also be found that the average size of  $\text{TiO}_2$  nanocrystals is of about 10 nm, and the anatase is the dominant phase, which is of benefit for the photocatalytic activity (Park *et al.* 2000).

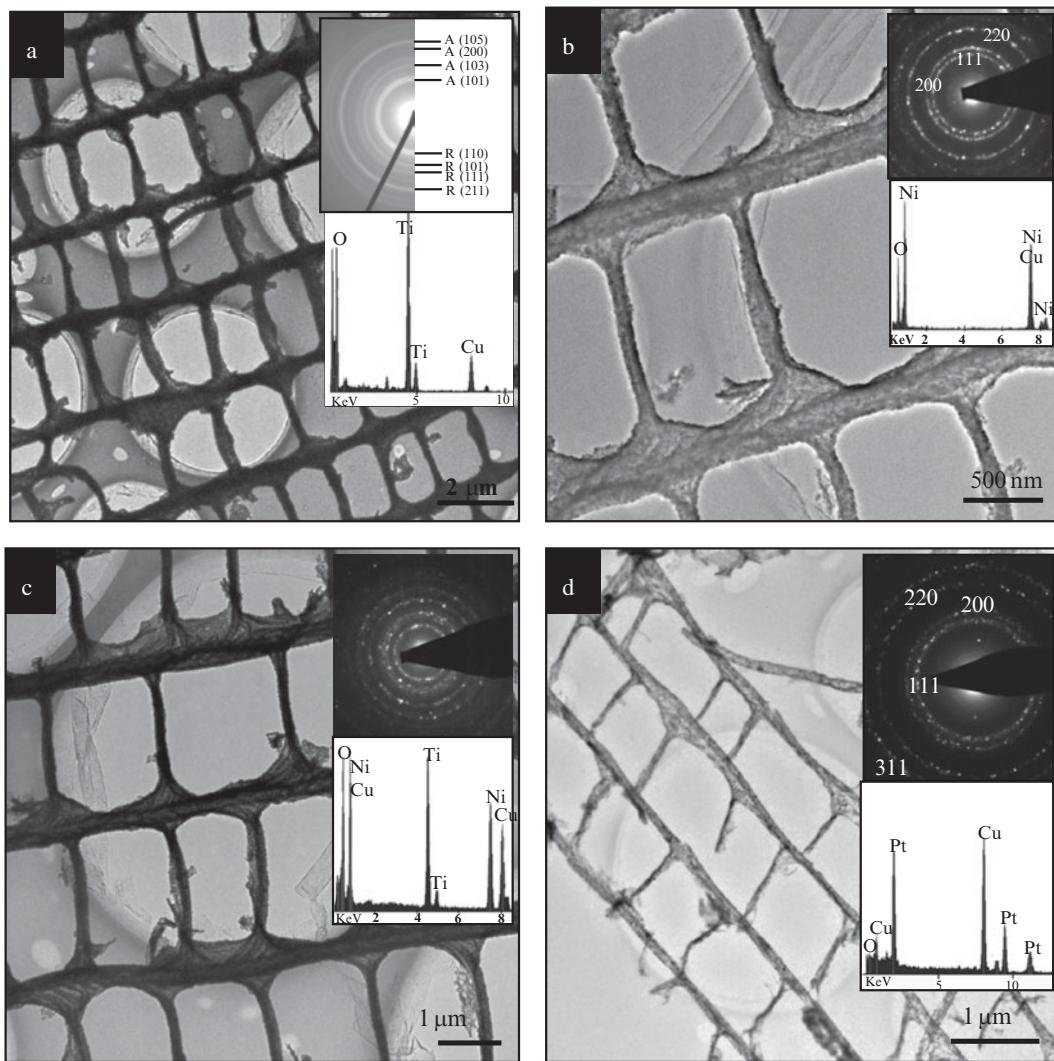
The biotemplating approach we developed shows a sophisticated and effective way of surface replicating. We also employed this novel technique to

fabricate a series of functional materials with a free-standing and uniform grid structure, including n-type semiconducting  $\text{SnO}_2$ , p-type semiconducting  $\text{NiO}$ , ferromagnetic  $\text{Fe}_2\text{O}_3$ , and noble-metal catalytic Pt and Pd, etc. The processing approach is similar to fabricate nanocrystalline  $\text{TiO}_2$  grids. The experimental conditions are dependent on the objective functional materials. For example, Pt grid, we coated a Pt thin film on the butterfly-wing scale templates with a thickness of 20–50 nm. Consequently, the samples were calcinated in a furnace at 650 °C for 30 min in air. The underlying grid-like biotemplates were removed, while the Pt grid-like nanostructures were preserved. For composited grids, two-step replicating techniques are employed. We can fabricate a kind of free-standing grid first, and transfer them onto a cleaning substrate, and then coat another metal thin film on the grids and oxidize at around 500 °C in air. For example, we have fabricated  $\text{TiO}_2/\text{NiO}$  composited grids with core-shell structure. Some typical grids of functional materials we developed are shown in Fig. 22.2.

The free-standing semiconducting grids are novel 2D building blocks and can be easily transferred onto any substrate, tailored and stacked to create 3D ordered nanostructures for nanodevices. The  $\text{TiO}_2$  grids stacked into 3D lattices array would be expected to enhance the light-harvesting efficiency (LHE) by the fact of manipulating light by the strong Bragg scattering among the three-dimensional lattice, provide a large surface area in the open architecture, and potential for the ideal photoactive electrodes for DSSCs. Other functional materials with uniform grid morphology, such as noble-metal Pt and Pd for catalysis, ferromagnetic  $\text{Fe}_2\text{O}_3$  for electromagnetic wave shielding, semiconducting  $\text{ZnO}$  for lasing, and  $\text{VO}_x$  grids for lithium capacity, etc., would enhance their performance dramatically due to their unique 2D grid-like nanostructures.

### **22.3 Assembly and photophysics of grid-like nanostructures into 3D open architectures for the photocatalytic electrodes**

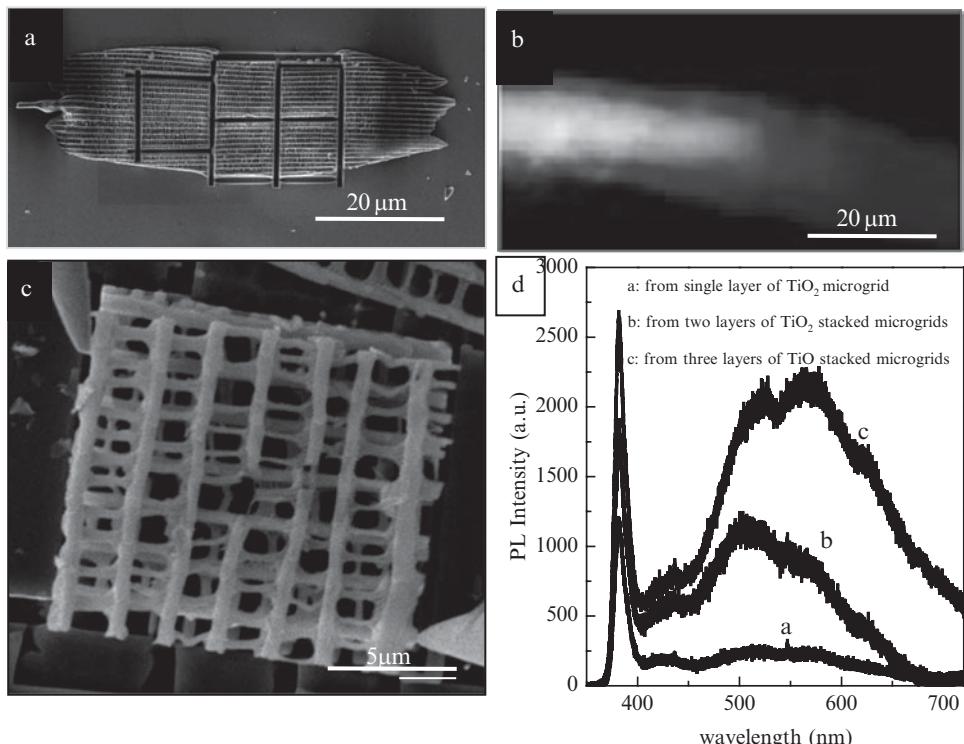
Creating 3D ordered nanostructures is of great interest in unique optoelectronic, magnetic, or catalytic properties that can be tuned by varying their size and/or interseparation distance for novel nanoelectronic and nanophotonic devices (Smith *et al.* 2004; Lin *et al.* 1998; Noda *et al.* 1999, 2000). Specifically, such periodic nanostructures can offer unprecedented optical properties in the visible light region (Cubukcu *et al.* 2003). However, traditional nanofabrication methods, such as e-beam lithography (Tetreault *et al.* 2006), nanoimprint (Chou *et al.* 2002), soft lithography (Xia and Whitesides 1998), fabricating periodic nanostructures with such dimensional scale and complexity as required with a high degree of control is very challenging. For fabricating 3D photocatalytic electrodes with unique optical properties, we propose herein to employ the layer-by-layer stacking method to achieve 3D nanostructures. It is unnecessary to involve the precise alignment and planarization processes between layers for the application of photoelectrodes. Thus, our strategy is to



**Fig. 22.2** A series of TEM images showing some functional materials with grid-like morphology replicated from the butterfly-wing scales. The inset images are the selected-area diffraction patterns as well as their EDX spectra corresponding to these functional materials: (a) n-type  $\text{TiO}_2$ , (b) p-type  $\text{NiO}$ , (c)  $\text{TiO}_2/\text{NiO}$  core/shell composited structure, (d) pure noble metal Pt.

fabricate the free-standing and uniform 2D grids of functional materials by a biotemplating approach, and stack them into 3D lattice arrays layer by layer using the Langmuir–Blodgett method or the ink-jet printing technique.

As described above, we fabricated nanocrystalline  $\text{TiO}_2$  2D grid-like nanostructures, replicated from individual butterfly-wing scales. The free-standing  $\text{TiO}_2$  grids can be easily transferred onto silicon substrate, for tailoring and stacking to create 3D ordered nanostructures. A free-standing leaf-like replica of  $\text{TiO}_2$  grid was transferred on silicon substrate, and tailored into square tiles ( $10 \times 10 \mu\text{m}^2$ ) by a focused ion beam (FIB) technique, shown in Fig. 22.3(a). The free-standing square grids can be stacked into a 3D lattice array layer by



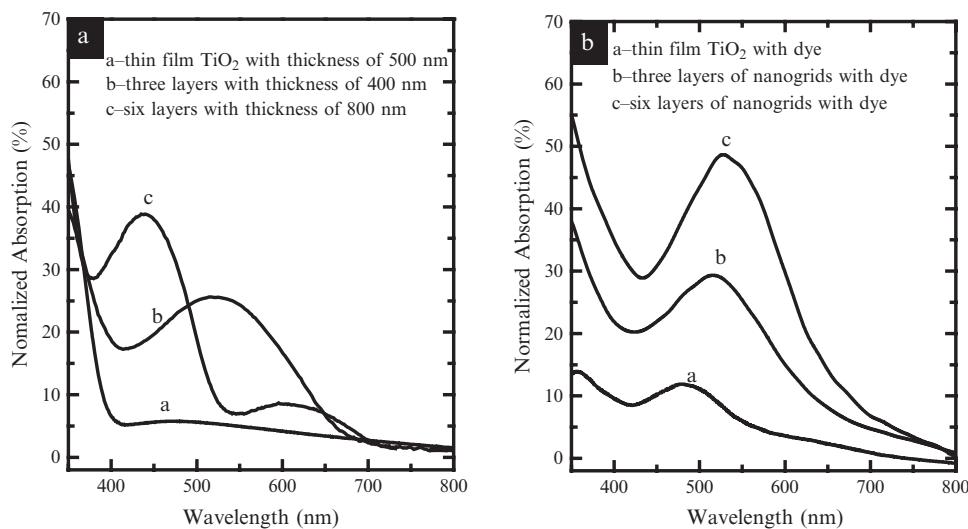
**Fig. 22.3** Assembly of 2D grids stacked into 3D lattice arrays and their optical behaviors. (a) A SEM image of the leaf-like replica of TiO<sub>2</sub> on silicon substrate, tailored into square tiles ( $10 \times 10 \mu\text{m}^2$ ) using a focused ion beam (FIB) technique. (b) A typical optical image of the stacked replica of TiO<sub>2</sub>, showing the iridescent color from the stacked three-dimensional (3D) grids array. (c) A typical SEM image of three layers of tailored grids ( $10 \times 10 \mu\text{m}^2$ ) stacked on silicon substrate by using a tungsten probe in an SEM chamber; (d) Photoluminescence (PL) spectra recorded from a single layer and several stacked layers of TiO<sub>2</sub> grids excited by the fourth harmonic of a Nd:YAG laser.

layer. We picked up the tailored tiles in ethanol and handled them, to stack on a substrate, by microprobes under an optical microscope. A typical optical image of the stacked TiO<sub>2</sub> grids showing the rainbow color is illustrated in Fig. 22.3(b). We also manipulated the tailored square tiles and stacked them layer by layer at right angles into 3D structures by the tungsten microprobes in SEM chamber. As a result, a square ordered array with three layers of tailored TiO<sub>2</sub> tiles is positioned on silicon substrate (Fig. 22.3(c)). The thickness of the stacked tiles is about 400 nm. In addition, the grids can also be manipulated in solution and stacked into 3D arrays with tens of layers by means of the Langmuir–Blodgett approach (Huang *et al.* 2001; Duan *et al.* 2003). It can produce scalable 3D nanostructures for the application of photocatalytic electrodes.

D ordered nanostructures, termed photonic crystals, can manipulate the localizations of photons, giving rise to the suppression and/or enhancement of light emission (Joannopoulos *et al.* 1997). We examined the photonic behaviors of the stacked-grid array with different layers as well as stacking styles. The photoluminescence (PL) spectra for TiO<sub>2</sub> grids excited by the fourth harmonic of a Nd:YAG laser (266 nm, 5-ns pulse width, 10-Hz pulse

repetition rate) are recorded at room temperature. PL spectra recorded from several stacked layers of TiO<sub>2</sub> grids are presented in Fig. 22.3(d). The outstanding emission peak at approximately 380 nm corresponds to the near-band-edge emission of anatase TiO<sub>2</sub>. For a single layer of a TiO<sub>2</sub> grid, there is a very weak spontaneous emission in the range 500–600 nm near the green band, implying the TiO<sub>2</sub> grids perform good quality crystallization and produce low oxygen vacancy concentration. In contrast, the PL spectra from stacked TiO<sub>2</sub> grids show remarkably enhanced emission, not only the strong near-band-edge emission at approximately 380 nm, but also the high intensity of spontaneous emission in the range 500–600 nm. It is believed that the stacked-grid array exhibits PBG-like behaviors, responsible for the remarkable enhancement of spontaneous emission. The strong multiple Bragg scattering and light localization would be coupled in the stacked-grid array, resulting in stimulating strong spontaneous emission (John 1987). This indicates that the spontaneous emission could be controlled by the stacked layers as well as the different stacking styles. Similar results are originated from III–V semiconductors and Si-based photonic-based photonic crystals with a “woodpile” structure fabricated by sophisticated lithographic techniques (Lin *et al.* 1998; Ogawa *et al.* 2004).

We also examined the optical absorption of the stacked-grid arrays on a UV-visible spectrophotometer. The incident light was perpendicular to the array plane. A series of UV-Vis spectra recorded from the stacked TiO<sub>2</sub> grids are shown in Fig. 22.4(a). As a reference, a clear titania film with the thickness of about 500 nm exhibited an absorption onset at 380–390 nm that corresponds directly to the characteristic 3.3-eV bandgap. In contrast to such a clear film, the spectra of the stacked-grids arrays with different layers or thickness show the broadening of the absorption band. A strong increase in absorption in the UV region is observed at wavelengths around 380 nm associated with the intrinsic absorption of titania, while a broaden absorption appears in the visible



**Fig. 22.4** A series of UV-Vis absorption spectra recorded from the stacked TiO<sub>2</sub> grids with different layer or thickness (a), and from the N719 dye adsorbed on the stacked TiO<sub>2</sub> grids (b).

and near-infrared region associated with the light scattering from 3D stacked-grid structures. There is a broad absorption throughout the spectrum with maxima at 435 and 620 nm for six-layered grids (about 800 nm in thickness), while the relatively weak and broaden absorption peak is around 540 nm and no other shoulder-broadened absorption band are observed for three-layered grids (about 400 nm in thickness). From the absorption intensities, it indicates that there is a difference in the adsorption properties of the different stacked layers. This phenomenon is attributed to the strong interaction of light with 3D stacked-grid structures, including the incident light scattered and strongly diffracted light of a specific wavelength.

When N719 dye was adsorbed on a 3D stacked-grid array of titania, the UV-visible spectra recorded from the stacked layers of  $\text{TiO}_2$  on conducting glass are shown in Fig. 22.4(b). The dye adsorbed on the reference (flat) titania film is characterized by one absorption peak at  $\lambda_{\max} = 490$  nm. In contrast, the dye adsorbed onto three-layer grids array shows a strong enhancement of absorbance shift to 526 nm, and the absorption onset could not be clearly identified. For the dye absorbed on the stacked six-layer grids, the broadened absorption band extends to the visible and near-infrared region and peaks around 550 nm. The absorption intensity is also enhanced dramatically, resulting from the increase of the dye-absorbed surface area with the thickness. From the observed increase in absorption efficiency, it is clear that the titania with 3D stacked-grid nanostructure would be the ideal photocatalytic electrode for DSSCs.

It is known that the optical path in a stacked-grid array becomes longer due to multiple scattering, when the light wavelength is comparable to the periodicity of the stacked-grid dimension (Yablonovitch 1987; Chow *et al.* 2000). The photoluminescence (PL) spectra from stacked  $\text{TiO}_2$  grids show the dramatically stimulated spontaneous emission, arising from the strong light localization and multiple Bragg scattering depending on periodically stacked layers and styles. 3D-ordered titania with lattice spacings of the order of wavelengths of light exhibits a stop band in the region of visible light. Near the stop-band edge, there exist low group velocities of light along with high electric-field intensity, which result in a strong interaction of light with titania, accounting for the strong stimulated spontaneous emission (Lodahl *et al.* 2004; Halaoui *et al.* 2005).

The longer optical path in the stacked grid also causes an enhancement of optical absorption, associated with a greater light-harvesting efficiency. Such enhancement occurs at resonant modes that are partially localized within the stacked-grid array. Partially confined photons travel slower through the stack-grid layer and are more likely to be absorbed. In principle, the enhancement takes place at wavelengths corresponding to the stop band of the photonic crystals involved. This effect is more pronounced in the dependence of the stacked-layer numbers or thickness. The amplified effect in optical absorption or light-harvesting efficiency (LHE) could also be accomplished by controlling the stacked structural configurations utilizing the surface-resonant modes, multiple path lengths from a Bragg reflection, and random light scattering in this grid-array structure. We may design this kind of 3D open architecture  $\text{TiO}_2$  with the stacked-grid structure as the photocatalytic electrode in DSSCs. It

is expected to achieve the efficiency of harvesting solar energy, increase the photocatalytic activity, and enhance the separation of the UV-excited electrons and holes simultaneously. We hypothesized that the enhancements arising from two fundamentally different origins can be united to give a cooperative effect for high photoconversion efficiency.

The technique of stacking grids into 3D ordered nanostructures we present here is also a new route to 3D photonic crystals. Due to various free-standing functional grids easily available from the replication of biotemplates, we can freely stack these 2D functional building blocks alternatively for the design of 3D photonic nanodevices. In comparison to the colloid photonic crystals, there are three advantages in our techniques: (1) various functional oxides or noble metals with 2D grid nanostructures as building blocks instead of colloidal nanoparticles; (2) These functional materials can be stacked alternatively into hybrid or composite structure. (3) The stacked style can be adjusted in various directions, while the structures of packed opal spheres are limited. It should be recognized that the principle of our technology can be extended to various functional materials with 3D nanostructures for the appropriate multifunctionality. We visualize our technique being used as a processing step in manufacturing components in photonics and optoelectronics. As a good example, we demonstrate here how to employ the open architecture with 3D stacked-grid nanostructure as photocatalytic electrode for DSSCs with high photoconversion efficiency.

## 22.4 Performance of DSSCs working with dye-sensitized $TiO_2$ stacked-grid array photoelectrodes

In general, efficient solar-energy conversion in a DSSC is connected with sufficient light absorption by the adsorbed dye molecules, efficient electron injection from the excited dye into the conduction band of semiconducting oxides, the transport of electrons through the photoelectrode without electron recombination, and the regeneration of the dye molecule in its ground state by electron injection from a reduced ion in the electrolyte into the HOMO of the adsorbed dye (Hodes *et al.* 1992; Fisher *et al.* 2000). However, the penetration of gel-like electrolytes or conjugated polymers into conventional mesoporous electrodes is problematic because of their small pores. Hence, the semiconducting photoelectrode with large surface area and high porosity is the core of a solar cell for high photocatalytic activity and efficiency (Montanari *et al.* 2002). Here, we employed the  $TiO_2$  grids stacked into 3D open architectures as photocatalytic electrodes in DSSCs. It can enhance the light-harvesting capability of the dye adsorbed on it and also allow the penetration of the electrolyte right up to maintaining good electrical connectivity to the substrate that aids in the charge-separation process, thus leading to high power conversion efficiency (Willig and Schwarzburg 1999).

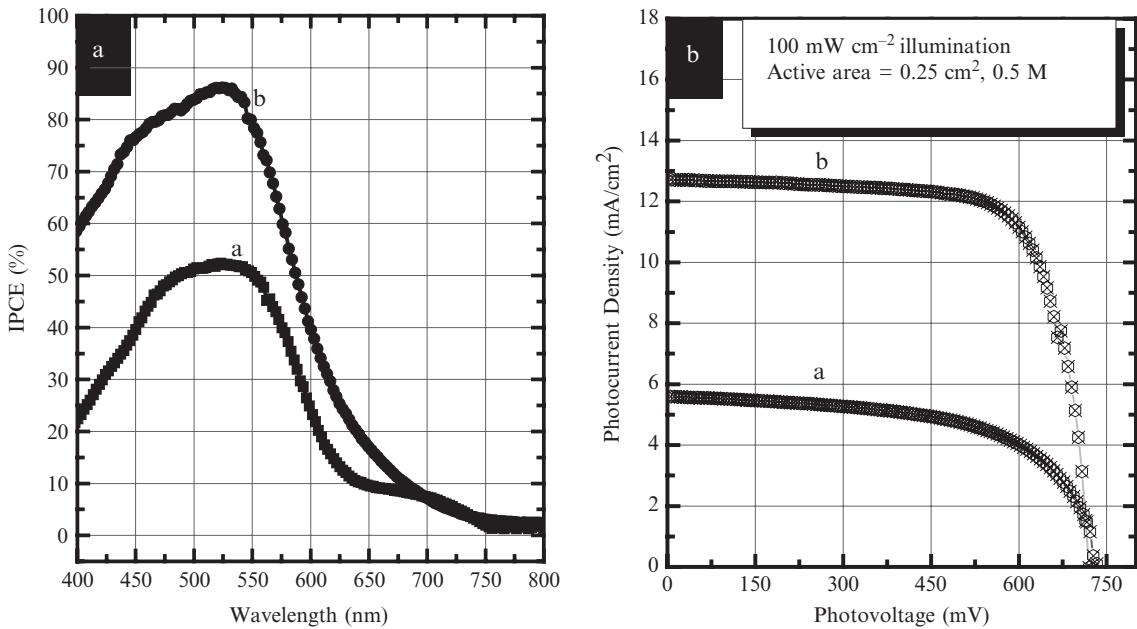
Since the dye-sensitized  $TiO_2$  grid-like electrode acts as a majority carrier device, electron-hole recombination in the “bulk” of the semiconductor is

excluded as a path for losing electrons injected into the TiO<sub>2</sub> grids. The appreciated acceptor in the electrolyte collecting holes rapidly from the semiconductor will increase the efficiency of the electrode. (Jongh *et al.* 1997; Franco 1999). Enhancing the light-energy conversion efficiency should therefore prove valuable in the field of photovoltaic cells. It has been shown that slow photons can significantly increase during the propagation of the light through the TiO<sub>2</sub> stacked-grid array. By reducing the group velocity of light at energies near the edge of the photonic stop band, a higher probability of absorption was achieved, which is responsible for the photoconversion efficiency of DSSCs.

The nanocrystalline TiO<sub>2</sub> grids can be cheaply produced in large scale by the biotemplating approach. We assembled the 2D grids into 3D open architectures on various conducting substrates by a Langmuir–Blodgett approach or an inkjet printing technique (Kong and Wang 2006). Thus, it is assumed that the TiO<sub>2</sub> stacked-grid arrays are a promising alternative to conventional TiO<sub>2</sub> nanoparticle electrodes, giving rise to the enhancement of conversion efficiency.

During the fabrication of DSSCs, the TiO<sub>2</sub> stacked-grid arrays are firstly prepared and annealed on FTO glasses at 350 °C for the good conduction between the grid layers and supporting substrate. The electrodes (area 1 cm<sup>2</sup>) were soaked in a 1:1 vol mixture of acetonitrile and *tert*-butanol with ruthenium dye (N719) for 12 h at room temperature. The soaked electrodes were washed with ethanol to remove non-anchored dye molecules and then dried in air. Pt-sputtered FTO glasses were used as a counterelectrode. The Pt counterelectrode and the dye-anchored TiO<sub>2</sub> grids electrode were assembled into a sealed sandwich-type cell. The electrolyte consisted of 0.5 M tetrabutylammonium iodide, 0.05 M I<sub>2</sub>, and 0.5 M 4-*tert*butylpyridine in acetonitrile, and was introduced between the electrodes by capillary forces. The active electrode area was typically 0.25 cm<sup>2</sup>. The thickness of the stacked-grid photoelectrodes is associated with the stacked layers of grids array. We examined the thickness using scanning electron microscopy (SEM).

The working characteristics of a DSSC based on the TiO<sub>2</sub> stacked-grid electrode sensitized by N719 were examined under irradiation of a 100 mW cm<sup>-2</sup> xenon lamp with simulated AM1.5 solarillumination. The incident photon-to-current conversion efficiency (IPCE) is defined as the number of generated electrons divided by the number of incident photons. Typically, the performance of the cells with the different thickness of about 1 μm and 4 μm was shown in Fig. 22.5. The curves a and b represent the cells with the thickness of about 1 μm and 4 μm, respectively. In Fig. 22.5(a), the spectra of incident monochromatic photon-to-current conversion efficiency (IPCE) exhibit a maximum of 52% at the wavelength of 525 nm for the TiO<sub>2</sub> grids array in the thickness of about 1 μm, shown in curve a. The curve b exhibits the IPCE maximum of 86% at the wavelength of 530 nm for the grid array with the thickness of 4 μm. The IPCE spectrum shows that the maximum efficiency was over 52% at the wavelength around 525 nm, and also indicates that efficient light harvesting by the TiO<sub>2</sub> nanorod electrode is occurring for wavelengths longer than 525 nm. For the thinner cell (1 μm), the short-circuit current density ( $J_{sc}$ ), the open-circuit photovoltage ( $V_{oc}$ ), and the fill factor (FF) are found to



**Fig. 22.5** The spectra of IPCE and the photocurrent density versus voltage characteristics of a DSSC based on the  $TiO_2$  stacked-grid electrode sensitized by N719, measured under standard global AM 1.5 solar conditions, the active area is about  $0.25\text{ cm}^2$ . The curves a and b represent the cells with the thickness of about  $1\text{ }\mu\text{m}$  and  $4\text{ }\mu\text{m}$ , respectively.

be  $5.8\text{ mA/cm}^2$ ,  $0.72\text{ V}$ ,  $0.58$  and yielding a conversion efficiency ( $\eta$ ) of  $2.4\%$ , respectively. In contrast, for the thicker cell ( $4\text{ }\mu\text{m}$ ), the curve b represents the characteristics of a DSSC with the photoelectrode of  $TiO_2$  stacked-grid array. Under similar test conditions,  $J_{sc}$ ,  $V_{oc}$ , and FF, are found to be  $12.8\text{ mA/cm}^2$ ,  $0.72\text{ V}$ ,  $0.75$  and yield a conversion efficiency ( $\eta$ ) of  $6.9\%$ , respectively.

Regarding the thickness of the photoelectrode, it shows that the cell with the increase of  $TiO_2$  grid layers has improved  $\eta$  and FF. Table 22.1 lists the photovoltaic properties such as short-circuit current density ( $J_{sc}$ ), open-circuit voltage ( $V_{oc}$ ), fill factor (FF) and efficiency ( $\eta$ ) of all four types of cells that we studied. The  $\eta$  and FF were obtained using standard equations. Of all four types, the thinner the electrode, the lower the  $\eta$  and FF of the cell. The  $\eta$  and FF were increased dramatically when a  $4\text{-}\mu\text{m}$  thick array of  $TiO_2$  was introduced. However, decreasing the thickness to less than  $500\text{ nm}$  can open up the pores for more electrolyte exposure. The results of this study, within the limits of the present experiment, indicate that a minimum layer thickness is preferred for better  $\eta$  and FF. Although no systematic study of the influence of the thickness of the  $TiO_2$  layer on the intensity dependence of the IPCE has been made, it is observed that cells with quite different thicknesses ranging from  $4$  to  $10\text{ }\mu\text{m}$  all showed no significant difference in their behavior, which shows that the thickness above  $4\text{ }\mu\text{m}$  of the cell has no critical influence on our results.

Furthermore, it is remarkable that our results are obtained using a stacked-grid array as photoelectrode, a fact that proves the advantages of the nanostructured grid-like geometry. The electron-transport mechanism may be altered and improved by changing the semiconductor morphology from a nanoparticle

**Table 22.1** Typical performance of DSSCs working with the stacked TiO<sub>2</sub> grids photoelectrodes with different stacked layers or thickness.

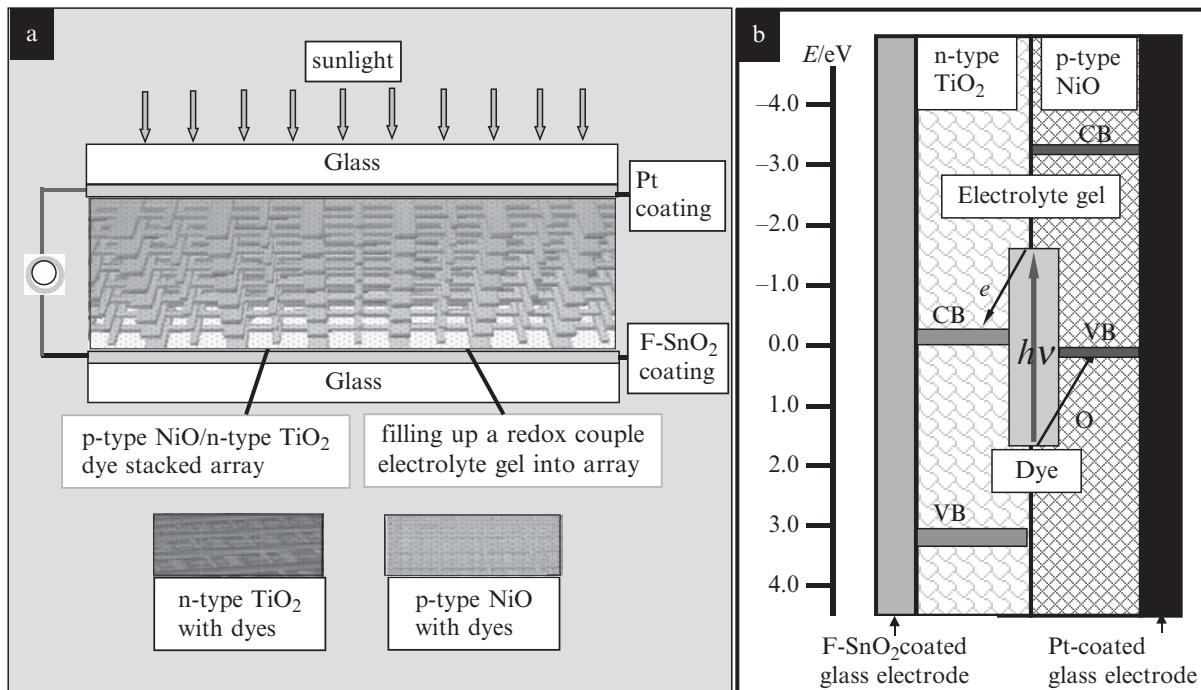
Thickness of electrodes	V <sub>oc</sub> (mV)	I <sub>sc</sub> (mA/cm <sup>2</sup> )	FF (%)	η (%)
~500 nm	712	3.92	0.51	1.42
~1 μm	720	5.81	0.58	2.42
~2 μm	720	9.36	0.71	4.78
~4 μm	720	12.82	0.75	6.92
~10 μm	720	13.44	0.72	6.96

film to a stacked-grid array. Electron transport in the grid-like network is surprisingly efficient, due to the larger internal surface area leading to an enhanced dye adsorption. It also provides the direct path to improve the efficient electron diffusion and rapid transport through the interpenetrated network to the conductive substrate, promoting the charge-separation process, thus giving rise to the increase of the output voltage or the photogenerated current. The stacked-grid array could produce the photonic-crystal-like behavior and amplify the effect in optical absorption or light-harvesting efficiency, accounting for the enhancement of IPCE, resulting in the significant photoconversion efficiency.

## 22.5 Characteristics and performance of DSSCs working with TiO<sub>2</sub>/NiO composite photoactive electrodes

DSSCs routinely work with a dye-sensitized n-type semiconductor electrode upon light excitation. The key step of these devices is the photoinjection of an electron into the conduction band of the cells from the excited state of dye sensitizers (Bedja *et al.* 1994). In contrast, there are only a limited number of studies of the sensitization of p-type semiconductors, for which the operation principle is just the inverse scheme and consists of the photoinjection of a hole into the valence band of the semiconductor. However, the dye-sensitized solid-state solar cells (DSSC) using p-type semiconductors often met problems of short-circuit and mass-transport limitations of the ions, resulting in low conversion efficiencies compared with the liquid electrolyte. In this respect, p-type oxide semiconductors have seldom been tested as solid hole collectors. Recently, there are considerable scientific and technological interests in developing nanostructured metal oxides with p-type semiconductivity. For NiO, one of the stable wide-bandgap materials, it has been demonstrated that thin-film NiO can be used as a transparent p-type semiconducting layer for DSSC. Studies of p-type systems may lead to the development of new types of devices and are particularly vital for the construction of tandem DSSCs in which both the cathode and the anode would be photoactive and for photocatalytic oxidations driven by visible light (He *et al.* 1999; Bandara *et al.* 2004, 2005).

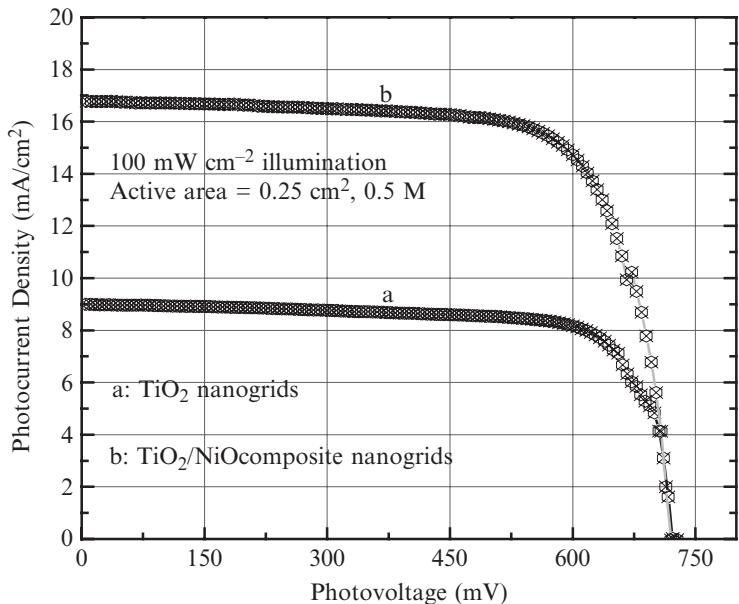
In our case, we fabricated the grid nanostructure of n-type TiO<sub>2</sub> and p-type NiO with good optical quality using biotemplating techniques. As described



**Fig. 22.6** The schematic of a new type of DSSC with composite structure (a) as well as the working principles (b). The composite solar cells employed two photoactive semiconductor electrodes, one using n-type  $TiO_2$  as a photoanode, and another using p-type NiO as a photocathode.

above, such DSSCs comprising two photoactive grid-like electrodes of p-type NiO as hole conductors and n-type  $TiO_2$  as electron conductors, allowing the large amount of dye absorbing on a 3D interpenetrated network, could provide a novel strategy to improve the performance of DSSCs. A 3D solar cell working with the composite electrodes is schematically represented in Fig. 22.6(a). In this case, a composite electrode comprises of two photoactive semiconductors as n-type  $TiO_2$  and p-type NiO stacked grids. The dye is adsorbed on both n-type and p-type semiconducting materials, and the 3D interpenetrated network would provide the direct path for transporting the electrons in n-type  $TiO_2$  while the hole in p-type NiO, leading to better charge separation and high output of photogenerated current. The working principle of the new type DSSCs is illustrated in Fig. 22.6(b).

To demonstrate the advantage of the 3D solar cell with the composite structures concept, we fabricated a series of DSSCs with  $TiO_2/NiO$  stacked photoelectrodes prepared by ink-jet printing on FTO glasses. The stacked-grid photoelectrodes on FTO glasses were annealed at 350 °C, and immersed in ruthenium dye (N719) for 12 h at room temperature. The soaked electrodes were washed with ethanol to remove non-anchored dye molecules and then dried in air. The composite photoelectrode is limited to about 2  $\mu m$  in thickness, and the interpenetrating network has a porosity of above 70%. To complete the devices, an electrical contact is applied on top. The top electrode was a conductive substrate with a 20 nm Pt layer deposited by e-beam evaporation. The composition of electrolyte is the same as before and was introduced



**Fig. 22.7** The photocurrent density versus voltage characteristics of a DSSC with the  $\text{TiO}_2/\text{NiO}$  stacked-grid composite photoelectrode sensitized by N719 dye (about  $2\ \mu\text{m}$  in thickness), measured under standard global AM 1.5 solar conditions, the active area is about  $0.25\ \text{cm}^2$ .

between the electrodes by capillary forces. We measured the 3D DSSC under standard global AM 1.5 solar conditions, the active area is about  $0.25\ \text{cm}^2$ .

The performances and characteristics of this 3D DSSC we fabricated are shown as curve b in Fig. 22.7. The short-circuit current density ( $J_{\text{sc}}$ ), the open-circuit photovoltage ( $V_{\text{oc}}$ ), and the fill factor (FF) are found to be  $16.5\ \text{mA}/\text{cm}^2$ ,  $0.75\ \text{V}$ , and  $0.68$ , yielding a conversion efficiency ( $\eta$ ) of  $8.4\%$ , respectively. As a reference, curve a represents the characteristics of a DSSC with the photoelectrode of only a  $\text{TiO}_2$  stacked-grid array. The photoelectrode is also limited to about  $2\ \mu\text{m}$  in thickness. Under the similar test conditions,  $J_{\text{sc}}$ ,  $V_{\text{oc}}$ , and FF, are found to be  $9.38\ \text{mA}/\text{cm}^2$ ,  $0.75\ \text{V}$ , and  $0.71$ , yielding a conversion efficiency ( $\eta$ ) of  $5.0\%$ , respectively. From the significant efficiency of this DSSC working with 3D  $\text{TiO}_2/\text{NiO}$  composite electrode, it is proposed that the electrons and holes are physically separated by n-type and p-type semiconductors, thus reducing the possibility of back electron transfer and suppressing the wasteful charge recombination. In addition, the 3D interpenetrated network may also contribute to better charge transportation because of their reduced grain boundaries (direct path) compared to those of the nanoparticles.

The reason for the high efficiency could be due to the 3D stacked-grid array electrode comprising of n-type  $\text{TiO}_2$  and p-type  $\text{NiO}$  semiconductors. As the conductor band (CB) of the  $\text{NiO}$  position is higher than that of  $\text{TiO}_2$ , the electrons injected from the excited dye molecules to the CB of  $\text{TiO}_2$  should surmount the  $\text{NiO}$  energy barrier and compete with the recombination at the interface (Farzad *et al.* 1999; Bisquert *et al.* 2002). The band positions of  $\text{NiO}$  and  $\text{TiO}_2$  allow charge transfer from  $\text{NiO}$  to  $\text{TiO}_2$  as shown in Fig. 22.6(b). Therefore, the relaxed electrons in the CB of  $\text{TiO}_2$  could not tunnel to the  $\text{NiO}$  layer efficiently and reach the conducting glass. The advantage of using a p-type  $\text{NiO}$  stacked grid electrode is that it introduces a potential gradient

across the 3D network as a charge-recombination barrier. Considering all these factors, it can be concluded that the charge transfer occurs from NiO to TiO<sub>2</sub>, traversing electrons through the NiO layer to the top electrode. According to this charge-transfer mechanism, the stacked-grid structure would be of benefit for the charge separation, electron and hole transfer rapidly from NiO to TiO<sub>2</sub>.

From the above results, it can be found that the significantly enhanced performance of solar cells with the TiO<sub>2</sub>/NiO composite electrode is comparable to that of solar cells using p-n junction semiconducting films. However, due to the limited thickness of stacked grids, the overall conversion efficiencies of the solar cells with composite electrode we report here up to 8.4%, could not compete with the available techniques. We believe that this would open up an alternative approach to get over the bottleneck and solve some of the problems encountered in the fabrication of traditional DSSCs.

## 22.6 Summary

We developed a technique to fabricate free-standing and uniform grids of a series of functional materials by replicating the biotemplates from butterfly-wing scales. The free-standing grids are novel 2D building blocks, and can be freely transferred, tailored and stacked into 3D ordered arrays for the design of nanodevices. Such stacked grids array can manipulate the light and stimulate, simultaneously amplifying the light-harvesting efficiency.

It is very important for nanotechnology to create 3D nanostructures from free-standing 2D nanoscale building blocks. The butterfly-wing scale we employed here is a kind of convenient example of template from biological structure. This biotemplating technique can also be extended to replicate the surface of other artificial templates obtained from top-down lithography, for achieving the free-standing and versatile building blocks. The versatile 2D building block can be assembled into 3D ordered structure for various applications. It can be extended to stack the free-standing grids with different materials alternatively for the 3D metamaterials.

We employed the grid-like nanostructure arrays as photocatalytic electrodes for a new type of dye-sensitized solar cells. The electrode comprising of the stacked-grid arrays, could provide more surface area to absorb the dye molecules rather than conventional mesoporous thin films. The stacked-grid array could produce photonic-crystal-like behaviors and amplify the effect in optical absorption or light-harvesting efficiency, accounting for the enhancement of IPCE, resulting in the significant photoconversion efficiency. It also provides the direct path to improve the efficient electron diffusion and rapid transport through the interpenetrated network to the conductive substrate, benefiting the charge-separation process, thus giving rise to the increase of the output voltage or the photogenerated current.

We present a new type of DSSC with two photoactive semiconductors composite electrodes. In this new type of DSSC using a TiO<sub>2</sub>/NiO stacked-grid array as the electrode, it shows significant photoconversion efficiency, comparable to that of solar cells using p-n junction semiconducting films. The main limitation of this approach is that DSSCs of large area could not

be achieved yet, and the thickness of the electrode is also limited by using a Langmuir–Blodgett approach or ink-jet printing technique. It is a challenge to fabricate the DSSCs with low cost, high efficiency and assured long-term performance. Thus, there is a long way to go for forging a breakthrough in the bottleneck of efficiency and solving the problems encountered in the state-of-the art DSSCs.

## Acknowledgments

This work was supported by the National Science Foundation of China under award number 50525208 and 10734020, and MOST (grants 2007CB936203 and 2006CB932602) and the project from the Ministry of Education of China and Shanghai Science and Technology Committee. We would also like to thank the facilities support from Prof. Z.F. Liu in Peking University.

## References

- Adachi, M., Murata, Y., Takao, J., Jiu, J.M.S., Wang, F. *J. Am. Chem. Soc.* **126**, 4943 (2004).
- Anderson, T.F., Richards, Jr., A.G. *J. Appl. Phys.* **13**, 748 (1942).
- Bandara, J., Divarathna, C., Nanayakkara, S. *Sol. Energy Mater. Sol. Cell* **81**, 429 (2004).
- Bandara, J., Pradeep, U.W., Bandara, R. *J. Ph. Chem. Boil. A: Chem.* **170**, 273 (2005).
- Baxter, J.B., Aydil, E.S. *Appl. Phys. Lett.* **86**, 053114 (2005).
- Bedja, I., Hotchandani, S., Kamat, P.V. *J. Phys. Chem.* **98**, 4133 (1994).
- Biró, L.P., Bálint, Zs., Kertész, K., Vértesy, Z., Márk, G.I., Horváth, Z.E., Balázs, J., Méhn, D., Kiricsi, I., Lousse, V., Vigneron, J.-P. *Phys. Rev. E* **67**, 021907 (2003).
- Bisquert, J., Garcia-Belmont, G., Fabregat. F. *J. Solid State Electrochem.* **3**, 337 (1999).
- Bisquert, J., Zaban, A., Salvador, A. *J. Phys. Chem. B* **106**, 8774 (2002).
- Caruso, R.A. *Angew. Chem. Int. Ed.* **43**, 2746 (2004).
- Chik, H., Xu, J.M. *Mater. Sci. Eng. R.* **43**, 103 (2004).
- Chou, S.Y., Keimel, C., Gu, J. *Nature* **120**, 435 (2002).
- Chow, E., Lin, S.Y., Johnson, S.G., Villeneuve, P.R., Joannopoulos, J.D., Wendt, J.R., Vawte, G.A., Zubrzycki, W., Hou, H., Alleman, A. *Nature* **407**, 983 (2000).
- Cook, G., Timms, P.L., Spickermann, C.G. *Angew. Chem. Int. Ed.* **42**, 557 (2003).
- Cubukcu, E., Aydin, K., Ozbay, E., Foteinopoulou, S., Soukoulis, C.M. *Nature* **423**, 604 (2003).
- Duan, X.F., Niu, C.M., Sahi, V., Chen, J., Parce, J.W., Empedocles, S., Goldman, J.L. *Nature* **425**, 274 (2003).
- Fisher, A.C., Peter, L.M., Ponomarev, E.A. *J. Phys. Chem. B* **104**, 949 (2000).
- Franco, G., Gehring, J., Peter, L.M., Ponomarev, E.A., Uhlendorf, I. *J. Phys. Chem. B* **103**, 692 (1999).
- Farzad, F., Thompson, D.W., Kelly, C.A., Meyer, G.J. *J. Am. Chem. Soc.* **121**, 5577 (1999).

- Gratzel, M. *Prog. Photovolt. Res. Appl.* **8**, 171 (2000).
- Gratzel, M. *Nature* **414**, 338 (2001).
- Hagfeldt, A., Gratzel, M. *Chem. Rev.* **95**, 49 (1995).
- Hagfeldt, A., Lindquist, S.E., Gratzel, M. *Sol. Energy Mater. Sol. Cell* **32**, 245 (1994).
- Halaoui, L.I., Abrams, N.M., Mallouk, T.E. *J. Phys. Chem. B* **109**, 6334 (2005).
- He, J., Lindstrom, H., Hagfeldt, A., Lindquist, S.E. *J. Phys. Chem. B* **103**, 8940 (1999).
- Hodes, G., Howell, I.D., Peter, L.M. *J. Electrochem. Soc.* **139**, 3136 (1992).
- Huang, J., Kunitake, T. *J. Am. Chem. Soc.* **125**, 11834 (2003).
- Huang, Y., Duan, X.F., Wei, Q.Q., Lieber, C.M. *Science* **291**, 630 (2001).
- Joannopoulos, J.D., Villeneuve, P.R., Fan, S. *Nature* **386**, 143 (1997).
- John, S. *Phys. Rev. Lett.* **58**, 2486 (1987).
- Jongh, D., Vanmaekelbergh, P.E. *J. Phys. Chem. B* **101**, 2716 (1997).
- Kay, A., Gratzel, M. *J. Phys. Chem.* **97**, 6272 (1993).
- Kay, A., Gratzel, M. *Chem. Mater.* **14**, 2930 (2002).
- Kong, X.Y., Wang, Y.C. Patent in China, file pending (2006).
- Law, M., Greene, L.E., Johnson, J.C., Saykally, R., Yang, P. *Nat. Mater.* **4**, 455 (2005).
- Lodahl, P., Van Driel, A.F., Nikolaev, I.V., Irman, A., Overgaag, K., Vanmaekelbergh, D., Willem, L.V. *Nature* **430**, 654 (2004).
- Lin, S.Y., Fleming, J.G., Hetherington, D.L., Smith, B.K., Biswas, R., Ho, K.M., Sigalas, M., Zubrzyski, S.R., Kurtz, B. *Nature* **394**, 251 (1998).
- Mihi, A., Miguez, H. *J. Phys. Chem. B* **109**, 15968 (2005).
- Montanari, I., Nelson, J., Durrant, J.R. *J. Phys. Chem. B* **106**, 12203 (2002).
- Mor, G.K., Shankar, K., Paulose, M., Varghese, O.K., Grimes, C.A. *Nano. Lett.* **6**, 215 (2006).
- Nazeerudin, M.K., Kay, A., Rodicio, I. *J. Am. Chem. Soc.* **115**, 6328 (1993).
- Nishimura, S., Abrams, N., Lewis, B.A., Halaoui, L.I., Mallouk, T.E., Benkstein, K.D., van de Lagemaat, J., Frank, A.J. *J. Am. Chem. Soc.* **125**, 6306 (2003).
- Noda, S., Tomoda, K., Yamamoto, N., Chutinan, A. *Science* **289**, 604 (2000).
- Noda, S., Yamamoto, N., Kobayashi, H., Okano, M., Tomoda, K. *Appl. Phys. Lett.* **75**, 905 (1999).
- Ogawa, S., Imada, M., Yoshimoto, S., Okano, M., Noda, S. *Science* **305**, 227 (2004).
- O'Regan, B., Gratzel, M. *Nature* **253**, 737 (1991).
- Park, N.G., van de Lagemaat, J., Frank, A.J. *J. Phys. Chem. B* **104**, 8989 (2000).
- Smith, D.R., Pendry, J.B., Wiltshire, M.C.K. *Science* **305**, 788 (2004).
- Song, M.Y., Ahn, Y.R., Jo, S.M., Kim, D.Y., Ahn, J.P. *Appl. Phys. Lett.* **87**, 113113 (2005).
- Tetreault, N., Freymann, G.N., Deubel, M., Hermatschweiler, M., Pérez-Willard, F., John, S., Wegener, M.A., Ozin, G. *Adv. Mater.* **18**, 457 (2006).
- Vukusic, P., Hooper, I. *Science* **310**, 1151 (2005).
- Willig, F., Schwarzburg, K. *J. Phys. Chem. B* **103**, 5743 (1999).
- Xia, Y.N., Whitesides, G.M. *Ann. Rev. Mater. Sci.* **28**, 153 (1998).
- Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059 (1987).

# 23

# Nanolithography using molecular films and processing

23.1 Introduction	808
23.2 Self- and directed patterning	814
23.3 Patterning via external tools	818
23.4 Directed self-masking via selective deposition on chemical patterns	831
23.5 Molecular rulers: A hybrid nanolithographic patterning method	842
23.6 Conclusion	848
References	848

## 23.1 Introduction

The demand for next-generation electronic devices, consisting of smaller electrical components, faster computational capabilities, and increased energy-efficient operations has driven the innovation of new device fabrication methods and materials. As device elements are now reaching to the tens of nanometers dimensions, limits in conventional “top-down” fabrication methods, where increasingly smaller features are carved from large pieces of semiconductors and metals, are rapidly being approached. These limitations are not only due to the resolution of the lithographic processes used to produce these features, but also from the expense required to build and operate the lithography tools. In order to overcome these limitations, many alternative device-fabrication procedures have been explored, including the construction of device features in a highly controlled, layer-by-layer growth scheme. Encompassing a variety of adsorbates and surfaces, this style of constructing electrical components from the “bottom up” has afforded much promise for extending the current limitations of lithography.

Organic molecules are particularly appealing building blocks for lithography processes due to the inherent precision in their size, shape, chemical functionality and relatively low production costs. Over the past twenty years a new nanolithography toolbox consisting of self- and directed organic molecular assembly (“bottom-up”) methods and their patterning techniques (“top-down”) has been developed. While conventional device-fabrication methods already utilize non-selective, spin-coated, polymeric films that are several micrometers thick for critical fabrication elements such as electron-beam resists or photoresists, this new toolbox allows device components to be constructed selectively, one monolayer at a time with dimensions controlled at the molecular level. This approach offers significant versatility, especially since both “bottom-up” and “top-down” approaches to device fabrication can be incorporated.

*C.L. McGuiness, R.K. Smith, M.E. Anderson, P.S. Weiss,  
and D.L. Allara*

Furthermore, since the specific chemical functionality of an organic molecule can be tailored, well-known chemical reactions can be used to couple patterned organic molecular assemblies with inorganic thin-film growth methods such as physical vapor deposition (PVD), chemical vapor deposition (CVD) or atomic layer deposition (ALD) to direct spatially constrained metal or semiconductor thin films. The ability to construct device elements through self-assembly with precise control at the molecular level is one of the benchmarks for nanolithography.

In this chapter, we review the efforts that have been made towards utilizing organic molecular assemblies in advancing the limitations of lithography. This includes methods of patterning self-assembled organic monolayer films through soft-lithographic methods such as microcontact printing ( $\mu$ CP) and nanoimprint lithography (NIL), through direct “write” or “machine” processes with a nanometer-sized tip and through exposure to electron or photon beams. We also review the efforts to pattern the organic assemblies via the physicochemical self-assembling interactions, including patterning via phase separation of chemically different molecules and insertion of guest adsorbates into host matrices. Ultimately, however, the successful incorporation of molecular assemblies into standard lithographic process depends on the ability to produce organic molecular films on both metal and semiconductor substrates with reproducible characteristics that maintain good mechanical, thermal and chemical stability throughout multiple fabrication steps, especially with regards to forming good diffusion barriers for directed thin-film growth. As such, we describe the efforts made to couple patterned molecular assemblies with inorganic thin-film growth methods to form spatially constrained, 3D thin films. In the final section, a hybrid self-assembly/conventional lithography (i.e. molecular rulers) approach to forming nanometer-sized structures is described. The molecular-ruler methodology emphasizes how patterned molecular assemblies combined with selective deposition of thin films can lead to one nanolithography goal: precise layer-by-layer control of nanoscale device features.

### 23.1.1 Molecular films used in nanolithography

The ability to modify and to control the surface properties of a variety of materials by the formation of dense, conformal self-assembled monolayer (SAM) coatings has become a significant scientific and technological tool for nanolithography in recent years. SAMs are typically composed of a single layer of amphiphilic molecules compactly and laterally organized across a surface; generally, this assembly process occurs on metal surfaces (Love *et al.* 2005), yet the surface chemistry of monolayers on semiconductors is an active area of research (Ulman 1996; McGuiness *et al.* 2007a,b). Molecules used in SAMs possess headgroups with chemical affinities for the surface that serve to pin them in place and tailgroups that interact through van der Waals forces to drive lateral organization across the surface. The use of SAMs in nanolithography has centered around two main classes of organic molecules: alkanethiolates and organosilanes.

Alkanethiolates chemisorb to a variety of transition-metal surfaces, including Au, Ag, Pd, Pt, and Cu (Love *et al.* 2005) as well as onto semiconductors such as GaAs (Sheen *et al.* 1992; McGuiness *et al.* 2007a,b), InP (Gu *et al.* 1995; Yamamoto *et al.* 1999; Schwartzman *et al.* 2003; Lim *et al.* 2004), and InAs (Tanzer *et al.* 1999), making alkanethiolates SAMs amenable to integration into many lithographic processes and device structures. In spite of the enormous technological importance of semiconductor materials, most nanolithographic applications have utilized alkanethiolates chemisorbed onto noble-metal surfaces and in particular Au surfaces (Love *et al.* 2005). This is in part due to the challenging surface chemistry of semiconductor native oxides and in part due to the ease with which alkanethiolates will self-assemble on Au surfaces.

Alkanethiolate self-assembly chemistry is not compatible, however, with native-oxide surfaces or silicon, the traditional device semiconductor. For native-oxide surfaces, self-assembled films consisting of trifunctional alkylsilanes molecules have been developed instead (Bigelow *et al.* 1946; Maoz and Sagiv 1984). Through condensation reactions, a cross-linked siloxane bond or Si–O–Si network results that shows remarkable stability to both thermal treatment and chemical attack (Parikh *et al.* 1994, 1995). To form molecular films on hydride-terminated semiconductor surfaces, organometallic grafting reactions with alkenes and alkynes have been developed (Buriak 2002). These reactions are compatible on both silicon and germanium surfaces and result in thermodynamically and kinetically robust semiconductor–C bonds (Buriak 2002). The self-assembling chemistry and properties of these organic molecular films on semiconductor surfaces have been the subject of a number of detailed reviews (see for example Ulman 1996; Buriak 2002; Love *et al.* 2005). In the following two sections we briefly summarize the self-assembly of the most prevalently used molecules, alkanethiolates and alkylsilanes, and the ensemble characteristics of their respective monolayers with a particular emphasis on the monolayer organization with regard to nanolithographic applications.

### 23.1.1.1 Alkanethiolate films on metal surfaces

Although alkanethiolate molecules self-assembled on metal surfaces have become significant scientific and technological tools in recent years (Joachim *et al.* 2000; Smith *et al.* 2004; Gates *et al.* 2005; Love *et al.* 2005), alkanethiolate SAMs formed on Au surfaces have been the most widely investigated and applied to nanolithographic processes (Smith *et al.* 2004; Love *et al.* 2005). In particular, alkanethiolate SAMs on Au surfaces have demonstrated excellent promise as ultrathin organic resists (Smith *et al.* 2004; Love *et al.* 2005). An alternative to thicker and less uniform layers of organics films and polymers (i.e. resists), SAMs can significantly reduce the etching of metal surfaces when exposed to oxidizing solutions. This allows for features to be patterned into the substrate that support the SAM, as the etch rates differ significantly between the bare metal and the SAM-covered metal (Kumar and Whitesides 1993; Wilbur *et al.* 1994; Kim *et al.* 1995; Gorman *et al.* 1995; Xia *et al.* 1996c; Delamarche *et al.* 1997; Geissler *et al.* 2002; Love *et al.* 2002). Thus, SAMs are used as sacrificial structures to create patterns into metal with a resolution

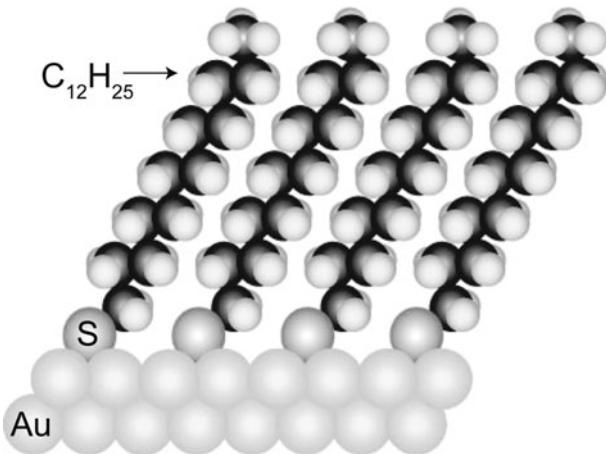
that is often difficult, time-consuming, or expensive to achieve with state-of-the-art, conventional lithographic tools.

One of the main reasons alkanethiolates self-assembled on gold surfaces have been the most widely SAM system used in nanolithographic applications is that the gold surface is relatively chemically inert; it does not readily form a surface oxide nor does it adsorb adventitious material, allowing for SAMs to be prepared with relative ease in ambient conditions. Additionally, the molecules are stable once adsorbed upon the surface, yet they also retain lateral mobility allowing for further selective processing after the adsorption events. Incorporating different terminal groups ( $\omega$ -functional groups) into the thiolate SAM allows for interfacial properties such as hydrophobicity and reactivity to be precisely manipulated and controlled. This self-assembling system of gold and alkanethiol is so very well studied that it is a model system for an abundance of new technologies, including serving as supports for thin-film growth and nucleation (see for example Nuzzo *et al.* 1990b; Seo *et al.* 2004), foundations for biomolecular assembly and capture (Mrksich 1998; Shuster *et al.* 2008) and as insulating films in molecular-scale electronic systems (Allara *et al.* 1998; Donhauser *et al.* 2001, 2003).

Since one of the demands of nanolithography is the ability to produce organic molecular films with reproducible characteristics and chemical stability throughout multiple fabrication steps, it is important to understand the dynamics of SAM formation, a topic of several excellent reviews (e.g. Dubois and Nuzzo 1992; Poirier 1997; Schreiber 2000). Briefly, upon exposure of a gold substrate to an *n*-alkanethiol in solution or in the gas phase, a bond between gold and sulfur ( $\sim 1.9\text{ eV}$ ) (Nuzzo *et al.* 1990a) forms rapidly, within seconds to minutes. Over the next few hours, the hydrocarbon chains closely pack into a predominantly all-*trans* configuration, which imparts significant order to the assembly. If the exposed thiol is in sufficiently high concentration (with a density of  $\sim 10^{15}$  molecules per  $\text{cm}^2$ ), the adsorption of the molecules extends laterally across the substrate. However, the film is restricted from growth in the *z*-dimension from the molecule's non-reactive, methyl-terminated tail; an insulating film of single-molecule thickness results, Fig. 23.1. This allows for a molecularly limited lithographic dimension in one direction.

At low surface coverage, the alkanethiolate molecules lie flat with their hydrocarbon backbones parallel to the gold surface and at higher surface coverages, they begin to stand up, tilting approximately  $30^\circ$  to the surface normal in an ideal all-*trans* configuration so as to maximize van der Waals interactions (Poirier and Pylant 1996). Adsorbate exchange occurs continuously with the solution phase in order to form a local, energetically favorable interaction through kinetically trapped surface structure. The addition of heat during the self-assembly process can aid the SAMs in annealing to a final, optimized, well-ordered conformation.

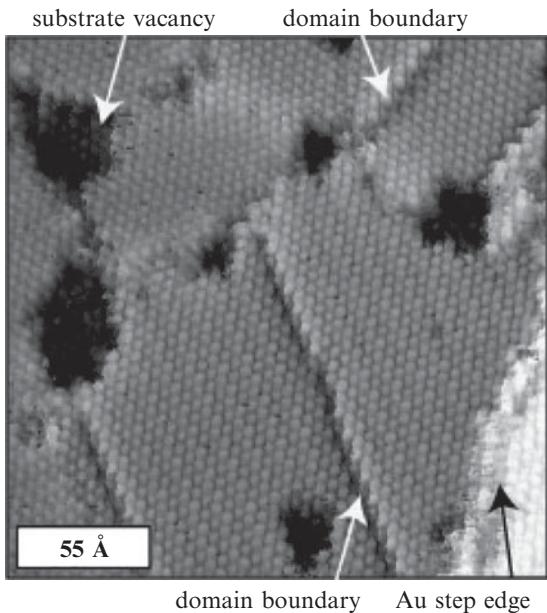
The physical structures and chemical properties of *n*-alkanethiolate SAMs adsorbed onto gold surfaces have been studied by many ensemble techniques to determine the macroscopic characteristics of the monolayer; complementary local probe techniques yield a comprehensive picture of the SAM. Such experimental methods to determine SAM properties include ellipsometry to



**Fig. 23.1** Schematic of an *n*-dodecanethiolate SAM on a gold substrate.

measure film thickness (Porter *et al.* 1987; Nuzzo *et al.* 1990a; Laibinis *et al.* 1991c); Fourier transform infrared spectroscopy (FT-IR) to examine tilt and order (Porter *et al.* 1987; Nuzzo *et al.* 1990a; Laibinis *et al.* 1991c); contact angle goniometry to measure the film's hydrophobic character (Bain *et al.* 1989b); electrochemistry to probe electron transport through the SAM and to examine structural defects such as pinholes (Porter *et al.* 1987; Chidsey 1991); quartz crystal microgravimetry to determine the kinetics of monolayer assembly (Schneider and Buttry 1993; Karpovich and Blanchard 1994); X-ray photoelectron spectroscopy (XPS) to evaluate the composition of the bound species (Nuzzo *et al.* 1990a); diffraction (electron/He/X-ray) to investigate the physical structures of assemblies (Camillone *et al.* 1994); and temperature programmed desorption to probe the thermodynamic aspects of adsorption and desorption and to determine bond strengths within the assembly (Nuzzo *et al.* 1987). From the early 1990s, scanning tunnelling microscope (STM) studies of *n*-alkanethiolate SAMs on gold have shown their organization on the nanometer scale (Anselmetti *et al.* 1994; Delamarche *et al.* 1994; Stranick *et al.* 1994; Poirier and Pylant 1996). Other scanning probes including the atomic force microscope (AFM) (Dürig *et al.* 1993; Piner *et al.* 1999; Liu *et al.* 2000) and the lateral force microscope (LFM) are continuing these studies as well as demonstrating the manipulation of assembled films to make patterned surface structures. SAMs formed at room temperature are governed by a complex mixture of thermodynamics and kinetics.

Even though SAMs find themselves at local thermodynamic minima, they are still kinetically trapped. When imaged with local probes such as the STM, a variety of local defects are seen, Fig. 23.2. Such defects include substrate vacancies, where "holes" that are one atomic layer deep of gold have been formed during the adsorption process; these are believed to be due to the ejection of individual gold atoms from the surface layer and subsequent rearrangement of the remaining gold adatoms, an adsorbate-mediated corrosion (Edinger *et al.* 1997). Monatomic step edges, where one atomic layer of gold separates the gold terraces from each other with a height difference of  $\sim 2.35$



**Fig. 23.2** STM image of an *n*-dodecanethiolate SAM on a gold surface. Several film defects are visible, including substrate vacancies, step edges, and domain boundaries (labelled with arrows). This image is a top-down view of Fig. 23.1; the terminal groups, here  $-\text{CH}_3$ , are uniformly spaced at 5 Å apart.

Å, are also present. Defects of the molecular lattice include domain boundaries of the SAM (mismatches in the tilts of the individual *n*-alkanethiolate adsorbates), vacancies within the crystalline lattice of the molecules, and larger grain boundaries.

The molecules are separated by  $\sim$ 5 Å, the nearest-neighbor distance between the sulfur headgroups bound to the gold substrate. While alkanethiolate SAMs are stable surface structures, the adsorption process is highly dynamic and molecules can continue to exchange with other thiol species in solution or in the vapor phase, with many exchange events occurring at structural defect sites. The structural features and defects shown by scanning probe microscopes such as the STM and the highly dynamic nature of the assemblies have shed light on the ability to pattern SAMs in specific locations for nanolithography.

### 23.1.1.2 Organosilane films on native-oxide surfaces

Alkylsilanes adsorbed on native-oxide surfaces are the other major class of molecular films used for nanolithography applications and the details of self-organization of the alkylsilane molecules have been the subject of a major review (Ulman 1996). Overall, the self-organization of alkylsilanes on native oxide surfaces is fundamentally different from the self-assembly of alkanethiolates on Au surfaces. While the conformational order within alkanethiolate monolayers on a Au surface is imparted by the van der Waals interactions from commensurately adsorbed nearest-neighbor adsorbate chains, highly organized silane monolayers result only when the native-oxide substrates are initially activated to allow adsorption of water layers. A Langmuir-type assembly at an air/water interface is proposed to explain the substrate independence with the water layer(s) acting to decouple any pinning lattice structure of the substrate from the alkyl chain structure (Parikh *et al.* 1994; Allara *et al.* 1995).

The lack of pinning lattice geometry restrictions is particularly important for forming conformationally ordered silane monolayers since substrate hydroxyl groups are randomly distributed across the surface and thus direct bonding of the alkylsiloxane groups to these surface groups would induce significant chain disorder relative to that observed for the monolayer chain assembly. The advantage of this substrate-independent adsorption mechanism is that alkylsiloxane films can be prepared on *any* native oxide surface including SiO<sub>2</sub>, TiO<sub>2</sub>, HfO<sub>2</sub>, aluminum oxide, glass, as well as many other surfaces, see for example (Ulman 1996).

Despite the chemical ubiquity of alkylsiloxane monolayers, the formation of reproducibly defect-free, densely packed and conformationally ordered alkylsiloxane films is far more challenging than for alkanethiolate monolayers, mainly due to the role of interfacial water and the role of temperature in the condensation mechanism and “completeness” (i.e. maximum packing density) of monolayer formation (Parikh *et al.* 1994, 1995). Forming reproducible monolayer structures of complete, densely packed alkylsilanes has been the foremost challenge in utilizing these monolayers in nanolithographic processes (Koide *et al.* 2000). Furthermore, upon formation of the thermally stable and chemically inert cross-linked Si–O–Si bonds, the monolayer structures are relatively static, limiting the extension of many of the self- and directed patterning methods developed for the comparably dynamic alkanethiolate monolayers. However, since there is significant interest in integrating molecular assembly nanolithography methods to current semiconductor device technologies, many efforts have been made to extend the patterning methods developed for alkanethiolates on Au surfaces to alkylsilanes on native oxide surfaces. In particular, microcontact printing and patterning with energetic beams have proved most successful; these applications of alkylsiloxane monolayers are reviewed in the subsequent sections.

## 23.2 Self- and directed patterning

As SAMs have demonstrated their utility across a wide range of scientific fields, it has only been a natural progression to increase the complexity of these thin organic films. As the mechanisms of their formation and their manipulation are increasingly understood, SAMs have begun to contain multiple adsorbates in pre-determined spatial distributions. This patterning has been engineered in a variety of ways, including self- and directed *coadsorption* of adsorbates by methods that allow spontaneous phase separation, insertion and exchange of multiple adsorbates within the monolayer, the selective *placement* of adsorbates by methods such as microcontact printing or dip-pen nanolithography, and by the selective *removal* of particular adsorbates via scanning probe lithography with a sharp tip (including mechanical scratching and indentation, local heating or electric field-enhanced oxidation) or via destruction with particle or photon beams whose incidence can be controlled. We review the implementation of each of the patterning methods and the monolayer chemistries with which they are compatible, beginning with self- and directed patterning.

In self- and directed patterning, different adsorbates are often separated into their individual domains once organized upon a surface, either by spontaneous assembly governed by thermodynamics and kinetics at the local scale, by their direct placement in a stepwise fashion, or by post-adsorption processing of the individual components. Although these processes of phase separation, insertion and exchange of multiple adsorbates require the formation of a kinetically stable monolayer structure, they also require the ability to form a dynamic equilibrium between the adsorbed monolayer structure and the adsorbate molecules in solution. As such, the applicability of self- and directed patterning methods has mainly been limited to alkanethiolate SAMs formed on Au surfaces from the solution phase. The following sections review the efforts to pattern alkanethiolate molecules on a surface starting from solution-phase coadsorption, from which it can be determined how molecules interact with both similar and dissimilar adsorbates on the surface at the local level.

### 23.2.1 Phase separation in multicomponent assemblies

When two different adsorbates are mixed in solution and then exposed to a substrate, both species will adsorb onto the surface. If the species are sufficiently different in molecular composition, they will form homogeneous domains so as to maximize self–self interactions, whether through van der Waals or hydrogen bonding within the chain or at the film interface. Many complex factors arise when attempting to pattern molecules upon a surface by coadsorption from the solution phase. It is important to note that the relative fractional surface coverage of the molecules is not necessarily that of the coadsorption solution; for example, an equimolar ratio of adsorbates *A* and *B* will not necessarily afford a surface composition of 1:1 *A:B*, an observation supported by contact angle goniometry, scanning probe microscopy and electrochemical studies (Bain and Whitesides 1988; Bain *et al.* 1989a; Folkers *et al.* 1992a; Hobara *et al.* 1999; Smith *et al.* 2001). Factors that affect the competition for binding to the surface include the relative solvation of the adsorbates, the sticking probability of each molecule, and the degree of interaction between the molecules once they are adsorbed. Well-solvated molecules may not adsorb as quickly to the surface, while comparatively less soluble molecules may either aggregate in solution or aggregate together on the surface. The enthalpies of these intermolecular interactions as well as the minimization of poor cross-interactions provide strong driving forces for phase separation.

Local probes such as the STM that are capable of imaging surfaces in real space have greatly assisted in understanding how molecules are spatially distributed on a surface, and so it is possible to understand how to manipulate once the component adsorbates are in place. From information at the nanoscale that has been gained using STM as an analytical tool, more complicated nanostructures have been created. Numerous spontaneously phase-separating SAM systems have been studied, using both local probes as well as ensemble measurements. Such systems include the coadsorption of short- and long-chain alkanethiols (Folkers *et al.* 1992a, 1992b, 1994), molecules that differ both in chain length and functional groups (i.e. 3-mercaptopropanol and

*n*-tetradecanethiol (Hobara and Kakiuchi 2001), molecules of similar length but with differing terminal groups [i.e. *n*-hexadecanethiol and its methyl ester analog (Stranick *et al.* 1994), *n*-undecanethiol and 11-mercaptopentadecanoic acid (Bain and Whitesides 1989)] and molecules of similar length but with differing, buried functional groups [i.e. 3-mercaptop-*N*-nonylpropionamide and *n*-decanethiol (Lewis *et al.* 2001; Smith *et al.* 2001)]. However, molecules of similar enough composition will not phase separate if formed at room temperature [*n*-decanethiol and *n*-dodecanethiol (Bumm *et al.* 1999; Anderson *et al.* 2002)]. This is primarily due to the fact that once the molecules randomly adsorb to the surface, they cannot readily diffuse to find similar adsorbates to form a separate phase. The likelihood of preferential exchange is decreased, as the energies of solvation and exchange will be similar between the two adsorbates.

Some of the first reports of phase separation in SAMs were formed from the coadsorption of *n*-alkanethiols that differed only in the alkyl chain lengths (Laibinis *et al.* 1991a, 1992; Folkers *et al.* 1992b). Using ensemble measurements and modelling of the adsorption events and intermolecular interactions, it was inferred that single-component SAMs were the lowest-energy structure of a SAM exposed to a solution of two adsorbates. Scanning probe microscopy data, however, has shown the presence of nanoscale domains of adsorbates separated into two components (Stranick *et al.* 1994). Both local probe and ensemble-averaging measurements agree that the adsorbate domains that exist on the surface tend towards homogeneity. The experiments of Whitesides and coworkers provided evidence for the fact that the ratio of adsorbates in solution does not mirror that of the surface-bound adsorbate composition. Scanning tunnelling microscopy indicated that these multicomponent films do not typically reach equilibrium (Stranick *et al.* 1994).

Separation of two adsorbates into their homogeneous domains on a surface has been observed when the molecules' tail groups are different in polarity. Methyl-terminated alkanethiolates have been shown to be essentially miscible with –OH-terminated alkanethiolates when the chain lengths are similar (Laibinis *et al.* 1991a). Many of the early phase separation studies by Whitesides and coworkers were performed using –OH-terminated and –CH<sub>3</sub>-terminated adsorbates in order to probe molecular cross-interactions and to analyze how each adsorbate responded to hydrophilic and hydrophobic contact-angle measurements (the terminal groups extending into or retracting from the probing solution). STM studies have shown that coadsorbed SAMs of *n*-hexadecanethiol and the corresponding methyl ester will phase separate into homogeneous, nanoscale domains (Stranick *et al.* 1994).

Phase separation has been reported for molecules possessing different functional groups that are buried near the film/metal interface. Phase-separated SAMs were reported for a mixture of *n*-alkanethiolate adsorbates and an adsorbate containing an amide group buried near the sulfur headgroup (Lewis *et al.* 2001; Smith *et al.* 2001). Surface-bonded amide-containing molecules are strongly hydrogen bonded to their nearest neighbors and provide a high enthalpic driving force for phase separation. The increasing interaction energies with molecules containing multiple amide groups further assists in imparting order to the assembly, and the directionality of the hydrogen bond assists

in aligning molecules (Clegg and Hutchison 1999; Clegg *et al.* 1999). These hydrogen bonds can assist in forming the sharpest of boundaries between domains, those that are only one molecule wide. This propensity for aggregation may be useful in exploiting intermolecular interactions for the precise placement of molecules in films.

### 23.2.2 Insertion and exchange in self-assembled monolayers

In SAM formation, equilibrium processes that are kinetically limited exist between species bound to the surface and free adsorbates. In the solution phase additional quantities such as the enthalpy and entropy of solvation play critical roles in the adsorption process. Substrate vacancy islands, adsorbate vacancies in the SAM lattice, and domain boundaries are critical sites in the monolayer that will allow for post-adsorption processing of the film. The degree of disorder at these sites is increased relative to the rest of the closely packed monolayer, and thus molecules adsorbed at these defect sites are postulated to be less constrained than the surrounding matrix and may have greater conformational mobility. Solvent molecules also have increased access to these defect sites, and their presence may promote the exchange of adsorbed molecules for new species. Step edges are also accessible to the solvent and are sites of high probability for adsorption and desorption events.

These defect sites within the “host” SAM are postulated to be the most susceptible to exchange to new “guest” adsorbates exposed to the films. The overall quality of the film allows the possibility for its further patterning, manipulation and postadsorption processing. Infrared spectroscopy studies have shown that multiple exchange events occur, with first a rapid exchange occurring at defect sites such as at grain boundaries, domain boundaries, or at the peripheries of substrate vacancy islands and then with a second, slow exchange, presumably occurring within the domains themselves (Dubois and Nuzzo 1992). STM has been used to examine at the nanoscale the insertion and assembly of molecules to be inserted into a surrounding, pre-formed alkanethiolate matrix (Bumm *et al.* 1996; Donhauser *et al.* 2001) Molecules can be brought into a well-defined SAM matrix in extremely low concentrations (typically 0.1–0.5 mM), using short exposure times (minutes to hours).

The fractional surface coverage of molecules inserted into host SAMs is not simply a function of the concentration of guest molecule exposed to the host SAM nor is it a function of exposure time. The defect density of the host SAM is of significant importance, for guest molecules tend to insert into the host SAMs at defect sites. It would then be expected that a greater defect density within the host matrix would lead to increased fractional surface coverage of guest molecules. Several studies have been performed to manipulate a film’s defect density, including (a) increasing the defect density by adsorbing the SAM for a short period of time, thus limiting slower adsorption processes, and (b) decreasing the defect density by backfilling the host matrix with adsorbates from the gas phase (Donhauser *et al.* 2002, 2003). When the defect density of a SAM is high (i.e. SAMs formed by exposure of the gold to *n*-alkanethiolate

for five minutes), molecules will insert at higher fractional surface coverage. Conversely, if guest adsorbates are inserted into the SAM and then the SAM is backfilled with additional adsorbate from the gas phase, the fractional surface coverage decreases. STM images show that the molecules inserted from the gas phase once the host–guest assembly is formed insert in the same locations as the guest molecules (step edges, domain boundaries, etc.), and provide a reinforced network of molecules surrounding the guest (Donhauser *et al.* 2003). The data also show that the molecules inserted from the vapor phase are phase separated from the host matrix.

### 23.3 Patterning via external tools

Spontaneous phase separation and insertion strategies have been useful in creating multifunctional surfaces, in particular allowing an understanding of how intermolecular interactions and film and substrate defects play an integral role in film formation and processing. However, these strategies are ones of negative control. A more active stance in the formation and patterning of SAMs has been undertaken in order to explicitly direct the placement of adsorbates to increase the utility of these assemblies. The following sections will discuss various methods of directly patterning SAMs on surfaces, or how the patterning of SAMs has been used to create patterns on the substrates underneath them.

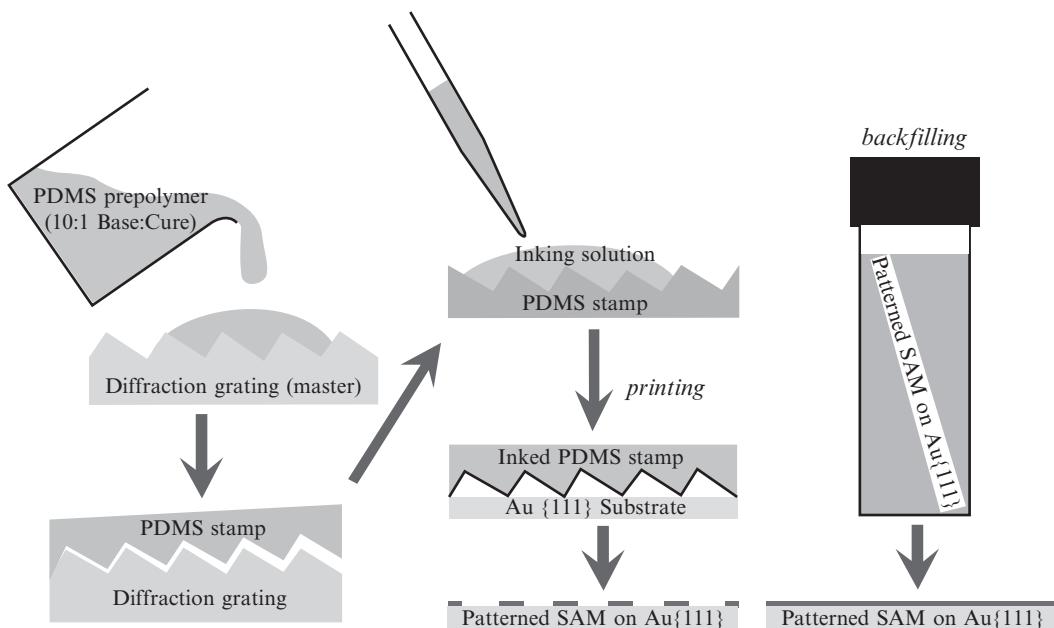
SAMs protect the surfaces underneath them, but when the alkanethiol has a reactive terminal group (i.e.  $-SH$ ,  $-NH_2$ ,  $-COOH$ ), the surface is poised to do chemistry, and a new “surface” emerges. This reactive interface can serve as a nucleation site for metal films deposited by physical or chemical methods, and as a tethering site for objects as diverse as proteins, cells, and nanoparticles. Large efforts have been undertaken to create patterns of molecules on surfaces, giving the films a spatial complexity and a patterned reactivity.

#### 23.3.1 Soft-lithographic contact patterning

In addition to forming SAMs by adsorbing molecules onto a surface from the solution phase, adsorbates can also be placed directly on a surface, both in homogeneous films as well as in patterns, in a technique referred to as soft lithography. Soft lithography describes how soft materials such as flexible, elastomeric polymers are used as the primary means of transferring and fabricating features into and onto substrates without the use of damaging energetic beams, such as photons, ions, or electrons. Planar surface contact patterning strategies, such as microcontact printing and displacement printing, are some of the most routine methods used for patterning molecular assemblies on solid surfaces.

##### 23.3.1.1 Molecular transfer on non-impressionable surfaces: Microcontact printing

“Microcontact printing” ( $\mu$ CP), part of a family of soft lithography techniques pioneered by Whitesides and coworkers in the mid-1990s (Kumar *et al.* 1992; Xia and Whitesides 1998), has become one of the most popular techniques of



**Fig. 23.3** Schematic of the microcontact printing process.

transferring molecules to surfaces. The chemical systems used in microcontact printing processes have included printing alkanethiols to metals such as Au, Ag, Cu, and Pd, as well as printing alkylsilanes to SiO<sub>2</sub> substrates (Wilbur *et al.* 1994; Xia *et al.* 1995, 1996a,b; Tien *et al.* 1998; Love *et al.* 2002). Akin to the concept of rubber stamping, microcontact printing transfers adsorbates (the molecules are the “ink”) at the points where the stamp contacts the substrate, Fig. 23.3. A “stamp” is formed by curing polydimethylsiloxane (PDMS) prepolymer against a “master” object that has a relief pattern in it (i.e. a patterned SiO<sub>2</sub> wafer or a diffraction grating). After curing, the PDMS stamp is peeled away from the master. A solution of molecules to be transferred to the substrate is directly applied to the stamp, and the solvent is removed. The stamp is then gently laid across the surface, and molecular transfer occurs—ideally—at the point of conformal contact, affording a substrate where alkanethiols only assemble at pre-determined spatial locations, and bare surface is exposed in the regions between.

SAMs formed in this way have often been used as ultrathin etch resists. These SAM-covered regions protect the surface against etching and oxidizing solutions; the first reports demonstrated that alkanethiols printed on Au substrates selectively protected the surface against etching to alkaline CN<sup>-</sup>/O<sub>2</sub> solutions (Kumar *et al.* 1992; Kumar and Whitesides 1993). Trenches on unprotected regions of the Au were formed, while the SAM-covered regions experienced comparatively little corrosion. A variety of etching conditions were subsequently explored and reported (Kumar *et al.* 1995).

Whitesides and coworkers developed the technique of topographically directed etching (TODE), in which they capitalized upon the varying degrees

of order in SAMs formed on surfaces with drastic changes in surface topography in order to pattern surfaces. This was put to use to create enhanced etch resists or supports for electroless metallization (Aizenberg *et al.* 1998; Black *et al.* 1999). Generally, a metal is evaporated onto a different metal substrate through a photoresist mask and is exposed to a solution of alkanethiols. The SAM that forms is less ordered at the transition region between the two metals and as such, is more labile to exchange with other alkanethiol molecules in solution, directed crystal nucleation, or etching processes. TODE has successfully etched  $\sim 100$  nm trenches on various substrate systems, including Ag/Ag (top metal/bottom metal), Au/Ag, SiO<sub>2</sub>/Si, and Al<sub>2</sub>O<sub>3</sub>/Al, as well as non-planar Ag surfaces. An additional advantage of this technique is that it is capable of forming both raised and entrenched features. Microcontact printing has been used to etch Au/Ti layers atop GaAs-based materials; layers of Ti and Au were evaporated atop GaAs/AlGaAs multiquantum-well structures, and then were selectively etched away using  $\mu$ CP-printed SAMs to protect particular areas of the surface (Kim *et al.* 1996). The exposed GaAs was subsequently etched away, transferring the pattern.

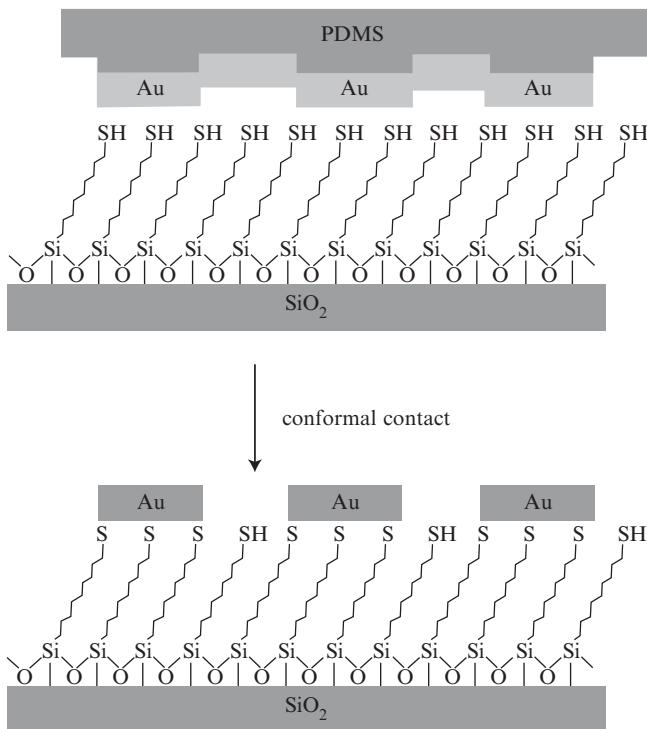
More often than not, the bare area of substrate that remains after the first stamping process is passivated by “backfilling” with a second alkanethiol, Fig. 23.3. The remaining Au is exposed to a second thiol, typically from the solution phase. The presence of these new thiols prevents the inevitable diffusion of the stamped thiol into the bare regions, that diffusion effectively erasing that any pattern that was created. However, an alkanethiol can be selected that has a terminal group that is different from its neighbor, which gives new functionality to the film.

The process of nanotransfer printing (nTP) to pattern features on surfaces is an elegant complement to  $\mu$ CP. Whereas  $\mu$ CP aims to print SAMs on surfaces where the configured elements are the alkanethiolate molecules attached to the gold substrate, the concept of nTP is to print metal features atop a SAM-functionalized substrate in a non-destructive fashion, where the SAM serves as an adhesion layer between substrate and patterned metal film (Loo *et al.* 2002a,b; Zaumseil *et al.* 2003), Fig. 23.4.

Nanotransfer printing aims to increase the edge resolution of patterned metal features further. Features printed by  $\mu$ CP are limited in resolution not only by the smearing of edges as a result of the deformation of an elastomeric stamp, but also by the diffusion (albeit limited) of alkanethiolate adsorbates. Nanotransfer printing experiments have been performed using a variety of materials for substrates and stamps; well-defined surface chemistry is required in order to build up to the complex-patterned features. The edge resolution of the patterned features is between 5 and 15 nm, which is comparable to the edge resolution of the PDMS stamp itself as well as the grain size of the evaporated Au metal.

### 23.3.1.2 Displacement patterning

A significant complication of the  $\mu$ CP process is that the adsorbates diffuse around the area of transfer, blurring the edge resolution of the group of molecules that was stamped. In particular, short-chain alkanethiols with sufficiently high vapor pressures such as *n*-dodecanethiol cannot be stamped at

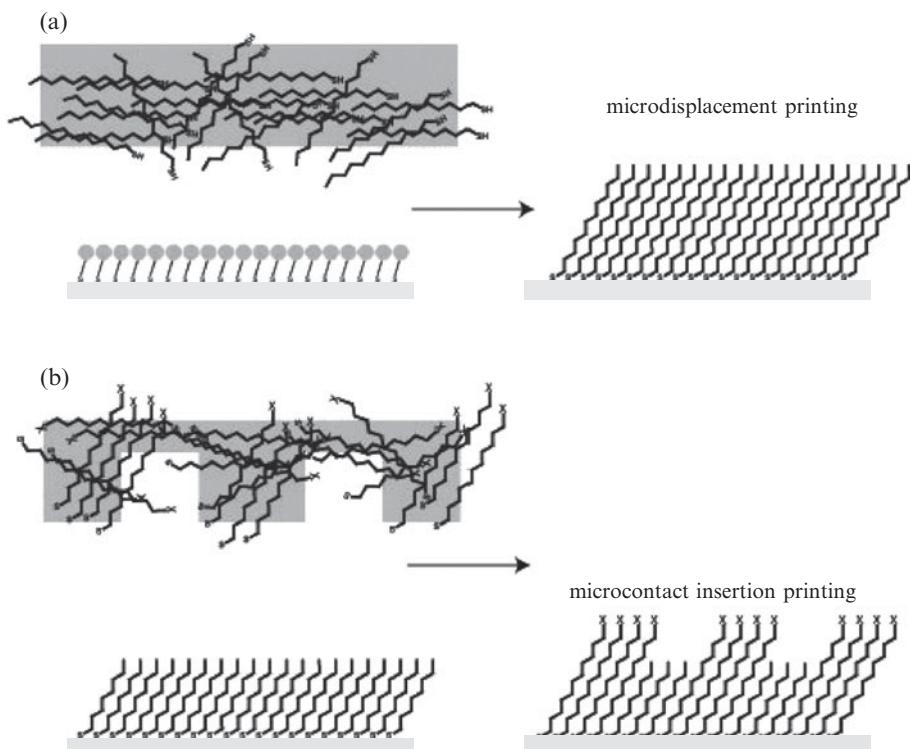


**Fig. 23.4** Schematic of the nanotransfer printing process. Here, gold that has been thermally evaporated on a piece of PDMS is conformally transferred to a SAM with reactive terminal groups.

all, as they diffuse quickly across the surface to bind, forming a homogenous SAM (Delamarche *et al.* 1998). Weiss and coworkers have shown that one method of circumventing the adventitious diffusion of thiols is to stamp them on top of a pre-existing SAM, rather than on a bare surface alone (Dameron *et al.* 2005; Mullen *et al.* 2007), Fig. 23.5. The presence of a SAM significantly impedes the lateral surface diffusion of the molecules to be patterned, and permits the use of molecules that are otherwise too mobile to pattern by other methods. This displacement process was first demonstrated by using SAMs of 1-adamantanethiolate as a “placeholder” for incident molecules, and is termed “microdisplacement printing.”

Adamantanethiolate on gold has an intermolecular spacing that is strained and expanded relative to the typical 5 Å spacing for *n*-alkanethiolates. Therefore, it is displaced with relative ease and locally at the points of contact, where the concentration of thiols to be printed is at its highest. The extent of displacement can be controlled by tuning the stamping duration and ink concentration during sample fabrication.

The phenomenon of full displacement of stamped *n*-alkanethiols on top of an adamantanethiolate SAM is shown in Fig. 23.5(a); (Dameron *et al.* 2005) the stamping of patterns of thiols atop adamantanethiolate SAMs is called microcontact insertion printing ( $\mu$ CIP) and is shown in Fig. 23.5(b) (Mullen *et al.* 2007). Microcontact insertion printing allows for increased engineering of phase separation of mixtures of adsorbates on a surface; it was shown that patterns of *n*-octanethiolate and 11-mercaptopoundecanoic acid could be formed by using  $\mu$ CIP (Mullen *et al.* 2007). Even though these molecules



**Fig. 23.5** (a) Schematic of the microdisplacement printing process, where molecules physisorbed to the PDMS stamp displace a pre-formed adamantanethiolate SAM. (b) Schematic of microcontact insertion printing, where patterns of molecules are delivered atop a pre-formed SAM.

phase separate on a surface when deposited from the solution phase based on differences in their chain length and their terminal group,  $\mu$ CIP imposes a spatial constraint on the phase separation, as the stamp determines the location of the insertion.

### 23.3.2 Scanning probe lithography

The high spatial resolution of scanning probe microscopes has led to their use as “direct-write” lithographic tools by creating patterns in self-assembled films at the nanometer-scale. This process, termed scanning probe lithography (SPL), has been the subject of several excellent reviews (Krämer *et al.* 2003; Garcia *et al.* 2006; Xie *et al.* 2006; Woodson and Liu 2007). The atomic force microscope and the scanning tunnelling microscope, tools with the ability to move at the nanometer scale in the  $x$ -,  $y$ -, and  $z$ -directions, have been able to pattern both molecules and substrates directly at these scales. Scanning probe lithography operates in several regimes: (a) where the patterning is additive and molecules are directly written onto substrates or reacted upon molecular interfaces; (b) where the patterning is subtractive, and molecules are directly removed from a substrate; and (c) where the probe tip is electroactive, either electrochemically desorbing molecules from substrates or directly modifying the substrate itself. These tools are particularly attractive as they can operate

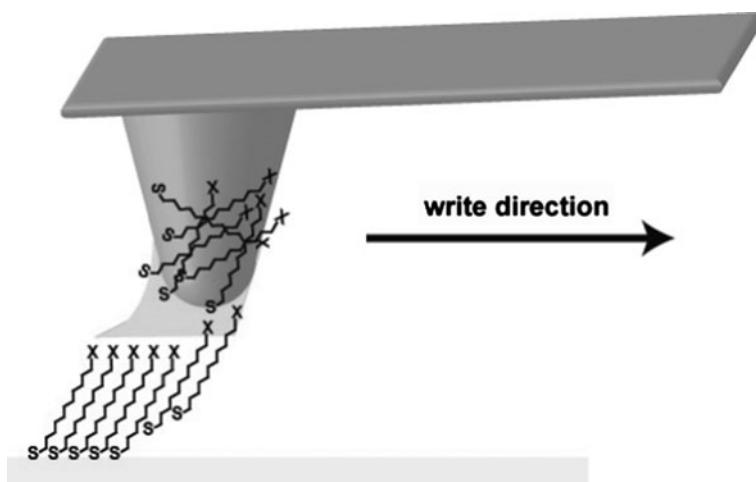
in a perturbative, lithographic mode as well a non-perturbative, imaging mode; the same tool can be used to characterize the very features it creates.

### 23.3.2.1 Dip-pen nanolithography

Dip-pen nanolithography (DPN) is a variety of scanning probe lithography developed by Mirkin and coworkers, where molecules to be “written” on a surface are transferred by means of an AFM tip to a substrate (Piner *et al.* 1999). As the scanning probe tip can be positioned or programmed to move in certain patterns, it is possible to use DPN to pattern SAMs with different molecular components; these patterns can then be used for the further design of nanoscale structures, by either selective molecular reaction or deposition in the patterned areas. The DPN printing mechanism is universal to a variety of molecules and substrates; thorough study and understanding of the transfer process has led to the DPN-printing of several molecules and novel materials, from alkanethiols to oligonucleotides to sol–gel precursors that afford mesoporous films upon processing (Su *et al.* 2002).

Typically, *n*-alkanethiols or their derivatives are suspended in droplets at the end of an AFM tip as a molecular ink. By rastering the probe tip close to a gold (or other metal) surface, the alkanethiol molecules are transported to the surface through a water meniscus that naturally occurs between the tip and sample in ambient conditions, Fig. 23.6.

An array of molecules is deposited that is a direct function of the rastering pattern and dwell time of the AFM tip over the surface. The S–Au bond affixes the molecules to the substrate, and thus their diffusion is limited by the enthalpic forces holding the assembly together (van der Waals between the chains, known as autophobic pinning (Biebuyck and Whitesides 1994)). The resolution of SAMs patterned by DPN is of the order of linewidths of ~15 nm (Mirkin *et al.* 2001), a result that is a complicated mixture of factors including relative humidity, scan speed, and the relative solubility of the molecule in the water meniscus; different molecules will have different transport rates through this meniscus. It is then possible to backfill the bare regions with other thiol-functionalized molecules simply by exposing the patterned substrate to



**Fig. 23.6** Schematic of “dip-pen nanolithography,” where a reservoir of molecules is applied to the tip of an atomic force microscope; the tip is brought close to the surface, and the molecules are transported to and bind to the surface through the presence of a water meniscus. The rastering capabilities of the AFM allow for patterns to be “written.”

a thiol solution, and these adsorbates backfill the bare regions. Lateral force microscopy, in which the AFM measures the relative frictional properties of adsorbates, is used to image the molecules patterned on the surface (Piner *et al.* 1999).

Dip-pen nanolithography has been developed for numerous applications, and with multiple substrate materials. For example, DPN has been used to pattern thiols on gold substrates where the partial SAMs were utilized as etch resists (Weinberger *et al.* 2000). Gold films were evaporated atop silicon substrates, and thiols were printed on the substrates by DPN. The assembly was exposed to wet-chemical etches to remove the Au and Ti adhesion layer, along with the underlying Si substrate; three-dimensional nanostructures with anisotropic features as well as those with isotropic features (pillars) were produced by this technique. While most SAM resists printed by DPN and microcontact printing are negative in nature, a positive resist has been recently reported (Wei and Ginger 2007). Octadecylamine (ODA) was printed onto a gold substrate by DPN, and the unpatterned areas were passivated by octadecanethiol. Upon exposure to an etching solution, the areas coated by ODA were preferentially etched due to the fact that the interaction of the amine headgroups with the surface is much weaker than the Au–S interaction. An area of depressions in the shape of the initial ODA patterns remained in the substrate after etching.

The major limitation of these “direct-write” methods is that they are serial in nature. Here, throughput is sacrificed for spatial resolution. One method of increasing the throughput of this type of patterning has been to interface it with computer-assisted design programs (Cruchon-Dupeyrat *et al.* 2001). A method designed to make this process more parallel has been to use microfabricated arrays of AFM tips. Nominally, each AFM tip is individually addressable and can have its own unique “ink,” thus giving the surface multidimensional functionality. Early versions of tip arrays included the “Millipede,” a two-dimensional,  $32 \times 32$  (1024) AFM cantilever array that was used to indent 30–40 nm features (Vettiger *et al.* 2000). Most recently, arrays of 55 000 tips have been reported (Salaita *et al.* 2006).

### 23.3.2.2 Scanning probe embossing: Nanografting

DPN is an excellent example of an additive form of SPL. Another common application of SPL is “nanografting,” which is a subtractive form of SPL. Here, an AFM tip is used to directly carve out selected molecules of an alkanethiolate SAM by applying a force sufficient to scrape the tip along the surface and remove the adsorbates (Liu *et al.* 2000). This method has been applied to SAMs of alkanethiolates on gold substrates (Liu *et al.* 2002; Yu *et al.* 2006; Liang and Scoles 2007), for alkylsilanes on  $\text{SiO}_2$  (Headrick *et al.* 2005; Lee *et al.* 2006), and for molecules grafted to H-terminated Si surfaces (Lee *et al.* 2006), among others. It is also possible to passivate the areas where the adsorbates have been removed by exposing the surface to a second alkanethiol, and it will fill in the bare areas left in the path of the tip (Amro *et al.* 2000). As the probe tip carves away a selected area of the monolayer, part of the surface is exposed, and can be selectively etched as the remaining SAM serves as a protective resist to the surface. Careful studies have been performed to

optimize the reproducibility and precision of features printed by these methods (Ngunjiri *et al.* 2008). This method has also been used to create patterns into biologically functional substrates; for example, an AFM probe tip was used to carve away patterns in a monolayer of bovine serum albumin protein coated on a borosilicate glass substrate. Supported lipid bilayers were formed in the trenches, which were 50–100 nm in width (Shi *et al.* 2008).

Scanning probe lithography is not merely specific to systems of *n*-alkanethiolate SAMs on gold. Sugimura and coworkers have performed similar experiments for organosilane SAMs assembled on silicon surfaces as resists for SPL (Sugimura *et al.* 1996, 1999a). Using an AFM with a conductive probe tip, a series of lines was patterned into the SAMs, and then the substrates were exposed to an ammonium fluoride and peroxide etch, which isotropically etched into the silicon underneath the SAM that had been electrochemically removed. Additionally, they demonstrated that upon exposing the patterned and etched Si substrate to a gold-plating solution, Au could be deposited selectively in the regions where the organosilane SAM was removed, and the rest of the SAM-covered substrate remained featureless and unperturbed (Sugimura *et al.* 1999b). Upon electrochemical oxidation of the silicon substrate with the tip, they were able to backfill the regions of destroyed SAM with an organosilane of differing chemical functionality (removing a methyl-terminated silane and replacing it with an amine-terminated silane) (Sugimura and Nakagiri 1997).

Scanning tunnelling microscopes have also been used to electrochemically desorb adsorbates from surfaces, from Au to GaAs. The STM delivers a local and controlled dose of low-energy electrons to the substrate, and the energy and number of electrons delivered to the substrate can be controlled by both the applied bias voltage and the magnitude of the tunnelling current. One of the first demonstrations of this technique was the desorption of SAMs of octadecanethiolate on Au and GaAs (Lercel *et al.* 1994). Patterns generated in this way were faithfully reproduced with wet-etching techniques. This patterning method can be used under both ultrahigh-vacuum and ambient conditions. Heinzmann and coworkers demonstrated that both alkyl- and arylthiolate SAMs on Au, as well as alkylsilane films on the Si(100) surface, could be patterned using the tunnelling electrons from a UHV-STM, and the features could be transferred via wet-etching techniques (Kleineberg *et al.* 2001). They also demonstrated that the STM could be used to write successfully into a PMMA resist down to linewidths of 75 nm, and that the pattern could be transferred to the Mo/Si substrate beneath it via reactive ion etching (Hartwich *et al.* 1998).

### 23.3.3 Patterning with energetic electron beams

While soft lithography has moved the field of patterning SAMs forward at an enormous rate, the field of using energetic beams to pattern SAMs has also advanced considerably. Energetic beams such as ions, electrons, and photons are routinely used to destroy films of polymer or inorganic resists. However, the effectiveness of these resists is limited by their thickness. Ultrathin organic

films such as SAMs are attractive alternatives to conventional resists. For instance, inorganic resists require high electron-beam dosages, as the mechanism for their destruction is based on damage to the material by the electron beam. In addition, high-energy electrons are also needed to permeate thick layers of material, yet they can cause a cascade of backscattered electrons that damage the surrounding areas, hence broadening the features (Whelan *et al.* 1996).

Ultrathin film resists comprised of SAMs, both of thiol-containing molecules as well as organosilanes on silicon surfaces, have assisted in advancing conventional lithography by serving as sacrificial resists (i.e. chemically modifying the SAM so that different regions will vary in their susceptibility to attack by wet etches) (Calvert 1993; Dressick and Calvert 1993). Despite the thickness advantage, there still remain many challenges to using SAMs as resists, including limitations such as homogeneous (i.e. non-isotropic) etching at the nanometer scale. The following sections summarize how SAMs have been applied as resists to energetic beams comprised of electrons and photons.

### 23.3.3.1 *Electron-beam lithography*

The earliest work demonstrating that SAMs could be used as patterning resists involved the use of electron beams “writing” into SAMs of alkanethiolates on GaAs (Lercel *et al.* 1993; Tiberio *et al.* 1993), and SAMs of octadecyltrichlorosilane (OTS) on SiO<sub>2</sub> (Carr *et al.* 1997). In all examples, the SAMs serve as positive, self-developing resists, as they degrade upon exposure to the electron beam; wet-chemical etching in the exposed regions leads to features entrenched in the substrate. AFM imaging confirmed the degradation of the SAMs upon exposure to the electron beam and before chemical etching, indicating their exposure (Lercel *et al.* 1993).

The pattern transferred to the SAM is a function of the mask that is placed in the path of the beam and the sample. Grunze and coworkers have adsorbed 4'-nitro-1, 1'-biphenyl-4-thiol (NBT) on a gold surface and then irradiated the sample with 50-eV electrons through a copper transmission electron microscope grid (a pattern of squares, 20 μm per side) (Eck *et al.* 2000). Initially, the NBT molecules form an ordered SAM as shown by IR, XPS, and NEXAFS (Geyer *et al.* 1999)—exposed nitro groups are reduced to amino groups, and the chemical bonds of the underlying aromatic layer break and cross-link to those of nearest-neighboring molecules (Eck *et al.* 2000). A non-irradiated SAM was exposed to *n*-dodecanethiol, which completely displaced all of the NBT. After the SAM was exposed to radiation, the alkanethiol was not capable of displacing the NBT molecules. In addition, the irradiated SAM was resistant to KCN etching solution (a known etchant of gold), making the irradiated SAM serve as a negative resist. It was further demonstrated that the chemical species of the irradiated monolayer were amine groups by reacting the pendant terminal with trifluoroacetic acid anhydride, rendering the -NH<sub>2</sub> groups -CF<sub>3</sub> terminated, as measured with lateral force microscopy (LFM) (Eck *et al.* 2000). Geyer *et al.* also immobilized rhodamine dyes to the terminal amine groups, and with laser-scanning confocal fluorescence microscopy and LFM demonstrated the presence of amine groups (Geyer *et al.* 2001).

This ultrathin resist system is successful both because of the thinness of the resist and the low mean-free path of the low-energy electrons. Even at high doses of  $\sim 50$  eV electrons, the SAM is not destroyed. The irradiation of the NBT molecule gave a minimum feature linewidth of 20 nm (Gölzhäuser *et al.* 2000, 2001; Geyer *et al.* 2001).

Gold nanostructures have been fabricated using positive- and negative-tone resists (*n*-alkanethiolate and arylthiolate SAMs, respectively) (Weimann *et al.* 2001). Weimann *et al.* claim that the resolution of electron-beam lithography is limited to the size of the molecules in the resist. The authors used aliphatic SAMs as positive-tone resists, as the radiation beam induces damage and disorder of the chains (Au–S bond cleavage, irregular cross-linking, etc.); they used arylthiolate SAMs as negative-tone resists, as the organic layer is strengthened by cross-linking rigid aromatic units (similar to the work by C–H bond cleavage by radiation, followed by cross-linking). Using a scanning transmission electron microscope as an ultrasharp source of electrons, Weimann *et al.* have achieved features that have minimum linewidths of 20 nm (for both types of SAMs). Pattern transfer and anisotropic etching are the limitations in these studies; electron-beam energies used were 200 keV and 2.5 keV, depending on the substrate (Weimann *et al.* 2001).

Grunze and coworkers have also performed studies on the irradiation of *n*-alkanethiolate SAMs for use as positive-tone resists; they examined the fundamental chemistries that occur to the SAM as a function of beam irradiation in order to improve its use as a resist (Zharnikov *et al.* 2000). The most noticeable processes that occur with damage are the loss of the SAM's orientational and conformational order, partial dehydrogenation with C=C double-bond formation, desorption of the layer fragments resulting in reduced film thicknesses, and reduction of the thiolate species, as shown by the appearance of new sulfur species. The longer the length of the alkyl chain, the slower the desorption of the thiolates. Whereas the irradiation-induced processes in the alkyl matrix are found to be essentially independent of the alkyl chain length and the substrate material, the extent and rate of the thiolate species reduction and new sulfur species formation are mainly determined by the strength and character of the thiolate–substrate bond. Grunze and coworkers studied *n*-dodecanethiol, *n*-octadecanethiol (ODT), and perdeuterated eicosanethiol on Au, and ODT on Ag. The isotopic effect was shown to be minimal, barely slowing the rate of molecular conversion (Zharnikov *et al.* 2000). The authors note that the damage to SAMs caused by ionizing, energetic beams is related to low-energy secondary electrons arising as a result of the inelastic scattering of the primary electrons created within the photoemission process (Zharnikov *et al.* 1999). The damage to the film by the incident electron beam causes the desorption, breakdown, and chemical rearrangement of the molecules, yielding a film of new molecular composition.

Craighead and coworkers have performed low-energy electron-beam studies of NH<sub>2</sub>-terminated organosilane SAMs on SiO<sub>2</sub>, irradiating with < 5 keV energy so that the primary damage of the SAM occurs at the terminal amine group (Harnett *et al.* 2000). They found that the lower the energy of the electron beam, the higher the resolution of the feature; with the beam at low accelerating voltages, most of the energy is left in the top part of the SAM

and the terminus is selectively damaged. The goal of these studies was to produce patterned amine-functionalized SAMs where upon radiation with the electron beam, the amino groups were destroyed. One can then pattern the unexposed, remaining  $\text{--NH}_2$  groups. Craighead and coworkers demonstrated the selective adherence of palladium nanoparticles, aldehyde-functionalized polystyrene spheres, and NeutrAvidin-coated polystyrene spheres (Harnett *et al.* 2000). These authors demonstrated several schemes to pattern amine-terminated SAMs, i.e. irradiating a methyl-terminated SAM with low-energy electrons followed by backfilling the destroyed regions with cysteamine generated phase-separated SAMs, with the binary component SAM presenting both  $\text{CH}_3\text{--}$  and  $\text{NH}_2\text{--}$  termini (Harnett *et al.* 2001). They reacted the terminal amines in the phase-separated domains so as to present biotin moieties that would selectively react with NeutrAvidin-coated polystyrene spheres. By varying the electron dosages, molecular gradients could be generated. Bard *et al.* exposed alkanethiolate SAMs to metastable, excited noble-gas atoms (both helium and argon atoms) in order to damage the resist; after the damage, the gold substrates were exposed to wet-chemical etches and the patterns were examined by AFM and the reflectivity of the substrate was measured (Bard *et al.* 1997). This “neutral-atom lithography” is highly beneficial due to the fact that neutral atoms have an extremely short de Broglie wavelength ( $<0.1\text{ nm}$ ) unlike their photon counterparts, and they only interact with the outermost layer of a surface (concentrated damage with no further penetration), unlike other species, which can scatter and broaden a feature on contact with the surface.

Organosilane SAMs have also been patterned with electron beams; the silicon was then etched with HF, and nickel was electroplated atop the exposed silicon (nickel is an excellent mask for RIE processing) (Carr *et al.* 1997). Additionally, one can pattern an organosilane SAM with photons ( $\lambda = 157\text{ nm}$ ) via a photomask (Sugimura *et al.* 2001). Sugimura *et al.* demonstrated that they could electrolessly plate nickel in the exposed regions, followed by an additional etch with a plasma in the exposed regions. Low-energy electrons ( $\sim 500\text{ eV}$ ) have also been used to pattern trimethylsilyl SAMs. The irradiated regions were destroyed in striped patterns, and an amine-terminated silane was backfilled in those regions (Krupke *et al.* 2002). Bundles of single-wall carbon nanotubes could be selectively deposited in the positively charged striped patterns.

### 23.3.3.2 Optical lithography

Optical lithography remains the mainstay of the semiconductor device industry. In conventional lithography, light is passed through a mask to illuminate a light-sensitive polymer or photoresist. Upon illumination, the photochemistry in the photosensitive molecules in the resist is initiated, resulting in either cross-linking or photodegradation of the illuminated areas. The lithographed size features are ultimately dependent on the wavelength of light. Historically, the wavelengths of light used in photolithography have progressed from blue wavelengths (436 nm) to UV (365 nm) to deep-UV (248 nm) to today’s mainstream 193 nm process wavelength. Despite the fundamental optical resolution limits, the use of double or immersion lithography processes have allowed chip makers to achieve 45 nm and 32 nm process technology (Hand 2006). To

drastically extend these dimension limits, next-generation lithography methods that use smaller wavelengths of light (such as extreme UV, 13.5 nm) and new materials will have to be implemented. Molecular, self-assembled films used as a sacrificial layer are one example of a new material under investigation for use in developing high spatially resolved patterns with photons on surfaces. Two approaches to using photons to pattern molecular assemblies, photooxidation patterning with UV and X-ray wavelengths of light and near-field nanolithography are detailed below. By using a spectrum of wavelengths for lithography, patterns with micrometer (UV light) down to nanometer (near-field and X-ray) dimensions can be achieved.

#### 23.3.3.2.1 UV photopatterning nanolithography

One of the most straightforward methods to pattern molecular assemblies with photons is through exposure to UV light. The first report of patterning molecular assemblies with UV light was by Terlov and coworkers (Terlov *et al.* 1993) who used the observations by Huang and Hemminger that, upon UV irradiation in air, alkanethiolates on Au are photooxidized to their corresponding alkanesulfonates (Huang and Hemminger 1993), which can subsequently be displaced by a second thiol adsorbate upon immersion in solution (Terlov and Newman 1992). Molecular assemblies are now routinely patterned by shining UV light through a mask onto a pre-formed SAM, allowing for the formation of patterned regions of different chemical functionality across a surface, which can be used for further chemical tailoring. Although photooxidation patterning is limited to patterned dimensions of several micrometers (Cooper *et al.* 1997; Cooper and Leggett 1999), it has proved to be a versatile patterning technique, and has been extended to other surfaces functionalized with alkanethiolates including Ag (Terlov *et al.* 1993; Gillen *et al.* 1994; Hutt *et al.* 1998) and more recently, GaAs (Zhou and Walker 2007).

Even though micrometer-sized features can be routinely patterned into SAMs using photooxidation patterning, little is known regarding the mechanism by which the films are photooxidized. For example, it has been suggested that ozone generation from the 184 nm line from the high-pressure Hg lamp (one of the most common UV light sources used for photopatterning) plays a pivotal role in the photooxidation chemistry (Norrod and Rowlen 1998; Zhang *et al.* 1998), but other results utilizing only the 254 nm line have also been shown to be effective in deriving patterns in the SAM (Brewer *et al.* 2001). Furthermore, the photochemistry is complex; the oxidation of the SAM is strongly influenced by the chemistry of the terminal functional group, the chain length of the molecule, the nature of the light source, and substrate upon which the SAM is adsorbed (Hutt and Leggett 1996; Cooper and Leggett 1998; Hutt *et al.* 1998; Rieley *et al.* 1998; Brewer *et al.* 2005). For example, SAMs of dodecanethiolate molecules oxidize more rapidly compared to SAMs comprised of mercaptoundecanoic acid (Zhang *et al.* 1998). Regardless of the precise mechanism for SAM photooxidation, UV photopatterning is a reliable and routine method for forming patterns in molecular assemblies.

#### 23.3.3.2.2 Near-field nanolithography

Traditionally, the resolution of photolithography is the Rayleigh limit, however, if a detector is placed near the illuminated surface ( $\ll \lambda$ ) then the

resolution is no longer limited to the wavelength of light but to the aperture of the detector. Beating the Rayleigh limit would be an inherently important alternative to the limits of traditional photolithography limit, especially since changing the photolithographic wavelength requires the development of new, and often very expensive, optical elements. With the development of scanning near-field lithography (SNOM), Betzig *et al.* demonstrated the construction of 60 nm features using visible (488 and 514 nm) light (Betzig *et al.* 1991). Unfortunately, the SNOM lithographic method was plagued with inconsistent results and one reason suggested for the aberrant results was the inability to control the divergence of the illuminated electric field by controlling the thin-film thickness precisely (Leggett 2006). The inherent thickness of a SAM, defined by the molecular length, makes molecular assembly films ideal photoresist candidates for near-field nanolithography applications. By coupling the photooxidation patterning chemistry described in the previous section with SNOM, written linewidths of 40 nm are now routinely achieved with 244 nm light (Sun *et al.* 2002).

Near-field nanolithography has been shown to be compatible with many different SAM chemistries and substrates. Thus far, near-field nanolithography has been used for the thiols-on-gold system to pattern lines of  $\omega$ -terminated  $-COOH$  alkanethiolates in a  $CH_3$ -terminated alkanethiolate host matrix (Sun *et al.* 2002), SAMs of mercaptopropanoic acid (Sun and Leggett 2002) and oligo(ethylene glycol) SAMs (Montague *et al.* 2007). Near-field nanolithography has also been applied to H-terminated silicon surfaces, passivated by decene (Sun *et al.* 2005), used to initiate the conversion of chloromethylphenyltrichlorosilane assembled on silicon dioxide to hydroxymethylphenyltrichlorosilane molecules at low exposure and to induce desorption of the silane molecules at high exposure (Sun *et al.* 2006). Other groups have written features into other molecular assembly films including silanes self-assembled on glass (Kobayashi *et al.* 2008). Variations on this technique have included using femtosecond near-infrared light that was suitable for replacing a deep-UV laser source (Chang *et al.* 2005).

While writing individual features in a device with a nanometer-sized probe is impractical for industrial applications, the use of technologies such as the ‘Millipede’ (Binnig *et al.* 1998; Vettiger *et al.* 2000) could allow a realistic implementation of near-field nanolithography in industrial applications by increasing the “write” feature throughput.

### 23.3.3.2.3 X-ray nanolithography

Using X-ray wavelengths for lithography is, in effect, one limit of optical lithography. Their diffraction from their short wavelengths (1–20 nm for “soft” X-rays, < 1 nm for “hard” X-rays) would ideally allow patterning linewidths to be achieved at the molecular and even atomic scale. Historically, X-rays have been used as a diagnostic tool to probe the SAM composition (surface techniques including XPS, grazing-incidence X-ray diffraction (GIXRD), and near-edge X-ray adsorption fine structure (NEXAFS) have become routine analytical methods). However, several studies have reported that X-rays also induce damage and modification of the SAM structure (Laibinis *et al.* 1991b; Graham *et al.* 1993; Rieke *et al.* 1993; Frydman *et al.* 1997; Kim *et al.* 2000;

Zharnikov and Grunze 2002). This damage and modification of the SAM structure by X-ray irradiation can either limit the compositional analysis, or be used to the advantage of lithography, to pattern surfaces. Although there are many theories as to the mechanism of how X-rays induce damage to the SAM. It is generally thought that it is the primary and secondary electrons that are emitted from the surface, and not the incident X-ray beam itself that induces damage to the film. This was verified by Whitesides and his collaborators who found that the amount of X-ray-induced damage in fluorinated SAMs adsorbed on Si, Au, Al, Ti and Cu was dependent on the efficiency of the substrate to eject electrons (Laibinis *et al.* 1991b). Further investigations revealed that specific chemical functionalities within the SAM molecule were particularly susceptible to X-ray irradiation and underwent specific X-ray-induced chemical reactions. Halide, nitro and carboxylic functional groups in  $\omega$ -terminated SAMs were observed to cleave from the parent molecule, while  $\text{CH}_3$ -terminated SAMs were damaged via dehydration, cross-linking, and re-arrangement of the carbon chain (Suh *et al.* 1993; Moon *et al.* 1998; Dressick *et al.* 1999; Kim *et al.* 2000; La *et al.* 2002, 2003a,b, 2004; Moon *et al.* 2000; Perry *et al.* 2002; Mendes *et al.* 2003; Mendes and Preece 2004).

By taking advantage of the selective X-ray-induced chemistry, feature sizes of 70–80 nm have been patterned into SAMs utilizing masks with similar dimensions (La *et al.* 2003b, 2004). While the precise mechanisms of forming selective patterns in SAMs with X-ray beams are still not understood, the technique and applications of X-ray lithography are appealing since they are not limited to a specific substrate/SAM combination. With further development, X-rays could be used to achieve sub-100 nm feature sizes reproducibly.

## 23.4 Directed self-masking via selective deposition on chemical patterns

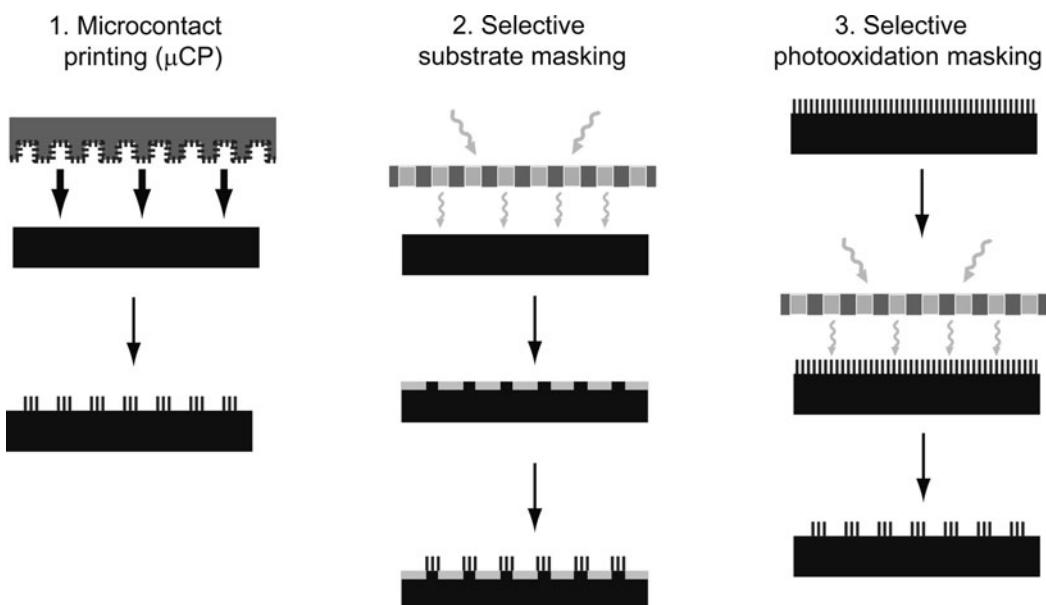
Directed growth of thin films on surfaces is one of the ultimate goals of nanolithography. The ability to control precisely the spatial location, in lateral and vertical thickness, and chemical composition of a thin film is one of the main challenges for surface scientists. This ability opens up the opportunity to tailor the surface chemistry such that complex two- and even three-dimensional nanometer-sized structures can be constructed.

There have been several approaches taken towards controlling and directing thin-film growth, each of which is highly specific to the chemical composition of the layer and include vapor-deposited or physical-vapor-deposited (PVD) metal films, chemical-vapor-deposited (CVD) inorganic films, atomic-layer-deposited (ALD) inorganic films, and electroless-deposited metal films. The first three methods, PVD, CVD and ALD, are *in-situ* vacuum techniques that use elemental metal vapors or volatile inorganic complexes to control and direct thin-film growth. In contrast, the fourth method, electroless metal deposition, is achieved with aqueous inorganic solutions.

While these methods for depositing thin films are neither novel nor mechanistically laterally constrained, the ability to couple these methods for thin-film

growth with chemically patterned molecular assemblies to selectively grow a thin film in a directed area of the surface is novel. In general, three methods have been used to form patterned molecular assemblies for directed thin-film growth. These include microcontact printing ( $\mu$ CP) where the molecules are stamped in pre-determined patterns, selective substrate masking where the underlying substrate is patterned such that only specific areas of the substrate are reactive to the organic molecule and selective photooxidation masking where the patterned areas of a pre-formed monolayer are removed via photooxidation through a mask, see Fig. 23.7. The choice of patterning method is made based on the organic molecule/substrate assembly combination needed.

These methods take advantage of combining chemically reactive organic molecules patterned in one surface location to nucleate thin growth and/or chemically inert organic molecules patterned in other surface locations to block thin-film growth. Through this approach, subsequent lithography steps such as etching away deposited material can be simplified or even eliminated. Furthermore, by using molecular assemblies that have a known surface density of uniformly organized organic groups at the surface, the fundamental reaction pathways of the often complex reaction pathways of the thin-film deposition method can be probed. The unifying challenge among all these approaches and methods of achieving selective-area thin-film growth, is minimizing damage to the organic molecules. The following sections review each of the four methods used for area-selective thin-film growth, highlighting their advantages and



**Fig. 23.7** Three methods that have been used to form patterned molecular assemblies for directed thin-film growth: 1. Microcontact printing ( $\mu$ CP) where the molecules are stamped in pre-determined patterns, 2. Selective substrate masking where the underlying substrate is patterned such that only specific areas of the substrate are reactive to the organic molecule and 3. Selective photooxidation masking, where the patterned areas of a pre-formed monolayer are removed via photooxidation through a mask.

drawbacks in application to the chemically patterned molecular assemblies discussed in the previous sections.

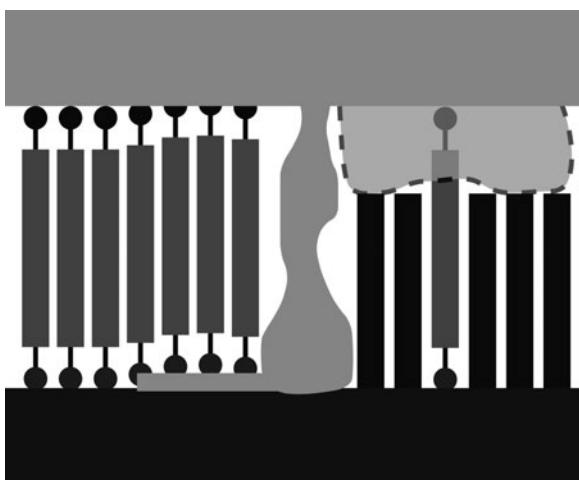
### 23.4.1 Metal-vapor deposition

The formation of a thin film via condensation of metal vapor atoms onto organic assemblies has long been of scientific and technological interest, especially since the growing prevalence of metal/organic interfaces has increasingly become a critical component in microelectronic fabrication and electrical contacts (Swalen *et al.* 1987; Xia *et al.* 1999; Gates *et al.* 2005). This field originated in the metallization of polymer films (Mittal 2008) but more recently the fundamental interactions of metal vapor atoms with self-assembled monolayers (SAMs) has received attention since the SAMs provide a model organic surfaces with a known surface density of uniformly organized organic groups at the vacuum interface (Ulman 1996; Love *et al.* 2005). Not only are SAMs of use as generalized models for elucidating metalorganic surface interactions, but there has been considerable interest in using organic molecules as electrical components (Aviram and Ratner 1974; Mirkin and Ratner 1992; Allara *et al.* 1998; Joachim *et al.* 2000; Service 2003; Kushmerick 2005). However, the ability to contact the molecules electrically, such as through evaporation of a top metal contact, without inducing any damage, has remained a significant challenge (Haick and Cahen 2008).

Condensation of a metal or any vapor atom on a surface is also commonly referred to as physical vapor deposition (PVD) (Mattox 1998). The three most common methods of producing a vapor of atoms are thermal evaporation where a material is resistively heated to its melting temperature, effectively increasing the material's vapor pressure, electron-beam (e-beam) evaporation where a material is heated through bombardment of a beam of high-energy electrons, and sputtering, wherein atoms are knocked out of a target with energetic ions created from a plasma created from a noble gas. The PVD process is versatile; almost any type of inorganic material and even some organic materials can be deposited. However, the physical deposition process has some significant disadvantages, especially with regards to growing thin metal films on organic molecular assemblies. It is a line-of-sight technique, making it difficult to coat non-planar features, and the incident particles have a high kinetic energy, which can induce structural defects and damage to the organic molecules (Jung and Czanderna 1994a). For example, Ti vapor atoms react vigorously with the organic molecules to form a complex-structured inorganic composite of carbides, hydrides, and oxides intermingled with residual elements of the original molecule (de Boer *et al.* 2004; Tighe *et al.* 2005; Walker *et al.* 2004b, 2005).

The general strategy to forming a thin metal film on organic molecular assemblies has been to incorporate chemical functionalities at the surface that serve to activate the molecule toward a chemical interaction, nucleating the metal atoms and forming a continuous, non-penetrating metal film. However, despite the many different metal atoms studied, for example Ag (Tarlov 1992;

Herdt and Czanderna 1994; Herdt *et al.* 1995; Ohgi *et al.* 2001a; Walker *et al.* 2003, 2004a), Au (Ohgi *et al.* 1999; de Boer *et al.* 2004; Walker *et al.* 2004a,b; Zhu *et al.* 2006; Haick *et al.* 2007, 2005), Ca (Walker *et al.* 2005) Ti (Konstadinidis *et al.* 1995; Chang *et al.* 2003; de Boer *et al.* 2004; Walker *et al.* 2004b, 2005; Tighe *et al.* 2005), Al (Fisher *et al.* 2000, 2002; de Boer *et al.* 2004; Walker *et al.* 2003, 2004a), Mg (Zhou *et al.* 2005; Walker *et al.* 2007) Cu (Herdt *et al.* 1995; Walker *et al.* 2003, 2004a), Pd (Haick *et al.* 2005, 2007), Cr (Jung and Czanderna 1994b; Herdt *et al.* 1995), and Fe (Carlo *et al.* 2000) deposited on a variety of functional and structural groups, such as  $-COOH$  (Herdt *et al.* 1995; Fisher *et al.* 2000; Zhu *et al.* 2006),  $-OCH_3$  (Konstadinidis *et al.* 1995; Hooper *et al.* 1999; Fisher *et al.* 2002; Walker *et al.* 2003, 2004a, 2005, 2007; Zhu *et al.* 2006),  $-SH$  (Ohgi *et al.* 2001b; de Boer *et al.* 2004),  $-CF_3$  (Carlo *et al.* 2000),  $-CH_3$  (Jung and Czanderna 1994b; Herdt *et al.* 1995; Hooper *et al.* 1999; Ohgi *et al.* 1999; Ohgi *et al.* 2001a; de Boer *et al.* 2004; Tighe *et al.* 2005; Zhu *et al.* 2006),  $-OH$  (Herdt *et al.* 1995; Konstadinidis *et al.* 1995; Fisher *et al.* 2002), and  $-CN$  (Jung and Czanderna 1994b; Konstadinidis *et al.* 1995; Herdt *et al.* 1995), investigations of the fundamental reaction pathways of the metal atoms with the organic molecules have revealed a far more complicated scenario, see Fig. 23.8. Regardless of the metal atoms deposited, it has been found that the metal atoms not only react with the organic functional groups, but will often diffuse and penetrate through the molecules to the substrate through structural pinhole defects to form metal filaments, discontinuous morphologies, and subsurface metal films (Jung and Czanderna 1994a; Fisher *et al.* 2002; Haick *et al.* 2005; Zhu *et al.* 2006; Haick and Cahen 2008). These phenomena have been found even when great care has been taken to reduce the interaction energetics between the impinging metal-vapor atoms and the organic molecules such as cooling the sample during evaporation (Haick *et al.* 2007) or transferring metal-vapor momentum through multiple collisions with an inert gas (Haick *et al.* 2007).



**Fig. 23.8** Schematic detailing the realistic characteristics of a thin metal film grown from condensed metal vapor atoms on a self-assembled monolayer. These can include a sharp interfacial layer of chemically bonded metal, metal filaments, discontinuous morphologies, subsurface metal films and complex-structured inorganic composites of carbides, hydrides, and oxides intermingled with residual elements of the original molecule.

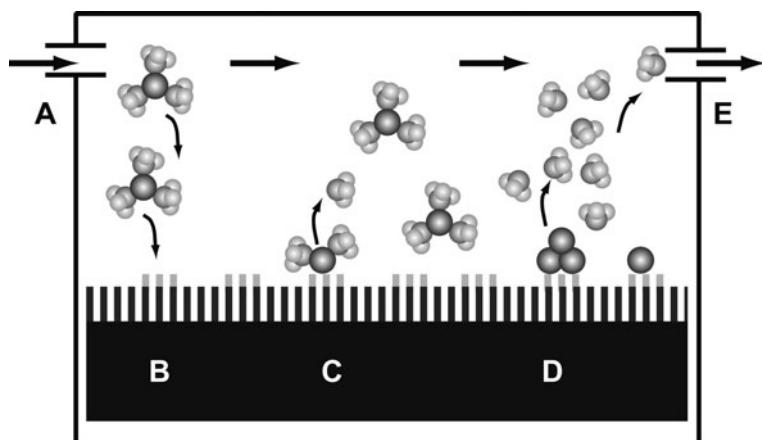
Compounding the challenge of forming an intact, high-quality thin metal film on self-assembled monolayers using metal-vapor atoms is the challenge of using this method to form area-selective thin films with patterned molecular assemblies, especially given the natural tendency of the metal atoms to diffuse across the substrate (Zhu *et al.* 2006). As such, there has only been one report of area-selective growth on SAMs using metal-vapor atoms, where patterned regions of  $\omega$ -terminated  $-COOH$  and  $-CH_3$  alkanethiolates were formed on a Au(111) surface through photooxidation masking (Zhou *et al.* 2005). Utilizing the fact that vapor-deposited Mg only chemically reacts with  $-OH$  (Zhou *et al.* 2005),  $-COOH$  (Zhou *et al.* 2005), and  $-OCH_3$  (Walker *et al.* 2007) functional groups to form a metalorganic complex, Mg was deposited over the entire surface. It was found that Mg could be patterned on the surface through reaction with the  $\omega$ -terminated  $-COOH$  molecules, and accumulating on top of the SAM. It was found that the Mg atoms did not react with the  $\omega$ -terminated  $-CH_3$  molecules but penetrated to the Au/S interface.

Since using highly reactive vapor metal atoms to form smooth, thin films that have good adhesion properties and do not penetrate through, react with, or induce structural damage to the organic molecules is a significant challenge, another approach toward forming area-selective thin films is to use volatile inorganic complexes that are initially less reactive and damaging to the molecular assemblies. The following two sections review the chemical vapor deposition and atomic-layer deposition processes that take this approach.

### 23.4.2 Chemical vapor deposition

One of the most widely used and technologically relevant methods for depositing thin films is chemical vapor deposition (CVD). Chemical vapor deposition (CVD) is a process whereby a thin solid film is synthesized from the gaseous phase by a chemical reaction or decomposition of inorganic molecular precursors. In an ideal deposition, the by-products and remnants from the reactions and decomposition are cleanly removed to the gas phase (Tagge *et al.* 1996; Crane *et al.* 2000), see Fig. 23.9. It is this chemical process that distinguishes CVD from physical deposition processes such as evaporation, sputtering and sublimation (Hitchman and Jensen 1993). Limited only by the chemistry available for reactions, virtually any type of thin film can be deposited via the CVD process including, metals, superconductors, semiconductors, insulators and dielectrics, carbide and corrosion resistant coatings, and optical layers (Hitchman and Jensen 1993).

Due to the synthesized thin film's superior conformality, CVD is the industry standard choice for many technological processes including microelectronics, optoelectronics, protective and decorative coatings and optical coatings (Hitchman and Jensen 1993). These films can have significantly different properties from thin films deposited from physical deposition processes due to the fact that specific chemical reactions occur at the surface. For example, evaporation of silicon onto a substrate produces a defective, amorphous non-conductive film. In comparison, CVD deposition of a silicon film from silane can create



**Fig. 23.9** Schematic of *ideal* area-selective CVD growth of thin films from organometallic precursors on chemically patterned organic molecular assemblies. (A) Gas-phase of organometallic precursors are flowed into the reaction chamber. (B) Through mass transport and diffusion, the precursors reach the surface. (C) The precursors selectively react and decompose with chemically patterned functionalities at the surface, and the organic by-products are removed to the gas phase. (D) Thin-film growth occurs via selective nucleation and selective-area growth. (E) The gas-phase by-products cleanly leave the reaction chamber.

a hydrogenated film with vastly improved electrical properties, including the ability to form p- and n-doped regions (Dobkin and Zuraw 2003).

Most chemical reactions in the CVD process are thermodynamically endothermic or have a kinetic energy of activation associated with them. As a result, thin films deposited by conventional CVD processes are usually initiated and controlled via thermal energy of a highly heated substrate. It is not uncommon for a CVD reaction to require substrate temperatures in the range of 800–1000 °C (Pierson 1999). As a result, several lower-temperature variations of CVD have been developed including plasma-enhanced CVD (PECVD), photoinitiated CVD, and organometallic CVD (OMCVD).

Conventional CVD or any of the lower-temperature variations have limited applicability to growing area-selective thin films with organic molecular assemblies. This is due to either the high temperatures ( $T > 200$  °C), highly energetic and damaging neutral, charged particles, or high-energy photons that are required to initiate the thin-film growth (Hitchman and Jensen 1993) which can easily damage the organic assembly resists (Nuzzo *et al.* 1990b; Zharnikov and Grunze 2002). Above room temperature, *gauche* defects within the organic monolayer increases, and defects created above 77 °C remain even upon sample cooling (Nuzzo *et al.* 1990b). Of these CVD variations, thin films grown via low-temperature decomposition reactions of organometallic precursors have proven to be the most promising for growing area-selective CVD thin films on organic molecular assemblies.

While most thin-film growth by CVD is uniform and non-selective (Hitchman and Jensen 1993; Jeon *et al.* 1997), there are several precursors that have been shown to be effective in surface selective deposition (Jeon *et al.* 1997; Fischer *et al.* 2001). Through careful selection of the organometallic precursors, namely the chemical nature and lability of the organic ligands surrounding the central metal atom(s), in conjunction with chemically patterned molecular assemblies, surface-selective nucleation and subsequent directed selective-area growth can occur via CVD (Jeon *et al.* 1995, 1997; Jeon and Nuzzo 1995).

There have been several reports of successful area-selective thin-film growth using the OMCVD approach. Thin films comprised of metals, metal oxides and polymers have all been deposited. These include Pt (Jeon *et al.* 1997), Pd (Jeon *et al.* 1997; Fischer *et al.* 2001; Weckenmann *et al.* 2004), Cu (Jeon *et al.* 1995; Jeon and Nuzzo 1995), Au (Winter *et al.* 2000; Fischer *et al.* 2001), Al (Lu *et al.* 2008), and Hg (Aliganga *et al.* 2004) as well as a bilayer system of Cu/Co (Yang *et al.* 2006a), metal oxides including  $\text{Al}_x\text{O}_y$  (Weiss *et al.* 1998; Wohlfart *et al.* 1999a; Lu *et al.* 2008),  $\text{HfO}_2$  (Kang *et al.* 2006),  $\text{TiO}_2$  (Kang *et al.* 2003) and even small organic/inorganic molecules including, poly(isobenzofuran) polymers (Choi *et al.* 2007) and alane (Weiss *et al.* 1998). Jeon *et al.* achieved selective-area growth of Pt, Pd and Cu CVD-grown metal films by utilizing microcontact printing ( $\mu\text{CP}$ ) patterned regions on SAMs of alkylsilanes (Jeon *et al.* 1995; Jeon *et al.* 1997; Fischer *et al.* 2001; Lu *et al.* 2008). The silanes were patterned on a variety of surfaces including TiN/TiO<sub>2</sub>, indium tin oxide (ITO), Al/Al<sub>2</sub>O<sub>3</sub>, Si/SiO<sub>2</sub>, borosilicate glasses and plasma-modified polyimide, and shown to be effective resists to deposition of the metal films. The organometallic precursors used to deposit the films were hexafluoroacetylacetone complexes that are among the few to be shown as effective toward surface-selective depositions. Furthermore, these organometallic complexes can be deposited at modest temperatures (175–275 °C) to minimize damage to the SAMs (Jeon *et al.* 1995, 1997).

Other groups have also reported area-selective CVD growth of noble-metal thin films of Au and Pd from different organometallic precursors, including trimethylphosphinemethylgold ( $\text{Me}_3\text{PAuMe}$ ) and (cyclopentadienyl) allyl-palladium ( $\text{Cp}(\text{allyl})\text{Pd}$ ), respectively (Wohlfart *et al.* 1999b; Winter *et al.* 2000; Fischer *et al.* 2001; Weckenmann *et al.* 2004). The organometallic precursors were deposited onto  $\mu\text{CP}$  patterned –SH, –OH, and –CH<sub>3</sub>  $\omega$ -functionalized long-chain alkanethiolates and 4-functionalized biphenyliolates (Wohlfart *et al.* 1999b; Winter *et al.* 2000; Fischer *et al.* 2001; Weckenmann *et al.* 2004). Through examination with XPS and AFM, nucleation of Au was shown to only occur on the thiol-terminated surfaces, while being inhibited on the methyl and hydroxyl-terminated SAMs, even at mild conditions of 100 °C (Winter *et al.* 2000). In the case of Pd, the reaction was shown to be non-selective at long deposition times. In order to make the deposition selective, trimethylaminealane (TMAA) was first deposited on –OH-terminated SAMs followed by exposure to the  $\text{Cp}(\text{allyl})\text{Pd}$  (Weckenmann *et al.* 2004). Subsequent CVD depositions of TMMA (Wohlfart *et al.* 1999a; Weckenmann *et al.* 2004) revealed that both alane and aluminum oxide could be selectively grown on –OH and –COOH  $\omega$ -terminated SAMs (Weiss *et al.* 1998). In contrast to the –OH- and –COOH-terminated SAMs, no Al deposition was observed on the –CH<sub>3</sub>-terminated regions. Despite the selectivity of the CVD growth, the thin-film growth was found to be non-linear; after 2 s of exposure to TMAA a 7-Å thick layer was deposited on a –OH-terminated SAM, but after 28 h, a film thickness of only 30 Å was achieved (Wohlfart *et al.* 1999a).

More recently, it was shown that chemically patterned thin films of aluminum and aluminum oxides can be achieved using room-temperature CVD

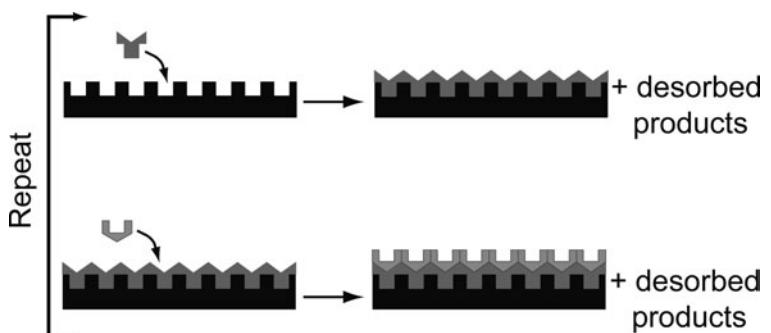
of trimethylaluminum (TMA) (Lu *et al.* 2008). TMA was deposited on  $-CH_3$ -,  $-OH$ - and  $-COOH$ -terminated alkylthiolate monolayers self-assembled on Au(111) surfaces. Using a UV photopatterning method (Zhou and Walker 2006), regions of chemically different terminated SAMs could be generated. TMS was used to deposit alumina and aluminum oxide ( $AlO_x$ ) films on  $-CH_3$ -,  $-OH$ - and  $-COOH$ -terminated SAMs on the  $-OH$ - and  $-COOH$ -terminated SAMs, the TMS reacted to form a dimethylaluminum complex and form alumina and aluminum oxide ( $AlO_x$ ) films. Since there was no reaction between TMA and  $-CH_3$ -terminated SAMs, any deposited alumina could be removed from the surface by rinsing in hexane, resulting in area-selective CVD.

### 23.4.3 Atomic-layer deposition

Deposition of thin films via atomic-layer deposition (ALD) is a strategy that utilizes self-limiting surface reactions to build up material one atomic layer at a time. Using sequential pulses of chemical precursor vapors, each chemical species reacts with the exposed surface chemistry to form one atomic layer. Through careful selection of the half-reactions, the cyclical growth of the film is carried out by regenerating the starting surface chemistry and allows precise control of the material thickness, see Fig. 23.10.

This technique, originally known as atomic-layer epitaxy, was pioneered in the late 1970s by Suntola and coworkers who submitted a patent for depositing thin layers of ZnS grown epitaxially by atomic layer growth (Suntola and Antson 1974; Suntola and Antson 1977; Suntola *et al.* 1983; Suntola and Hyvarinen 1985). Atomic-layer epitaxy proved to be applicable to the growth of other single-crystal thin films including GaAs, CdTe, and AlAs (Goodman and Pessa 1986). Over the past 20 years, however, this method for depositing thin films one atomic layer at a time has expanded to include metal oxides, II–VI and III–V semiconductors, nitrides, sulfides and elemental thin films that are not necessarily epitaxial and is now more commonly known as atomic-layer deposition (George *et al.* 1996).

Thin films deposited by ALD have proven to be enormously useful in technological applications because films grown via this method are not only precisely controlled to an atomic-layer thickness, but they also have strong chemical adhesion and large-area conformality with minimal defects. For



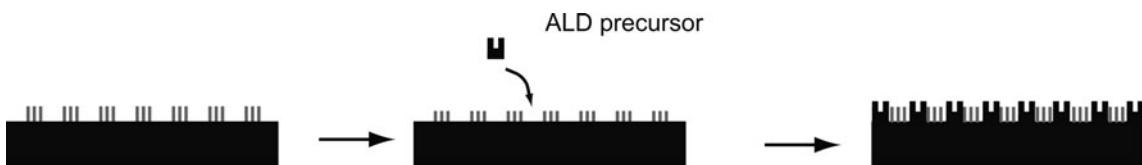
**Fig. 23.10** Schematic representation of a binary atomic layer controlled deposition using self-limiting surface chemistry, modelled after the work of George *et al.* 1996.

example, the “breakthrough” 45 nm transistor technology developed by Intel Corp. relied on a thin film of only 2.5–3 nm (approximately one monolayer) of the high-*k* dielectric material HfO<sub>2</sub> deposited by atomic layer deposition for its low leakage transistor gate electrode (Borgh *et al.* 2007).

Given the successful application of the ALD process in forming one-dimensional thin films, there has been recent interest in developing processes to form laterally controlled, two-dimensional ALD films. The critical challenge in achieving area-selective ALD is devising materials and methods for modifying selected regions of a substrate surface to prevent ALD reactions from occurring, thus preventing film growth (Sinha *et al.* 2006). One approach to forming spatially constrained ALD films has been to utilize patterned molecular assemblies to selectively and chemically block areas of the surface where thin-film growth is not desired.

Spatially constraining ALD films with patterned molecular assemblies is a relatively new method, having been developed only within the last six years (Yan *et al.* 2001), and there have been three general approaches taken to form the patterned molecular assemblies: microcontact printing ( $\mu$ CP) which stamps the molecules into pre-determined patterns (Yan *et al.* 2001; Lee and Sung 2007) patterning of the underlying substrate such that only specific areas of the surface react with the molecules (Chen *et al.* 2005b; Chen and Bent 2006a; Hong *et al.* 2007), and selectively removing monolayer regions through photooxidation masking (Lee *et al.* 2003c; Lee and Sung 2004), see Fig. 23.11.

Due to the stringent surface chemistries required to form the ALD films combined with the molecular assembly attachment chemistry requirements, reports of area-selective growth using molecular assemblies as masks have thus far been limited to only a few surface-molecular assembly attachment chemistry combinations. These include alkyltrichlorosilanes self-assembled on SiO<sub>2</sub> (Yan *et al.* 2001; Lee *et al.* 2003b; Chen *et al.* 2004, 2005a, 2005b, Dubé *et al.* 2006, 2007; Lee and Sung 2007), HfO<sub>2</sub> (Park *et al.* 2005), TiO<sub>2</sub> (Lee *et al.* 2003c), and yttria-stabilized zirconia (YSZ) surfaces (Jiang and Bent 2007; Jiang *et al.* 2007); 1-alkenes and 1-alkynes self-assembled on Si and Ge surfaces (Chen and Bent 2006a,b; Hong *et al.* 2007); and alkanethiols self-assembled on Au surfaces (Seo *et al.* 2004). Of these surfaces, by far the most studied surface-molecular assembly attachment chemistry for selective area ALD thin-film growth has been the SiO<sub>2</sub>-alkyltrichlorosilane system. For all of these surface-molecular assembly systems successful growth of spatially constrained ALD thin films has been mainly limited to metal oxides such as ZnO (Yan *et al.* 2001; Lee and Sung 2007), ZrO<sub>2</sub> (Chen *et al.* 2004; Lee and



**Fig. 23.11** Schematic of *ideal* selective-area ALD growth on a patterned molecular assembly. The patterned molecular assembly is exposed to the ALD precursor for selective-area ALD growth.

Sung 2007), TiO<sub>2</sub> (Seo *et al.* 2004; Park *et al.* 2004; Lee and Sung 2007), and HfO<sub>2</sub> (Chen *et al.* 2004, 2005a; Chen *et al.* 2005b; Jiang and Bent 2007; Jiang *et al.* 2007; Hong *et al.* 2007), but there has been several reports of elemental metal films including Pt (Chen and Bent 2006b; Jiang *et al.* 2007; Hong *et al.* 2007), and Ru (Park *et al.* 2005).

In most of the cases, the patterned molecules act as resists to prevent thin-film growth (Yan *et al.* 2001; Lee *et al.* 2003c; Lee and Sung 2004; Chen and Bent 2006a; Chen and Bent 2006b; Sinha *et al.* 2006; Hong *et al.* 2007), but more recently there have been several cases where the molecules have served as anchors for initiating thin-film growth (Lee *et al.* 2003b; Lee *et al.* 2003c; Seo *et al.* 2004; Dube *et al.* 2006, 2007). Regardless of the method of molecular assembly patterning, there are two critical parameters to forming spatially constrained ALD thin films: the density of molecular conformational order and the selective inertness (or selective reactivity in the case of initiating thin-film growth) of the organic molecules, and in particular the end-functional groups, to the ALD precursors. In general, poorly formed, loosely packed monolayers result in exposure of the ALD precursor molecules to nucleation sites on the substrate, resulting in uncontrolled growth within the masking region. Given the relative small size of the ALD precursors, even small inconsistencies in the molecular array, such as pinhole defects, can lead to nucleation sites for the ALD precursors leading some researchers to investigate the use of lithography-defined patterned polymers with no defects, such as poly(methylmethacrylate) (Sinha *et al.* 2006), for selective-area ALD growth.

While the majority of reports on spatially constrained ALD growth has focused on the qualitative ability of the molecules to block or initiate thin-film growth, there have been a few reports that have focused on how the fundamental self-assembly characteristics, including the effect of chain length, conformational packing, and backbone functionality affect the ability of the molecular assemblies to prevent (or initiate) thin-film growth (Hong *et al.* 2007). For example, Hong and coworkers found that using a fluorinated hydrocarbon chain resulted in better blocking capabilities compared to the same hydrogenated hydrocarbon chain and that as the conformational order of the monolayer increased with time, the blocking capabilities of the SAM to ALD thin-film growth also increased.

### 23.4.4 Electroless deposition

The previous three sections focused on utilizing molecular assemblies for area-selective thin-film growth from gas-phase atoms or reactants. In this final section, area-selective thin-film growth in aqueous solutions via electroless deposition is reviewed. The process of electroless deposition describes a method of thin-film growth where a metal ion or its complex is chemically reduced on a substrate in a controlled fashion (Park *et al.* 2004). Although electroless deposition broadly describes all processes of metal and alloy deposition without an external source of current including displacement, contact and autocatalytic deposition, it is most commonly used to describe the autocatalytic process (Conway and White 2002) and will be the focus of the electroless processes described here. In the autocatalytic process, deposition is achieved

by the incorporation of a reducing agent in the aqueous solution, resulting in thin-film growth only on catalytically active surfaces. Electroless deposition has mainly been used for Ni or Cu thin-film growth (Conway and White 2002), however, many other electrolessly depositable metals and/or alloys have been studied such as Ag, Au, Co, Sn, AuSn, NiWP, Pd, Pt, Ir and Rh (Conway and White 2002; Rao and Trivedi 2005).

Electroless thin-film growth has had significant practical importance in modern technologies including electronics, wear and corrosion-resistant materials, medical applications and battery technologies because it allows plating on complex shapes and plating through holes (Conway and White 2002). For example, damascence Cu electroplating has been used to form Cu chip interconnects in ultralarge-scale integration (ULSI) technologies since the early 1990s because of a decreased resistance and selective, void-free deposits compared with Al wiring (Andricacos *et al.* 1998). Furthermore, it offers a method of depositing Cu thin films that, through other traditional deposition methods such as vapor-metal deposition, easily diffuse into the dielectric substrate and degrade electrical performance (Osaka 1984).

Through selective seeding, electroless deposition offers a method of metallizing insulating substrates at room temperature, making an ideal metallization process for easily damaged materials such as organic molecular assemblies. Coupling electroless deposition with organic molecular assemblies is only a recent implementation and most approaches have utilized the ability of the  $\omega$ -functional groups to complex with metal ions such that the dielectric self-assembled monolayer is sensitized or activated to forming selective sites for seeding the zero-valent thin film (Kind *et al.* 1998; Lee *et al.* 2003a; Zangmeister and van Zee 2003; Zhu *et al.* 2004; Ivanova *et al.* 2005; Sawada *et al.* 2006; Speets *et al.* 2006; Yang *et al.* 2006b; Garno *et al.* 2007; Hsu *et al.* 2007; Lu and Walker 2007). However, only a few of these reports have demonstrated successful two dimensionally constrained, area-selective electroless metal growth via complexation with the organic molecules (Lee *et al.* 2003a; Zangmeister and van Zee 2003; Sawada *et al.* 2006; Lu and Walker 2007). In addition, there have also been a few reports that utilized the molecular assembly to specifically block growth of the electroless deposited thin film (Bittner *et al.* 2002; Wu *et al.* 2002, 2004; Hsu *et al.* 2007). Most of these reports have focused on the qualitative ability of the molecular assemblies to form electroless patterned metal depositions without specific attention to the fundamental interactions between the metal ions and the organic molecules or mechanisms of thin-film growth.

Zangmeister and van Zee were the first to demonstrate the use of selective-area electroless deposition of a metal thin film on a self-assembled organic layer (Zangmeister and van Zee 2003). By exposing  $\mu$ CP patterned films of 4-mercaptopbenzoic acid on Au(111) surface to an aqueous solution of  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , sodium hydrogen tartrate and formaldehyde reducing agent, the copper(II) ions formed a carboxyl-copper complex with the 4-mercaptopbenzoic acid molecules that was able to seed selective-area Cu metal crystallite growth.

Selective-area growth of electrolessly deposited Cu metal films has been demonstrated by three other groups. Lu and Walker examined the electroless

deposition of Cu thin films on  $\mu$ CP patterned films (Lu and Walker 2007). They found a temperature-dependent coverage of the Cu selectively grown  $\omega$ -terminated  $-\text{CH}_3$  and  $-\text{COOH}$  monolayers. The  $\omega$ -terminated  $-\text{OH}$  monolayers remained impervious to Cu thin-film growth due to the  $-\text{OH}$  functional group reacting with the formaldehyde reducing agent to form an acetal. Garno *et al.* utilized scanning probe lithography to pattern regions of dodecanethiol ( $\omega$ -terminated  $-\text{CH}_3$ ), 6-mercaptophexanol and 11-mercaptoundecanol ( $\omega$ -terminated  $-\text{OH}$ ) monolayers by displacement with 16-mercaptophexadecanoic acid ( $\omega$ -terminated  $-\text{COOH}$ ) molecules (Garno *et al.* 2007). Consistent with Lu and Walker's results, Garno *et al.* also found that the  $\omega$ -terminated  $-\text{COOH}$  monolayers had a high spatial selectivity for Cu deposition (Garno *et al.* 2007). Sawada *et al.* and Zhu *et al.* used UV patterning of 3-mercaptopropyltrimethoxysilane ( $\omega$ -terminated  $-\text{SH}$ ) and 3-aminopropyltrimethoxysilane ( $\omega$ -terminated  $-\text{NH}_2$ ) monolayers chemisorbed on SiO<sub>2</sub> derivatized substrates to form electroless, selective-area thin films of Cu on only the  $-\text{SH}$  or  $-\text{NH}_2$  terminated regions (Zhu *et al.* 2004; Sawada *et al.* 2006).

Demonstrations of using a non-Cu metal for selective-area electroless deposition has included the selective growth of Ni thin films (Lee *et al.* 2003a) and Ag metal films (Hsu *et al.* 2007). Lee *et al.* used  $\mu$ CP 16-mercaptophexadecanoic acid ( $\omega$ -terminated  $-\text{COOH}$ ) monolayer on Au surfaces, seeded with a polyelectrolyte Pd catalyst, the selective-area growth of Ni films was demonstrated. In contrast, Hsu *et al.* used  $\mu$ CP octadecyltrichlorosilane on glass substrates to block Ag metal growth. However, the bare glass substrate had to be preferentially seeded in the regions that did not contain the molecular assemblies through deposition of Sn to form covalent bonds with the glass surface before the Ag metal films could grow.

### 23.5 Molecular rulers: A hybrid nanolithographic patterning method

The molecular-ruler methodology is a hybrid nanolithographic strategy that integrates chemical self-assembly (bottom-up) with conventional lithographic processing techniques (top-down) for the creation of structures whose spacing and edge resolution reach nanometer-scale precision (Lee *et al.* 1988a,b; Evans *et al.* 1991; Ansell *et al.* 1996, 2000 2000; Hatzor and Weiss 2001; Anderson *et al.* 2002, 2003; Anderson 1996). The basic scheme for the molecular-ruler process utilizes multilayer molecular films self-assembled on lithographically defined structures as precision spacers to molecularly define device features. One of the primary advantages of this method is providing precise layer-by-layer control of patterning dimensions defined by the molecular-ruler stack thickness. In addition, the fusion of bottom-up and top-down approaches allows parallel creation of hierarchical structures that interface components on the sub-100 nm range simultaneously with micrometer-scale structures, another long-standing lithographic goal.

These final sections describe how precise, nanometer-sized features can be patterned across a surface utilizing the molecular rulers formed from both

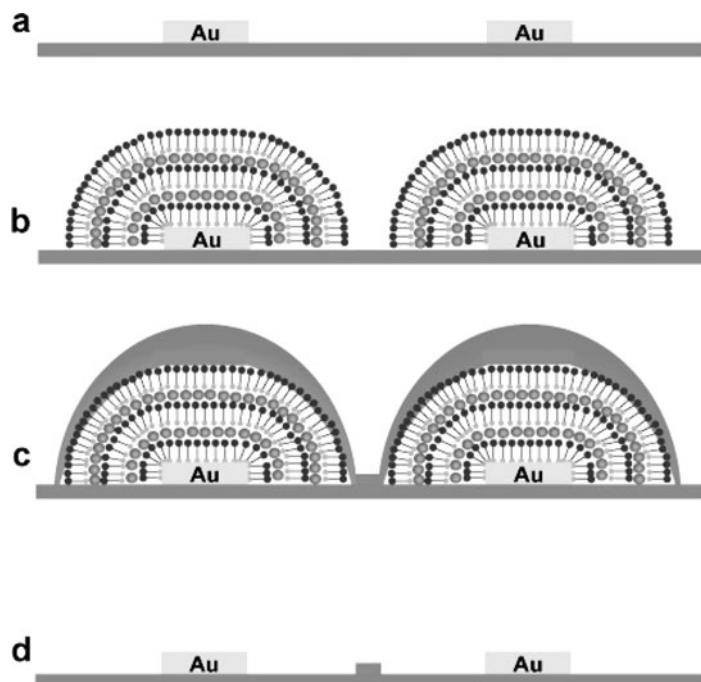
multilayer molecular assemblies and conventional lithographic processes; the integrity of these structures, based on the evaluation of their electrical characteristics is also reviewed. Finally, the outlook for the integration and adoption of this hybrid nanolithographic patterning method into standard lithographic processes is discussed.

### 23.5.1 Molecularly defined structures from multilayer molecular assemblies

Central to the molecular-ruler process is the formation of robust, multilayer molecular films as precision spacers. These multilayer films are built from the bottom-up using a base self-assembled monolayer, comprised of  $\alpha$ - $\omega$ -bifunctional organic molecules as the initial anchor unit. In the initial molecular layer, one functional end of the molecule (“head”) selectively binds to a lithographically defined metal structure. The exposed end-functional group (“tail”), coordinated with a metal ion, forms the reactive site for subsequent multilayer growth. Iterative exposure of the foundation monolayer to bifunctional molecules and metal ions results in a self-assembled layer-by-layer growth of the multilayer film. Several ligated organic molecule–metal ion chemistries have been used to form multilayer films; some examples of these systems include  $\alpha$ ,  $\omega$ -mercaptoalkanoic acid/ $\text{Cu}^{2+}$ ,  $\alpha$ ,  $\omega$ -alkylbisphosphonic acids/ $\text{Zr}^{4+}$  and dicyanobenzene/ $\text{Co}^{2+}$  (Lee *et al.* 1988a,b; Evans *et al.* 1991; Ansell *et al.* 1996, 2000). Of these chemistries, the  $\alpha$ ,  $\omega$ -mercaptoalkanoic acid/ $\text{Cu}^{2+}$  system is the most routinely used. Layer-by-layer growth of the multilayers resulting from these organic–metal ion chemistries is generally regarded as providing precise, highly tuned and uniform films, though recent work on the details of the mercaptoalkanoic acid– $\text{Cu}^{2+}$  (Daniel *et al.* 2007) show that the self-organization of the chains is inherently not complete due to the low stoichiometry (50% attachment of the top layer) of the coupling reaction.

Since the thickness of the self-assembled multilayers is defined by the number of iterative layer growth steps, these films (“molecular rulers”) can be used to molecularly define the size of the junctions/gaps between lithographically defined metallic structures. It is the precise, proximal placement of these metal structures that is the strength of the process. This fusion of established patterning methods with novel self-assembly processes is an approach that holds promise for the further miniaturization of electronic devices.

The molecular-ruler process begins with a lithographically defined metal structure, or “parent” structure (Fig. 23.12(a)). The choice of metal depends critically on the initial base layer self-assembly chemistry. For example, gold is typically used with bifunctional alkanethiolate monolayers. Subsequently, self-assembled multilayers (typically composed of alternating, coordinated  $\alpha$ ,  $\omega$ -mercaptoalkanoic acid molecules and  $\text{Cu}^{2+}$  ions) are grown selectively on the parent structure to form a type of lift-off resist (Fig. 23.12(b)). After metal deposition (Fig. 23.12(c)), this resist is removed chemically, leaving behind only the newly created “daughter” structure and the lithographic parent structure (Fig. 23.12(d)). The width of the gap between the created daughter



**Fig. 23.12** Schematic of the ‘‘molecular ruler’’ process. (a) Gold ‘‘parent’’ features are patterned upon a silicon wafer using either a chromium or titanium adhesion layer using photolithography or electron-beam lithography; (b) a self-assembling multilayer, composed of alternating layers of  $\alpha, \omega$ -mercaptoalkanoic acid and  $\text{Cu}^{2+}$  ions, is formed on top of the gold parent features; (c) a metal is then evaporated atop the entire assembly; (d) the organic layers are then lifted off, with the final metal ‘‘daughter’’ structure remaining.

structure and the parent structure is measured out precisely by the tailored thickness of the multilayer resist.

Gaps of differing sizes between parent and daughter structures can be generated by selecting both the number of molecular-ruler layers deposited as well as by altering the length of the molecular ruler itself. For example, when utilizing the component molecule of 16-mercaptopentadecanoic acid (MHDA,  $\text{HS}(\text{CH}_2)_{15}\text{COOH}$ ), each layer of MHDA/ $\text{Cu}^{2+}$  is  $\sim 2\text{ nm}$  thick. By using different numbers of this MHDA/ $\text{Cu}^{2+}$  precise building block, a variety of gap sizes have been manufactured. The length of the molecular ruler itself can be altered, for example, 11-mercaptopentadecanoic acid (MUDA,  $\text{HS}(\text{CH}_2)_{10}\text{COOH}$ ) differs in length by  $\sim 0.5\text{ nm}$ , so the MUDA/ $\text{Cu}^{2+}$  building block is  $\sim 1.5\text{ nm}$  thick. This permits high-resolution tuning of the molecular film thickness and hence the resulting nanogap between the metal structures. Gap sizes have been demonstrated routinely in the 10–50 nm range (Lee *et al.* 1988a,b; Evans *et al.* 1991; Ansell *et al.* 1996, 2000; Hatzor and Weiss 2001; Anderson *et al.* 2002, 2003; Anderson 2006).

This research highlights the strength of self-assembly for the fabrication of precise nanostructures with multilayers tailored from the bottom-up to compose a type of lift-off resist, which can withstand metal deposition to define critical nanoscale dimensions. One advantage that multilayers offer over the use of single monolayers as resists is that common pinhole defects in an initial monolayer (which allow metal deposition to penetrate the film and make contact with the underlying metal substrate) are self-healed by the subsequent fluid multilayers. The robustness of the multilayer films is impressive, as they

can be produced and stored for weeks without degradation. They are also stable in a variety of chemical solvents, a characteristic vital for incorporation with conventional lithographic processing (Anderson *et al.* 2002; 2003; Anderson 2006).

### 23.5.2 Molecular-ruler patterning: Merging “top-down” and “bottom-up” nanolithography methods

Molecular rulers have demonstrated their versatility and effectiveness as a means for routinely producing sub-100 nm features and are a true hybrid of “bottom-up” and “top-down” patterning methods. Although molecular rulers are routinely applied to parent structures patterned by conventional photolithographic methods, they have also been applied to nonconventional lithographic patterning methods such as nanosphere lithography. As a second level of lithography control, the molecular rulers formed on a specific parent structure can be subsequently patterned using energetic beams as a means of constructing intricate structures. Finally, since the self-assembled multilayers and molecular-ruler resists are robust to lithographic processing both before and after multilayer deposition, isolated daughter structures can be selectively positioned across a surface. Through this hybrid patterning approach, hierarchical nanostructures with controlled orientation and spacing necessary for nanoscale device fabrication can be precisely constructed.

#### 23.5.2.1 A non-conventional method of patterning the parent structure: Integrating molecular rulers with nanosphere lithography

The molecular-ruler process has been combined with nanosphere lithography (Anderson *et al.* 2003), a novel surface patterning technique advanced by Van Duyne and coworkers (Deckman and Dunsmuir 1982; Hulteen and Van Duyne 1995; Hulteen *et al.* 1999; Haynes and Van Duyne 2001; Haynes *et al.* 2002). In this work, triangular parent structures are generated using a close-packed monolayer of polystyrene spheres as a hard mask. The aim of this work is to generate nanometer-scale features without using conventional lithography. The molecular-ruler process was utilized in combination with parent structures created by nanosphere lithography to produce arrays of metal features with 10 nm spacings, which have potential application in the field of plasmonics. The integrity of the rulers was maintained having excellent conformity around the sharp edges of the triangular parents, displaying the versatility of the molecular-ruler method on non-planar structures. Thus, without using any form of conventional lithography, precise 10 nm structures have been formed reproducibly (Anderson *et al.* 2003).

#### 23.5.2.2 Position selective patterning of molecular rulers using energetic beams

To construct intricate lithographic structures, such as position-selected gaps, a method for selectively isolating a molecular-ruler functionalized parent structure was developed (Tanaka *et al.* 2004). In the typical molecular-ruler process, multilayer growth results on all parent structures. Selective removal of the molecular rulers from a specific parent structure was achieved using

the focused electron beam of a field emission scanning electron microscope (FESEM). Upon exposure to the electron beam, the molecular-ruler-modified parent structure is coated with a hydrocarbon film that serves as a resist to the subsequent etching step. In this way, only the molecular-ruler multilayer in the non-irradiated area is removed, which allows for selective-area positioning of nanogaps; this process is referred to as position-selected molecular rulers (PS-MR). After deposition of metal for daughter structures, the PS-MR are chemically removed, and tailored spacings are generated *only* where the PS-MR were present. By using energetic beams and selective chemical etching, position control has been imparted to the molecular-ruler process of generating nanogaps with controlled width and position (Tanaka *et al.* 2004).

#### 23.5.2.3 *Integrating molecular rulers with lithographically defined daughter structures*

Molecular-ruler nanolithography shows great promise for high-resolution patterning with precise, proximal placement of metallic structures fabricated in a parallel manner. These molecular-ruler resists have been used to define nanometer-scale spacings between initial parent structures and the photolithographically patterned daughter structures. Two flow processes have been demonstrated; in one process, the molecular rulers are deposited before the second level of photolithography, in the second process, the molecular rulers are deposited after the second level of photolithography. Both processes create metal structures with precise, controllable, nanometer-scale spacings, demonstrating the robustness of this hybrid method because the lithographic processing conditions did not disrupt the integrity of the multilayer resist. The molecular-ruler process is advanced by employing a second level of photolithography selectively to determine the orientation of the daughter structure and to fabricate structures that interface nanometer- and micrometer-scale components (Anderson 2006).

The same concept of using photolithography to define the secondary daughter structure placement has been employed with electron-beam lithography. This combination employs “the best of both worlds” because electron-beam lithography allows for the creation of sub-micrometer structures connecting to the outside world, while the molecular-ruler technique allows critical dimensions in the sub-20 nm regime to be defined in a parallel manner. Electron-beam lithography faces serious challenges in defining precise junction structures in this regime due to proximity effects, beam size, and stringent processing conditions (Anderson 2006).

#### 23.5.3 **Electrical characterization of molecular-ruler-defined patterns**

Testing the electrical integrity of metallic junctions, created by combining multiple levels of photolithography with the molecular-ruler process, is foundational work required to proceed forward toward the fabrication of devices that incorporate molecules or nanomaterials, such as nanowires, nanoparticles, and

fullerenes. The evaluation of device structures is a critical test for determining the reproducibility and robustness of the molecular-ruler process.

The integrity of the nanoscale gaps produced by the molecular rulers was systematically investigated by electrical and structural characterization, yielding the identification of a primary failure mode. The failure mode was not a result of the multilayer molecular film, but intrinsic to problems with a conventional lithographic resist defining both the parent and daughter structure. Parent structures with precise edges are critical because they define the junction quality that can be produced by the molecular-ruler process. Poor quality of the daughter structures' edges resulted from untethered metal edges that contacted the parent structure forming a short. Development of improved lithographic processing techniques optimized the process and improved the quality of the parent and daughter structures generated (McCarty 2004, 2006; Anderson *et al.* 2005; Srinivasan *et al.* 2006).

### 23.5.4 Molecular rulers: Outlook

The molecular-ruler process is advantageous for patterning precise, nanometer-scale spacings because of its high resolution and precision, selective deposition, compatibility with different forms of lithographic processing, low production cost, parallel nature, and processing conditions (ambient temperature and pressure). The basic procedure detailed here for the molecular-ruler process has the advantage of precise, proximal, parallel placement of metallic structures with nanometer resolution. The molecular-ruler method, in combination with other processing techniques, has the potential to create complex hierarchical structures and device architectures. The molecular-ruler concept is not limited to the materials mentioned here and diversification to different multilayer systems and metals would enable better CMOS compatibility. Molecular rulers have been established as a parallel technique for the accurate definition of tailored sub-100 nm spacings between metallic structures with prospects for the fabrication of nanoscale electronics.

Recently, the potential for generating quartz imprint lithography templates using this molecular-ruler technique has been realized. Using reactive ion etching, these nanogaps can be transferred into the underlying substrate to create precise recessed features. The parent and daughter structures act as hard etch masks permitting the creation of high aspect ratio trenches. The molecular-ruler process has been demonstrated as a parallel and economical method for obtaining nanoscale features on large-area quartz substrates for use in nanoimprint lithography template fabrication (Srinivasan *et al.* 2008).

The compatibility and robustness of hybrid strategies employing chemical self-assembly with conventional lithographic fabrication schemes and processes has been demonstrated through the development of the molecular-ruler process. These methods were the direct result of interdisciplinary work overlapping chemistry and engineering science. By merging bottom-up and top-down methodologies, chemically advanced nanolithography techniques can be developed that have the advantage of molecular-scale resolution with the benefit of building on well-established technologies.

## 23.6 Conclusion

Self-assembly strategies of organic molecules that have been developed for metals, semiconductors, and native-oxide surfaces show tremendous promise for extending the limits of conventional lithography techniques. Unlike conventional lithography techniques, which require almost an exclusive “top-down” approach to forming patterns, structures, and device features across a surface, molecular assemblies can be patterned via exclusive “top-down” methods where the molecular assemblies are used as sacrificial resists to machining, electron and photon beams and via “bottom-up” methods where specific physicochemical self-assembling interactions between molecules or between substrate–molecule adsorption chemistries dictate the final pattern. By tailoring the chemical functionalities of the organic molecules, the patterned assemblies can even be used to grow spatially constrained thin films. Molecular assemblies, and specifically molecular rulers, can maintain good mechanical, thermal and chemical stability throughout multiple fabrication steps and are an example of successful incorporation into standard lithographic processes.

These molecular assemblies can be formed from multiple adsorbate–substrate chemistries (for example Si–CR, Si–O–R, Au–SR, Ag–SR, GaAs–SR) and thus patterns with specific chemical functionality can be constructed on virtually any technologically relevant surface. Furthermore, since these molecular assemblies are formed through self-limiting chemistries, the thickness and chemical functionality of the monolayer can be precisely controlled. The inherent thickness of a SAM, defined by the molecular length, makes molecular-assembly films ideal photoresist candidates especially for reproducibly controlling feature dimensions in applications such as near-field nanolithography. Although most of these SAMs are plagued by intrinsic defects such as pinholes and grain boundaries, degrading their ability to form good diffusion barriers for directed thin-film growth, the use of multilayers of these molecules (such as the molecular-rulers method) have shown to be effective resists to metal-vapor atoms upon when depositing the daughter structure.

This chapter highlighted many of the different processes to forming nanometer-sized structures with molecular films. The numerous approaches to constructing patterns and spatially constrained thin films reflect the diverse chemistries that can be harnessed to achieve the goals of nanolithography. It is the layer-by-layer control of these molecular assemblies that makes them ideal candidates for constructing the next-generation of electronic devices.

## References

- Aizenberg, J., Black, A.J., Whitesides, G.M. *Nature* **394**, 868 (1998).
- Aliganga, A.K.A., Wang, Z.H., Mittler, S. *J. Phys. Chem. B* **108**, 10949 (2004).
- Allara, D.L., Dunbar, T.D., Weiss, P.S., Bumm, L.A., Cygan, M.T., Tour, J.M., Reinerth, W.A., Yao, Y., Kozaki, M., Jones, L. *Ann. NY Acad. Sci.* **852**, 349 (1998).

- Allara, D.L., Parikh, A.N., Rondelez, F. *Langmuir* **11**, 2357 (1995).
- Amro, N.A., Xu, S., Liu, G.-Y. *Langmuir* **16**, 3006 (2000).
- Anderson, M.E. *Chemically Advanced Nanolithography*. PhD Thesis, The Pennsylvania State University (2006).
- Anderson, M.E., Mihok, M., Tanaka, H., Tan, L.P., Horn, M.W., McCarty, G.S., Weiss, P.S. *Adv. Mater.* **18**, 1020 (2006).
- Anderson, M.E., Smith, R.K., Donhauser, Z.J., Hatzor, A., Lewis, P.A., Tan, L.P., Tanaka, H., Horn, M.W., Weiss, P.S. *J. Vac. Sci. Technol. B* **20**, 2739 (2002).
- Anderson, M.E., Srinivasan, C., Jayaraman, R., Weiss, P.S., Horn, M.W. *Microelectron. Eng.* **78–79**, 248 (2005).
- Anderson, M.E., Tan, L.-P., Tanaka, H., Mihok, M., Lee, H., Horn, M.W., Weiss, P.S. *J. Vac. Sci. Technol. B* **21**, 3116 (2003).
- Andricacos, P.C., Uzoh, C., Dukovic, J.O., Horkans, J., Deligianni, H. *IBM J. Res. Devel.* **42**, 567 (1998).
- Ansell, M.A., Cogan, E.B., Page, C.J. *Langmuir* **16**, 1172 (2000).
- Ansell, M.A., Zeppenfeld, A.C., Yoshimoto, K., Cogan, E.B., Page, C.J. *Chem. Mater.* **8**, 591 (1996).
- Anselmetti, D., Baratoff, A., Güntherodt, H.-J., Delamarche, E., Michel, B., Gerber, C., Kang, H., Wolf, H., Ringsdorf, H. *Europhys. Lett.* **27**, 365 (1994).
- Aviram, A., Ratner, M.A. *Chem. Phys. Lett.* **29**, 277 (1974).
- Bain, C.D., Evall, J., Whitesides, G.M. *J. Am. Chem. Soc.* **111**, 7155 (1989a).
- Bain, C.D., Troughton, E.B., Tao, Y.T., Evall, J., Whitesides, G.M., Nuzzo, R.G. *J. Am. Chem. Soc.* **111**, 321 (1989b).
- Bain, C.D., Whitesides, G.M. *J. Am. Chem. Soc.* **110**, 6560 (1988).
- Bain, C.D., Whitesides, G.M. *Langmuir* **5**, 1370 (1989).
- Bard, A., Berggren, K.K., Wilbur, J.L., Gillaspy, J.D., Rolston, S.L., McClelland, J.J., Phillips, W.D., Prentiss, M., Whitesides, G.M. *J. Vac. Sci. Technol. B* **15**, 1805 (1997).
- Betzig, E., Trautman, J.K., Harris, T.D., Weiner, J.S., Kostelak, R.L. *Science* **251**, 1468 (1991).
- Biebuyck, H.A., Whitesides, G.M. *Langmuir* **10**, 4581 (1994).
- Bigelow, W.C., Pickett, D.L., Zisman, W.A. *J. Colloid Sci.* **1**, 513 (1946).
- Binnig, G., Rohrer, H., Vettiger, P. *Mass-storage Applications of Local Probe Arrays*, US Patent 5,835,477 (1998).
- Bittner, A.M., Wu, X.C., Kern, K. *Adv. Funct. Mater.* **12**, 432 (2002).
- Black, A.J., Paul, K.E., Aizenberg, J., Whitesides, G.M. *J. Am. Chem. Soc.* **121**, 8356 (1999).
- Borh, M., Chau, R.S., Ghani, T., Mistry, K. *IEEE Spectrum* **29** (2007).
- Brewer, N.J., Janusz, S., Critchley, K., Evans, S.D., Leggett, G.J. *J. Phys. Chem. B* **109**, 11247 (2005).
- Brewer, N.J., Rawsterne, R.E., Kothari, S., Leggett, G.J. *J. Am. Chem. Soc.* **123**, 4089 (2001).
- Bumm, L.A., Arnold, J.J., Charles, L.F., Dunbar, T.D., Allara, D.L., Weiss, P.S. *J. Am. Chem. Soc.* **121**, 8017 (1999).

- Bumm, L.A., Arnold, J.J., Cygan, M.T., Dunbar, T.D., Burgin, T.P., Jones II, L., Allara, D.L., Tour, J.M., Weiss, P.S. *Science* **271**, 1705 (1996).
- Buriak, J.M. *Chem. Rev.* **102**, 1271 (2002).
- Calvert, J.M. *J. Vac. Sci. Technol. B* **11**, 2155 (1993).
- Camillone, N., Eisenberger, P., Leung, T.Y.B., Schwartz, P., Scoles, G., Poirier, G.E., Tarlov, M.J. *J. Chem. Phys.* **101**, 11031 (1994).
- Carlo, S.R., Wagner, A.J., Fairbrother, D.H. *J. Phys. Chem. B* **104**, 6633 (2000).
- Carr, D.W., Lercel, M.J., Whelan, C.S., Craighead, H.G., Seshadri, K., Allara, D.L. *J. Vac. Sci. Technol. A* **15**, 1446 (1997).
- Chang, S.C., Li, Z.Y., Lau, C.N., Larade, B., Williams, R.S. *Appl. Phys. Lett.* **83**, 3198 (2003).
- Chang, W.S., Kim, J., Cho, S.H., Choi, M. *J. Korean Phys. Soc.* **47**, S233 (2005).
- Chen, R., Bent, S.F. *Adv. Mater.* **18**, 1086 (2006a).
- Chen, R., Bent, S.F. *Chem. Mater.* **18**, 3733 (2006b).
- Chen, R., Kim, H., McIntyre, P.C., Bent, S.F. *Appl. Phys. Lett.* **84**, 4017 (2004).
- Chen, R., Kim, H., McIntyre, P.C., Bent, S.F. *Chem. Mater.* **17**, 536 (2005a).
- Chen, R., Kim, H., McIntyre, P.C., Porter, D.W., Bent, S.F. *Appl. Phys. Lett.* **86**, 191910 (2005b).
- Chidsey, C.E.D. *Science* **251**, 919 (1991).
- Choi, H.G., Amara, J.P., Swager, T.M., Jensen, K.F. *Langmuir* **23**, 2483 (2007).
- Clegg, R.S., Hutchison, J.E. *J. Am. Chem. Soc.* **121**, 5319 (1999).
- Clegg, R.S., Reed, S.M., Smith, R.K., Barron, B.L., Rear, J.A., Hutchison, J.E. *Langmuir* **15**, 8876 (1999).
- Conway, B.E., White, R.E. *Modern Aspects of Electrochemistry* (New York, Kluwer Academic, 2002).
- Cooper, E., Leggett, G.J. *Langmuir* **14**, 4795 (1998).
- Cooper, E., Leggett, G.J. *Langmuir* **15**, 1024 (1999).
- Cooper, E., Wiggs, R., Hutt, D.A., Parker, L., Leggett, G.J., Parker, T.L. *J. Mater. Chem.* **7**, 435 (1997).
- Crane, E.L., Girolami, G.S., Nuzzo, R.G. *Acc. Chem. Res.* **33**, 869 (2000).
- Cruchon-Dupeyrat, S., Porthun, S., Liu, G.-Y. *Appl. Surf. Sci.* **175**, 636 (2001).
- Dameron, A.A., Hampton, J.R., Smith, R.K., Mullen, T.J., Gillmor, S.D., Weiss, P.S. *Nano Lett.* **5**, 1834 (2005).
- Daniel, T.A., Uppili, S., McCarty, G., Allara, D.L. *Langmuir* **23**, 638 (2007).
- De Boer, B., Frank, M.M., Chabal, Y.J., Jiang, W.R., Garfunkel, E., Bao, Z. *Langmuir* **20**, 1539 (2004).
- Deckman, H.W., Dunsmuir, J.H. *Appl. Phys. Lett.* **41**, 377 (1982).
- Delamarche, E., Hoole, A.C.F., Michel, B., Wilkes, S., Despont, M., Welland, M.E., Biebuyck, H. *J. Phys. Chem. B* **101**, 9263 (1997).
- Delamarche, E., Michel, B., Gerber, C., Anselmetti, D., Güntherodt, H.-J., Wolf, H., Ringsdorf, H. *Langmuir* **10**, 2869 (1994).
- Delamarche, E., Schmid, H., Bietsch, A., Larsen, N.B., Rothuizen, H., Michel, B., Biebuyck, H. *J. Phys. Chem. B* **102**, 3324 (1998).

- Dobkin, D.M., Zuraw, M.K. *Principles of Chemical Vapor Deposition* (Kluwer Academic Publishers, Dordrecht, 2003).
- Donhauser, Z.J., Mantooth, B.A., Kelly, K.F., Bumm, L.A., Monnell, J.D., Stapleton, J.J., Price II, D.W., Rawlett, A.M., Allara, D.L., Tour, J.M., Weiss, P.S. *Science* **292**, 2303 (2001).
- Donhauser, Z.J., Mantooth, B.A., Pearl, T.P., Kelly, K.F., Nanayakkara, S.U., Weiss, P.S. *Jpn. J. Appl. Phys.* **41**, 4871 (2002).
- Donhauser, Z.J., Price II, D.W., Tour, J.M., Weiss, P.S. *J. Am. Chem. Soc.* **125**, 11462 (2003).
- Dressick, W.J., Calvert, J.M. *Jpn. J. Appl. Phys.* **32**, 5829 (1993).
- Dressick, W.J., Dulcey, C.S., Brandow, S.L., Witschi, H., Neeley, P.F. *J. Vac. Sci. Technol. A* **17**, 1432 (1999).
- Dube, A., Sharma, M., Ma, P.F., Engstrom, J.R. *Appl. Phys. Lett.* **89**, 164108 (2006).
- Dube, A., Sharma, M., Ma, P.F., Ercius, P.A., Muller, D.A., Engstrom, J.R. *J. Phys. Chem. C* **111**, 11045 (2007).
- Dubois, L.H., Nuzzo, R.G. *Annu. Rev. Phys. Chem.* **43**, 437 (1992).
- Dürig, U., Zuger, O., Michel, B., Häussling, L., Ringsdorf, H. *Phys. Rev. B* **48**, 1711 (1993).
- Eck, W., Stadler, V., Geyer, W., Zharnikov, M., Gölzhäuser, A., Grunze, M. *Adv. Mater.* **12**, 805 (2000).
- Edinger, K., Grunze, M., Wöll, C. *Phys. Chem. Chem. Phys.* **101**, 1811 (1997).
- Evans, S.D., Ulman, A., Goppert-Berarducci, K.E., Gerenser, L.J. *J. Am. Chem. Soc.* **113**, 5866 (1991).
- Fischer, R.A., Weckenmann, U., Winter, C., Kashammer, J., Scheumann, V., Mittler, S. *J. Phys. IV* **11**, 1183 (2001).
- Fisher, G.L., Hooper, A.E., Opila, R.L., Allara, D.L., Winograd, N. *J. Phys. Chem. B* **104**, 3267 (2000).
- Fisher, G.L., Walker, A.V., Hooper, A.E., Tighe, T.B., Bahneck, K.B., Skriba, H.T., Reinard, M.D., Haynie, B.C., Opila, R.L., Winograd, N., Allara, D.L. *J. Am. Chem. Soc.* **124**, 5528 (2002).
- Folkers, J.P., Laibinis, P.E., Whitesides, G.M. *Langmuir* **8**, 1330 (1992a).
- Folkers, J.P., Laibinis, P.E., Whitesides, G.M. *J. Adhes. Sci. Technol.* **6**, 1397 (1992b).
- Folkers, J.P., Laibinis, P.E., Whitesides, G.M., Deutch, J. *J. Phys. Chem.* **98**, 563 (1994).
- Frydman, E., Cohen, H., Maoz, R., Sagiv, J. *Langmuir* **13**, 5089 (1997).
- Garcia, R., Martinez, R.V., Martinez, J. *Chem. Soc. Rev.* **35**, 29 (2006).
- Garno, J.C., Zangmeister, C.D., Batteas, J.D. *Langmuir* **23**, 7874 (2007).
- Gates, B.D., Xu, Q.B., Stewart, M., Ryan, D., Willson, C.G., Whitesides, G.M. *Chem. Rev.* **105**, 1171 (2005).
- Geissler, M., Schmid, H., Bietsch, A., Michel, B., Delamarche, E. *Langmuir* **18**, 2374 (2002).
- George, S.M., Ott, A.W., Klaus, J.W. *J. Phys. Chem.* **100**, 13121 (1996).
- Geyer, W., Stadler, V., Eck, W., Gölzhäuser, A., Grunze, M., Sauer, M., Weimann, T., Hinze, P. *J. Vac. Sci. Technol. B* **19**, 2732 (2001).
- Geyer, W., Stadler, V., Eck, W., Zharnikov, M., Gölzhäuser, A., Grunze, M. *Appl. Phys. Lett.* **75**, 2401 (1999).

- Gillen, G., Bennett, J., Tarlov, M.J., Burgess, F. *Anal. Chem.* **66**, 2170 (1994).
- Gölzhäuser, A., Eck, W., Geyer, W., Stadler, V., Weimann, T., Hinze, P., Grunze, M. *Adv. Mater.* **13**, 806 (2001).
- Gölzhäuser, A., Geyer, W., Stadler, V., Eck, W., Grunze, M., Edinger, K., Weimann, T., Hinze, P. *J. Vac. Sci. Technol. B* **18**, 3414 (2000).
- Goodman, C.H.L., Pessa, M.V. *J. Appl. Phys.* **60**, R65 (1986).
- Gorman, C.B., Biebuyck, H.A., Whitesides, G.M. *Chem. Mater.* **7**, 252 (1995).
- Graham, R.L., Bain, C.D., Biebuyck, H.A., Laibinis, P.E., Whitesides, G.M. *J. Phys. Chem.* **97**, 9456 (1993).
- Gu, Y., Lin, Z., Butera, R.A., Smentkowski, V.S., Waldeck, D.H. *Langmuir* **11**, 1849 (1995).
- Haick, H., Cahen, D. *Acc. Chem. Res.* **41**, 359 (2008).
- Haick, H., Ghabboun, J., Cahen, D. *Appl. Phys. Lett.* **86**, 042113 (2005).
- Haick, H., Niitsoo, O., Ghabboun, J., Cahen, D. *J. Phys. Chem. C* **111**, 2318 (2007).
- Hand, A. *Semicon. Int.* **29**, 34 (2006).
- Harnett, C.K., Satyalakshmi, K.M., Craighead, H.G. *Appl. Phys. Lett.* **76**, 2466 (2000).
- Harnett, C.K., Satyalakshmi, K.M., Craighead, H.G. *Langmuir* **17**, 178 (2001).
- Hartwich, J., Dreeskornfeld, L., Heisig, V., Rahn, S., Wehmeyer, O., Kleineberg, U., Heinzmamn, U. *Appl. Phys. A* **66**, S685 (1998).
- Hatzor, A., Weiss, P.S. *Science* **291**, 1019 (2001).
- Haynes, C.L., McFarland, A.D., Smith, M.T., Hulteen, J.C., Van Duyne, R.P. *J. Phys. Chem. B* **106**, 1898 (2002).
- Haynes, C.L., Van Duyne, R.P. *J. Phys. Chem. B* **105**, 5599 (2001).
- Headrick, J.E., Armstrong, M., Cratty, J., Hammond, S., Sheriff, B.A., Berrie, C.L. *Langmuir* **21**, 4117 (2005).
- Herdt, G.C., Czanderna, A.W. *J. Vac. Sci. Technol. A* **12**, 2410 (1994).
- Herdt, G.C., Jung, D.R., Czanderna, A.W. *Prog. Surf. Sci.* **50**, 103 (1995).
- Hitchman, M.L., Jensen, K.F. *Chemical Vapor Deposition* (Academic Press, London, 1993).
- Hobara, D., Kakiuchi, T. *Electrochem. Commun.* **3**, 154 (2001).
- Hobara, D., Sasaki, T., Imabayashi, S., Kakiuchi, T. *Langmuir* **15**, 5073 (1999).
- Hong, J., Porter, D.W., Sreenivasan, R., McIntyre, P.C., Bent, S.F. *Langmuir* **23**, 1160 (2007).
- Hooper, A., Fisher, G.L., Konstadinidis, K., Jung, D., Nguyen, H., Opila, R., Collins, R.W., Winograd, N., Allara, D.L. *J. Am. Chem. Soc.* **121**, 8052 (1999).
- Hsu, C.H., Yeh, M.C., Lo, K.L., Chen, L.J. *Langmuir* **23**, 12111 (2007).
- Huang, J.Y., Hemminger, J.C. *J. Am. Chem. Soc.* **115**, 3342 (1993).
- Hulteen, J.C., Treichel, D.A., Smith, M.T., Duval, M.L., Jensen, T.R., Van Duyne, R.P. *J. Phys. Chem. B* **103**, 3854 (1999).
- Hulteen, J.C., Van Duyne, R.P. *J. Vac. Sci. Technol. A* **13**, 1553 (1995).
- Hutt, D.A., Cooper, E., Leggett, G.J. *J. Phys. Chem. B* **102**, 174 (1998).
- Hutt, D.A., Leggett, G.J. *J. Phys. Chem.* **100**, 6657 (1996).

- Ivanova, V., Baunach, T., Kolb, D.A. *Electrochim. Acta* **50**, 4283 (2005).
- Jeon, N.L., Lin, W.B., Erhardt, M.K., Girolami, G.S., Nuzzo, R.G. *Langmuir* **13**, 3833 (1997).
- Jeon, N.L., Nuzzo, R.G. *Langmuir* **11**, 341 (1995).
- Jeon, N.L., Nuzzo, R.G., Xia, Y.N., Mrksich, M., Whitesides, G.M. *Langmuir* **11**, 3024 (1995).
- Jiang, X., Bent, S.F. *J. Electrochem. Soc.* **154**, D648 (2007).
- Jiang, X.R., Chen, R., Bent, S.F. *Surf. Coat. Technol.* **201**, 8799 (2007).
- Joachim, C., Gimzewski, J.K., Aviram, A. *Nature* **408**, 541 (2000).
- Jung, D.R., Czanderna, A.W. *Crit. Rev. Solid State Mater. Sci.* **19**, 1 (1994a).
- Jung, D.R., Czanderna, A.W. *J. Vac. Sci. Technol. A* **12**, 2402 (1994b).
- Kang, B.C., Jung, D.Y., Boo, J.H. *IEEE Trans. Nanotechnol.* **5**, 701 (2006).
- Kang, B.C., Lee, J.H., Chae, H.Y., Jung, D.Y., Lee, S.B., Boo, J.H. *J. Vac. Sci. Technol. B* **21**, 1773 (2003).
- Karpovich, D.S., Blanchard, G.J. *Langmuir* **10**, 3315 (1994).
- Kim, E., Kumar, A., Whitesides, G.M. *J. Electrochem. Soc.* **142**, 628 (1995).
- Kim, E., Whitesides, G.M., Freiler, M.B., Levy, M., Lin, J.L., Osgood, R.M. *Nanotechnology* **7**, 266 (1996).
- Kim, T.K., Yang, X.M., Peters, R.D., Sohn, B.H., Nealey, P.F. *J. Phys. Chem. B* **104**, 7403 (2000).
- Kind, H., Bittner, A.M., Cavalleri, O., Kern, K., Greber, T. *J. Phys. Chem. B* **102**, 7582 (1998).
- Kleineberg, U., Brechling, A., Sundermann, M., Heinzmann, U. *Adv. Funct. Mater.* **11**, 208 (2001).
- Kobayashi, Y., Sakai, M., Ueda, A., Maruyama, K., Saiki, T., Suzuki, K. *Anal. Chem.* **24**, 571 (2008).
- Koide, Y., Wang, Q.W., Cui, J., Benson, D.D., Marks, T.J. *J. Am. Chem. Soc.* **122**, 11266 (2000).
- Konstadinidis, K., Zhang, P., Opila, R.L., Allara, D.L. *Surf. Sci.* **338**, 300 (1995).
- Krämer, S., Fuierer, R.R., Gorman, C.B. *Chem. Rev.* **103**, 4367 (2003).
- Krupke, R., Malik, S., Weber, H.B., Hampe, O., Kappes, M.M., Von Lohneysen, H. *Nano Lett.* **2**, 1161 (2002).
- Kumar, A., Abbott, N.L., Kim, E., Biebuyck, H.A., Whitesides, G.M. *Acc. Chem. Res.* **28**, 219 (1995).
- Kumar, A., Biebuyck, H.A., Abbott, N.L., Whitesides, G.M. *J. Am. Chem. Soc.* **114**, 9188 (1992).
- Kumar, A., Whitesides, G.M. *Appl. Phys. Lett.* **63**, 2002 (1993).
- Kushmerick, J.G. *Mater. Today* **8**, 26 (2005).
- La, Y.H., Jung, Y.J., Kang, T.H., Ihm, K., Kim, K.J., Kim, B., Park, J.W. *Langmuir* **19**, 9984 (2003a).
- La, Y.H., Jung, Y.J., Kim, H.J., Kang, T.H., Ihm, K., Kim, K.J., Kim, B., Park, J.W. *Langmuir* **19**, 4390 (2003b).
- La, Y.H., Jung, Y.J., Kim, H.J., Kang, T.H., Ihm, K.W., Kim, K.J., Kim, B.S., Park, J.W. *Langmuir* **20**, 6964 (2004).

- La, Y.H., Kim, H.J., Maeng, I.S., Jung, Y.J., Park, J.W., Kim, K.J., Kang, T.H., Kim, B. *Langmuir* **18**, 2430 (2002).
- Laibinis, P.E., Fox, M.A., Folkers, J.P., Whitesides, G.M. *Langmuir* **7**, 3167 (1991a).
- Laibinis, P.E., Graham, R.L., Biebuyck, H.A., Whitesides, G.M. *Science* **254**, 981 (1991b).
- Laibinis, P.E., Nuzzo, R.G., Whitesides, G.M. *J. Phys. Chem.* **96**, 5097 (1992).
- Laibinis, P.E., Whitesides, G.M., Allara, D.L., Tao, Y.T., Parikh, A.N., Nuzzo, R.G. *J. Am. Chem. Soc.* **113**, 7152 (1991c).
- Lee, B.H., Sung, M.M. *J. Nanosci. Nanotechnol.* **7**, 3758 (2007).
- Lee, H., Kepley, L.J., Hong, H.-G., Akhter, S., Mallouk, T.E. *J. Phys. Chem.* **92**, 2597 (1988a).
- Lee, H., Kepley, L.J., Hong, H.-G., Mallouk, T.E. *J. Am. Chem. Soc.* **110**, 618 (1988b).
- Lee, I.S., Hammond, P.T., Rubner, M.F. *Chem. Mater.* **15**, 4583 (2003a).
- Lee, J.P., Jang, Y.J., Sung, M.M. *Adv. Funct. Mater.* **13**, 873 (2003b).
- Lee, J.P., Kim, H.K., Park, C.R., Park, G., Kwak, H.T., Koo, S.M., Sung, M.M. *J. Phys. Chem. B* **107**, 8997 (2003c).
- Lee, J.P., Sung, M.M. *J. Am. Chem. Soc.* **126**, 28 (2004).
- Lee, M.V., Hoffman, M.T., Barnett, K., Geiss, J.M., Smentkowski, V.S., Linford, M.R., Davis, R.C. *J. Nanosci. Nanotechnol.* **6**, 1639 (2006).
- Leggett, G.J. *Chem. Soc. Rev.* **35**, 1150 (2006).
- Lercel, M.J., Redinbo, G.F., Craighead, H.G., Sheen, C.W., Allara, D.L. *Appl. Phys. Lett.* **65**, 974 (1994).
- Lercel, M.J., Tiberio, R.C., Chapman, P.F., Craighead, H.G., Sheen, C.W., Parikh, A.N., Allara, D.L. *J. Vac. Sci. Technol. B* **11**, 2823 (1993).
- Lewis, P.A., Smith, R.K., Kelly, K.F., Bumm, L.A., Reed, S.M., Clegg, R.S., Gunder-son, J.D., Hutchison, J.E., Weiss, P.S. *J. Phys. Chem. B* **105**, 10630 (2001).
- Liang, J., Scoles, G. *Langmuir* **23**, 6142 (2007).
- Lim, H., Carraro, C., Maboudian, R., Pruessner, M.W., Ghodssi, R. *Langmuir* **20**, 743 (2004).
- Liu, G.-Y., Xu, S., Qian, Y.L. *Acc. Chem. Res.* **33**, 457 (2000).
- Liu, J.F., Cruchon-Dupeyrat, S., Garno, J.C., Frommer, J., Liu, G.-Y. *Nano Lett.* **2**, 937 (2002).
- Loo, Y.L., Willett, R.L., Baldwin, K.W., Rogers, J.A. *Appl. Phys. Lett.* **81**, 562 (2002a).
- Loo, Y.L., Willett, R.L., Baldwin, K.W., Rogers, J.A. *J. Am. Chem. Soc.* **124**, 7654 (2002b).
- Love, J.C., Estroff, L.A., Kriebel, J.K., Nuzzo, R.G., Whitesides, G.M. *Chem. Rev.* **105**, 1103 (2005).
- Love, J.C., Wolfe, D.B., Chabinyc, M.L., Paul, K.E., Whitesides, G.M. *J. Am. Chem. Soc.* **124**, 1576 (2002).
- Lu, P., Demirkan, K., Opila, R.L., Walker, A.V. *J. Phys. Chem. C* **112**, 2091 (2008).
- Lu, P., Walker, A.V. *Langmuir* **23**, 12577 (2007).
- Maoz, R., Sagiv, J. *J. Colloid Interface Sci.* **100**, 465 (1984).
- Mattox, D.M. (*Handbook of Physical Vapor Deposition (PVD) Processing*) (Noyes Publications, Park Ridge, NJ, 1998).

- McCarty, G.S. *Nano Lett.* **4**, 1391 (2004).
- McCarty, G.S. *J. Appl. Phys.* **99**, 064701 (2006).
- McGuiness, C.L., Blasini, D., Masejewski, J.P., Uppili, S., Cabarcos, O.M., Smilgies, D., Allarat, D.L. *ACS Nano* **1**, 30 (2007a).
- McGuiness, C.L., Shaporenko, A., Zharnikov, M., Walker, A.V., Allara, D.L. *J. Phys. Chem. C* **111**, 4226 (2007b).
- Mendes, P., Belloni, M., Ashworth, M., Hardy, C., Nikitin, K., Fitzmaurice, D., Critchley, K., Evans, S., Preece, J. *ChemPhysChem* **4**, 884 (2003).
- Mendes, P.M., Preece, J.A. *Curr. Opin. Coll. Interfac. Sci.* **9**, 236 (2004).
- Mirkin, C.A., Hong, S.-H., Demers, L.M. *ChemPhysChem* **2**, 37 (2001).
- Mirkin, C.A., Ratner, M.A. *Annu. Rev. Phys. Chem.* **43**, 719 (1992).
- Mittal, K.L. *Metallized Plastics: Fundamentals and Applications* (Marcel Dekker, Inc, Hopewell Junction, NY, 2008).
- Montague, M., Ducker, R.E., Chong, K.S.L., Manning, R.J., Rutten, F.J.M., Davies, M.C., Leggett, G.J. *Langmuir* **23**, 7328 (2007).
- Moon, J.H., Kim, K.J., Kang, T.H., Kim, B., Kang, H., Park, J.W. *Langmuir* **14**, 5673 (1998).
- Moon, J.H., La, Y.H., Shim, J.Y., Hong, B.J., Kim, K.J., Kang, T.H., Kim, B., Kang, H., Park, J.W. *Langmuir* **16**, 2981 (2000).
- Mrksich, M. *Cell. Mol. Life Sci.* **54**, 653 (1998).
- Mullen, T.J., Srinivasan, C., Hohman, J.N., Gillmor, S.D., Shuster, M.J., Horn, M.W., Andrews, A.M., Weiss, P.S. *Appl. Phys. Lett.* **90**, 063114 (2007).
- Ngunjiri, J.N., Kelley, A.T., Lejeune, Z.M., Li, J.R., Lewandowski, B.R., Serem, W.K., Daniels, S.L., Lusker, K.L., Garno, J.C. *Scanning* **30**, 123 (2008).
- Norrod, K.L., Rowlen, K.L. *J. Am. Chem. Soc.* **120**, 2656 (1998).
- Nuzzo, R.G., Dubois, L.H., Allara, D.L. *J. Am. Chem. Soc.* **112**, 558 (1990a).
- Nuzzo, R.G., Korenic, E.M., Dubois, L.H. *J. Chem. Phys.* **93**, 767 (1990b).
- Nuzzo, R.G., Zegarski, B.R., Dubois, L.H. *J. Am. Chem. Soc.* **109**, 733 (1987).
- Ohgi, T., Fujita, D., Deng, W., Dong, Z.C., Nejoh, H. *Surf. Sci.* **493**, 453 (2001a).
- Ohgi, T., Sheng, H.Y., Dong, Z.C., Nejoh, H. *Surf. Sci.* **442**, 277 (1999).
- Ohgi, T., Sheng, H.Y., Dong, Z.C., Nejoh, H., Fujita, D. *Appl. Phys. Lett.* **79**, 2453 (2001b).
- Osaka, T. *Denki Kagaku* **52**, 438 (1984).
- Parikh, A.N., Allara, D.L., Azouz, I.B., Rondelez, F. *J. Phys. Chem.* **98**, 7577 (1994).
- Parikh, A.N., Liedberg, B., Atre, S.V., Ho, M., Allara, D.L. *J. Phys. Chem.* **99**, 9996 (1995).
- Park, K.J., Doub, J.M., Gougousi, T., Parsons, G.N. *Appl. Phys. Lett.* **86**, (2005).
- Park, M.H., Jang, Y.J., Sung-Suh, H.M., Sung, M.M. *Langmuir* **20**, 2257 (2004).
- Perry, C.C., Wagner, A.J., Fairbrother, D.H. *Chem. Phys.* **280**, 111 (2002).
- Pierson, H.O. *Handbook of Chemical Vapor Deposition, Second Edition: Principles, Technologies and Applications* (Noyes Publications, New York, 1999).
- Piner, R.D., Zhu, J., Xu, F., Hong, S.H., Mirkin, C.A. *Science* **283**, 661 (1999).
- Poirier, G.E. *Chem. Rev.* **97**, 1117 (1997).

- Poirier, G.E., Pylant, E.D. *Science* **272**, 1145 (1996).
- Porter, M.D., Bright, T.B., Allara, D.L., Chidsey, C.E.D. *J. Am. Chem. Soc.* **109**, 3559 (1987).
- Rao, C.R.K., Trivedi, D.C. *Coord. Chem. Rev.* **249**, 613 (2005).
- Rieke, P.C., Baer, D.R., Fryxell, G.E., Engelhard, M.H., Porter, M.S. *J. Vac. Sci. Technol. A* **11**, 2292 (1993).
- Riley, H., Kendall, G.K., Zemicael, F.W., Smith, T.L., Yang, S.H. *Langmuir* **14**, 5147 (1998).
- Salaita, K., Wang, Y.H., Fragala, J., Vega, R.A., Liu, C., Mirkin, C.A. *Angew. Chem. Int. Edit.* **45**, 7220 (2006).
- Sawada, S., Masuda, Y., Zhu, P.X., Koumoto, K. *Langmuir* **22**, 332 (2006).
- Schneider, T.W., Buttry, D.A. *J. Am. Chem. Soc.* **115**, 12391 (1993).
- Schreiber, F. *Prog. Surf. Sci.* **65**, 151 (2000).
- Schwartzman, M., Sidorov, V., Ritter, D., Paz, Y. *J. Vac. Sci. Technol. B* **21**, 148 (2003).
- Seo, E.K., Lee, J.W., Sung-Suh, H.M., Sung, M.M. *Chem. Mater.* **16**, 1878 (2004).
- Service, R.F. *Science* **302**, 556 (2003).
- Sheen, C.W., Shi, J.X., Martensson, J., Parikh, A.N., Allara, D.L. *J. Am. Chem. Soc.* **114**, 1514 (1992).
- Shi, J.J., Chen, J.X., Cremer, P.S. *J. Am. Chem. Soc.* **130**, 2718 (2008).
- Shuster, M.J., Vaish, A., Szapacs, M.E., Anderson, M.E., Weiss, P.S., Andrews, A.M. *Adv. Mater.* **20**, 164 (2008).
- Sinha, A., Hess, D.W., Henderson, C.L. *J. Electrochem. Soc.* **153**, G465 (2006).
- Smith, R.K., Lewis, P.A., Weiss, P.S. *Prog. Surf. Sci.* **75**, 1 (2004).
- Smith, R.K., Reed, S.M., Lewis, P.A., Monnell, J.D., Clegg, R.S., Kelly, K.F., Bumm, L.A., Hutchison, J.E., Weiss, P.S. *J. Phys. Chem. B* **105**, 1119 (2001).
- Speets, E.A., Riele, P.T., Van Den Boogaart, M.A.F., Doeswijk, L.M., Ravoo, B.J., Rijnders, G., Brugger, J., Reinhoudt, D.N., Blank, D.H.A. *Adv. Funct. Mater.* **16**, 1337 (2006).
- Srinivasan, C., Anderson, M.E., Jayaraman, R., Weiss, P.S., Horn, M.W. *Microelectron. Eng.* **83**, 1517 (2006).
- Srinivasan, C., Hohman, J.N., Anderson, M.E., Weiss, P.S., Horn, M.W. *Appl. Phys. Lett.* **93**, 083123 (2008).
- Stranick, S.J., Parikh, A.N., Tao, Y.T., Allara, D.L., Weiss, P.S. *J. Phys. Chem.* **98**, 7636 (1994).
- Su, M., Liu, X.G., Li, S.Y., Dravid, V.P., Mirkin, C.A. *J. Am. Chem. Soc.* **124**, 1560 (2002).
- Sugimura, H., Hanji, T., Takai, O., Masuda, T., Misawa, H. *Electrochim. Acta* **47**, 103 (2001).
- Sugimura, H., Nakagiri, N. *J. Vac. Sci. Technol. B* **15**, 1394 (1997).
- Sugimura, H., Okiguchi, K., Nakagiri, N., Miyashita, M. *J. Vac. Sci. Technol. B* **14**, 4140 (1996).
- Sugimura, H., Takai, O., Nakagiri, N. *J. Vac. Sci. Technol. B* **17**, 1605 (1999a).
- Sugimura, H., Takai, O., Nakagiri, N. *J. Electroanal. Chem.* **473**, 230 (1999b).

- Suh, D., Simons, J.K., Taylor, J.W., Koloski, T.S., Calvert, J.M. *J. Vac. Sci. Technol. B* **11**, 2850 (1993).
- Sun, S., Leggett, G.J. *Nano Lett.* **2**, 1223 (2002).
- Sun, S.Q., Chong, K.S.L., Leggett, G.J. *J. Am. Chem. Soc.* **124**, 2414 (2002).
- Sun, S.Q., Chong, K.S.L., Leggett, G.J. *Nanotechnology* **16**, 1798 (2005).
- Sun, S.Q., Montague, M., Critchley, K., Chen, M.S., Dressick, W.J., Evans, S.D., Leggett, G.J. *Nano Lett.* **6**, 29 (2006).
- Suntola, T., Antson, J. *Method and Apparatus for Producing Compound Thin Film Compound*, Finland, FI347374 (1974).
- Suntola, T., Antson, J. *Method for Producing Compound Thin Films* (United States of America, 4058430, 1977).
- Suntola, T., Hyvarinen, J. *Annu. Rev. Mater. Sci.* **15**, 177 (1985).
- Suntola, T., Pakkala, A., Lindfors, S. *Method for Performing Growth of Compound Thin Films* (United States of America, 4413022, 1983).
- Swalen, J.D., Allara, D.L., Andrade, J.D., Chandross, E.A., Garoff, S., Israelachvili, J., McCarthy, T.J., Murray, R., Pease, R.F., Rabolt, J.F., Wynne, K.J., Yu, H. *Langmuir* **3**, 932 (1987).
- Tagge, C.D., Simpson, R.D., Bergman, R.G., Hostetler, M.J., Girolami, G.S., Nuzzo, R.G. *J. Am. Chem. Soc.* **118**, 2634 (1996).
- Tanaka, H., Anderson, M.E., Horn, M.W., Weiss, P.S. *Jpn. J. Appl. Phys.* **2**, 43, L950 (2004).
- Tanzer, T.A., Bohn, P.W., Roshchin, I.V., Greene, L.H., Klem, J.F. *Appl. Phys. Lett.* **75**, 2794 (1999).
- Tarlov, M.J. *Langmuir* **8**, 80 (1992).
- Tarlov, M.J., Burgess, D.R.F., Gillen, G. *J. Am. Chem. Soc.* **115**, 5305 (1993).
- Tarlov, M.J., Newman, J.G. *Abst. Papers. Am. Chem. Soc.* **204**, 76 (1992).
- Tiberio, R.C., Craighead, H.G., Lercel, M., Lau, T., Sheen, C.W., Allara, D.L. *Appl. Phys. Lett.* **62**, 476 (1993).
- Tien, J., Xia, Y., Whitesides, G.M. Microcontact Printing of SAMs. In Ulman, A. (ed.), *Thin films* (San Diego, Academic Press, 1998).
- Tighe, T.B., Daniel, T.A., Zhu, Z.H., Uppili, S., Winograd, N., Allara, D.L. *J. Phys. Chem. B* **109**, 21006 (2005).
- Ulman, A. *Chem. Rev.* **96**, 1533 (1996).
- Vettiger, P., Despont, M., Drechsler, U., Durig, U., Haberle, W., Lutwyche, M.I., Rothuizen, H.E., Stutz, R., Widmer, R., Binnig, G.K. *IBM J. Res. Devel.* **44**, 323 (2000).
- Walker, A.V., Tighe, T.B., Cabarcos, O., Haynie, B.C., Allara, D.L., Winograd, N. *J. Phys. Chem. C* **111**, 765 (2007).
- Walker, A.V., Tighe, T.B., Cabarcos, O.M., Reinard, M.D., Haynie, B.C., Uppili, S., Winograd, N., Allara, D.L. *J. Am. Chem. Soc.* **126**, 3954 (2004a).
- Walker, A.V., Tighe, T.B., Haynie, B.C., Uppili, S., Winograd, N., Allara, D.L. *J. Phys. Chem. B* **109**, 11263 (2005).
- Walker, A.V., Tighe, T.B., Reinard, M.D., Haynie, B.C., Allara, D.L., Winograd, N. *Chem. Phys. Lett.* **369**, 615 (2003).
- Walker, A.V., Tighe, T.B., Stapleton, J.J., Haynie, B.C., Upilli, S., Allara, D.L., Winograd, N. *Appl. Phys. Lett.* **84**, 4008 (2004b).

- Weckenmann, U., Mittler, S., Kramer, S., Aliganga, A.K.A., Fischer, R.A. *Chem. Mater.* **16**, 621 (2004).
- Wei, J.H., Ginger, D.S. *Small* **3**, 2034 (2007).
- Weimann, T., Geyer, W., Hinze, P., Stadler, V., Eck, W., Gölzhäuser, A. *Microelectron. Eng.* **57–58**, 903 (2001).
- Weinberger, D.A., Hong, S., Mirkin, C.A., Wessels, B.W., Higgins, T.B. *Adv. Mater.* **12**, 1600 (2000).
- Weiss, J., Himmel, H.J., Fischer, R.A., Woll, C. *Chem. Vap. Depos.* **4**, 17 (1998).
- Whelan, C.S., Lercel, M.J., Craighead, H.G., Seshadri, K., Allara, D.L. *Appl. Phys. Lett.* **69**, 4245 (1996).
- Wilbur, J.L., Kumar, A., Kim, E., Whitesides, G.M. *Adv. Mater.* **6**, 600 (1994).
- Winter, C., Weckenmann, U., Fischer, R.A., Kashammer, J., Scheumann, V., Mittler, S. *Chem. Vap. Depos.* **6**, 199 (2000).
- Wohlfart, P., Weiss, J., Kashammer, J., Kreiter, M., Winter, C., Fischer, R., Mittler-Neher, S. *Chem. Vap. Depos.* **5**, 165 (1999a).
- Wohlfart, P., Weiss, J., Kashammer, J., Winter, C., Scheumann, V., Fischer, R.A., Mittler-Neher, S. *Thin Solid Films* **340**, 274 (1999b).
- Woodson, M., Liu, J. *Phys. Chem. Chem. Phys.* **9**, 207 (2007).
- Wu, X.C., Bittner, A.M., Kern, K. *Langmuir* **18**, 4984 (2002).
- Wu, X.C., Bittner, A.M., Kern, K. *Adv. Mater.* **16**, 413 (2004).
- Xia, Y., Kim, E., Whitesides, G.M. *J. Electrochem. Soc.* **143**, 1070 (1996a).
- Xia, Y., Kim, E., Mrksich, M., Whitesides, G.M. *Chem. Mater.* **8**, 601 (1996b).
- Xia, Y., Mrksich, M., Kim, E., Whitesides, G.M. *J. Am. Chem. Soc.* **117**, 9576 (1995).
- Xia, Y., Rogers, J.A., Paul, K.E., Whitesides, G.M. *Chem. Rev.* **99**, 1823 (1999).
- Xia, Y., Whitesides, G.M. *Angew. Chem., Int. Ed.* **37**, 551 (1998).
- Xia, Y., Zhao, X.M., Whitesides, G.M. *Microelectron. Eng.* **32**, 255 (1996c).
- Xie, X., Chung, H.J., Sow, C.H., Wee, A.T.S. *Mater. Sci. Eng. R* **54**, 1 (2006).
- Yamamoto, H., Butera, R.A., Gu, Y., Waldeck, D.H. *Langmuir* **15**, 8640 (1999).
- Yan, M., Koide, Y., Babcock, J.R., Markworth, P.R., Belot, J.A., Marks, T.J., Chang, R.P.H. *Appl. Phys. Lett.* **79**, 1709 (2001).
- Yang, H.J., Lee, J., Kim, S., Ko, Y.K., Shin, H.J., Lee, J.G., Kim, C., Sung, M.M., Bang, H.J., Cho, B.S., Bae, Y.H., Lee, J.H., Kim, D.H., Jeong, C.O., Kim, S.Y., Lim, S.K. *J. Appl. Phys.* **100** (2006a).
- Yang, Y.C., Yau, S.L., Lee, Y.L. *J. Am. Chem. Soc.* **128**, 3677 (2006b).
- Yu, J.J., Tan, Y.H., Li, X., Kuo, P.K., Liu, G.Y. *J. Am. Chem. Soc.* **128**, 11574 (2006).
- Zangmeister, C.D., Van Zee, R.D. *Langmuir* **19**, 8065 (2003).
- Zaumseil, J., Meitl, M.A., Hsu, J.W.P., Acharya, B.R., Baldwin, K.W., Loo, Y.-L., Rogers, J.A. *Nano Lett.* **3**, 1223 (2003).
- Zhang, Y.M., Terrill, R.H., Tanzer, T.A., Bohn, P.W. *J. Am. Chem. Soc.* **120**, 2654 (1998).
- Zharnikov, M., Frey, S., Heister, K., Grunze, M. *Langmuir* **16**, 2697 (2000).
- Zharnikov, M., Geyer, W., Gölzhäuser, A., Frey, S., Grunze, M. *Phys. Chem. Chem. Phys.* **1**, 3163 (1999).
- Zharnikov, M., Grunze, M. *J. Vac. Sci. Technol. B* **20**, 1793 (2002).

- Zhou, C.Z., Nagy, G., Walker, A.V. *J. Am. Chem. Soc.* **127**, 12160 (2005).
- Zhou, C.Z., Walker, A.V. *Langmuir* **22**, 11420 (2006).
- Zhou, C.Z., Walker, A.V. *Langmuir* **23**, 8876 (2007).
- Zhu, P.X., Masuda, Y., Koumoto, K. *J. Mater. Chem.* **14**, 976 (2004).
- Zhu, Z.H., Daniel, T.A., Maitani, M., Cabarcos, O.M., Allara, D.L., Winograd, N. *J. Am. Chem. Soc.* **128**, 13710 (2006).

# 24

# Laser applications in nanotechnology

*M.H. Hong*

24.1 Introduction	860
24.2 Pulsed laser ablation for nanomaterials synthesis	860
24.3 Laser as a heat source for device nanoprocessing	865
24.4 Laser surface nanopatterning with near-field and light-enhancement effects	868
24.5 Large-area parallel laser nanopatterning	879
24.6 Conclusions	884
References	885

## 24.1 Introduction

The laser is one of many mature and reliable manufacturing tools, which has been extensively applied in modern industries, such as for surface cleaning (Hong *et al.* 2002), 3D microstructure formation (Tsukada *et al.* 2005), IC package marking (Blokken 2007), die via-hole drilling (Okada 2000), CSBGA package singulation (Shi *et al.* 2001), and thin-film deposition (Chrisey *et al.* 1994). To cater for the ever-increasing demand from industries to fabricate smaller, faster and more functional micro-/nanodevices, nanoengineering technology is needed to push device feature size to less than 100 nm. Compared to electron-beam and focused ion beam processing, laser nanoengineering has the advantages of being a low-cost, high-speed process in air, vacuum or chemical environments and most importantly, it has the capability to fulfill flexible integration control (Metev 1998; Ready *et al.* 2001). In this chapter, laser nanotechnology in the following areas is featured: (1) pulsed laser ablation for nanomaterials synthesis; (2) laser nanoprocessing to make nanobumps for disk media nanotribology and anneal ultrashort PN junctions; (3) surface nanopatterning with near-field, and light-enhancement effects, and (4) large-area parallel laser nanopatterning by laser interference lithography and laser irradiation through a microlens array (MLA). It shows that the laser will continue to be one of the highly potential nanoengineering means in next-generation manufacturing.

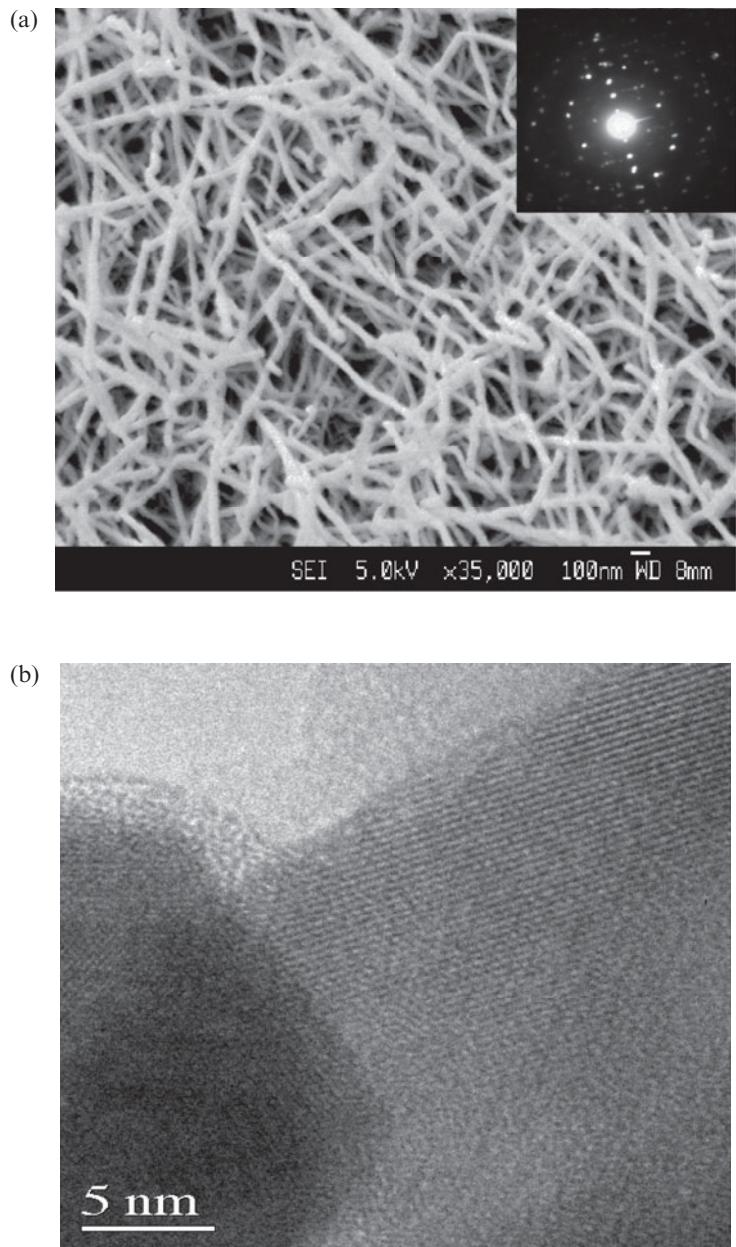
## 24.2 Pulsed laser ablation for nanomaterials synthesis

Nanomaterials have unique physical and chemical properties as the sizes are less than 20 nm. They are building blocks of nanodevices in the bottom-up fabrication approach (Hu *et al.* 1999). Different from the conventional lithography top-down approach, it is flexible to create a wider variety of functional nanodevices with the self-assembly of the same nanomaterials

or a combination of different nanomaterials. The key technical challenge is how to control the process for a narrow size distribution. Laser ablation is a flexible and straightforward approach in nanomaterials synthesis (Siegal *et al.* 2002; Chen *et al.* 2004a). Nanomaterials can be synthesized by pulsed laser ablation in vacuum or liquid environments. A laser beam with its pulse duration in femtosecond (fs) or nanosecond (ns) scales irradiates on the target surface. High-power and short-pulse laser interactions with target materials induce superheating of the materials during the pulse duration from solid state, through liquid state, gas state to plasma state because of the evaporated atoms' ionization. Key to the technology for the nanomaterials synthesis by laser ablation is to have proper tuning of laser processing parameters, such as laser fluence, target-to-substrate distance, substrate temperature and chamber pressure, to manipulate the nanomaterial size and control the size distribution.

### 24.2.1 Synthesis in vacuum

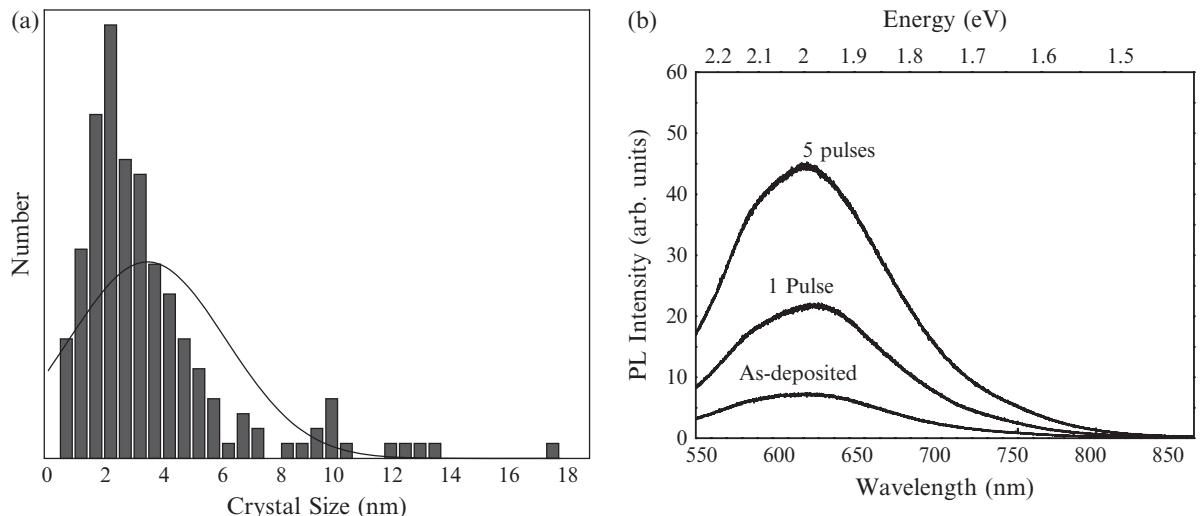
For the nanomaterials synthesis in a vacuum, the plasma species generated are in fast dynamics and collide with each other in an expanding high-temperature and high-pressure plasma. When the laser pulse turns off, the plasma expands freely inside a high-vacuum environment. Rapid cooling and nucleation of these atoms form nanomaterials with their sizes and distributions controlled by processing parameters, such as the laser fluence and vacuum pressure (Luk'yanchuk *et al.* 1999; Willmott *et al.* 2000). Figure 24.1 shows SEM and TEM images of GaN nanowires synthesized by 248 nm/30 ns KrF excimer laser ablation in a vacuum of  $10^{-6}$  Torr. They were grown uniformly in an area of  $2\text{ cm}^2$  with the diameter of  $\sim 20\text{ nm}$  and a length of  $\sim 5\text{ }\mu\text{m}$ . The lattice spacing is  $2.8\text{ \AA}$ . Pulsed laser ablation for the nanomaterials synthesis has the advantages of synthesizing pure nanomaterials and a wide scope of nanomaterial compositions as target materials can be changed and combined easily. Figure 24.2(a) shows Si nanoparticle sizes and their distributions after KrF excimer laser ablation inside a vacuum. The Si nanomaterials synthesized have a narrow distribution with the mean size of  $3.5\text{ nm}$  and peaks at  $2.5$  and  $10\text{ nm}$ . Since it is smaller than the Si exciton Bohr radius of  $4.3\text{ nm}$ , there is a strong quantum confinement effect for the Si nanomaterials. It can be observed from Fig. 24.2(b) that the photoluminescence (PL) spectrum peak moves to the visible region ( $\sim 620\text{ nm}$ ). It is a novel optical property of Si nanomaterials, different from its bulk substrate, which only emits infrared light with the energy bandgap of  $1.12\text{ eV}$ . It implies that Si nanomaterials could be used to make visible micro-/nanolasers on Si substrates. It can also be applied to achieve Si energy bandgap engineering to enhance Si solar-cell absorption capabilities (as sunlight is in a wide spectrum from  $300\text{ nm}$  to  $3\text{ }\mu\text{m}$ ) to increase solar-cell energy conversion efficiency. Figure 24.2(b) also shows that further laser annealing of the Si nanomaterials synthesized can enhance PL intensities but at the same peak position. It shows that this laser processing does not change the size distribution but activates more Si nanomaterials for stronger light emission.



**Fig. 24.1** (a) SEM and (b) TEM images of GaN nanowires fabricated by KrF excimer laser ablation in a vacuum pressure of  $10^{-6}$  Torr.

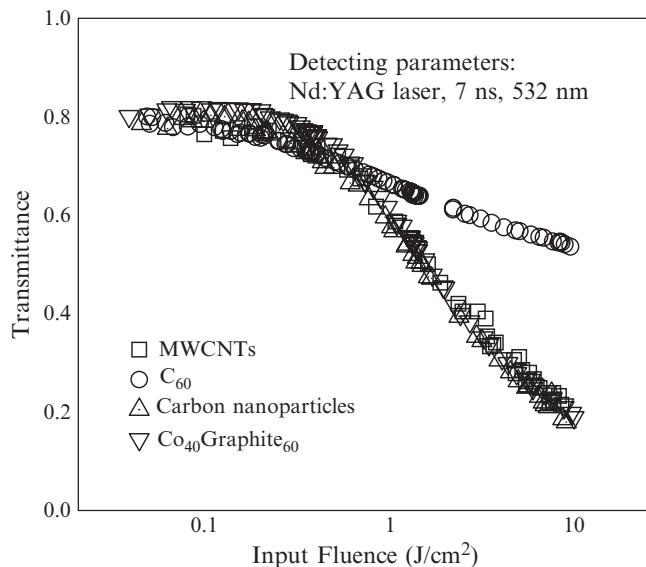
### 24.2.2 Synthesis in liquid

For the nanomaterials synthesis in a liquid environment, high-power and short-pulse laser-ablation-induced plasma cannot move freely with the strong interactions between plasma species and water molecules. High-temperature and high-pressure plasma is confined inside a small volume and increases the collisions among species atoms and liquid molecules. Typical nanomaterial



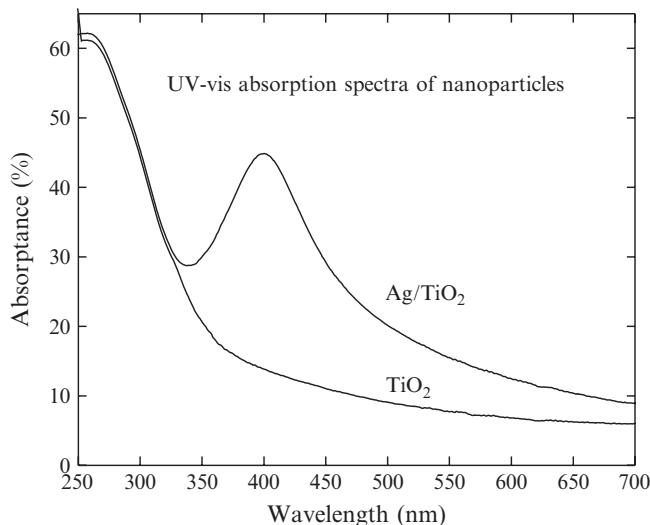
**Fig. 24.2** (a) Size distribution of Si nanoparticles fabricated by KrF excimer laser ablation in a vacuum and (b) PL spectra before and after 1 and 5 pulses of laser annealing at a laser fluence of  $50 \text{ mJ/cm}^2$  and pulse repetition rate of 1 Hz.

size and their distribution are much bigger than those grown in vacuum. However, pulsed laser ablation in the liquid for nanomaterials synthesis has the advantages of simple setup and low system cost with higher nanomaterials yield. To avoid the nanomaterials oxidation and aggregation, a chemical surfactant is generally added into the liquid during the nanomaterials synthesis (Chen *et al.* 2004b). Carbon nanoparticles were synthesized by 532 nm/7 ns Nd:YAG laser ablation of a glassy (vitreous) carbon plate ( $\sim 99\%$  pure) in water at a laser fluence of  $0.8 \text{ J/cm}^2$ . The low laser fluence selection is to avoid the target crack damage due to the water hammer effect for the laser ablation in the liquid (Zhu *et al.* 2001). It can be observed that the size of the nanoparticles ranges from 10 to 20 nm. It is very interesting to find that the non-linear optical-limiting properties of laser-synthesized carbon nanoparticles in liquid from the glassy carbon plate are very strong, using  $C_{60}$  as reference (Sun *et al.* 2000; Vivien *et al.* 2002). As shown in Fig. 24.3, at the same linear transparency of 80%, a more rapid decrease in transparency towards 532 nm laser light is observed for the synthesized carbon nanoparticles than that of  $C_{60}$  when the incident laser fluence is above  $0.6 \text{ J/cm}^2$ . Due to its excellent optical-limiting responses,  $C_{60}$  is generally taken as a benchmark in the research on the optical-limiting effect. However,  $C_{60}$  solution gives a colored appearance to the liquid. This effect introduces visual loss for the eyes. For carbon nanoparticle colloids with the same transmittance (80%) as that of  $C_{60}$ , the visual loss is notably reduced due to its wide spectrum property. On the other hand, carbon nanoparticle colloids have broadband transmittance. Furthermore, carbon nanoparticles show optical-limiting responses as strong as those of MWCNTs and  $\text{Co}_{40}\text{Graphite}_{60}$ , which are among the most widely studied highly efficient optical-limiters. It means that laser-synthesized carbon nanoparticles can be used as good optical-limiting materials.



**Fig. 24.3** Comparison of optical-limiting performances of carbon nanoparticles in water with C<sub>60</sub>, MWCNTs and Co<sub>40</sub>Graphite<sub>60</sub> towards 532 nm/7 ns laser beam (Chen *et al.* 2004b, Copyright American Institute of Physics).

Pulsed laser ablation synthesis of nanocomposites in the liquid can also be achieved with different material targets immersed in water. Ag-deposited TiO<sub>2</sub> nanoparticles were synthesized by 1064 nm/7 ns Nd:YAG laser ablation. Ag and Ti metal plates (~99.5% pure) were used for the laser ablation in the liquid one after the other. It can be found that the nanocomposites with ~30 nm TiO<sub>2</sub> and ~10 nm Ag nanoparticles can be formed after several minutes of laser irradiation. TiO<sub>2</sub> is one of the most widely studied semiconductors for heterogeneous photocatalysts. It is also one of the promising candidates for photoanode materials, which can split water in solar energy-conversion cells. However, the effect of photocatalysts is reduced for the photogenerated electron–hole recombination and the TiO<sub>2</sub> electrodes suffer from low solar energy-conversion efficiencies of less than 4%. Deposition of noble metals, such as Ag, Au and Pt, on TiO<sub>2</sub> surface is one of the effective methods to reduce electron–hole recombination in the photocatalytic process. It is because of the formation of Schottky barriers at each Ag/TiO<sub>2</sub> contact region. With this approach, the absorption capability of photocatalysts is enhanced. In the meantime, the noble metal creates conditions for additional absorption bands in visible-light spectra due to the surface plasmon resonance (SPR) effect (Liu *et al.* 2007). It gives an opportunity to extend the spectral range of the response of a TiO<sub>2</sub> photoanode to a much longer wavelength. Figure 24.4 shows the absorption spectra comparison of TiO<sub>2</sub> nanoparticles and Ag-deposited TiO<sub>2</sub> nanoparticles in water. It can be observed that Ag-deposited TiO<sub>2</sub> nanoparticles possess a special peak at 400–450 nm due to surface plasmon resonance (SPR) in the free-electron metal inclusion. Ag-deposited TiO<sub>2</sub> nanoparticles synthesis with pulsed laser ablation in water can extend the absorption range to the visible wavelength as compared with that of pure TiO<sub>2</sub> nanoparticles, which promotes the TiO<sub>2</sub> photoanode property greatly.



**Fig. 24.4** Absorption spectra for TiO<sub>2</sub> nanoparticles and Ag-deposited TiO<sub>2</sub> nanocomposites synthesized in liquid environment by 1064 nm/7 ns Nd:YAG laser ablation (Liu *et al.* 2007, Copyright Institute of Physics Publishing).

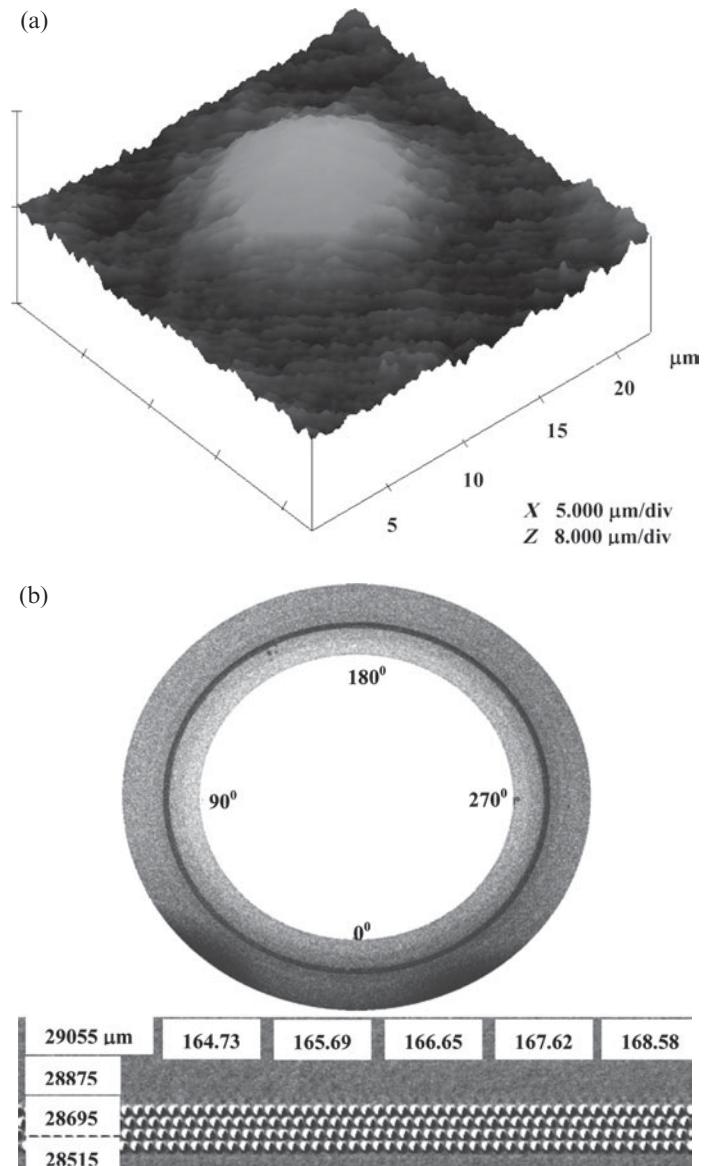
## 24.3 Laser as a heat source for device nanoprocessing

### 24.3.1 Laser nanobumping for disk media nanotribology

With strong demand from the consumer electronics market, information storage density is maintaining growth at a rapid rate, which requires many technology breakthroughs along the industry roadmap. Laser disk texturing and disk tagging have been running in the production lines for landing-zone fabrication and non-contamination marking of media surfaces. Ultrahigh density storage leads to the reduction of the spacing between read/write head and disk media. Surface roughness is critical with flying-height characterization of nanobumped disks. At an areal density above 100 Gbits/in<sup>2</sup>, disk bumping faces two technical challenges: (1) change of substrate materials from Al-Mg to glass as glass has superior mechanical properties in high-speed disk operation, and (2) fabrication of sub 10 nm height nanobumps. Figure 24.5 shows AFM and OSA images of nanobumps created on a glass disk by a CO<sub>2</sub> laser irradiation. Uniform bumps with their heights down to 3.5 nm were fabricated by laser heating. With fine tuning of the laser power, light absorption by glass induces plastic deformation and volume increase for the nanobump formation. Flying-height performance was tested by a phase metrics dynamic flying-height tester, which is based on a white-light interferometry technique. The light density changes as a head slider flies over the glass disk. Distinct peak signals are correlated to the bump array fabricated. The flying height was determined from the peak amplitude and width (Wang *et al.* 2007).

### 24.3.2 Laser annealing of ultrashort pn junction

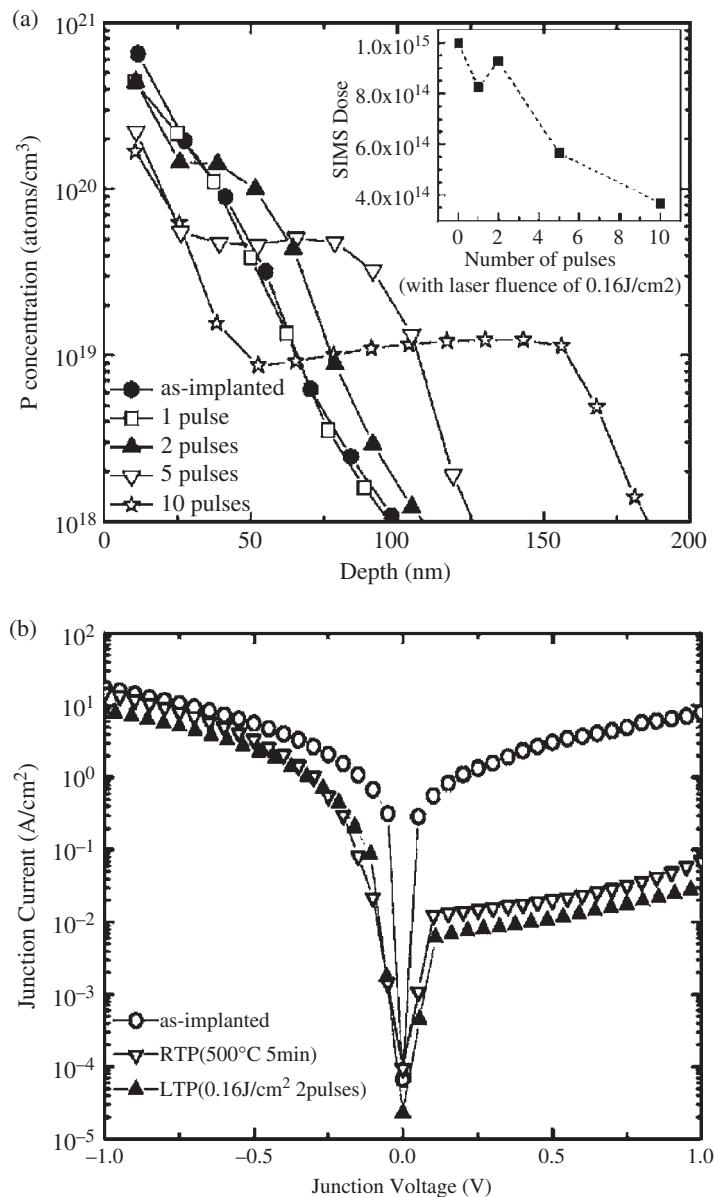
In the semiconductor industry, continual scaling of device dimensions presents great challenges in device fabrication to ensure its functionality. The pn junction is made shallower to prevent the short-channel effect. High electri-



**Fig. 24.5** AFM and OSA images of nanobumps on a glass disk by CO<sub>2</sub> laser irradiation (Wang *et al.* 2007, Copyright Institute of Physics Publishing).

cal activation is needed for maximum possible current through the device. Increase of junction lateral abruptness provides the desired performance. Heat generated by laser thermal annealing (LTA) causes a thin layer of silicon to melt. When a laser pulse terminates, molten silicon cools and recrystallizes back to crystalline silicon. Since dopants diffuse preferentially in liquid silicon, they redistribute uniformly within the melt region for an abrupt junction formation. LTA was applied to process P-implanted Ge-n<sup>+</sup>/p junctions. The P concentration distribution is closely related to processing parameters. As shown in Fig. 24.6(a), two pulses of LTA at a laser fluence of 160 mJ/cm<sup>2</sup>

achieve an optimal dopant distribution inside the junction. Compared with As-implantation and rapid thermal annealing (RTA), good rectifying diode characteristics were obtained by LTA, as shown in Fig. 24.6(b) (*Huang et al.* 2005).



**Fig. 24.6** (a) SIMS profile of P-implanted Ge substrates at a laser fluence of  $160 \text{ mJ/cm}^2$  and different pulses and (b)  $I$ - $V$  characteristics without and with RTA or LTA (*Huang et al.* 2005, Copyright American Institute of Physics).

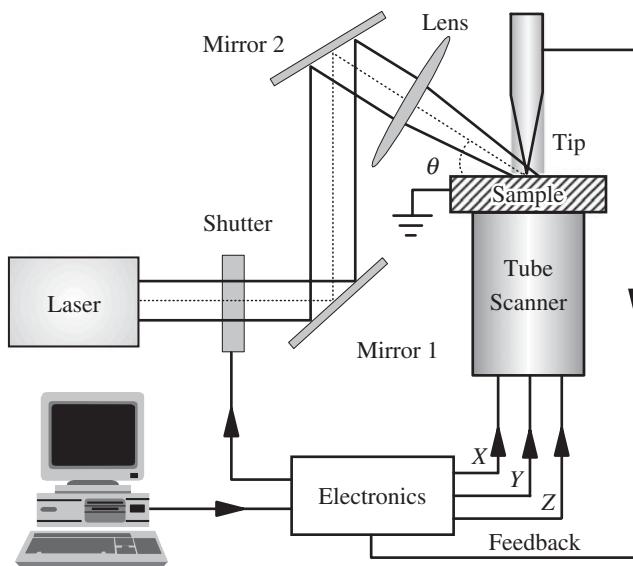
## 24.4 Laser surface nanopatterning with near-field and light-enhancement effects

### 24.4.1 Challenge for laser nanofabrication

Generally there are two approaches, top-down and bottom-up, to create functional nanodevices. The bottom-up approach is based on self-assembly of synthesized nanomaterials, which serve as building blocks to fabricate nanodevices. As the nanomaterials with sizes less than 5 nm can be fabricated by chemical or physical means, they can be used to make nanodevices with sizes less than 20 nm. However, it is extremely difficult in engineering to grow uniformly sized nanomaterials and also poor controllability to make devices with the size up to millimeter scale in a monolayer. Its application in nanomanufacturing is very limited. This is why industries prefer the top-down approach, such as surface lithography in wafer production lines with high NA optics and short-wavelength light sources. Since the smallest achievable resolution is limited by the light wavelength, deep ultraviolet (DUV, KrF 248 nm and ArF 193 nm) laser sources are used in production lines for 130 nm and 90 nm lithography (45 nm with the optics immersed liquid approach). To achieve smaller surface patterning feature size, many researches have recently been carried out on F2 157 nm and EUV 13 nm nanolithography. However, these VUV and EUV laser sources have technical challenges of high cost, low output and unstable light intensity. Strong absorption of the light by air molecules requires the nanolithography to be in a vacuum or a dry high-purity N<sub>2</sub> gas protection chamber. How to break through the optical diffraction limit with mature, stable UV or visible light sources is attracting much research interest worldwide. In this chapter, our recently developed laser nanofabrication techniques will be introduced. They include: (1) combination of pulsed laser irradiation with a scanning probe microscopy (SPM) tip for 10 nm resolution nanopatterning; (2) femtosecond laser combination with near-field scanning optical microscopy (NSOM) for 20 nm resolution surface nanostructuring; (3) Laser irradiation through transparent microparticles for nanohole array creation, which leads to laser nanoimprinting as the next-generation nanolithography technology for industries and (4) laser interference lithography for high-speed and large-area periodic nanoline or nanodot array fabrication.

### 24.4.2 Laser combination with SPM

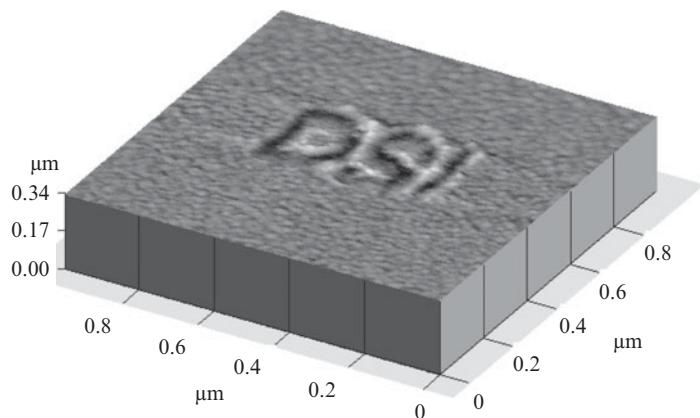
Pulsed laser irradiation is combined with a scanning probe microscopy (SPM) to achieve the surface nanopatterning. A nanosecond pulsed laser beam is introduced to a gap between the SPM tip and substrate surface. With the tip scanning over the substrate surface, 10 nm resolution nanolines and nanocharacters can be obtained on metal and photoresist surfaces. Figure 24.7 shows an experimental setup for laser surface nanopatterning combined with SPM. An experiment was carried out with a commercial scanning probe microscopy system (Autoprobe CP, Park Scientific Instruments). Its probe head has an open architecture, which allows an external laser-beam irradiation onto the tip and



**Fig. 24.7** Experimental setup for laser surface nanopatterning combined with SPM.

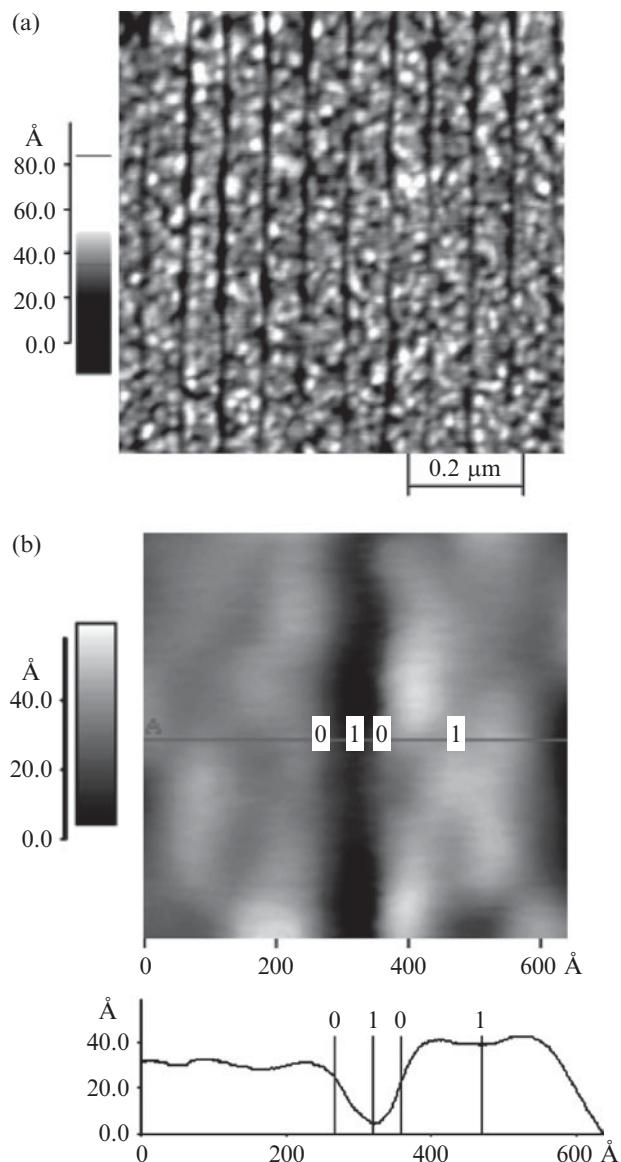
sample directly. A 532 nm Nd:YAG laser (B&I 5022 D.NS 10) with a pulse duration (FWHM) of 7 ns was applied as a light source. Laser pulse energy was adjusted with an attenuator installed inside the laser to control the laser fluence so that there is no damage on both the tip and substrate surfaces under the laser irradiation. During SPM imaging and surface nanopatterning, its tip was fixed and the sample was moved via a tube scanner. The laser beam hits the sample at an incidence angle around 80° to the normal of the substrate surface. SPM was operated in AFM mode and a boron-doped silicon tip (Park Scientific Instruments, 0.001 Ω cm) was used in contact mode.

An electronic shutter was applied to control laser pulses onto the system. The laser/SPM surface nanopatterning is flexible in the system control. With simple software programming to change the scanning path, different nanopatterns can be created on the substrate surfaces. Figure 24.8 presents an AFM image of 400 nm × 400 nm “DSI” characters created on a magnetic media surface.



**Fig. 24.8** AFM image of 400 nm × 400 nm “DSI” characters written on a magnetic media surface by laser/SPM surface nanopatterning.

Figure 24.9(a) shows an AFM image of a nanoline array fabricated by laser combination with SPM at a laser fluence of  $56 \text{ mJ/cm}^2$  and a scanning speed of  $0.02 \mu\text{m/s}$ . There are several uniform nanolines created on the surface with a period of  $100 \text{ nm}$ . With a higher scanning speed, nanodot arrays were observed on the surface due to less overlapping of nanopatterns generated by the laser nanoprocessing. Figure 24.9(b) presents the height profile of a nanoline fabricated. It demonstrates that laser/SPM surface nanopatterning can create a nanoline  $10 \text{ nm}$  wide and  $3.5 \text{ nm}$  deep. Depth and width of the nanolines

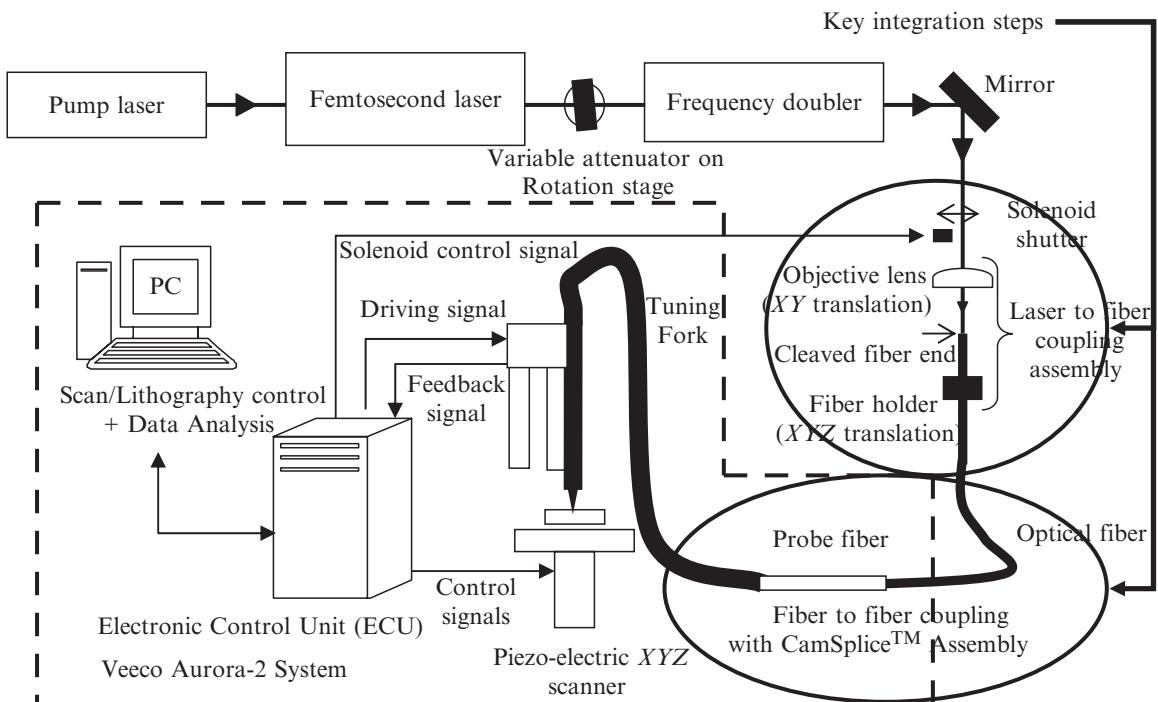


**Fig. 24.9** (a) AFM image and (b) height profile of nanolines fabricated by laser combination with SPM at a laser fluence of  $56 \text{ mJ/cm}^2$  and a scanning speed of  $0.02 \mu\text{m/s}$  (Huang *et al.* 2002a, Copyright American Institute of Physics).

depend on many factors, such as tip sharpness, scanning speed, processing time and laser fluence. The mechanism of laser/SPM surface nanopatterning is quite complicated (Huang *et al.* 2002a). One possible reason behind this result is that during the laser irradiation, the SPM tip and substrate surface absorb laser energy and induce the thermal expansion to process the substrate surface. The other possible mechanism is during the laser irradiation, the tip functions as an antenna. It enhances the electric field under the tip so as to etch away surface materials to form the nanopatterns on the surface. With a lower scanning speed, more scanning time and higher laser fluence, the nanostructure at a depth up to 10 nm can be obtained. This technique can be used to create different nanostructures on Si, Al and photoresists for device fabrication.

#### 24.4.3 Femtosecond laser combination with NSOM

Figure 24.10 shows the schematic drawing of NSOM and femtosecond laser integration for surface nanolithography. A laser light from a frequency-doubled Ti:sapphire femtosecond laser (Spectra-Physics, wavelength of 400 nm, pulse duration of 100 fs, pulse repetition rate of 80 MHz, and linearly polarized) was coupled into an NSOM (Model: Aurora-2, Veeco) tip through a 15-m fiber (Einst, PCS fiber). The NSOM tip (Vecco, Model 1640-00) used has an approximately 50 nm aperture ( $20^\circ$  cone angle) at the end. The transmission



**Fig. 24.10** NSOM and femtosecond laser integration for surface nanolithography.

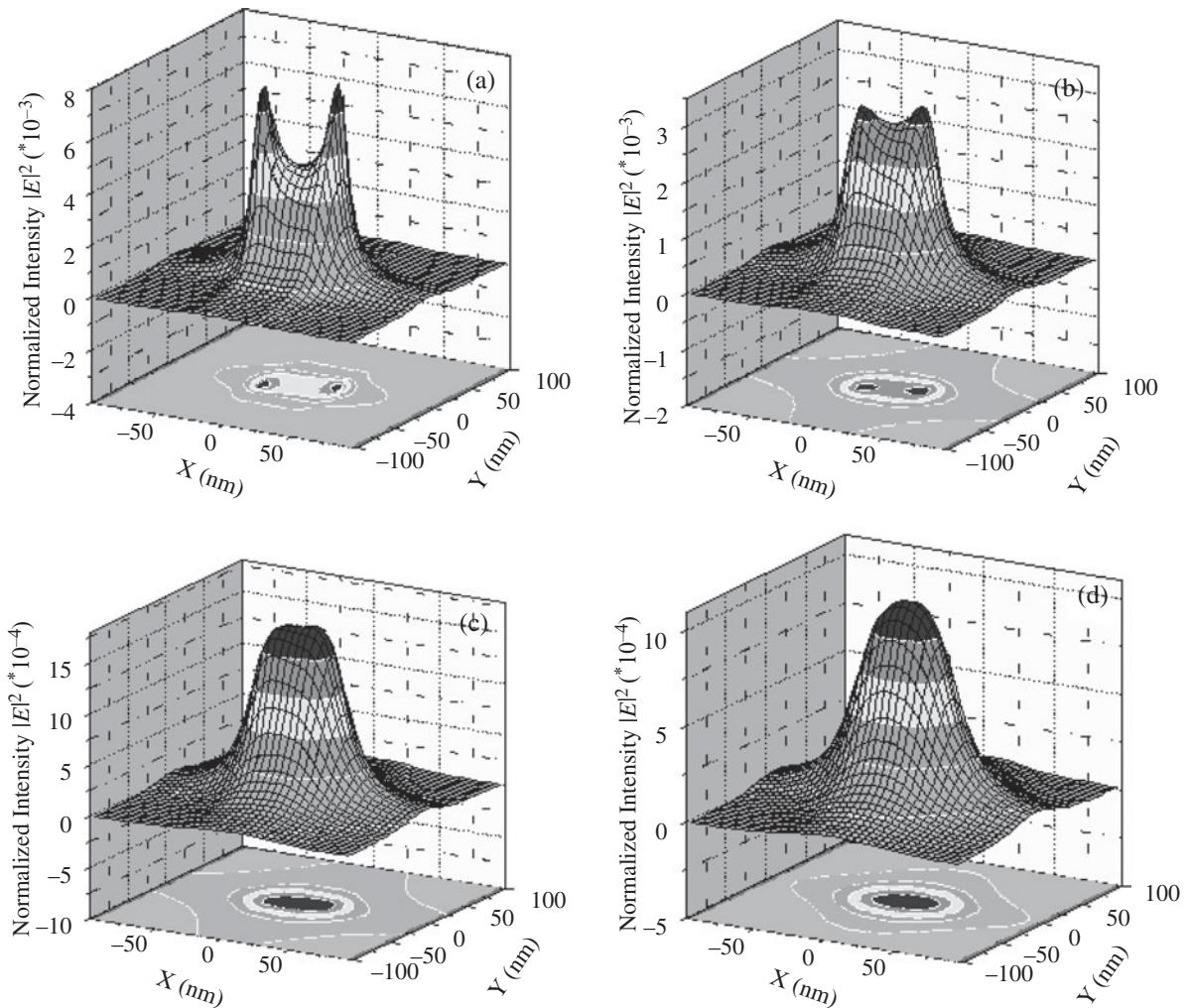
efficiency of such an aperture system is of the order of  $10^{-4}$ – $10^{-5}$ . The tip-sample distance was regulated by a tuning-fork-based shear-force detection feedback. One solenoid shutter was placed in the optical path to block off light, controlled by an 1-V signal from the control unit of the NSOM system. The output power from the fiber was measured by a thermopile broadband power meter (Melles Griot, 13PEM001/J) and tuned by an energy attenuator. After the laser writing, the sample was put in developer ma-D 331 for 20 s, washed in DI water, and then dried by nitrogen gas. The sample was characterized with scanning electron microscopy (SEM: Hitachi S4100) and the profiles were measured by the NSOM system operating in the AFM mode. The laser coming out of the NSOM tip is in the form of an evanescent wave, which decays quickly with increasing distance but contributes significantly to the sample process in near-field. In order to study its effect on the nanopattern fabrication, the evanescent-field distribution out of the tip is simulated according to the Bethie–Bouwkamp model for the field profile in the vicinity of the near-field probe.

Figure 24.11 presents the light-intensity distributions across an NSOM probe at probe-to-sample distances of (a) 5 nm, (b) 10 nm, (c) 15 nm and (d) 20 nm. It is clear that the light-intensity distribution varies greatly with different probe-to-sample distances. At a small distance (smaller than 10 nm), double-peak intensity is generated. As the distance increases, the double peak is replaced by a single-peak profile, but the peak intensity decreases (from  $10^{-3}$  to  $10^{-4}$ ). At a certain distance  $> 20$  nm, a Gaussian profile distribution is obtained. It indicates that the probe-to-sample distance is one of the critical parameters to control nanofeature size and even the shape.

At small aperture size and probe-to-substrate distance, the NSOM overcomes the traditional far-field diffraction limit and can be used to obtain subwavelength-size patterns. Figure 24.12 shows the nanoline arrays created at different incident laser powers. The scanning speed is  $6 \mu\text{m}/\text{s}$ . With a small laser power (around 0.1 mW), the linewidth can be reduced to less than 40 nm, which is smaller than the probe aperture size of 50 nm. Figure 24.13 is an integrated circuit metal-oxide-semiconductor (MOS) structure created at an incident laser power of 0.3 mW and a writing speed at  $2 \mu\text{m}/\text{s}$ . The gate width of the MOS structure is around 80 nm (Hong *et al.* 2007).

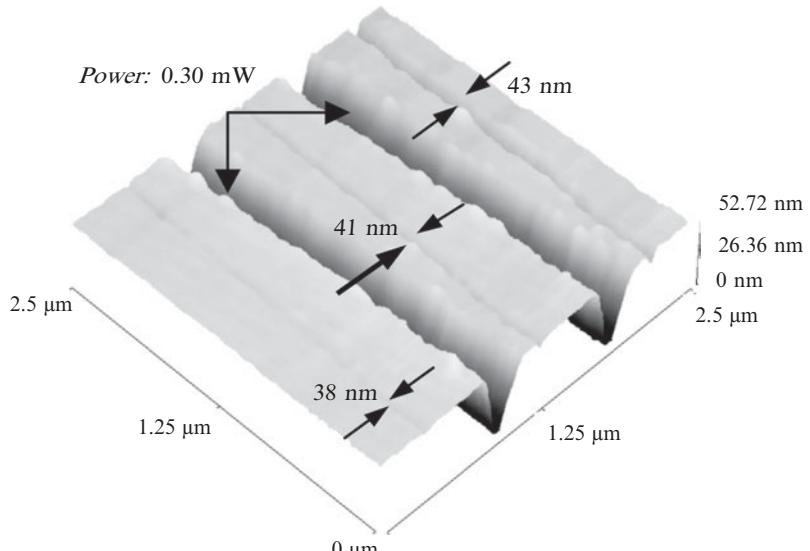
At a fixed probe-to-substrate distance, the exposure energy dose can be controlled by varying laser input power  $P$ , which was measured before coupling into the NSOM probe and laser writing speed  $V$ . Figures 24.14(a) and (b) show the dependence of the lithography depth and width on laser power and writing speed, respectively. At a fixed writing speed (such as  $8 \mu\text{m}/\text{s}$ ), both the depth and width increase with laser power, as shown in Fig. 24.14(a).

For the photoresist exposure process, exposure energy dose is another important parameter, which is determined by exposure energy and exposure time. When the laser power coupled to the NSOM probe is higher, the exposure energy dose (evanescent energy has effects in the near-field) absorbed by the photoresist film is higher. As a result, the area absorbing energy becomes larger along the surface and deeper along the propagating direction. Therefore, lithography structure size increases with laser power. The depth increases

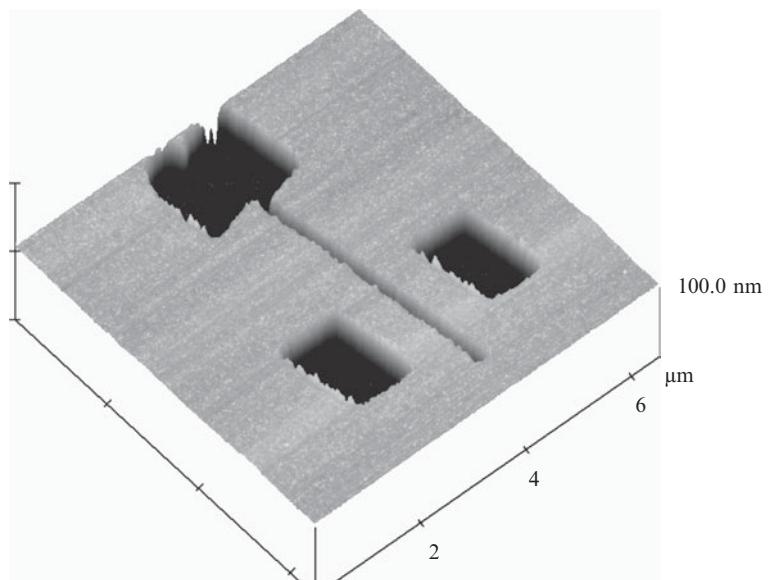


**Fig. 24.11** 3D and bottom contour images of light-intensity distributions across an NSOM probe at different probe-to-sample distances of (a) 5 nm, (b) 10 nm, (c) 15 nm and (d) 20 nm.

with power almost linearly when  $P < 0.3$  mW, but tends to saturate at about 120 nm, this is because of the thickness limit of the photoresist film. The width of the curve in Fig. 24.14(a), however, gives a more complicated evolution: Firstly, a linear dependence is observed at  $P < 0.15$  mW. At  $0.15 < P < 0.45$  mW, the slope of the increase slows down. But at  $P > 0.45$  mW, the slope increases faster and is similar to that in the first region ( $P < 0.15$  mW). On the other hand, the exposure time also affects the linewidth and depth because it affects the exposure energy dose as well. Actually, the exposure time is controlled by the writing speed in the experiment, except when fabricating dot patterns (in dot fabrication, the exposure time is determined by the writing delay, one parameter in the control software). At a fixed laser power of 0.35 mW, both the width and depth decrease with increasing writing speed (as shown in Fig. 24.14(b)). Higher writing speed leads to shorter exposure time

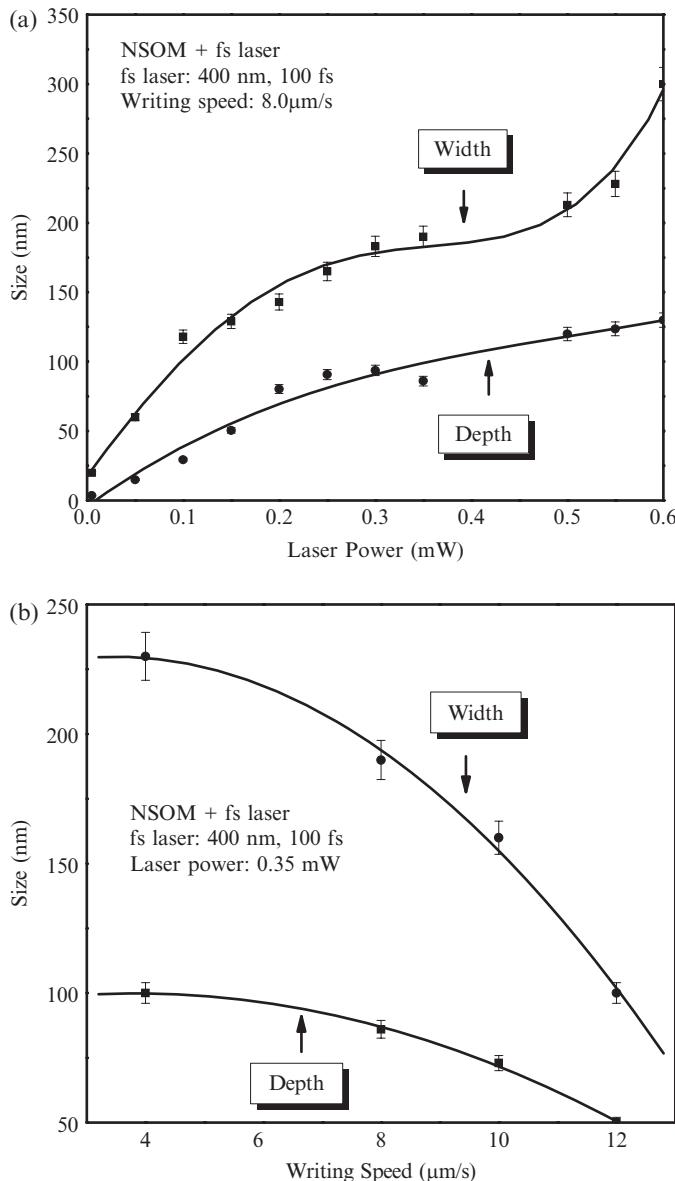


**Fig. 24.12** Nanolines created at different incident laser powers (Hong *et al.* 2007, Copyright Institute of Physics Publishing).



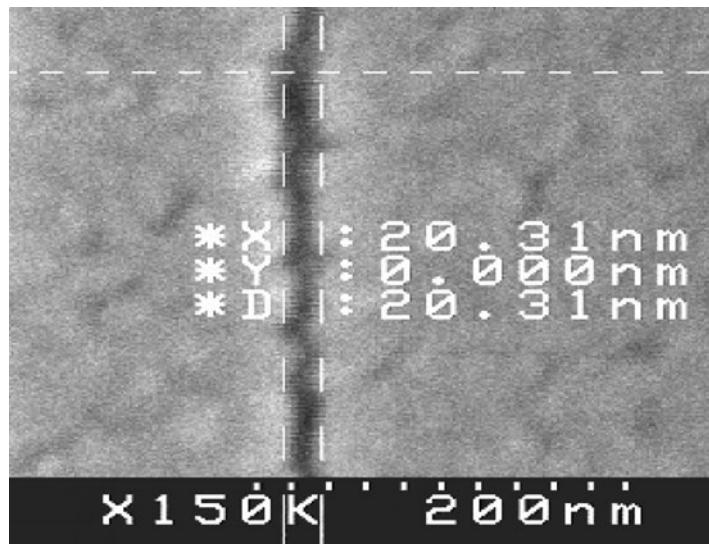
**Fig. 24.13** AFM image of a MOS nanostructure (Hong *et al.* 2007, Copyright Institute of Physics Publishing).

and then lower exposure dose, resulting in smaller linewidth and shallower depth. The different trends of depth and width with laser power and writing speed may be due to the laser field distribution out of the tip (Yin *et al.* 2002). According to Riehn *et al.* (2003), the steep profile of the electric field implies that the photoresist on the surface absorbs energy dose about a few tens of times of the photoresist under the surface. Therefore, the width of the patterns increases more quickly than the depth at a higher laser input



**Fig. 24.14** Dependence of lithography width and depth on (a) laser input power coupled into NSOM probe at a writing speed  $8.0 \mu\text{m/s}$  and (b) writing speed at a laser input power of  $0.35 \text{ mW}$  (Lin *et al.* 2005, Copyright Springer).

power. High exposure energy and long exposure time both increase exposure dose. When the exposure time is long enough, the same exposure dose can be achieved even at a lower input energy. Different energies and writing speeds can achieve similar effects on linewidth and depth. Considering there is a melting threshold of the NSOM tip metal coating, low energy is recommended in the nanofabrication to avoid damaging the NSOM tip. Through varying the writing speed, nanopatterns with different widths and depths can be achieved at a low laser input power. With the fine tuning of laser power and writing speed, sub-30 nm feature size can be achieved. Figure 24.15 shows a nanoline with

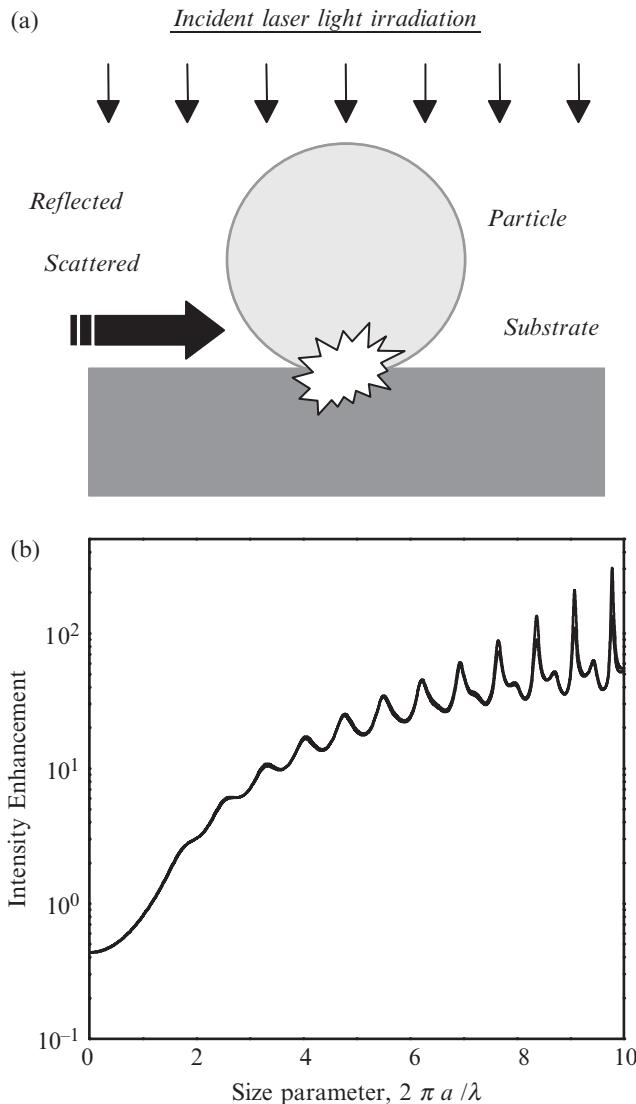


**Fig. 24.15** SEM image of a nanosize line created by the femtosecond laser and NSOM nanolithography (Lin *et al.* 2005, Copyright Springer).

a width of  $20 \pm 5$  nm created by the fs laser and NSOM integration system. It is at a resolution of  $\lambda/20$  ( $\lambda$ : laser wavelength) and  $d/2$  ( $d$ : NSOM probe aperture diameter), respectively. Such high resolution can be comparable with electron-beam lithography and be used for ultrahigh capacity data storage and new functional nanodevice fabrication. Furthermore, this small nanopattern was fabricated in air with only the light irradiation in near-field. It is not affected by electromagnetic wave interference, which degrades the electron-beam lithography resolution greatly. With a multi-NSOM fiber tip design, parallel nanolithography can be fulfilled for high-speed surface nano-structuring by the femtosecond laser integration with NSOM.

#### 24.4.4 Laser irradiation through transparent particles

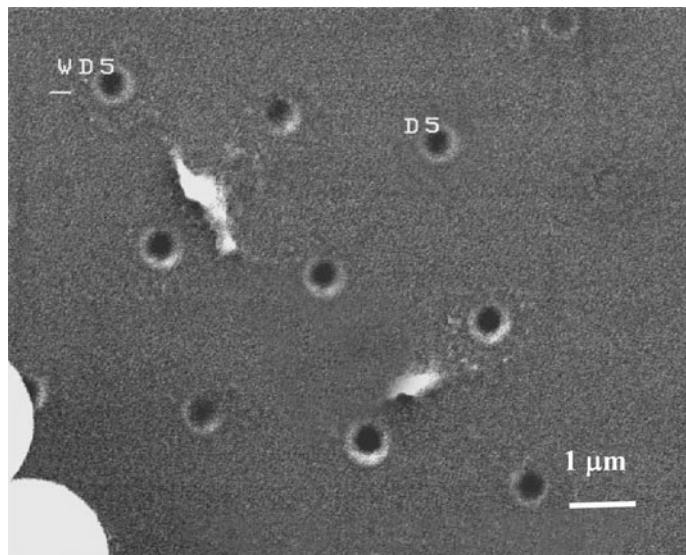
Another approach to break through the light diffraction limit is laser surface nanopatterning with transparent particle self-assembly as the lenses. In our previous study for laser cleaning of 500 nm particles away from Si substrate, it was found that there were many 100 nm hillock structures created on the substrate surface after the laser irradiation (Lu *et al.* 2000). This is due to laser light intensity enhancement near the contact area between the particle and substrate. Figure 24.16 shows (a) a diagram of laser irradiation on a transparent particle and (b) laser light intensity distribution calculated based on Mie theory. Since the distance between the particle and substrate is much smaller than the light wavelength (the particle sits on the surface) and particle size is smaller than or of the order of light wavelength, laser irradiation of the particle on the substrate is different from the situation as a sphere lens focusing in far-field. It is the optical resonance effect in the near-field (Kerker 1969). Besides the laser-processing parameters, surface-patterning results are also depended on particle size and light wavelength. Figure 24.16(b) indicates light enhancement under



**Fig. 24.16** (a) Diagram of laser irradiation on a transparent particle and (b) laser light intensity distribution under the particle calculated from Mie theory (Huang *et al.* 2002b, Copyright American Institute of Physics).

the particle during laser irradiation. For a 266 nm laser and 500 nm particle, light intensity can be enhanced up to 50 times at the particle center.

For particle nanopatterning, monodisperse polystyrene (PS) spheres were used to form a particle mask by self-assembly. They were dissolved into liquid and then disposed onto the sample, which was slightly titled, kept in a refrigerator at 10 °C for drying and finally baked in a vacuum oven at 80 °C for 10 min to remove water molecules. An hexagonally closed-packed colloidal monolayer of 1 μm PS particles (refractive index 1.6, 2% size deviation) was deposited on a 100 nm Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub> (GST)/Al thin-film surface. A KrF excimer laser (Lambda Physik LPX100, 248 nm, 23 ns FWHM) was applied to irradiate the sample. A beam homogenizer was used to provide a uniform light distribution over a spot of 25 × 5 mm. Each sample was irradiated with

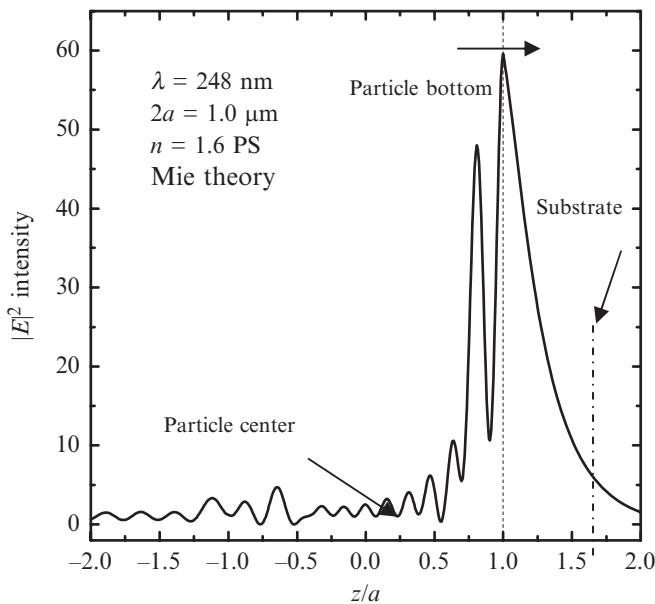


**Fig. 24.17** SEM image of nanohole arrays formed on a GST thin-film surface after one pulse of a KrF excimer laser irradiation at a laser fluence of  $5.8 \text{ mJ/cm}^2$  (Wang *et al.* 2004, Copyright Springer).

one laser pulse only. Figure 24.17 shows a SEM image of nanohole arrays formed on GST (melting point  $616^\circ\text{C}$ ) thin film after one pulse of KrF excimer laser irradiation at a laser fluence of  $5.8 \text{ mJ/cm}^2$ . It can be observed that bowl-shaped nanoholes have a uniform diameter of 120 nm. It is attributed to light intensity enhancement near the contact area between the transparent particles and substrate. It is also found that light enhancement increases as particle diameter increases and light wavelength decreases (Huang *et al.* 2003). However, the increase in particle size results in a larger light enhancement area and the fabricated nanoholes become bigger. It is not a good option for surface nanostructuring.

Meanwhile, due to the nature of the particle material, there is a limited selection for shorter-wavelength lasers. A 157 nm F<sub>2</sub> excimer laser is a shorter-wavelength light source next to 193 nm and 248 nm excimer lasers. But this laser light is strongly absorbed by most optical materials and even air. Therefore, in order to get a smaller light-enhancement region, small particle size needs to be selected with the sacrifice of a strong light-enhancement effect. A femtosecond laser is a potential light source to get much smaller nanohole sizes with the combination of light enhancement and multiphoton absorption effects. Nanoholes can also be fabricated on Al thin film and glass substrate with pulsed laser irradiation through transparent particles. It is clear that this technique can be applied to achieve surface nanopatterning on different functional material substrates for nanostructure fabrication. Meanwhile, there are no cracks generated along the nanoholes on the glass surface under the laser irradiation (Zhou *et al.* 2006).

Self-assembly of the particles on substrate is not a practical way in nanoengineering. It is difficult to form a mask with its size up to the millimeter scale. The other issue is during laser irradiation, transparent particles are removed from the surface. It can only be used for single-pulse irradiation.



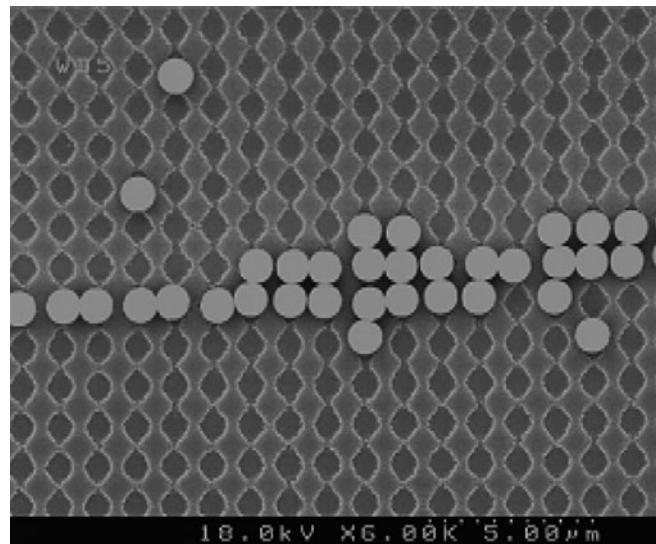
**Fig. 24.18** Laser light intensity along laser irradiation direction,  $z$ -axis ( $z = 1.0$ : particle–substrate contact point) during KrF excimer laser irradiation of  $1.0\mu\text{m}$  PS particle (Hong *et al.* 2007, Copyright Institute of Physics Publishing).

Meanwhile, redeposited particles contaminate the surface as well. A transparent mask fabricated on a quartz substrate is a feasible way to solve these problems. A bottom mask surface would be fabricated into ball shapes by lithography and chemical etching. As laser light goes through it, ball-shape transparent structures induce light enhancement and transfer the mask design to the substrate, which is a laser nanoimprinting process. One of the key advantages is that this fixed transparent mask can be used for multiple pulse irradiation to achieve deeper nanostructures. How to control the distance between the transparent mask and surface is another technical issue. Figure 24.18 depicts the light-intensity distribution along the laser irradiation direction ( $z$ -axis) during KrF excimer laser normal irradiation of  $1\mu\text{m}$  PS particle. It can be found that light intensity is enhanced up to 60 times under the particle ( $z/a = 1.0$ ) and decays very fast to zero at  $z/a = 2.0$  (500 nm from the contacting point). To ensure light enhancement at 25 times, the distance between the mask and substrate is set at 125 nm ( $z/a = 1.25$ ). To create this near-field environment, proper nanostage control is required to tune the mask-to-substrate distance. Meanwhile, to make the transparent mask, laser interference lithography was applied to form a pattern array for PS particles self-assembly, as shown in Fig. 24.19.

## 24.5 Large-area parallel laser nanopatterning

### 24.5.1 Laser-interference lithography

This technique is based on the interference of coherent lights to form a horizontal standing-wave pattern that can be recorded on photoresist (Farhoud *et al.* 1998; Cheng *et al.* 2002). For two-beam interference, the standing wave forms a grating pattern. The period  $PD$  is dependent on the wavelength of the

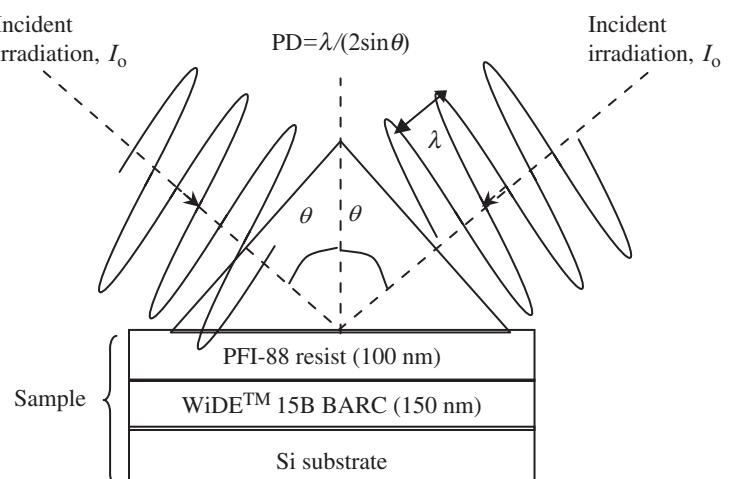


**Fig. 24.19** Transparent mask by the self-assembly of PS particles on pre-patterned surface structures created by laser interference lithography (Hong *et al.* 2007, Copyright Institute of Physics Publishing).

light  $\lambda$  and the half-angle where two incident beams intersect,  $\theta$ . By a 90° sample rotation and a second exposure, a grid pattern on the photoresist can be defined. After the nanopatterns are transferred by chemical etching or lift-off, this nanostructure can be used for the patterned media in high-density data storage, microsieves for microfiltration and submicrometer perforated membranes, nanotemplate for self-assembly and field emission flat-panel displays. Figure 24.20 shows the schematic drawing of two-beam interference. The horizontal standing-wave period can be calculated from:

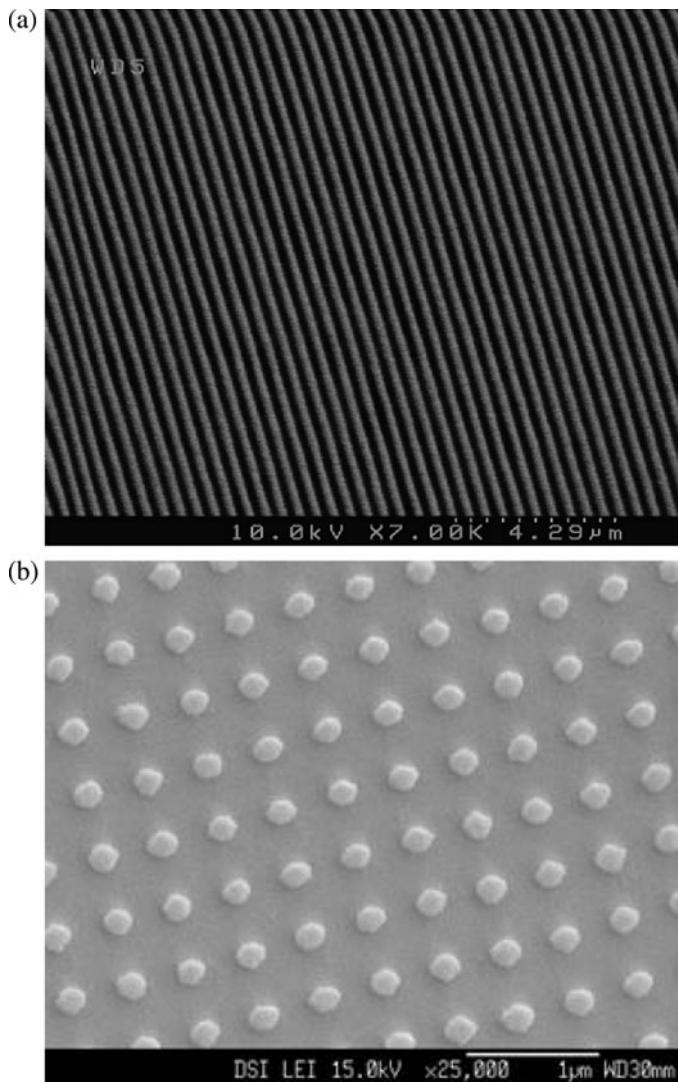
$$PD = \lambda / (2 \sin \theta),$$

where  $\theta$  is the half-angle of the two incident beams and  $\lambda$  the laser wavelength.



**Fig. 24.20** Standing wave of two-beam interference for pattern recording on photoresist.

Two layers of photoresist stack (PFI-88 and WiDE<sup>TM</sup> 15B) were coated on Si (100) substrate. A grating pattern was recorded on PFI-88 resist with a single exposure and a nanodot array pattern was recorded with the double exposure by a 90° sample rotation after the first exposure. A 15-mW He-Cd laser (Kimmon IK3151R-E, TEM00 mode) at 325 nm was used as the light source to irradiate into a Lloyd's interferometer. A spatial filter was applied to allow the high-frequency noise to be removed from the beam to provide a clean Gaussian profile. The centers of the mirror and substrate assembly remain on the optical axis. Half of the expanded beam is reflected by a mirror back onto itself to serve as two beams for recording patterns on the sample with the coated photoresist. Figure 24.21 shows the SEM images of nanograting and nanodot arrays fabricated on photoresist with laser-interference lithography. The dot



**Fig. 24.21** SEM images of (a) nanograting and (b) nanodot array fabricated on photoresist.

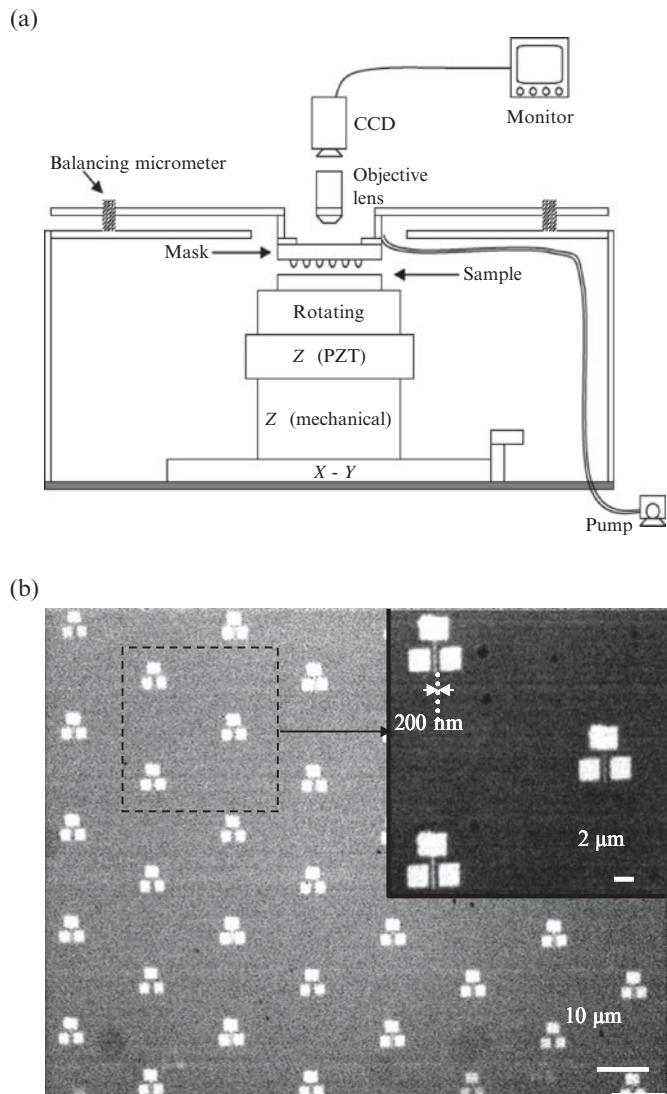
and line sizes can be reduced down to 100 nm as a shorter-wavelength laser is used (e.g. 248 nm/30 ns KrF excimer laser) and the trench width increased at a fixed period of the nanostructures. There is one critical issue in laser-interference lithography: to apply a back antireflective coating (BARC) and eliminate the vertical standing wave for the steep sidewall edge of high-quality nanostructures (Xie *et al.* 2008).

#### **24.5.2 Ultrafast laser-induced phase-change nanolithography**

Though sub-50 nm processing feature sizes can be obtained by laser combination with SPM, NSOM and transparent particles, the processing speed is slow (on the  $\mu\text{m}/\text{s}$  scale). This limits its applications in the electronic industry. Laser nanoimprinting by light irradiation through a microlens array (MLA) is explored. Microlenses can convert the laser beam into thousands of small focal points, which act as an array of “light pens” to fabricate tiny structures uniformly over a large area at a high speed (Wu *et al.* 2002; Kato *et al.* 2005). Figure 24.22(a) shows the experimental setup. This is a non-contact parallel laser nanopatterning process in far-field. Figure 24.22(b) is the optical image of field emission transistor (FET) structure fabricated on GeSbTe thin film by 800 nm/100 fs femtosecond laser irradiation through a  $23\mu\text{m}$  diameter MLA (laser power: 200 mW, scanning speed:  $300\mu\text{m}/\text{min}$ ). The inset is the enlarged image of the features. It shows that 160 000 FETs with the gate width of 200 nm can be created in 3 min on an area of  $1\text{ cm} \times 1\text{ cm}$  (Lin *et al.* 2006).

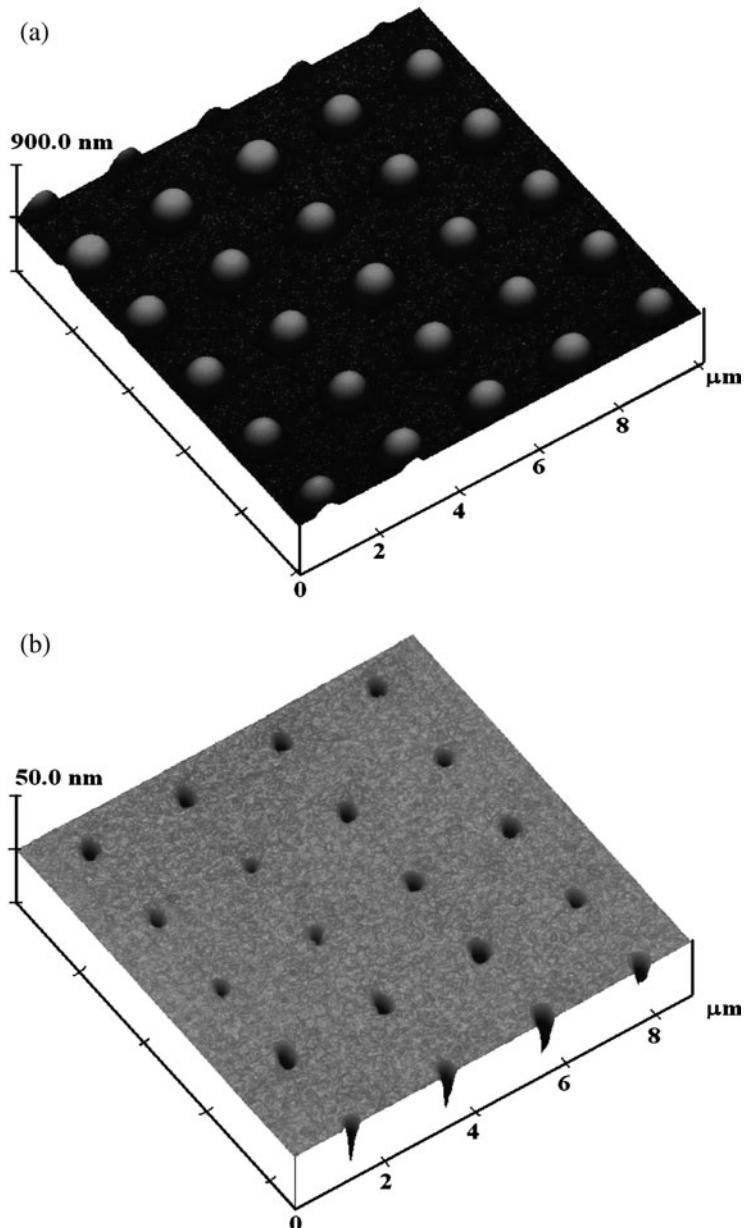
#### **24.5.3 Super-resolution laser nanoimprinting through microlens array (MLA)**

To obtain smaller feature sizes, a smaller-diameter MLA was fabricated for a shorter focal length. S1805 PR (Shipley) was spin coated on quartz at a speed of 2000 rpm to obtain a thickness of 800 nm. The sample was exposed by the interference of two 325 nm He-Cd laser beams, resulting in the generation of a periodic structure. The structure period is determined by the light wavelength and angle of two beams interfered. A line array periodic structure was formed after a single exposure. The subsequent cross-exposure by rotating the sample by 90 degrees enables the formation of a dot array. Exposed PR was then passing through a reflow process at a temperature of  $150^\circ\text{C}$ . Due to surface tension in melted PR, sample surface and surrounding air, the PR dots changed the shapes into hemispheres to minimize the surface energy. Finally, the patterns on the PR were transferred into quartz substrate by reactive ion etching in  $\text{CF}_4$  gas. Figure 24.23(a) shows an AFM image of a MLA fabricated by laser-interference lithography (LIL), reflow and chemical etching (Lim *et al.* 2006). Microlenses are aligned in parallel (diameter:  $1\mu\text{m}$ , height:  $330\text{ nm}$ , spacing:  $2\mu\text{m}$ ). Figure 24.23(b) shows an AFM image of dot arrays patterned with one pulse of 248 nm/30 ns KrF excimer laser irradiation of the PR on Si substrate through the MLA. By tuning the sample at the focal plane, uniform



**Fig. 24.22** (a) Laser nanoimprinting setup and (b) optical image of FET structures fabricated on GeSbTe thin film by fs laser and a 23 μm diameter MLA. The inset is the enlarged image of the features (Lin *et al.* 2006, Copyright American Institute of Physics).

dot arrays with a spacing of 2 μm were formed. Spot sizes are different. This is due to different sample-to-MLA distances for the microlenses. As the focal length is 1.1 μm, a small difference in sample-to-microlens distance makes the microlenses out of focus. High-resolution sample-alignment and position-checking systems are needed to ensure all the focused spots at a same focal plane. An interesting result is that some spot sizes are 78 nm with a depth of 10 nm. This is  $\lambda/3$  super-resolution. It is because only the tip of the laser profile has enough photon dose to trigger laser-induced photochemical reaction of the photoresist.



**Fig. 24.23** AFM images of (a) MLA fabricated by LIL and (b) dot arrays patterned with one pulse of KrF excimer laser irradiation of the photoresist (Lim *et al.* 2006, Copyright American Institute of Physics).

## 24.6 Conclusions

Pulsed laser ablation can be applied to synthesize sub-20 nm nanomaterials in either vacuum or a liquid environment. It has the advantages of simple setup and pure nanomaterials composition are produced. Nanomaterials obtained can be used as the building blocks for future nanodevices. Meanwhile, a laser can also be used as a unique heat source to form sub-10 nm high

nanobumps on glass substrate for magnetic media tribology and can confine the heat dissipation as well in a sub-20 nm range for ultrashort PN junction annealing. Among the nanoscale patterning techniques, the non-contact nature of laser beam irradiation makes it a potentially low-cost operation method over electron-beam and focused ion beam irradiation. The other key factor is that laser processing does not need a high-vacuum environment, which can provide fast speed and large-area nanoscale patterning. Laser-interference lithography is a fast and large-area nanoengineering approach with the two laser beam interference generating a standing wave on the photoresist surface. In order to break the optical limit in far-field, ultrafast laser-induced multiphoton absorption and laser combined with SPM, NSOM and transparent particles techniques were investigated to push the feature size down to 10 nm. Laser irradiation through a microlens array can split a laser beam into hundreds of thousands of tiny beams. It functions as a pen array for parallel direct writing to make complicated geometry nanodevices on the substrate surface in a short time. All these results show that lasers will find extensive applications from current industrial microfabrication to nanomanufacturing in the near future.

## References

- Blokken, E., *Laser Focus World* **43**(7) (2007).
- Chen, G.X., Hong, M.H., Ong, T.S., Lam, M., Chen, W.Z., Elim, H.I., Ji, W., Chong, T.C. *Carbon* **42**, 2735 (2004a).
- Chen, G.X., Hong, M.H., Chong, T.C., Elim, H.I., Ma, G.H., Ji, W. *J. Appl. Phys.* **95**, 1455 (2004b).
- Cheng, J.Y., Ross, C.A., Thomas, E.L., Smith, H.I., Lammertink, R.G.H., Vancso, G.J. *IEEE Trans. Magn.* **38**, 2541 (2002).
- Christey, D.B., Hubler, G.K. *Pulsed Laser Deposition of Thin Films* (John Wiley & Sons, New York, 1994).
- Farhoud, M., Huang, M., Smith, H.I., Schattenburg, M.L., Bae, J.M., Youcef-Toumi, K., Ross, C.A. *IEEE Trans. Magn.* **34**, 1087 (1998).
- Hong, M.H., Song, W.D., Lu, Y.F., Luk'yanchuk, B., Chong, T.C. *Laser Cleaning*, (ed.) Luk'yanchuk, B. (World Scientific Publishing, Singapore, 2002) 433.
- Hong, M.H., Lin, Y., Chen, G.X., Tan, L.S., Xie, Q., Luk'yanchuk, B., Shi, L.P., Chong, T.C. *J. Phys.: Conf. Ser.* **59**, 64 (2007).
- Hu, J., Odom, T.W., Lieber, C.M. *Acc. Chem. Res.* **32**, 435 (1999).
- Huang, J.D., Wu, N., Zhang, Q.C., Zhu, C.X., Tay, A.A.O., Hong, M.H. *Appl. Phys. Lett.* **87**, 173507 (2005).
- Huang, S.M., Hong, M.H., Luk'yanchuk, B.S., Chong, T.C. *J. Appl. Phys.* **91**, 3268 (2002a).
- Huang, S.M., Hong, M.H., Luk'yanchuk, B.S., Chong, T.C. *J. Appl. Phys.* **91**, 2495 (2002b).
- Huang, S.M., Hong, M.H., Luk'yanchuk, B.S., Chong, T.C. *Appl. Phys. A* **77**, 293 (2003).
- Kato, J., Takeyasu, N., Adachi, Y., Sun, H.B., Kawata, S. *Appl. Phys. Lett.* **86**, 044102 (2005).

- Kerker, M. *The Scattering of Light* (Academic Press, New York, 1969).
- Lim, C.S., Hong, M.H., Lin, Y., Xie, Q., Lukyanchuk, B., Kumar, A.S., Rahman, M. *Appl. Phys. Lett.* **89**, 191125 (2006).
- Lin, Y., Hong, M.H., Wang, W.J., Law, Y.Z., Chong, T.C. *Appl. Phys. A* **80**, 461 (2005).
- Lin, Y., Hong, M.H., Lim, C.S., Chen, G.X., Tan, L.S., Wang, Z.B., Shi, L.P., Chong, T.C. *Appl. Phys. Lett.* **89**, 041108 (2006).
- Liu, C.H., Hong, M.H., Zhou, Y., Chen, G.X., Sam, M.M., Hor, A.T.S. *Phys. Scr.* **T129**, 326 (2007).
- Lu, Y.F., Zhang, L., Zheng, Y.W., Song, W.D., Goh, Y.W., Luk'yanchuk, B. *JETP* **72**, 457 (2000).
- Luk'yanchuk, B.S., Marine, W., Anisimov, S.I., Simakina, G.A. *Proc. SPIE* **3618**, 434 (1999).
- Metev, S.M. *Laser Assisted Microtechnology* (Springer-Verlag, New York, 1998).
- Okada, T. *Proc. SPIE* **4088**, 148 (2000).
- Ready, J.F., Farson, D.F. *Handbook of Laser Materials Processing* (Laser Institute of America, Orlando, FL, 2001).
- Riehn, R., Cgaras, A., Morgado, J., Cacialli, F. *Appl. Phys. Lett.* **82**, 526 (2003).
- Siegel, M.P., Overmyer, D.L., Kottenstette, R.J., Tallant, D.R., Yetton, W.G. *Appl. Phys. Lett.* **80**, 3940 (2002).
- Shi, D., Weisshaus, I. *Adv. Packag.* **10**(1), (2001).
- Sun, Y.P., Riggs, J.E., Henbest, K.B., Martin, R.B. *J. Nonlin. Opt. Phys. Mater.* **9**, 481 (2000).
- Tsukada, N., Nakao, T., Higuchi, T. *Proceedings of the 18th IEEE International Conference on Micro Electro Mechanical Systems*, p. 576 (2005).
- Vivien, L., Riehl, D., Hache, F., Anglaret, E. *Physica B* **323**, 233 (2002).
- Wang, W.J., Lim, G.H., Song, W.D., Ye, K.D., Zhou, J., Hong, M.H., Liu, B. *J. Phys.* **59**, 177 (2007).
- Wang, Z.B., Hong, M.H., Huang, Lukyanchuk, B.S., Wang, Q.F., Shi, L.P., Chong, T.C. *Appl. Phys. A* **79**, 1603 (2004).
- Willmott, P.R., Huber, J.R. *Rev. Mod. Phys.* **72**, 315 (2000).
- Wu, M.H., Paul, K.E., Whitesides, G.M. *Appl. Opt.* **41**, 2575 (2002).
- Xie, Q., Hong, M.H., Tan, H.L., Chen, G.X., Shi, L.P., Chong, T.C. *J. Alloys Comp.* **449**(1), 261 (2008).
- Yin, X.B., Fang, N., Zhang, X., Martini, I.B., Schwartz, B.J. *Appl. Phys. Lett.* **81**, 3663 (2002).
- Zhou, Y., Hong, M.H., Fuh, J.Y.H., Lu, L., Luk'yanchuk, B.S., Wang, Z.B., Shi, L.P., Chong, T.C. *Appl. Phys. Lett.* **88**, 023110 (2006).
- Zhu, S., Lu, Y.F., Hong, M.H. *Appl. Phys. Lett.* **79**, 1396 (2001).

# Evaluating the risks associated with nanomaterials

*K. Thomas, N. Monteiro-Riviere, D. Warheit, and N. Savage*

# 25

25.1 Introduction	887
25.2 Nanomaterials in consumer products	888
25.3 Characterization of nanomaterials	890
25.4 Hazard evaluation	891
25.5 Pulmonary exposure assessment	893
25.6 Dermal exposure assessment	894
25.7 Evaluating the risks associated with exposure to nanomaterials	896
25.8 Research priorities for the development of more refined estimates of nanomaterial risk	899
25.9 Conclusion	902
References	903

## 25.1 Introduction

The use of nanomaterials in consumer products and industrial applications is becoming more prevalent as the benefits of these materials are becoming more widely known. Nanomaterials have broad utility in a number of consumer and industrial applications including energy production, home appliances, water treatment, novel therapeutic delivery techniques and dietary supplements, consumer electronics, and sports equipment.

While considerable attention has been given to the likely commercial advantages associated with nanomaterials, less emphasis has been placed on the development of a systematic approach for characterizing the human health and environmental risks from exposure to nanomaterials. Essential to the development of a comprehensive framework for evaluating the safety of nanomaterials is a better understanding of the potential human and environmental exposure pathways of these materials from occupational exposures as well as exposures from consumer products.

Although some of the hazards, particularly those from inhalation exposures, associated with pure nanomaterials have been well documented in animal studies, the hazards of nanomaterials that are constituents of consumer products are not well known. In addition, there is also very little information available regarding potential human exposures to pure forms of these materials in commercial applications. In addition, the long-term behavior of nanomaterials in complex matrices, in which they are often present in consumer products, has not been comprehensively evaluated. An extensive assessment of the inherent stability of nanomaterials, as well as a life-cycle assessment of nanomaterials in the matrices in which they exist in consumer products is also a necessity for assessing the safety of these materials.

There is also a need to better characterize the likely degradation products and human-exposure pathways to gain a more fundamental understanding of the risks associated with human exposure to nanomaterials contained in consumer products. This would also include an assessment of product-use patterns and

a qualitative assessment of potential unintended uses of products containing nanomaterials.

Although existing data may preclude the development of a comprehensive safety evaluation for nanomaterials, sufficient information does exist to determine whether a given nanomaterial merits concern with regard to safety or is not likely to be a concern for human health or ecological risk.

## **25.2 Nanomaterials in consumer products**

Nanomaterials currently being used in a number of products could result in widespread consumer exposure to these materials as a result of their pervasive use. Some of these products include sunscreens, cosmetics, electronic components, textiles, cigarette filters and sporting equipment. There are a variety of benefits of using nanomaterials in consumer products. For textiles, the benefits include stain and wrinkle resistance. For sporting equipment, the benefits include greater strength and structural integrity, and for cosmetics the benefits include greater bioavailability for the active components. It is also noteworthy that some of the products currently available that contain nanomaterials, cosmetics and sporting equipment in particular do not require pre-market approval. Consequently, there is no government mandate to demonstrate their safety before they are made commercially available. Many products are required to undergo substantial evaluations to demonstrate their safety, prior to receiving regulatory approval and subsequent market access.

Consumer exposure to products containing nanomaterials may occur via various routes, including oral, intravenous, dermal, ocular, and inhalation. It is also likely that the population currently exposed to nanoscale materials in consumer products would be representative of the entire population, and not exclusive to a particular age, sex, ethnic background or disease condition, because many of these products appeal to a diverse range of individuals.

### **25.2.1 Cosmetics**

Cosmetics are being formulated with nanomaterials to incorporate smaller-size ultraviolet (UV) radiation-blocking particles. This provides the advantage of improved efficacy relative to traditional cosmetics. In addition, a number of biotechnology companies are developing functionalized nanomaterials that facilitate localized topical delivery, controlled release, and stabilization of cosmetic ingredients. Liposomes, emulsions, and encapsulated particles have all been developed as nanomaterials to improve the effectiveness of cosmetic products. Some of the potential impacts from skin exposure to nanomaterials include the following: (1) enhanced amount and depth of penetration of active ingredients in cosmetics into the skin resulting in increased activity; (2) ingredients that are chemically unstable in air and light such as retinol and Vitamin E may be more readily used in topical products following encapsulation in nanoparticles; and, (3) timed release of ingredients may become more feasible in topical products and could allow for improved effectiveness equivalent to current controlled-release orally administered drugs.

### 25.2.2 Sporting equipment

Nanomaterials are being used in a variety of different types of sporting equipment. The use of nanotechnology in sporting equipment allows the development of lighter materials that can be stronger than their traditional counterparts. Frequently, the nanomaterials in sporting equipment are embedded in a matrix within the product. For example, nanomaterials are being used in tennis rackets and baseball bats to improve their strength and stability. The incorporation of silicon-oxide particles into the voids in the graphite frame of a tennis racket make it stronger and the use of a nanocomposite material in the yoke of the racket minimizes the extent to which the racket bends when it comes in contact with the ball. This provides more speed for the ball after contact with the racket. In addition, carbon nanotubes are injected into the resin in carbon-fiber baseball bats to improve their strength, and soccer and tennis balls are being modified with nanomaterials to achieve a number of benefits including, improvement of air retention, providing more consistent bounce, and extending their useful life. Also, the inner lining of some tennis balls have been modified to incorporate a nanocomposite-coating technology composed of an aqueous suspension of nanodispersed silicates in a polymer matrix. The human health and environmental impacts of nanomaterials in sporting equipment is not well known because the availability of the underlying nanomaterials for human and environmental exposure has not been comprehensively evaluated. While the human-health impact would be expected to be minimal because the nanomaterials are generally imbedded in a matrix as a constituent of the product, the potential for these materials to be released from the products over time and the expected use conditions for each product should be evaluated carefully to fully understand the likely disassociation characteristics of the nanomaterials. In addition, a full assessment of the life-cycle of these products would allow the development of a comprehensive assessment of the complete human-health and environmental impacts of these materials.

### 25.2.3 Textiles

Consumer textile products are among the most widely used materials in the world. Over the past decade, the textile industry has become a significant user of nanotechnology. Novel uses for nanotechnology have infused the U.S. textile industry with a variety of innovative niche products, or improved versions of existing ones (Popowitz 2003). These new systems are increasingly replacing traditional textile finishing processes that typically involve surface chemical applications. The benefits of using nanotechnology for textiles used for fabrics include stain, water and wrinkle resistance. Clothing with wrinkle- and stain-resistant properties is a trend that was made possible by nanotechnology (Agins 2004). In addition, mattress manufacturers have also used fabrics that incorporate nanotechnology into their bedding products (Popowitz 2003). There are a variety of systems and chemistries used to confer the benefits of nanotechnology to textile applications. Nanomaterials commonly applied to textiles include fibers that range from 50 to 100 nm in length (Cole 2004). These materials, often referred to as nanofibers or “nanowhiskers”, are attached

to natural (such as cotton) or synthetic fibers (such as polyester) by means of polymer chemistry applications. Textiles are not coated with a stain- or water-resistant chemical, the change to the fabric occurs at the molecular level, and the nanomaterials can be configured to give the finished fabric a particular desirable attribute (Rodie 2001). Nanofibers are created through electrospinning, a textile manufacturing process that dates back to the 1930s. Using electrospinning, manufacturers apply electrical charges to water-based polymer solutions containing nanoparticles. When sufficient electrical charge is applied to the solution, an unstable jet of solution and nanoparticles is formed. While, the whipping motion elongates the jet, the solvent evaporates, producing a tiny fiber containing the nanoparticles (Doshi and Reneker 1995). The potential benefits to consumers from the use of nanotechnology in textiles include stain and water resistance, antistatic properties, moisture wicking, wrinkle resistance, as well as antibacterial and UV absorbance with no discernible alteration of aesthetics. Textile end products, including clothing and home furnishings, are in close contact with consumers and allow a substantial opportunity for exposure. Consequently, the human-health and environmental implications of exposure to these products should be fully assessed. Textiles undergo a variety of processes from fiber production to finishing. Given the degree of processing for finished textiles, it is unlikely that the original fibers contained in these products would cause any health concerns to consumers (Wakelyn 1994). However, given the limited data available regarding the behavior of nanomaterials in textiles, more thorough testing should be performed to fully characterize the fate of nanoscale materials in textiles over time to assess the consequences of human and environmental exposures to these materials.

## 25.3 Characterization of nanomaterials

Nanomaterials are generally defined as particles that have at least one dimension in the 1–100 nm size range. An essential part of evaluating the risks associated with nanomaterials is an accurate characterization of the material. Nanoscale particles; whether termed ultrafine, nano-, engineered, intentional, or incidental; pose challenges for physical, chemical and biological characterization.

Nanoparticles have a tendency to agglomerate, or form multiple single-particle clusters, in biological fluids. The formation of agglomerates can hinder dissolution (increasing persistence) by both reducing the average equilibrium solubility of the particle system, as well as by kinetic hindrance to the diffusion process. When characterizing a nanoparticle, one should be aware of the potential role of agglomerates in dissolution, since the state of agglomeration is largely impacted by the mode of respiratory-tract administration (i.e. internasal/intertracheal inhalation, insufflation, or instillation). The probability for agglomeration increases with increased localized dose and for doses given over shorter timeframes.

Toxicity studies on engineered nanomaterials such as fullerenes, single and multiwalled carbon nanotubes, metal oxides such as TiO<sub>2</sub>, and nanometer-diameter low-solubility particles support the need to carefully consider how

these materials are characterized when evaluating potential biological activity (Yamago *et al.* 1995; Brown *et al.* 2000, 2001; Shvedova *et al.* 2003; Warheit *et al.* 2004; Monteiro-Riviere *et al.* 2005). Essential parameters to consider for material characterization include physicochemical properties such as size distribution, agglomeration state, crystalline structure, chemical composition and three-dimensional configuration. It is also important to use appropriate controls such as bulk (micrometer) sized materials of like chemistry as well as benchmark materials. This will allow a relative comparison of the nanomaterial to materials that have previously been well described.

## 25.4 Hazard evaluation

A systematic evaluation of the hazards associated with nanomaterials would need to assess those target-organ effects that would be important for ensuring the safety of products containing nanomaterials. At a minimum, exposures that could reasonably be expected to result from products in commerce should be evaluated. As a result of their size, inhalation exposure would be an obvious priority for assessing the risk associated with nanomaterials. In addition, previous studies in laboratory animals have demonstrated that inhalation exposures to these materials can be associated with cytokine-mediated inflammatory responses and granuloma formation. Consequently, pulmonary exposures should be evaluated when assessing the potential human health hazards of nanomaterials. In addition, because a number of personal-care products such as skin creams and sun blocks have been formulated with nanomaterials, primarily metal oxides, to enhance their absorbance, and to improve other practical considerations such as application and aesthetics, dermal exposures would also be very important to assess when attempting to assess the risks associated with nanomaterials.

The increased reactivity of nanomaterials that arises as a consequence of their larger surface area has created considerable interest in the development of a better understanding of the effects of nanomaterials on biological systems. Nanomaterials have been evaluated in both *in-vitro* and *in-vivo* systems to explore effects from dermal and inhalation exposures. In a broad sense, these studies were designed to characterize the extent to which these materials interact with organ systems and cellular organelles, or to understand the mechanisms of cellular interactions, or the biological impacts of those interactions, and the potential health effects from acute exposure to nanomaterials.

Fullerenes, a type of nanomaterial that could have utility in several areas including in the development of novel drug-delivery systems, have been shown to induce oxidative stress in juvenile largemouth bass (Oberdorster *et al.* 2004). Exposure to uncoated C<sub>60</sub> fullerenes caused lipid peroxidation in the brain tissue of juvenile largemouth bass. While the specific mechanism(s) responsible for this effect was not clear, it was postulated to be associated with a selective transport mechanism from the olfactory nerve into the olfactory bulb. In another study, a time-dependent increase in nanoscale elemental <sup>13</sup>C was observed in the olfactory bulb of rodents following inhalation exposure

(Oberdorster *et al.* 2004). The increase was measured subsequent to exposure to  $^{13}\text{C}$  and was attributed to neuronal transport from the olfactory nerve to the olfactory bulb.

Due to the poor solubility of fullerenes their biological interactions have been limited. Functionalization of fullerenes with complexes such as amino acids have the potential to provide greater interaction between the fullerene and the biological environment, yielding potential new medical and pharmaceutical applications. Fullerene-based amino acid solutions were assessed in human epidermal keratinocytes at different concentrations at different time points showing the localization and proinflammatory response of fullerene-derived amino acids (Rouse *et al.* 2006). Derivatized fullerenes have also been used to study the effects of mechanical flexion on the penetration of fullerenes through intact skin (Rouse *et al.* 2007).

These data underscore the need to develop a more comprehensive understanding of the translocation specificity of fullerenes, and other lipophilic nanoscale materials, into lipid-rich tissues and cellular organelles and any subsequent effects on biological function. Another important consideration for evaluating the biological effects of nanomaterials is the presence and types of surface coatings on the material. Although uncoated material was used for the largemouth bass study, a significant amount of the material that will be used in commerce is likely to be coated with one of a variety of materials that have varying levels of biological availability and persistence. The characteristics of the underlying nanoscale material therefore must be understood in combination with the relevant surface coating or treatment.

Respiratory exposure to nanomaterials such as carbon nanotubes has also received considerable interest. Because of difficulties associated with generating aerosols of carbon nanotubes to facilitate a respiratory evaluation, and the high cost of homogeneous well-characterized material, several *in-vivo* pulmonary studies have employed intratracheal instillation as the exposure methodology. Single-walled carbon nanotubes (SWCNT) have been evaluated in rodents using this technique. SWCNTs instilled into the lungs of mice produced granulomas in the pulmonary interstitial space of the lungs (Lam *et al.* 2004). Rats exposed to SWCNTs developed multifocal granulomas in the absence of any pulmonary inflammation or cellular proliferation (Warheit *et al.* 2004) which suggests that SWCNTs may act via a different mechanism of toxicity than other inhaled toxicants such as crystalline silica. There are ongoing studies to assess the aerosolization and inhalation toxicity of SWCNTs (Baron *et al.* 2008).

These studies highlight the need for further experimentation to fully elucidate the mechanisms that are responsible for the behavior of carbon fullerenes and nanotubes. Specifically, to develop a comprehensive evaluation of the human-health implications from exposure to nanoscale materials, more data are needed on the adsorption, distribution, metabolism, and excretion, of these materials in biological systems. In addition, more data are needed to assess the extent to which the fundamental properties of these materials, such as particle size and size distribution, affect the distribution and elimination of these materials in the body. More studies are also needed to characterize the toxicity

of a broader range of commercial nanoscale materials as a function of chemical composition, shape, surface characteristics and method of production. Several of these properties, particularly surface coatings, could have a substantial effect on the biological activity of these materials. While only a few studies have been noted, other recently published articles capture a broader range of studies that have been performed to evaluate the human health implications of exposure to nanoscale materials (The Royal Society 2004; Oberdorster *et al.* 2005; Monteiro-Riviere and Tran 2007).

Information regarding the potential adverse effects of nanomaterials is currently limited. Although engineered nanoparticles have not been systematically tested, a few inhalation and epidemiology studies using ambient ultrafine particles have yielded some results from which preliminary conclusions can be drawn (Oberdorster and Utell 2002). Additionally, there are limited toxicology studies that have addressed the effects of nanomaterials in a variety of organisms and environments (Yamakoshi *et al.* 1999; Oberdorster *et al.* 2004). Pharmaceutical applications have also been a source of information regarding the potential translocation of nanoparticles from the site of exposure to distal areas of the body (Weber 1999; Cui and Gao 2003). Therefore, several studies using nanoparticles have shed some light on the biokinetics and biodistribution of nanomaterials once inhaled or ingested.

## 25.5 Pulmonary exposure assessment

Apart from their specific chemical surface reactivity, a particle load or burden in the lung can induce toxicological response(s) that differ principally from soluble or non-particulate toxicants. For the interpretation of inhaled particle effects, dose, deposition, dimension, durability and defense are important determinants of response. First, the dose at a specific site (in the lungs) determines the potential toxicity of particles. Obviously, this deposited dose is dependent on the inhaled concentration as well as the dimensions of the particle. Due to diffusion processes, the deposition probability of nanoparticles in the respiratory tract markedly increases for ultrafine particles when compared to larger particles where convective flow dominates deposition. The dose metric for nanoscale particles (NSP) is an additional complexity, as particle number, surface area, shape and other factors may play a role in addition to the traditional mass-based metric (Donaldson *et al.* 2002; Oberdorster *et al.* 2005).

The lung has extensive, location-specific defense systems such as mucociliary clearance in the upper airways and macrophage clearance in the lower, non-ciliated portion of the respiratory tract. Particle transport by macrophages from the alveolar region towards the larynx is rather slow in humans, even under normal conditions, and eliminates only a fraction of the deposited particles in the peripheral lung. The remainder may accumulate unless the particles are biodegradable or cleared by simple chemical dissolution in lung surfactant or interstitial fluid. Therefore, the same deposition of NSP with potentially a different durability can lead to a different cumulative dose.

Existing experimental evidence supports the notion that nanoparticles can cross biological membranes and gain access to tissues for which comparable bulk-scale materials would not be expected to easily access. Preferential translocation of nanoscale titanium dioxide ( $\text{TiO}_2$ ) particles into lung interstitium has been reported (Ferin *et al.* 1992), in addition to translocation of iridium to secondary organs (Semmler *et al.* 2004). For many nanoparticles, physical interactions such as van der Wals forces would create an environment that favors the formation of aggregates that would minimize the likelihood that the particles would exist as discrete primary particles. Aggregates of nanoparticles would likely be subject to normal macrophage clearance mechanisms in mammalian respiratory systems. Indeed, there is experimental evidence (Oberdorster *et al.* 1992) that ultrafine  $\text{TiO}_2$  particles are phagocytized by alveolar macrophages following intratracheal instillation, reducing both the pulmonary inflammatory reaction and the interstitial access of the ultrafine particles. However, dose-dependent inflammation has been observed with some nanoparticles in the respiratory tract, likely reflecting the relatively large surface area. Accentuated inflammation with ultrafine  $\text{TiO}_2$  has been demonstrated (Bermudez *et al.* 2004): inhalation of  $10 \text{ mg/m}^3$  for 13 weeks resulted in pulmonary overload in rats and mice with inflammation similar to higher mass doses of fine  $\text{TiO}_2$ . There is also evidence that nanomaterials can elicit effects beyond the respiratory tract. Research on ultrafine particles associated with air pollution (Dockery *et al.* 1993) has shown that increased mortality was most strongly associated with fine particulates. Subsequent studies have focused on respiratory and cardiovascular impacts. An overview of the fate of nanomaterials in the lungs in addition to potential pro-inflammatory and oxidative stress-related cellular responses has been published (Oberdorster *et al.* 2005).

Understanding the exposure and effects of discrete nanoparticles will be extremely important for future studies. However, with the broad applications of nanomaterials for nanotechnology, a simple focus on discrete nanoparticles is not adequate. Cohesive forces maintain nanoparticles as aggregates and agglomerates, markedly affecting their propensity to become airborne, as well as their aerodynamic diameter. Uses of nanomaterials in liquids and composites may severely limit or preclude airborne exposure. Disaggregation, deagglomeration, and dissolution in biological fluids are important factors potentially contributing to a complete understanding of nanoparticle fate. Although inhalation will be a critical route of exposure in some cases, a full understanding of exposure and fate requires consideration of the technology applications and physical state of the nanomaterials.

## 25.6 Dermal exposure assessment

Because of their ability to enhance the absorbance and biological availability of many personal-care products, nanomaterials are being used in sunscreens, skin creams, and a number of other cosmetic applications. Thus, it is important to evaluate the dermal exposure to nanomaterials contained in these applications via penetration of the skin.

Skin is the largest functional organ of the human body, which renders it susceptible to penetration by nanomaterials subsequent to consumer, environmental, or occupational exposure. To fully assess skin exposure of a given nanomaterial, a quantitative assessment of the permeability of the nanomaterial through the successive layers of skin is required. Currently, little is known regarding the permeability of skin to nanomaterials because the diversity of skin among species and across anatomical locations makes it extremely difficult to quantitatively estimate penetration into the skin, and adsorption through the skin.

Skin is composed of three layers including the epidermis, dermis, and hypodermis. The epidermis is the outermost layer of skin, and is most responsible for protecting the body from exposure to foreign substances. The epidermis consists of five layers including the stratum corneum, stratum lucidum (thick skin only), stratum granulosum, stratum spinosum, and the stratum basale. The outermost superficial epidermal layer is the stratum corneum that consists of several dead keratinized cells bathed within a lipid matrix that provides the most effective barrier against absorption and provides the greatest protection to the lower cell layers. Beneath this layer is the dermis, which is composed of connective tissue, fibers, ground substance, arterial and venous blood vessels, lymph vessels, receptors, hair follicles, smooth muscles and connecting capillaries to nourish the skin.

The extent to which materials can cross the stratum corneum is the rate-limiting step for nanomaterial absorption through the skin, and is directly proportional to concentration gradients across the membrane, the lipid/water partition coefficient of the material, and the diffusion coefficient of the nanomaterial under consideration if the mechanism of nanoparticle movement is diffusion. The partition coefficient represents the ability of a material to pass through the lipid membrane. Molecular size and three-dimensional configuration will also influence the ability of a material to passively diffuse across a cell membrane. Should nanoparticles move by a mechanism other than diffusion, different parameters may be important. The exposed surface area of the membrane and its functional and chemical constituents will also impact the passage of a material across the membrane. For skin toxicological evaluations, dose is determined on the basis of the amount of toxicant per area of skin, as opposed to the amount of toxicant per unit of body weight used in oral or parenteral studies.

Multiwalled carbon nanotubes (MWCNTs) have been examined for effects that could result from skin exposure to workers and consumers. *In-vitro* tests using human epidermal keratinocytes (HEK) indicate that MWCNTs could induce cytokine-mediated inflammatory responses and localize within cytoplasmic vacuoles in the skin cells (Monteiro-Riviere *et al.* 2005). *In-vitro* tests are used to understand the mechanism of cellular interactions of nanomaterials. If *in-vivo* tests are used, it would be difficult to localize these very small particles by transmission electron microscopy. Currently, there are no methods available to easily detect these small particles in tissues, a situation that makes interpretation of *in-vivo* studies problematic. It is possible that some clinical manifestation may occur but most probably, these particles will agglomerate within the body and not be a good measure of clinical diagnosis.

Flow-through diffusion cells are an established *in-vitro* model for measuring the permeability of skin to penetrants and have been used frequently to study the penetration of several types of nanomaterials (Ryman-Rasmussen *et al.* 2006; Monteiro-Riviere and Inman 2008; Zhang *et al.* 2008; Zhang and Monteiro-Riviere 2008). When quantum dots (QD) coated with carboxylic acid were studied in flexed, tape-stripped and abraded skin, no penetration occurred with the non-flexed control, flexed and tape-stripped skin, but minimal penetration occurred in abraded skin (Zhang and Monteiro-Riviere 2008). Quantum dots are easy to detect because of their intense and photostable fluorescence, commercial availability in various sizes and shapes with diverse surface coatings making them useful tools by which a nanotoxicologist can study their interactions with cells. This type of research provides a better understanding of absorption and penetration of damaged skin. Nanoparticle penetration not only occurs on the surface of the stratum corneum or within the stratum corneum layers, but may penetrate deeper with skin flexion. This research also suggests that there is risk for potential health effects to medical personnel exposed to QD during medical applications with damaged or abnormal skin. In addition, this study also provided information on QD nanoparticles absorption that could occur only in abraded skin and this information could be relevant in certain occupation exposure scenarios and could serve as a method of drug delivery.

Intradermal injections with QD in SKH-1 hairless mice, QD migration occurred from the injection site to regional lymph nodes through the lymphatic duct system and then to the liver and other organs (Gope *et al.* 2007). The biodistribution of intra-arterially infused QD621 was studied in a perfused skin flap model demonstrated that the QDs can migrate out of the capillaries into the surrounding tissue (Lee *et al.* 2007).

There are a few studies that have been conducted with titanium dioxide and zinc oxide, which are common ingredients in cosmetics and sunscreens that did not show penetration of these much larger agglomerates through skin. Many of the skin-penetration studies conducted to date suggests that the skin is permeable to some nanomaterials with diverse physicochemical properties and could serve as a portal of entry for localized, and systemic delivery.

## 25.7 Evaluating the risks associated with exposure to nanomaterials

Risk assessment, the science of evaluating the hazard along with the dose-response relationship and the exposure scenario to understand associated risks, is critical for engineered nanomaterials. Evaluating the risks associated with nanomaterials should be considered within the classical construct of a risk assessment with an emphasis on evaluating both the hazards and exposures associated with these materials. This can be difficult for nanomaterials as these materials utilize a variety of functional moieties, coatings, and morphology that will impact physical-chemical properties and consequently have an effect on the underlying hazard of the nanomaterial. In addition, the range of applications being explored for nanomaterials continues to expand, which necessitates

the need to assess the unique exposure considerations associated with each application. The hazards associated with these materials are currently being evaluated, and have not been fully explored for the range of nanomaterials being considered for commercial use. For those nanomaterials whose hazards have been well studied, it is clear that the toxicity of these materials is influenced by the physical and chemical characteristics (i.e. morphology, surface coating, diameter, agglomeration/aggregation state, size, functionalization, etc.) of the material. In addition, consumer and workplace exposure scenarios have not been extensively assessed for these emerging materials, although there have been a few studies of occupational exposures (Baron *et al.* 2003; Maynard *et al.* 2004). It is also not clear which point during the product life-cycle for nanomaterials represents the most substantial potential for human exposure. Finally, the dose-response relationship for many commercially important nanomaterials has not been determined. Consequently, the task of assessing risks and subsequently managing and minimizing these risks is challenging.

Nanomaterials are likely to have properties that are substantially different from their bulk counterparts. These properties could influence the hazards associated with nanomaterials and exposures to them. The unique properties of nanomaterials should be assessed with the goal of determining the extent to which they can influence the hazards, exposure, and risks associated with nanomaterials. The hazards associated with nanomaterials should be evaluated relative to target-organ impacts and the impact of these materials in terms of effects on metabolic and cell-signalling pathways. Commonly employed human and environmental toxicity testing approaches have been applied to assess nanomaterials. As our understanding of the toxicology of nanoparticles continues to grow, these standard toxicology tests may not be appropriate for use with nanomaterials. In consideration of testing for nanomaterials, emphasis needs to be given to characterization of the material.

The appropriate route of exposure and the appropriate endpoints need to be considered in the design of toxicity studies for nanomaterials. Other important portals of entry for nanoparticle exposure could include the gastrointestinal tract, or ocular exposure. Uptake of nanoparticles via the GI tract has been documented in oral-feeding studies and gavage studies using particles ranging from 10 nm to 500 nm (Jani *et al.* 1994; Hillyer and Albrecht 2001). As previously discussed, inhalation and dermal are important routes of exposure for many nanomaterials. Effects from inhalation of micrometer-sized particles are generally restricted to the lung, or portal of entry, with no systemic distribution. Therefore, traditional inhalation toxicology studies have been typically restricted to studying effects directly related to the lung when studying the effects of particles. However, the lung is a major route of exposure for gases, vapors and liquid aerosols that can produce systemic exposures, so it is not uncommon for inhalation studies to include systemic evaluations. The skin could also serve as a route of exposure to some nanomaterials especially in diseased skin. Because more is known today regarding the kinetics and distribution of nanoparticles, functional endpoints can be expanded beyond the traditional route of entry effect to include systemic effects.

Determining the appropriate dose is crucial in evaluating the true risk of these materials. More research and information is needed regarding environmental exposures and exposure assessments. Overload conditions or exceeding the maximum tolerated dose should be avoided. However, we cannot make the decisions at this time because of the lack of true exposure assessments. Currently, we have no idea as to the amount of nanomaterials humans will be exposed to until more exposure assessments studies have been conducted in the workplace or in the environment.

A tiered approach may be implemented for evaluating nanomaterials. *In-vitro* testing, followed by escalation to more complex testing models may provide useful information in the evaluation of these materials *in lieu* of chronic bioassays. Short-term mechanistic studies, *in-vitro* studies and ultrafine-particle epidemiological studies can provide important enhancements to traditional inhalation toxicity assays.

Tiered toxicological screening strategies have been proposed to evaluate the hazards associated with nanomaterials. Tiered strategies provide an opportunity to assess the potential impact of a material in a way that prioritizes those that are critical for further evaluation. A typical toxicological hazard screening approach developed for hazard identification includes a physicochemical evaluation followed by *in-vitro* and *in-vivo* assessments (Oberdorster *et al.* 2005). The strategy uses *in-vivo* studies to consider the potential effects of nanomaterials following inhalation, dermal, oral, and injection exposures. The strategy also includes an assessment of mechanisms of toxicity using *in-vitro* techniques.

The fundamental assumption of the physicochemical evaluation in the toxicological hazard screening approach proposed by Oberdorster *et al.* is that the underlying hazards associated with nanomaterials will be substantially influenced by the material's physicochemical properties including particle-size distribution, agglomeration state, shape, crystal structure, surface area, surface chemistry, surface charge, and porosity. The approach also noted the importance of measuring dose against a physical metric of mass, surface area, or particle number for enabling quantitative interpretation of the data. The basis for determining the most appropriate dose metric should be the parameter that is thought to be most closely associated with the anticipated response.

Filter-based off-line mass-concentration methods are routinely used for nanomaterial inhalation studies. On-line mass concentration measurements are also frequently used for inhalation studies. Gravimetric and/or chemical analysis of filter samples provides a more accurate characterization of exposure than off-line surface area and number-concentration assessments. These methods can also be used to facilitate estimates for aerosol surface area or number concentration. For assessments of size distribution, shape, and agglomeration state, transmission electron microscopy (TEM) analysis provides comprehensive information. Reasonable surface-area evaluations can be achieved with isothermal gas adsorption, although better sample filtering techniques need to be developed. Techniques for evaluating crystal structure and surface charge are readily available and include X-ray diffraction (XRD) and zeta potential, respectively. Measurements of surface area, surface chemistry, and porosity are sometimes difficult, although in some cases TEM can be used.

A variety of *in-vitro* systems are available for evaluating nanomaterials. To determine the specific *in-vitro* system(s) that will be helpful for evaluating the nanomaterial of interest, consideration should be given to the specific portal of entry and/or target organ systems for which effects should be evaluated. In addition, it is important to note that *in-vitro* systems have a range of challenges that make it difficult to assess their relevance for *in-vivo* biological responses including, but not limited to, whether and to what extent the *in-vitro* effects contribute to measurable biological responses and challenges for harmonizing *in-vitro* dosimetry with *in-vivo* dose. Portals of entry that are important for evaluating the impact of nanomaterials include the lung, skin, and mucosa. To evaluate the lung for potential nanomaterial impacts epithelial cells, macrophages, immune cells, and fibroblasts allow an examination of endpoints important for lung response such as inflammation, immunopathology, fibrosis, genotoxicity, microbial defense and clearance. To evaluate skin for potential nanomaterial impacts keratinocyte culture, Flow-through diffusion cell, and isolated perfused porcine skin flap (IPPSF) can be used. A variety of *in-vitro* systems are available for evaluating potential impacts to mucosa tissue.

The *in-vivo* evaluation in the toxicological hazard screening approach proposed by Oberdorster *et al.* proceeds in tiers where the first-tier *in-vivo* assessments evaluate pulmonary, oral, dermal, and injection exposures, and the second tier involves a more comprehensive pulmonary evaluation. In the first tier, markers of damage, oxidant stress, and cell proliferation are assessed, and in tier 2 pulmonary deposition, translocation, biopersistence, multiple exposure impacts, and potential effects on the reproductive system, placenta, and fetus are evaluated. Additional studies conducted as part of tier 2 include whole-animal models using a variety of exposure methodologies including intratracheal instillation and pharyngeal aspiration, and mechanistic studies.

The Organization for Economic Cooperation and Development (OECD) has also developed a list of endpoints for evaluating the hazards associated with nanomaterials (OECD 2008). OECD is an intergovernmental organization in which representatives of 30 industrialized countries in North America, Europe, and the Asia and Pacific region, as well as the European Commission coordinate and harmonize policies in which they have mutual interest. The OECD Chemicals Committee and Working Party on Chemicals have jointly initiated an effort to identify human-health and environmental concerns associated with manufactured nanomaterials. The list of endpoints for evaluating the hazards associated with manufactured nanomaterials has been proposed as part of a tiered system for identifying nanomaterials hazards. The endpoints that have been included in this system are listed in Table 25.1.

## 25.8 Research priorities for the development of more refined estimates of nanomaterial risk

Research efforts are underway in several parts of the world to develop priorities for improving risk-based safety evaluations for nanomaterials. Each of these

**Table 25.1** OECD suggested endpoints for evaluating nanomaterial hazards.

Nanomaterial Information/Identification
○ Nanomaterial name (from list)
○ CAS number
○ Structural formula/molecular structure
○ Composition of nanomaterial being tested (including degree of purity, known impurities or additives)
○ Basic morphology
○ Degree of surface chemistry (e.g. coating or modification)
○ Major commercial uses
○ Known catalytic activity
○ Method of production (e.g. precipitation, gas phase)
Physical-Chemical Properties and Material Characterization
○ Agglomeration/aggregation
○ Water solubility
○ Crystalline phase
○ Dustiness
○ Crystalline size
○ Representative TEM picture(s)
○ Particle-size distribution
○ Specific surface area
○ Zeta potential (surface charge)
○ Surface chemistry (where appropriate)
○ Photocatalytic activity
○ Pour density
○ Porosity
○ Octanol–water partition coefficient, where relevant
○ Redox potential
○ Radical formation potential
○ Other relevant information (where available)
Environmental Fate
○ Dispersion stability in water
○ Biotic degradability
○ Ready biodegradability
○ Stimulation testing on ultimate degradation in surface water
○ Soil simulation testing
○ Sediment simulation testing
○ Sewage-treatment simulation testing
○ Identification of degradation product(s)
○ Further testing of degradation products and required
○ Abiotic degradability and fate
○ Hydrolysis, for surface-modified nanomaterials
○ Adsorption-desorption
○ Adsorption to soil or sediment
○ Bioaccumulation potential
○ Other relevant information (when available)
Environmental Toxicology
○ Effects on pelagic species (short term/long term)
○ Effects on sediment species (short term/long term)
○ Effects on soil species (short term/long term)
○ Effects on terrestrial species
○ Effects on microorganisms
○ Other relevant information (when available)

(cont.)

**Table 25.2** (Continued)

Mammalian Toxicology
◦ Pharmacokinetics (ADME)
◦ Acute toxicity
◦ Repeated dose toxicity
◦ If available
• Chronic toxicity
• Reproductive toxicity
• Developmental toxicity
• Genetic toxicity
• Experience with human exposure
• Other relevant test data
Material Safety (where available)
◦ Flammability
◦ Explosivity
◦ Incompatibility

efforts is intended to provide better data for evaluating the environmental and human-health effects of nanomaterials.

Many parts of the world have devoted considerable resources to promoting nanotechnology for economic, commercial, and societal benefits by allocating significant resources to identifying and developing promising applications for the technology. There are substantial research efforts underway to generate data to facilitate the development of scientifically defensible risk assessments for nanomaterials in an effort to better assess their safety. There are many common elements to this research. In particular, these programs intend to evaluate the human health and environmental impacts from exposure to nanomaterials. With regard to human exposure, they are attempting to identify the risks associated with workplace and consumer exposure to nanomaterials, and to develop dose-response models for engineered nanomaterials. They are also seeking to develop monitoring techniques to assess workplace and environmental concentrations of nanomaterials, and to develop techniques for measuring nanomaterials in a variety of media, including biological fluids and human tissues.

International cooperation for standardization of assessment methods and harmonization of risk-evaluation techniques will be critical for the development of scientifically rational standards for public-health decision making. Given the similarities of the goals among many of these programs, the development of formal collaborations and consortia could facilitate the generation of data for risk assessments in a more efficient manner, and minimize duplication of effort. Government regulatory and research agencies should encourage these collaborations in a way that is transparent, and that allow input from a broad spectrum of stakeholders.

There is a strong interest among government research organizations to evaluate the utility of existing testing and measurement methods for evaluating nanomaterials. This was identified as a priority for developing strategies for appropriately characterizing nanomaterials. There is also a need to evaluate the production, use, and disposal of products containing nanomaterials from

a life-cycle perspective. It does not appear that sufficient data is currently available to accommodate a life-cycle evaluation of nanomaterials. Development of a multidisciplinary, international collaboration to identify and prioritize the critical elements for developing life-cycle assessments for nanomaterials is the requisite initial step for evaluating the full impact of these materials.

There are also research needs that would benefit from additional attention from government, academic, and industry research. Specifically, a coordinated effort to determine the most appropriate personal-protective equipment (PPE) and handling considerations, (e.g. laboratory-hood requirements) for those nanomaterials that are the focus of current research efforts would be especially helpful to those that are routinely handling these materials. In addition, much of the current research focuses on inhalation and dermal exposure. While these routes are certainly important, particularly given what is known about the systemic hazards associated exposures to fine-sized particles, it is also important to develop comprehensive programs that would also include an assessment of oral exposure. While there is considerable activity in these areas underway in the private sector, government, and academia, it is unclear whether the information generated from these efforts is being communicated or coordinated internationally.

## 25.9 Conclusion

Although a considerable number of experimental studies have been conducted to develop hazard information for nanomaterials, sufficient information does not exist to comprehensively evaluate the risks associated with these materials. While *in-vitro* and *in-vivo* hazards have been conclusively linked to exposure to nanomaterials, it is difficult to quantitatively assess human exposure to nanomaterials in the forms in which the materials are present in commercial products. Many of the nanomaterials contained in commercial products are present in small quantities, and/or embedded in a matrix in which the nanomaterial may be one of many constituents. The presence of a substance with known hazards in a consumer product does not necessarily constitute a human health or environmental hazard if the product design or use prevents the consumer from being exposed to the component that has known hazards. For those nanomaterials that are contained within a matrix of constituents in a consumer product, it is important to have data that characterizes the stability of nanomaterial in the matrix and the likelihood that the nanomaterial could become separated from the matrix and create an opportunity for exposure. In addition, it is important to identify the specific form of the nanomaterial that is contained within the product. Nanomaterials may be incorporated in products as discrete nanoscale entities, having unique size and compound-specific properties. However, it is probably more likely that the nanoparticles in consumer products will agglomerate into larger particles or longer fiber chains that may change their properties and influence their behavior in biological systems as well as the environment. The conditions of use for the product may also impact the form of the compound to which exposure would be

expected. During typical or expected consumer use, the nanomaterial may be released in one form, while under different conditions; the form to which exposure would be greatest could change. Sufficient data exist to evaluate the hazards associated with nanomaterials. In addition, tiered frameworks have been proposed to facilitate prioritization of those data in the context of a hazard evaluation. The critical gap for evaluating the risks posed by nanomaterials is the need for better exposure data for these materials in consumer applications. As better data are generated to quantitatively assess human and environmental exposure to nanomaterials, the development of scientifically defensible risk assessments for nanomaterials will become a reality.

## References

- Agins, T. *The Wall Street Journal*. **22**, November (2004).
- Baron, P.A., Deye, G.J., Chen, B.T., Schwegler-Berry, D.E., Shvedova, A.A., Castranova, V. *Inhal. Toxicol.* **20**(8), 751–60 (2008).
- Bermudez, E., Mangum, J.B., Wong, B.A., Asgharian, B., Hext, P.M., Warheit, D.B., Everitt, J.I. *Toxicol. Sci.* **77**, 347 (2004).
- Brown, D.M., Wilson, M.R., MacNee, W., Stone, V., Donaldson, K. *Toxicol. Appl. Pharmacol.* **175**, 191 (2001).
- Brown, D.M., Stone, V., Findlay, P., MacNee, W., Donaldson, K. *Occup. Environ. Med.* **57**, 685 (2000).
- Cole, M. *Int. Fiber J.* **19**, 12–15 (2004).
- Cui, D., Gao, H. *Biotechnol. Prog.* **19**, 683 (2003).
- Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., Speizer, F.E. *N. Engl. J. Med.* **329**, 1753 (1993).
- Donaldson, K., Brown, D., Clouter, A., Duffin, R., Macnee, W., Renwick, L., Tran, L., Stone, V. *J. Aerosol Med.* **15**, 213 (2002).
- Doshi, J., Reneker, D.H. *J. Electrostat.* **35**, 151 (1995).
- Ferin, J., Oberdorster, G., Penney, D.P. *Am. J. Respir. Cell. Mol. Biol.* **6**, 535 (1992).
- Gope, N.V., Roberts, D.W., Webb, P., Cozart, C.R., Siitonen, P.H., Warbritton, A.R., Yu, W.W., Colvin, V.L., Walker, N.J., Howard, P.C. *Toxicol. Sci.* **98**, 249 (2007).
- Hillyer, J.F., Albrecht, R.M. *J. Pharm. Sci.* **90**, 1927 (2001).
- Jani, P.U., McCarthy, D.E., Florence, A.T. *Int. J. Pharm.* **105**, 157 (1994).
- Lam, C.W., James, J.T., McCluskey, R., Hunter, R.L. *Toxicol. Sci.* **77**, 126 (2004).
- Lee, H.A., Imran, M., Monteiro-Riviere, N.A., Colvin, V.L., Yu, W.W., Riviere, J.E. *Nano. Lett.* **9**, 2865 (2007).
- Maynard, A.D., Baron, P.A., Foley, M., Shvedova, A.A., Kisin, E.R., Castranova, V. *J. Toxicol. Environ. Health A* **67**, 87–107 (2004).
- Monteiro-Riviere, N.A., Nemanich, R.J., Inman, A.O., Wang, Y.Y., Riviere, J.E. *Toxicol. Lett.* **155**, 377 (2005).
- Monteiro-Riviere, N.A., Tran, C.L. (eds). *Nanotoxicology: Characterization, Dosing and Health Effects*, Taylor and Francis, Informa Healthcare (New York, NY, 2007).
- Monteiro-Riviere, N.A., Inman, A.O.Z. *The Toxicologist CD-An Official Journal of the Society of Toxicology*, 102:S-1, **1029**, 211 (2008).

- Oberdorster, G., Utell, M.J. *Environ. Health Perspect.* **110**, A440 (2002).
- Oberdorster, G., Ferin, J., Gelein, R., Soderholm, S.C., Finkelstein, J. *Environ. Health Perspect.* **97**, 193 (1992).
- Oberdorster, E. *Environ. Health Perspect.* **112**, 1058 (2004).
- Oberdorster, G., Oberdorster, E., Oberdorster, J. *Environ. Health Perspect.* **113**, 823 (2005).
- Oberdorster, G., Maynard, A., Donaldson, K., Castranova, V., Fitzpatrick, J., Ausman, K., Carter, J., Karn, B., Kreyling, W., Lai, D., Olin, S., Monteiro-Riviere, N., Warheit, D., Yang, H. *Particle Fibre Toxicol.* **2**, 8 (2005).
- OECD.
- Popowitz, N. *BAEP 557: Technology Commercialization, Spring* **98** (2003).
- Rodie, J.B. *Textile World*, March (2001).
- Rouse, J.G., Yang, J., Barron, A.R., Monteiro-Riviere, N.A. *Toxicol. In Vitro* **20**, 1313 (2006).
- Rouse, J.G., Yang, J., Ryman-Rasmussen, J.P., Barron, A.R., Monteiro-Riviere, N.A. *Nano Lett.* **7**, 155 (2007).
- Ryman-Rasmussen, J., Riviere, J.E., Monteiro-Riviere, N.A. *Toxicol. Sci.* **91**, 159 (2006).
- Royal Society and the Royal Academy of Engineering. *Royal Society and Royal Academy of Engineering*, London, UK, 116 (2004).
- Semmler, M., Seitz, J., Erbe, F., Mayer, P., Heyder, J., Oberdörster, G., Kreyling, W.G. *Inhal. Toxicol.* **16**, 453 (2004).
- Shvedova, A.A., Castranova, V., Kisin, E.R., Schwegler-Berry, D., Murray, A.R., Gandelsman, V.Z., Maynard, A., Baron, P. *J. Toxicol. Environ. Health A* **66**, 1909 (2003).
- Wakelyn, P.J. Cotton Yarn Manufacturing. In *ILO Encyclopedia of Occupational Health and Safety*, (eds) A.L. Ivester, J.D. Neefus, 4th edn (International Labour Office, Geneva, Switzerland, 1994), pp. 89.9–89.11.
- Warheit, D.B., Laurence, B.R., Reed, K.L., Roach, D.H., Reynolds, G.A., Webb, T.R. *Toxicol. Sci.* **77**, 117 (2004).
- Weber, D.O. *Health Forum J.* **42**, 36 (1999).
- Yamago, S., Tokuyama, H., Nakamura, E., Kikuchi, K., Kananishi, S., Sueki, K., Nakahara, H., Enomoto, S., Ambe, F. *Chem. Biol.* **2**, 385 (1995).
- Yamakoshi, Y., Sueyoshi, S., Miyata, N. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku*. **117**, 50 (1999).
- Zhang, L.W., Yu, W.W., Colvin, V.L., Monteiro-Riviere, N.A. *Toxicol. Appl. Pharmacol.* **228**, 200 (2008).
- Zhang, L.W., Monteiro-Riviere, N.A. *Skin Pharmacol. Physiol.* **21**, 166 (2008).

# Subject Index

- Admittance loci design method, 643–646  
Adsorption, 479–481  
AFM (atomic force microscope), 244, 247, 249, 253, 257, 258, 259, 260, 357, 372, 389, 390, 392, 406, 516, 905, 907, 851, 954, 955, 970  
Alkanethiolate films on metal surfaces, 810–813  
Allosteric modulation, 573  
Anisotropic surface energy, 779–780  
Anodic aluminum oxide (AAO), 387–390, 392–398, 405, 412  
Antibody binding, 664  
  kinetic analysis, 664–669  
Anticancer activity of gold nanoparticles, 494–495  
Antigens, 507  
Applications of atomic switch, 306–309  
Application specific integrated circuits (ASICs), 309, 310  
Applications of SWNTs and MWNTs, 484  
Arrays of quantum dots, 234–236  
Atomic-layer deposition (ALD), 809, 831, 838–840  
Atomistic Si-oxidation mechanism, 5–16  
Attenuated total reflection (ATR), 546, 648  
Auger electron spectroscopy (AES), 420
- Band structure of CNT, 370–374  
Band structure of heterodevices, 186–189  
Basis of hydrogen sorbents, 701–703  
Berry phase, 138  
Bifunctional nanomaterials for imaging and treatment of cancer, 474–497  
Bioconjugated quantum dots for tumour molecular imaging and profiling, 612–636  
Biodistribution of CNTs, 487–488  
Biological barriers, 515–516  
Biomolecular layer effect, 661–669  
Biophysical implications of protein-based nanoelectronics, 588–591  
Bonding energy of hydrogen in HSM (hydrogen-storage materials), 701–703  
Bottlenecks of metallic field emitters, 738–739  
Bottom-up and top-down, 353  
Breast tomosynthesis, 686–688  
  digital, 686
- C<sub>60</sub>, fullerenes, 891  
Carbon nanopearls (CNPs), 762–764  
Carbon nanoparticle field emitters, 749–761  
Carbon nanotubes (CNTs), 42, 317, 343, 358, 367–369, 481, 676, 712, 716, 743, 750–753, 757–761, 889  
  anticancer activity of, 539
- in tumour imaging, 491–492  
  physical properties, 481–484  
Carbon nano-test-tubes (NTTs), 387–403, 405, 412  
  synthesis of, 397–399  
Cell-based integrated circuits (CBICs), 308, 310  
Channel engineering, 38–39  
Characterization of nanomaterials, 890–891  
Charge-neutrality level (CNL), 20  
Chemical lattice image, 206  
Classical and quantum light, 230–231  
Classical approach for quantum regime, 147–152  
Clearance of nanoparticles from human body, 517–518  
CNT-based X-ray source, 674  
CNT cathode, 684, 690, 691  
CNT cathode array, 690  
CNT cavities as a reaction field of hydrothermal synthesis, 403–412  
CNT field emission cellular microbeam system, 693, 694  
CNT field emission electron and X-ray technology, 673–695  
CNT field emission technology, 689, 693  
CNT field emitters, 749–761  
  catalytic processes of CNT growth, 750–753  
CNT pixels, 689–690  
CNTs and CNPs as cold cathodes, 765–768  
Complementary metal oxide semiconductor (CMOS) technology, 2, 307, 328, 331  
  CMOS device structure, 3  
Conducting metal oxide (CMO), 642, 659–660  
Conduction-band minimum (CBM), 38  
Conduction electron spin resonance (CESR), 90, 97  
Configuration interaction (CI) approach, 50  
Controlled filling of magnetic materials into CNTT, 392–397  
Conventional understanding of Si oxidation, 3–5  
Coordinatively unsaturated sites (CUS), 438  
Coupled circular dots, 57–62  
Coupled Gaussian potential model, 57–67  
Coupled quantum dots, 232–234  
Covalent modulation, 573  
Cross-linkers and blockers, 724  
Current-driven magnetic excitations, 114–118  
Current-induced magnetization switching (CIMS), 108–114  
  time to switch experiments, 108–109  
Current problems with use of nanoparticles in medicine, 513–519  
CVD, 354, 355, 387, 388, 390, 391, 750–751, 809, 831, 835–838  
  growth process of carbon nanopearls, 762–764  
  plasma enhanced CVD (PECVD), 753–756

- Defects properties from first-principles calculations, 16–20  
 Dendrimers, 511–512  
 Dermal exposure assessment, 894–896  
 Device scaling for future MOSFET, 38  
 DFT, 56, 426, 440, 441, 443, 453–455, 463  
 Diffusion theory, 344–347  
 Dip-pen nanolithography (DPN), 823–824  
     concept of, 343–344  
 Directed assembly of nanostructures, 357–369  
     mechanism, 361–369  
     procedure, 357–360  
 Directed self-masking, 831–842  
 Direct patterning of nanostructures, 343–357  
     dip-pen nanolithography (DPN), 343–348  
     microcontact printing (MCP), 348–353  
 Dispersive X-ray spectroscopy, 421  
 Displacement patterning, 820–822  
 DNA, 348, 367, 477, 507, 512, 614  
     heat-induced alterations, 477  
 DNA linkers, 546  
 Domain-wall motion, 112–114  
 Double molecular layer (DML), 344, 345  
 Double-parabola confinement potential, 53  
 Double-parabola potential model, 52–57  
 Double-walled nanotube, 373  
 Double-triple point (DTP) separation, 60–63  
 Dye-sensitized solar cells (DSSCs), 789, 791, 799, 800, 802, 805  
     performance of, 802–805  
 Dynamic random access memory (DRAM), 234  
 Effect of magnetic interaction on the water dispersibility of CNTTs, 400–403  
 Electrochemical deposition (ECD), 393–396, 398  
 Electroless deposition, 840–842  
 Electromagnetic scattering, 551–554  
 Electron-beam lithography (EBL), 826–828  
 Electron double layer, 392  
 Electron energy-loss spectroscopy (EELS), 421, 459, 460, 545  
 Electron field emission from CNTs, 674–677  
 Electronic and transport properties of CNTs, 370–374  
 Electronic transport properties of network devices, 374–380  
     model, 375–377  
     Monte Carlo results, 377–378  
     nanotube network, 378–380  
 Electron paramagnetic resonance (EPR), 141, 146, 156  
 Electrostatic interactions between hydrogen molecule and charges, 710–711  
 Elliptical QD, 64  
 ELISA, 615  
 Environmental decoherence effects in nanomagnets, 161–166  
 Environmental TEM (ETEM), 425, 459–463  
 Epitaxial growth, 210  
     methods of, 210–212  
 Epitaxial quantum dots, 236  
 Evaluating risks associated with nanomaterials, 887–903  
 Exact diagonalization vs. Hubbard model, 62–63  
 Exchange bias, 155  
 Experimental QD IR photodetectors, 260–269  
 Fermi level pinning (FLP), 28  
 Few-electron quantum-dot spintronics, 47–85  
 Few electrons in triple QDs, 76–81  
 Field effect transistor (FET), 355, 883  
 Field emission microscopy (FEM), 736  
 Field-enhancement factor, 783–785  
 Field-induced pyramidal nanotip, 781  
 Field-programmable gate array (FPGA), 307, 308  
 First-principles calculations, 425  
 Fock–Darwin energy spectrum, 53  
 Fowler–Nordheim (FN) plots, 757, 758  
 Fresnel projection microscope (FPM), 742–744  
 FTIR, 454  
 Fullerenes, 386  
 Fundamentals of atomic switch, 299–301  
 Future trends in Si nanotechnology, 42  
 GaAs nanowire, 41, 42  
 Giant magnetoresistance (GMR), 91, 96, 123  
 Giant spin model for nanomagnets, 141–152  
 Gold nanoclusters, 454–458  
 Gold nanoclusters on planar supports, 454–458  
 Gold nanoparticles (AuNP), 509–510  
 Gold nanoshells and nanorods, 492–495  
     physical properties of, 492  
 G-protein coupled receptors (GPCRs), 603  
 Grain-boundary grooving, 778  
 Graphite, 387  
 Grid-like nanostructures, 789  
     for photovoltaic devices, 789–806  
 Half-metallic ferromagnet (HMF), 91, 97  
 Hazard evaluation with nanomaterials, 891–893  
 Heat-induced alterations of the plasma membranes, 476–477  
 Heat-induced alterations to proteins, 477–478  
 Heat-shock proteins, 479–481  
 Heitler–London limit, 52–57  
 Heitler–London method, 50, 57, 85  
 Heitler–London states, 56  
 Heterostructure bipolar transistor (HBT), 182, 189–192, 202  
 High-performance detectors and arrays, 288–289  
 Hole-digging method, 163–165  
 HRTEM (high-resolution transmission electron microscopy), 206, 388, 406, 460, 743, 762  
 Hydrogen–HSM interaction, 700, 705–714  
 Hydrogen interaction with carbon-based sorbants, 711–714  
 Hydrogen storage in nanoscale materials, 699–732  
 Hydrogen-storage materials (HSMs), 699, 714, 717, 731  
     internal interactions in, 714–722  
 Hydrogen-storage properties, 725–731  
     binding energy, 728–729  
     capacity-design principles, 730  
     sorbent design principles, 727–728  
     stability-design principles, 730–731  
 Hyperfine interaction, 166  
 Infra-red photon absorption, 247–254  
 Infra-red spectroscopy, 826  
 Interference lithography, 576  
 Intermolecular dipole interaction, 165, 166  
 Internal interaction in boron-based materials, 720–722  
 Internal interaction in carbon–metal system, 717–720

- Internal interaction in pure carbon and hydrogenated carbon structures, 715–717
- Intersublevel QD infra-red photodetectors, 244–290
- Ion-beam lithography, 575
- Ion-scattering spectroscopy (ISS), 456
- Iron-oxide nanoparticles, 496
- Kinetic Monte Carlo (KMC) simulation, 215, 427
- Kondo effect, 167, 169, 170, 172, 208, 317, 330
- Kubus coordination, 708
- Landau–Zener tunnelling, 142–144
- Langmuir–Blodgett (LB) monolayers, 547
- Large-scale integrated (LSI) circuits, 307
- Laser applications in nanotechnology, 860–885
- Laser thermal annealing (LTA), 866  
for ultrashort pn junctions, 865–867
- Laser as heat source for device nanoprocessing, 865–867
- Laser combination with SPM, 868–871
- Laser-interference lithography (LIL), 879–882, 885
- Laser nanofabrication, 868
- Laser nanopatterning, 879–884
- Lateral force microscopy (LFM), 826
- LH (light hole)/SO states, 199, 201
- Light-harvesting efficiency (LHE), 790, 794, 798
- Liposomes, 512
- Low-energy electron diffraction (LEED), 420, 428
- Magnetic properties of Ni Fe-filled CNTTs, 396–397
- Mainstream nanoelectronic applications, 189–194
- Material issues for field emitters, 748–749
- MCIP (Microcontact insertion printing), 821
- MCP (Microcontact printing), 348, 349, 809, 818, 820
- Mesoscopic physics, 122
- Metalorganic vapor-phase epitaxy, 211
- Metal vapor deposition, 833–835
- Methods of nanostructure printing, 353–357
- Microcomputed tomography, 681–686
- Micro-CT scanners, 682
- Microfluidic parts, 659–661
- Microlens array (MLA), 882–884
- Micromolding in capillaries (MIMIC), 576
- Microradiotherapy system, 688–693
- Microscopic process of Si oxidation, 3–16
- Mid-wave and long-wave QD IR photodetectors, 261–269  
current–voltage characteristics, 270–271  
detector peak responsivity, 271–276  
device characteristics, 270–282  
noise characteristics of the device, 276–278
- M–H curves, 157, 161, 162
- Model catalysts, 438–443
- Modified Landau–Lifshitz domain walls, 106–108
- Modulation design of plasmonics for diagnostic and drug screening, 641–670
- Molecular beam epitaxy (MBE), 182, 185, 211, 224, 231
- Molecular electronics of SAMs, 312–331
- Molecular films used in nanolithography, 809–814
- Molecularly defined structures, 843–845
- Molecular multidot devices, 172–173
- Molecular nanomagnets towards molecular spintronics, 120–157
- Molecular rectifying diode, 323–325
- Molecular rulers, 842–847  
integrated with nanosphere lithography, 845  
outlook, 847  
 patterning, 845–846  
position-selective patterning, 845–846
- Molecular semiconducting wire, 321–323
- Molecular spin-transistor, 146–148
- Molecular spin-valve, 148–149
- Molecular spintronics with SMMs, 145–152
- Molecular switches and memories, 326–330  
charge-based memory, 328–329  
conformational memory, 326–328  
RTD-based memory, 146
- Molecular transistor, 330–331
- Molecular tunnelling barrier, 319–321
- MOSFET, 41, 193, 194, 356
- MOVPE, 222, 223, 225, 236
- Multilayer asymmetrical structure, 646
- Multilayer symmetrical structure, 645
- MWNT, 481, 483, 486, 496, 750, 863, 896
- Nanocatalysis, 416–465
- Nanofabrication of molecular devices, 313–318
- Nano FET (n-FET), 41, 42
- Nanofibers, 890
- Nanografting, 588–591
- Nanoimprint lithography (NIL), 809
- Nanoionic materials, 295–296
- Nanoionics and its device applications, 294–310
- Nano-LAMPs (nanolayered metal probes), 541, 542, 549, 550  
fabrication and applicability, 565–566  
designing, 564–565  
theoretical modelling, 548–564
- Nanolithography using molecular films and processing, 808–848
- Nanomaterial applications, 481–492
- Nanomaterials in consumer products, 888–890  
cosmetics, 888  
sporting equipment, 889
- Nanomaterials in thermal therapy, 481–492
- Nanoparticle–cell interactions, 519–525
- Nanoparticles (NP) (or nanoscale particles NSP), 437–458, 894, 896–899  
exposure and the risk evaluation, 896–899
- Nanoparticles for nanodiagnostic and nanotherapeutic use, 505
- Nanoparticles in medicine, 503–530
- Nanophotonics with QDs, 227–234
- Nanosorbents, 724
- Nano-SQUID, 146, 149
- Nanostructured probes to enhance optical spectroscopic imaging for biomedical applications, 539–566
- Nanotechnology contribution field emitters—electron cold sources, 736–785
- Nanotransfer printing (nTP), 820
- Nanotribology, 864, 868
- Nanowiskers, 889
- Nanowires (NWs), 94, 343, 358
- Nanowires (NWs) for spintronics, 94–96
- Néel–Brown model, 147–149
- Negative differential resistance (NDR), 326, 329, 331

- New catalysts from first principles, 433–437  
 Non-polar and polar covalent bonds, 707–708
- Object reconstruction from FPM images, 741–746  
 Optical lithography, 828  
 Optical matrix element for inter- and intraband transitions, 251–252  
 Optical properties of individual QDs, 554–559  
 Optical tunability, 554–564  
 Organically modified silica (ORMOSIL) nanoparticles, 511  
 Organic LED (OLED), 656  
 Organization for Economic Cooperation and Development (OECD), 899  
   their suggested endpoints for evaluating nanomaterial hazards, 900–901  
 Oscillations of tunnel splitting, 144–147  
 Oxidation-enhanced diffusion (OED), 10  
 Oxidation-induced stacking faults (OSF), 9  
 Oxidation-reduced diffusion (ORD), 10
- Patterning via external tools, 818–831  
 Peak detectivity of QD photodetector, 278–281  
 Peptides, 507  
 Personal protective equipment (PPE) for handling nanomaterials, 902  
 Phase separation in multicomponent assemblies, 815–817  
 Phonon bottleneck, 156, 158, 267  
 Photocatalytic electrodes, 794–799  
 Photoelectrodes with 2D grid-like nanostructures, 791–794  
 Photoluminescence (PL), 229, 796  
 Photon-assisted tunnelling in SMM, 140, 141  
 Physical vapor deposition (PVD), 809, 831  
 Poisson-CI approach, 52  
 Polarization modulation infra-red reflectance absorption spectroscopy (PM-IRAS), 424  
 Polymeric micelles, 507–508  
 Primary, secondary, tertiary and quaternary structures, 507  
 Prism coupler, 650  
 Protein-based nanodevices, 570–605  
 Proteins, 507  
   affinity proteins, 573  
   fundamentals, 572–574  
 Proton sponge effect, 523  
 Pseudogap, 374  
 Pulmonary exposures, 891  
   assessment of, 893–894  
 Pulsed laser ablation for nanomaterials, 860–865  
   synthesis in vacuum, 861–862  
   synthesis in liquid, 862–865
- QD-based FISH/immuno-histo-chemistry techniques, 636  
 QD configuration, 66  
 QD confinement plane, 52  
 QD device, 60  
 QD resistance, 616  
 QD size, 74  
 Quantum cascade emitters, 197–202  
 Quantum cascade lasers (QCL), 197, 199  
 Quantum-dot computer memories, 234–235  
 Quantum-dot crystals, 235–236  
   three-D, 235
- Quantum-dot focal plane array imagers, 282–288  
 Quantum dot—uniformity requirements, 284–288  
 Quantum dots (QDs), 47, 49, 53, 56, 57, 59, 61, 65, 66, 67, 72–76, 79, 81, 82, 147, 151, 508–509, 513, 529, 530, 614–618, 621–629, 896  
   and their epitaxial synthesis, 260–261  
   antibody conjugates, 627, 636  
   based multicolor flow cytometry, 630–631  
   bioprobe, 619, 629, 632–634  
   coupled, 63  
   elliptical deformations, 49  
   gated, 47  
   labelled biomarkers, 632  
   lateral (planar), 48  
   molecular profiling technology, 636  
   photo-physical properties, 615–621  
   science and technology, 208–209  
   streptavidin probes, 633  
   surface, 628  
   vertical, 47, 48
- Quantum-dot-based probes for biomedical applications, 621–629  
   QD bioconjugation, 626–629  
   QD synthesis, 621–622
- Quantum dots—self-organized and self-limiting assembly, 205–238  
 Quantum dynamics of a dimer of nanomagnets, 152–155  
 Quantum efficiency of QD photodetector, 282  
 Quantum Hall effect, 206  
   fractional, 206  
 Quantum nanostructures, 205  
 Quantum phase interference, 147  
 Quantum wire (QW), 172  
 Quantum-wire quantum-dot (QWQD), 48, 49  
   coupled QWQDs, 49
- Rabi oscillations, 159  
 Realistic double QDs, 67–72  
   Coulomb localization in coupled QDs, 67–72  
   hybrid multiscale approach, 67–69
- Redshift between absorbed and emitted photon, 619  
 Reflectance difference spectroscopy (RDS), 14–16  
 Reflection high-energy electron diffraction (RHEED), 211, 215  
 Replica moulding, (REM), 577  
 Research priorities for development of more refined estimates of nanomaterial risk, 899–902  
 Resonant current excitations, 118–123  
 Resonant excitation of domain walls, 121  
 Resonant excitation of spin-waves, 120–121  
 Resonant excitation of vortices, 121  
 Resonant photon absorption, 136–139  
 Resonant tunnelling diodes (RTDs), 194–197  
 RNA, 614  
   Mature RNA (mRNA), 572
- Role of computational sciences in Si-nanotechnology, 1–43
- SAMFET, 330, 331  
 SAMs (Self-assembled monolayers), 316, 317, 319, 321, 322, 326, 330, 344–346, 368–369, 582, 668, 809–813, 817–833, 848  
   insertion and exchange in, 817–818
- Scanning-anode field emission microscope (SAFEM), 756–759  
 Scanning probe lithography (SPL), 576, 822

- Scanning transmission electron microscope (STEM), 421  
 Schottky barrier heights and CMOS applications, 20–38  
 Schottky barrier heights at interfaces, 23–34  
 Schottky barrier—conventional theory, 20–23  
 Schottky barrier height—new theory, 34–37  
 Secondary ion mass spectroscopy, 10  
 SEIRA (surface-enhanced infra-red absorption), 546–548, 550  
 Selected-area electron diffraction (SAED), 395, 405, 408  
 Selectivity of nanoparticles for applications, 516–517  
 Self-assembled hybrid nanodevices, 369–380  
 Self-assembly strategy of nanomanufacturing of hybrid devices, 343–380  
 Self- and directed patterning, 814–818  
 Self-organization in Stranski–Krastanov systems, 212–220  
   general phenomenology, 213–216  
   multiplayer quantum dots, 219–220  
   statistics of quantum dot arrays, 216–219  
   thermodynamic considerations, 212–213  
 SEM, 226, 296, 297, 362, 388, 406, 576, 762, 796, 861, 862  
 Semiconductor photon detectors, 245–247  
   detection through band-to-band transitions, 246  
   detection via impurity level conduction band, 246–247  
 Sensitised TiO<sub>2</sub> stacked-grid array photoelectrodes, 799–802  
 SERS (surface-enhanced Raman spectroscopy), 542–546, 548, 550  
   single-molecule SERS (SM-SERS), 545, 546  
 SEVS (surface-enhanced vibrational spectroscopy), 542–548, 549  
   biomolecular applications of, 548  
 Si–Ge alloys, 181–183  
 Si nanotechnologies, 1  
   present trend, 2–3  
 Si nanowires (SiNWs), 38–41  
 Si nanowire (SiNW) MOSFET, 38  
 Single-atom emitters, 739–742  
 Single-atom nanotip and Fresnel projection microscope, 742–746  
 Single-atom nanotip—microgun, 746–748  
 Single-cell irradiation system, 693–694  
 Single-crystal surfaces, 427–433  
 Single-layer dielectric thin film, 644–645  
 Single-layer metallic film, 645  
 Single molecular magnets (SMMs), 136–174  
 Single-photon sources, 231–232  
 Si/SiGe heterostructures in nanoelectronics, 181–202  
 Site control of quantum dots on patterned substrate, 220–227  
   ordered arrays of QDs on patterned substrate, 224–227  
   thermodynamic considerations, 221–224  
 Size modulation of exchange energy, 74–75  
 Soft lithography, 577, 818–822  
 Solid electrochemical reaction, 296–299  
 Spin-degenerate state, 229  
 Spin diffusion, 96–100  
 Spin-parity effect, 146–147  
 Spin-polarized currents acting on magnetization, 101–108  
 Spin-transfer torque, 102–105  
 Spin-transistor, 167–170  
 Spintronics with metallic nanowires, 90–124  
 Spin-valves and tunnel junctions SPM (scanning probe microscopy), 419, 424, 425  
 SPR angle, 644, 650  
 SPR devices, 646, 657  
 SPR system, 655  
 SPR wave-guide measurement system, 661  
 Stability of nanomaterials, 887  
   safety evaluation, 888  
 Stability of nanoparticles, 513–514  
 STM, 92, 219, 294, 296–298, 316, 318, 329, 420, 421, 427, 433, 435, 438–443, 447–552, 580, 585, 586, 812, 825  
 Strain, 183–186  
 Stranski–Krastanov morphology, 213, 214, 216, 218  
 Strong coupling limit, 169, 170  
 Structure of hydrogen sorbents, 722–725  
 Sum-frequency generation (SFG), 424  
 Super-resolution laser nano-imprinting, 882–884  
 Surface characterization, 419–427  
 Surface chemical reactions, 432–433  
 Surface-enhanced infra-red absorption (SEIRA) spectroscopy, 540, 541  
 Surface-enhanced Raman (SER) spectroscopy, 540, 541  
 Surface-enhanced vibrational (SEV) spectroscopy, 540, 541, 542  
 Surface-plasmon coupled emission (SPCE), 648  
 Surface-plasmon grating coupled emission (SPGCE), 646–649  
 Surface-plasmon resonance (SPR), 509, 643, 649, 662, 667–670  
   long-range SPR (LRSPR), 642, 657–659  
 Surface-plasmon wave (SPW), 652, 657  
 Surface properties of nanoparticles, 514–515  
 Surface structure, 428–429  
 SWNT (single-walled nanotube), 350, 353, 354, 358, 368, 386–413, 496, 581, 597, 676, 683, 750, 892  
 Synthesis of CNTs and carbon nano-test-tubes (CNTTs), 387–392  
 Targeting strategies employing nanotubes, 490–491  
 Temperature-programmed desorption (TPD), 456  
 Templated CNTs and use of their cavities for nanomaterials synthesis, 386–413  
 Terminal events—apoptosis or necrosis, 478–479  
 The interaction–structure–property paradigm, 704–705  
 Thermal ablative therapy in cancer, 475–481  
 Thermally activated magnetization reversal, 147–149  
 Thermotoxicity, 476  
   Mechanisms of, 476–479  
 Three-D nanostructures, 805  
 Three-D open architecture, 790, 791, 794–799  
 Three-D photonic nanodevices, 799  
 Three-D quantum wire, 50  
 Three-D stacked-grid arrays, 790, 798  
 Tip-profile evolution, 774–781  
 Topographically directed etching (TODE), 819  
 Total-energy distribution (TED), 741, 757, 764–765, 773  
 Toxicity, 518–519, 890  
 Toxicology of nanotubes, 488–490  
 Transmission electron microscopy (TEM), 195, 201, 206, 223, 234, 236, 390, 395, 396, 398, 399, 401, 405, 406, 408–410, 420, 421, 462, 463, 601, 631, 632, 793, 898  
   studies of nanoclusters, 458–463  
 Triple-walled nanotube, 373  
 Triple QD (TQD), 76, 77, 81  
 Tumour molecular imaging and profiling, 629–636  
 Tuneable and size-dependent light emission, 616–618  
 Tunnel magnetoresistance (TMR), 91, 92  
 Two-D functional building blocks, 799  
 Two-D nanoscale building blocks, 805

- Two dimensional electron gas (2DEG), 49, 205  
Two-DEG-based QDs, 49  
Two electrons in double QDs, 50–72  
Two electrons in QWQDs, 72–76  
Tyramide signal amplification (TSA), 633
- UHV-STM, 316, 321, 322, 825  
Ultrafast laser-induced phase-change nanolithography, 882, 883  
Ultrahigh vacuum (UHV), 211, 418, 419, 422  
Universality of generalized charge neutrality level, 37–38  
Use of gold nanoshells and nanorods in cancer imaging, 495  
Use of nanoparticles in anticancer thermal therapy, 485  
UV absorbance, 870  
UV patterning nanolithography, 829
- V<sub>2</sub>O<sub>5</sub> nanowire, 358–360  
Vapor-phase epitaxy (VPE), 211–212  
Variational Heitler–London method, 72–74, 85  
Variational Monte Carlo (VMC) model, 56–57, 80, 82–85  
Vector spherical harmonics (VSH), 551, 552  
Vibrational spectroscopy, 423
- Volmer–Weber growth, 212  
Von Neumann–Wigner Theorem in coupled dots, 66–67
- Water-dispersible and magnetically responsive CNTTs, 397–403  
Weak-coupling limit, 168–169
- XAFS (X-ray absorption fine structure), 422, 434, 443  
XPS, 420, 423, 424, 826  
XPS high-pressure (HPXPS), 423  
X-ray (nano)lithography, 830–831  
X-ray photoemission spectroscopy, 420  
X-ray tubes, 678–694
  - CNT-based field emission microfocus, 678–679
  - CNT-based field emission multi-pixel X-ray source, 680–681, 690
  - CNT field emission X-ray source, 678–681
  - multibeam field emission X-ray (MBFEX) source, 688
  - multipixel X-ray, 690
- XRD, 405, 420, 422
- Zero-field oscillations, 117  
ZnO nanowires, 360