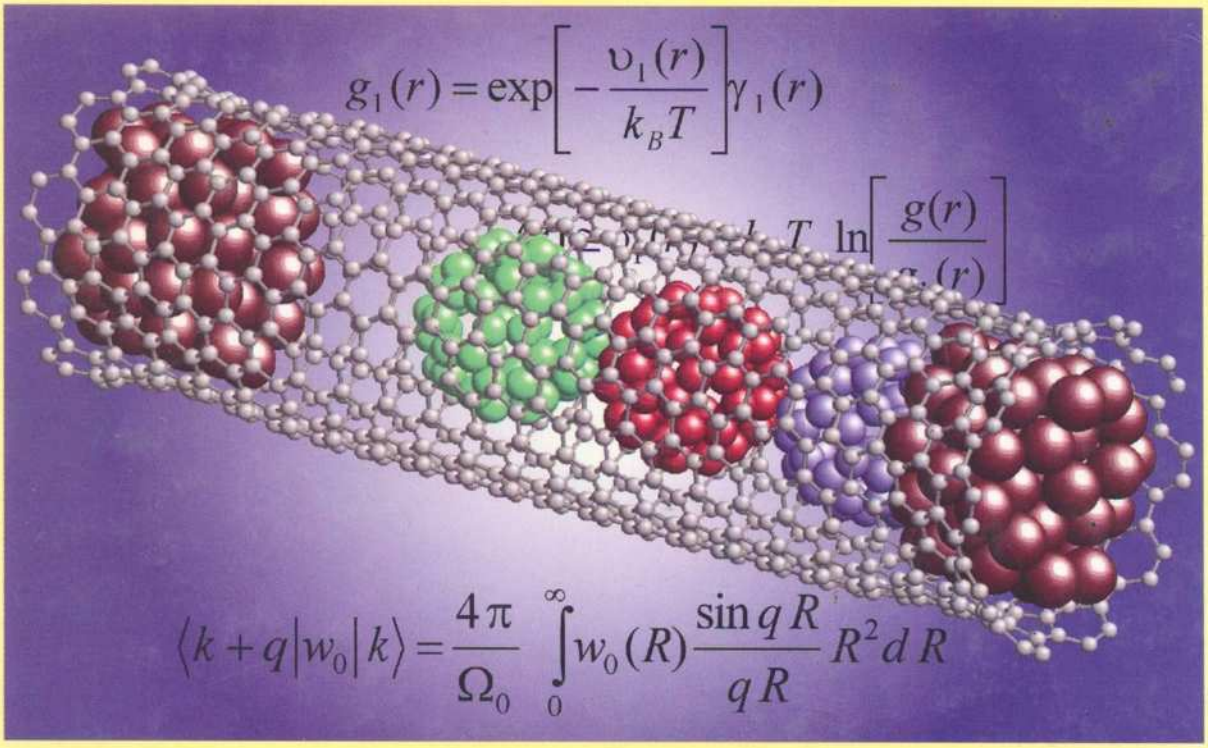
Handbook of

8

THEORETICAL and COMPUTATIONAL NANOTECHNOLOGY

Functional Nanomaterials, Nanoparticles, and Polymer Design



Edited by

M[illegible]olfram Schommers

Fo[illegible]

Pierre-Gilles de Gennes, Nobel Prize Laureate

ASP
AMERICAN SCIENTIFIC PUBLISHERS

Handbook of

THEORETICAL and COMPUTATIONAL NANOTECHNOLOGY

Titles in Nanotechnology Book Series

Founding Editor

Dr. Hari Singh Nalwa

1. **Encyclopedia of Nanoscience and Nanotechnology, 10-Volume Set**
   Edited by Hari Singh Nalwa
2. **Handbook of Theoretical and Computational Nanotechnology, 10-Volume Set**
   Edited by Michael Rieth and Wolfram Schommers
3. **Bottom-up Nanofabrication: Supramolecules, Self-Assemblies, and Organized Films, 6-Volume Set**
   Edited by Katsuhiko Ariga and Hari Singh Nalwa
4. **Handbook of Semiconductor Nanostructures and Nanodevices, 5-Volume Set**
   Edited by A. A. Balandin and K. L. Wang
5. **Handbook of Organic-Inorganic Hybrid Materials and Nanocomposites, 2-Volume Set**
   Edited by Hari Singh Nalwa
6. **Handbook of Nanostructured Biomaterials and Their Applications in Nanobiotechnology, 2-Volume Set**
   Edited by Hari Singh Nalwa
7. **Handbook of Electrochemical Nanotechnology, 2-Volume Set**
   Edited by Yuehe Lin and Hari Singh Nalwa
8. **Polymeric Nanostructures and Their Applications, 2-Volume Set**
   Edited by Hari Singh Nalwa
9. **Soft Nanomaterials, 2-Volume Set**
   Edited by Hari Singh Nalwa
10. **Functional Nanomaterials**
    Edited by Kurt E. Geckeler and Edward Rosenberg
11. **Synthesis, Functionalization and Surface Treatment of Nanoparticles**
    Edited by I. M. Baraton
12. **Quantum Dots and Nanowires**
    Edited by S. Bandyopadhyay and Hari Singh Nalwa
13. **Nanoclusters and Nanocrystals**
    Edited by Hari Singh Nalwa
14. **Molecular Nanoelectronics**
    Edited by Mark A. Reed and T. Lee
15. **Magnetic Nanostructures**
    Edited by Hari Singh Nalwa
16. **Nanoparticles for Pharmaceutical Applications**
    Edited by J. Domb, Y. Tabata, M. N. V. Ravi Kumar, and S. Farber
17. **Cancer Nanotechnology**
    Edited by Hari Singh Nalwa and Thomas Webster
18. **Biochips Nanotechnology**
    Edited by Nongyue He and Hari Singh Nalwa
19. **Nanotoxicology**
    Edited by Yuliang Zhao and Hari Singh Nalwa
20. **Polymer Nanocomposites and Their Applications**
    Written by Suprakas Sinha Ray and Mosto Bousmina
21. **Nanoscale Science and Engineering Education**
    Edited by Aldrin E. Sweeney and Sudipta Seal
22. **Hard Nanomaterials**
    Edited by Hari Singh Nalwa

Additional Volumes in Preparation

Visit: www.aspbs.com

Handbook of
THEORETICAL and COMPUTATIONAL NANOTECHNOLOGY

Volume 8

Functional Nanomaterials, Nanoparticles, and Polymer Design

Edited by

Michael Rieth and **Wolfram Schommers**

Forschungszentrum Karlsruhe, Karlsruhe, Germany

AMERICAN SCIENTIFIC PUBLISHERS

25650 North Lewis Way

Stevenson Ranch, California 91381-1439, USA

AMERICAN SCIENTIFIC PUBLISHERS
25650 North Lewis Way, Stevenson Ranch, California 91381-1439, USA
Tel.: (661) 254-0807
Fax: (661) 254-1207
E-mail: order@aspbs.com
WEB: www.aspbs.com

Handbook of Theoretical and Computational Nanotechnology
edited by Michael Rieth and Wolfram Schommers.

The image on the cover of this handbook was provided by Professor Jeong Won Kang, Chung-Ang University, Seoul, Korea. See Jeong Won Kang, Won Young Choi, and Ho Jung Hwang, *Journal of Computational and Theoretical Nanoscience*, Vol. 1(2), pp. 199–203 (2004). Copyright © 2004, American Scientific Publishers.

This book is printed on acid-free paper. ♾

Library of Congress Control Number: 2003111818

International Standard Book Number: 1-58883-042-X (Set)
International Standard Book Number: 1-58883-043-8 (Volume 1)
International Standard Book Number: 1-58883-044-6 (Volume 2)
International Standard Book Number: 1-58883-045-4 (Volume 3)
International Standard Book Number: 1-58883-046-2 (Volume 4)
International Standard Book Number: 1-58883-047-0 (Volume 5)
International Standard Book Number: 1-58883-048-9 (Volume 6)
International Standard Book Number: 1-58883-049-7 (Volume 7)
International Standard Book Number: 1-58883-050-0 (Volume 8)
International Standard Book Number: 1-58883-051-9 (Volume 9)
International Standard Book Number: 1-58883-052-7 (Volume 10)

Dr. Michael Rieth
Forschungszentrum Karlsruhe
Institute of Materials Research I
D-76021 Karlsruhe, GERMANY

Prof. Dr. Wolfram Schommers
Forschungszentrum Karlsruhe
Institute for Scientific Computing
D-76021 Karlsruhe, GERMANY

PRINTED IN THE UNITED STATES OF AMERICA
10 9 8 7 6 5 4 3 2 1

Foreword

Nanoscience is fashionable. All administrations in the Western world have stressed their interest in nanoobjects and nanotechnologies. As usual, this type of large scientific movement has its pluses and minuses. Many scientists join the crowd without necessarily changing anything in their actual work. Most chemists, for instance, build new molecules that may be called nanoobjects: but again, as usual, the movement does generate significant new content.

Let us, for instance, follow the role of nanostructures in *chemistry*. On one side, nature has provided us with beautiful, robust objects such as fullerenes and carbon tubes, which have some admirable properties. The current challenge is to obtain them in large amounts and at a reasonable price. Here is the real problem.

A completely different sector is obtained from *chemical nanomachines*, for which a molecular unit of nanometric size moves with respect to another one through a change in redox potential or pH. Some of these machines have been built. At the moment, I feel rather skeptical about them because they are extremely costly, extremely fragile (sensitive to poisons), and not easy to protect with a suitable coating—or by a local "antipoison" center. But, here again, there is a challenge.

Let us now turn to *biology*. Here we find an immense group of working nanomachines, enzymes, ionic channels, sensor proteins, adhesion molecules, and so on. They are extremely impressive, but of course they represent progressive construction by trial and error over more than a billion years. Should we try to mimic these machines or, rather, use them for technological purposes, *as they are*, for instance, to grow plants or create proteins at an industrial level according to the techniques of molecular genetics? This is a major question.

A third, open side is *quantum physics* and the (remote) possibility of quantum computers. In my youth, I had hopes for digital storage via quantized flux quanta: The corresponding technology, based on Josephson functions, was patiently built by IBM, but they ultimately dropped out. This shows the hardship of nanotechnologies even when they are handled by a large, competent group. But the cause is not lost, and it may well be that our children use some unexpected form of quantum computers.

Thus, we are facing real challenges, not just the vague recommendations of some anonymous boards. And, we need the tools. We need to know the behavior of materials at the nanlevel, the clever tricks of physical chemistry required to produce nanoparticles or nanopores, the special properties of small cooperative systems (nanomagnets, nanosuperconductors, nanoferroelectrics, etc.), the ability for assembling functional units, and so on.

The aim of the present handbook is to help us with the tools by suitable modelizations. It is written by leading experts, starting from general theoretical principles and progressing to detailed recipes.

In the second half of the 18th century, all the knowledge (fundamental and practical) of the Western world was condensed into an outstanding encyclopedia constructed energetically by Denis Diderot just after the industrial revolution started. Here, at a more modest level, we can hope for something similar. Soon after the first wave, including this handbook, a certain form of nanoindustry may be born.

The discussions started in this handbook will continue in a journal (*Journal of Computational and Theoretical Nanoscience*) launched by the present editors. I wish them the best.

Professor Pierre-Gilles de Gennes
Nobel Prize Laureate, Physics
Collège de France
Paris, France

Preface

This is the first handbook that deals with theoretical and computational developments in nanotechnology. The 10-volume compendium is an unprecedented single reference source that provides ideal introduction and overview of the most recent advances and emerging new aspects of nanotechnology spanning from science and engineering to neurogenetics. Many works in the field of theoretical and computational nanotechnology have been published to date, but no book or handbook has focused on all aspects in this field that deal with nanomachines, electronics, devices, quantum computing, nanostructured materials, nanorobotics, medicine, biology, biotechnology, and more.

There is no doubt that nanoscience will be the dominant direction for technology in this new century, and this science will influence our lives to an extent impossible in years past: Specific manipulations of matter at its ultimate level will open completely new perspectives on all scientific and technological disciplines. To be able to produce optimal nanosystems with tailor-made properties, it is necessary to analyze and construct such systems in advance by adequate theoretical and computational methods. The handbook gives a complete overview of the essential methods, models, and basic pictures.

But, as is well known, there are also threats connected with nanotechnology, specifically with respect to biological systems: Self-assembly can be an uncontrolled process, and the final state of a developing system is in general uncertain in such cases. To avoid undesirable developments, the theoretical (computational) analysis of such processes is not only desirable but also absolutely necessary. Thus, the computational and theoretical methods of nanoscience are essential for the prediction of new and custom nanosystems and can help keep nanoscience under control. There is basically no alternative. Therefore, one possible answer to the question, "Why a book on theoretical and computational nanotechnology?" is *to give nanotechnology a direction!*

In the design of macroscopic and microscopic systems, engineering is essentially based on *intuitive concepts*, which are tailored to observations in everyday life. Classical mechanics is also based on these macroscopic observations, and its notions have been chosen with respect to our intuitive demands for *visualizability*. However, when we approach the nanolevel, the tools used for the design of macroscopic and microscopic systems become more and more useless. At the nanolevel, *quantum phenomena* are dominant, and the main features in connection with quantum effects are not accessible to our intuitive concepts, which are merely useful at the macroscopic level; the framework of quantum theory is in striking conflict with our intuitive demands for visualizability, and we are forced to use abstract physical laws expressed by mathematical equations. In other words, effects at the nanolevel are (almost) not accessible to our usual engineering concepts. Therefore, here we rely on the abstract mathematical relations of theoretical physics. In nanotechnology functional systems, machines and the like cannot be adequately designed without the use of these abstract theoretical laws and the application of suitable computational methods. Therefore, in nanotechnology, theoretical and computational methods are centrally important: This makes the present handbook an indispensable compendium.

Nanometer-scale units are by definition very small atomic structures and functional systems; it is the smallest level at which functional matter can exist. We already learned to manipulate matter at this ultimate level: Atoms can be moved experimentally in a controlled manner from one position to another. This is astonishing because one nanometer only corresponds to one millionth of a millimeter. For example, an electrical nanogenerator could be designed consisting of various parts that included a very fast revolving rotator. One million of these generators could be arranged side by side on a length of two centimeters; it is remarkable that not only *static* nanostructures could in principle be produced and significantly manipulated but also artificial *dynamical* nanosystems. But, the downscaling of functional structures from the macroscopic to the nanometer scale is only one of the essential

points in connection with nanotechnology. In addition—and maybe much more important—nanosystems provide unique properties in comparison to those we observe at the macroscopic level. For example, a metal nanocluster shows a melting temperature that strongly deviates from that of a macroscopic piece of metal; its melting point is significantly lower. A decrease down to a fraction of only 20% is typical, depending, however, on the material and particle number.

A professional treatment of the various problems in nanoscience and nanotechnology makes the application and development of theoretical and computational methods in this field absolutely necessary. In other words, the discipline of theoretical and computational nanotechnology has to be considered as a key topic to be able to treat nanotechnology adequately and to reach optimal solutions for certain tasks. It is therefore desirable to get a timely overview about the specific topics presently relevant in this field. In this respect, the handbook gives a complete overview of the specific topics so far established in nanotechnology. Each chapter gives a certain overview of actual activities of the envisaged topic and in most cases an adequate description of the basics, so advanced students also can benefit from the handbook. It was our strategy to provide consistent and complete representations so the reader would be able to study each chapter without consulting other works. This of course leads to certain overlaps, which was also part of our strategy to enable an approach to the same topic from various points of view.

The handbook reflects the spectrum of questions and facts that are and could be relevant in the field of nanotechnology. Not only formal developments and methods are outlined, but also descriptions of a broad variety of applications in particular are typical for the handbook. All relevant topics have been taken into account, from functional structures—like an electrical nanogenerator—or quantum computing to questions that deal directly with basic physics. Almost all fields related to theoretical and computational nanotechnology could be covered, including *multiscale modeling*, which is important for the transition from microscale to nanoscale and vice versa.

All theoretical and computational methods used in connection with the various topics in nanoscience are directly based on the *same* theoretical physical laws. At the nanolevel, all properties of our world emerge at the level of the *basic* theoretical laws. In traditional technologies, engineers do not work at the ultimate level. They use more or less phenomenological descriptions that generally cannot be deduced from the basic physical theoretical laws. We have as many phenomenological descriptions as there are technological disciplines, and each is tailor-made to a specific topic. An exchange of concepts is either not possible or rather difficult. In contrast, at the ultimate nanolevel the world is based on only one theory for all disciplines, and this is expressed by basic theoretical physics. This situation opens the possibility for interconnections between the various topics in nanotechnology to bring about new effects and chances for further applications. In other words, nanotechnology and nanoscience can be considered interdisciplinary. Clearly, the handbook reflects the interdisciplinary character of this new science and technology.

The *Handbook of Theoretical and Computational Nanotechnology* includes 138 chapters written by hundreds of the world's leading scientists. Topics cover mainly the following areas:

(i) Computational biology: DNA, enzymes, proteins, biomechanisms, neurogenetic information processing, and nanomedicine
(ii) Computational chemistry: quantum chemistry, molecular design, chemical reactions, drugs, and design
(iii) Computational methods and simulation techniques from *ab initio* to multiscale modeling
(iv) Materials behavior at the nanolevel, such as mechanics, defects, diffusion, and dynamics
(v) Nanoscale processes: membranes, pores, diffusion, growth, friction, wear, catalysis
(vi) Nanostructured materials: metals, composites, polymers, liquid crystals, photonic crystals, colloids, and nanotubes
(vii) Nanostructures: fullerenes, nanotubes, clusters, layers, quantum dots, thin films, surfaces, and interfaces
(viii) Nanoengineering and nanodesign: nanomachines, nano-CAD, nanodevices, and logic circuits

(ix) Nanoelectronics: molecular electronics, nanodevices, electronic states, and nanowires
(x) Nanomagnetism: magnetic properties of nanostructures and nanostructured materials
(xi) Nanooptics: optical response theory, quantum dots, luminescence, and photonic crystals
(xii) Quantum computers: theoretical aspects, devices, and computational methods for simulating quantum computers and algorithms

The handbook provides broad information on all basic and applied aspects of theoretical and computational nanotechnology by considering more than two decades of pioneering research. It is the only scientific work of its kind since the beginning of nanotechnology, bringing together core knowledge and the very latest advances. The handbook is written for audiences of various levels while providing the latest up-to-date information to active scientists and experts in the field. This handbook is an indispensable source for research professionals and developers seeking the most up-to-date information on theoretical and computational nanotechnology among a wide range of disciplines, from science and engineering to medicine.

This handbook was written by leading experts, and we are highly grateful to all contributing authors for their tremendous efforts in writing these outstanding state-of-the-art chapters that altogether form a unified whole. K. Eric Drexler (designer of nanomachines, founder of the Foresight Institute, coiner of the term *nanotechnology*) gives an excellent introductory chapter about possible trends of future nanotechnology. We especially express our sincere gratitude to Dr. Drexler for his instructive and basic representation.

We cordially extend our special thanks to Professor Pierre-Gilles de Gennes for his valuable and insightful Foreword.

The editors are particularly thankful to Dr. Hari Singh Nalwa, President and CEO of American Scientific Publishers, for his continuous support of the project and the enthusiastic cooperation in connection with all questions concerning the development of the handbook. Furthermore, we are grateful to the entire team at Bytheway Publishing and especially to Kate Brown for copyediting.

Dr. Michael Rieth
Prof. Dr. Wolfram Schommers
Karlsruhe, Germany

Contents

Foreword v
Preface vii
About the Editors xxi
List of Contributors xxiii
Contents of Volumes in This Set xxvii

CHAPTER 1. Computational Studies of Nanomaterials: A Historical Perspective

Douglas L. Irving, Susan B. Sinnott

1. Introduction 1
2. Introduction to Computational Methods 2
   2.1. Calculating Interaction Energies and Forces 2
   2.2. Structure Optimization and Simulation 6
3. Application to Nanometer-Scale Systems 10
   3.1. Electronic Properties of Single-Walled Carbon Nanotubes 10
   3.2. Structure and Properties of Nanotube Junctions 14
   3.3. Mechanical Properties of Carbon Nanotubes 19
   3.4. Chemical Modification of Carbon Nanotubes 22
   3.5. Carbon Nanotube Gas Sensors 26
   3.6. Filling Carbon Nanotubes 29
   3.7. Nanoindentation 34
4. Summary and Future Outlooks 36
   References 37

CHAPTER 2. Density Functional Calculations of Clusters and Cluster Assembly

J. A. Alonso, M. J. Stott

1. Introduction 43
2. Theory 47
   2.1. Density Functional Theory 47
   2.2. Pseudopotentials 57
   2.3. Synthesis 60
3. Applications to Clusters and Cluster-Assembled Solids 63
   3.1. Electronic Shells 63
   3.2. Perturbed Clusters 70
   3.3. Clusters with Impurities 77
   3.4. Cluster Melting 87
   3.5. Cluster Assembly 92
4. Conclusions 97
   References 98

CHAPTER 3. Modeling the Structural Evolution, Equilibrium Morphology, and Macroscopic Behavior of Polymer/Nanoparticle Composites

Anna C. Balazs, Gavin A. Buxton

1. Introduction 104
2. Models for Determining the Structure of Nanocomposites 106
   2.1. Self-Consistent Field/Density Functional Theory Approach 106
   2.2. Cahn-Hilliard/Brownian Dynamics Model 108
3. Models for Determining Macroscopic Properties of Nanocomposites 111
   3.1. Mechanical Properties: Lattice Spring Model 111
   3.2. Electrical Properties: Finite Difference Model 113
   3.3. Optical Properties: Finite Difference Time Domain Method 114
4. Results and Discussion 116
   4.1. Properties of Mixtures of Diblock Copolymers and Spherical Nanoparticles 116
   4.2. Behavior of Nanorods in Binary Polymer Blends 126
   4.3. Properties of Mixtures of Nanorods and Diblock Copolymers 132
   4.4. Structure and Optical Properties of Nanoparticle-Filled Diblock Copolymers 142
5. Conclusions 146
   References 147

CHAPTER 4. Monte Carlo Simulations and Self-Consistent Field Theory for Thin Polymer Films

Marcus Müller

1. Introduction 152
2. Coarse-Grained Models 153
3. Monte Carlo Simulation and Self-Consistent Field Technique 156
   3.1. Simulation Techniques for One-Component Polymer Liquids 156
   3.2. Self-Consistent Field Technique for One-Component Polymer Liquids 160
4. Wetting of Polymer Liquids 170
   4.1. Locating the Wetting Transition in Monte Carlo Simulations 170
   4.2. Nanodroplets on Layered Substrates 175
   4.3. Wetting on a Polymer Brush 180
   4.4. Adsorption on Laterally Structured Substrates 183
5. Binary Polymer Brushes: Self-Assembled Structures with Switchable Wetting Properties 184
6. Incompressible Binary Polymer Films 189
   6.1. Generalization of the Computational Techniques to Binary Polymer Blends 189
   6.2. Wetting Transition in Incompressible Binary Polymer Blends 194
   6.3. Interplay Between Wetting and Phase Behavior in Confined Geometry 195
7. Outlook 198
   References 199

CHAPTER 5. Conjugated Organic Polymers: From Bulk to Molecular Wire

Ulrike Salzner

1. Introduction 204
   1.1. Properties of Conjugated Organic Polymers 204
   1.2. Theoretical Approach 206
2. Accuracy of Theoretical Excitation Energies for Polyenes in the Gas Phase 209

2.1. High Level *Ab Initio* Calculations ... 209
2.2. Single-Electron Methods ... 211
2.3. Calculation of Excitation Energies Employing Density-Functional Theory ... 213
2.4. Frozen Orbital Approximation ... 214
3. Extrapolation to Infinitely Long Chains ... 217
4. Excitation Energies of Oligomers in Liquid and Solid Solutions ... 221
4.1. Polyenes ... 222
4.2. Thiophene Oligomers ... 222
4.3. Pyrrole Oligomers ... 223
5. Interchain Interactions ... 224
5.1. Experiment ... 224
5.2. Theory ... 225
6. Comparison Between Theory and Experiment for Polymers ... 227
7. Conducting State of Conjugated Organic Polymers ... 229
8. Polarizability ... 235
9. Conductivity ... 237
10. Summary ... 239
References ... 241

CHAPTER 6. Nanomechanics of Nanoreinforced Polymers

Frank T. Fisher, L. Cate Brinson

1. Introduction ... 254
1.1. Motivation for Nanoreinforced Materials ... 254
1.2. Forms of Nanoscale Reinforcement ... 258
1.3. Introduction to Viscoelasticity ... 269
2. Nanoreinforced Polymers—Experimental Results ... 274
2.1. Equiaxed Nanoparticulate Reinforcement ... 274
2.2. Nanotube Reinforcement ... 277
2.3. Nanoplatelet Reinforcement ... 288
2.4. The Interphase Region and Viscoelastic Behavior in Nanoreinforced Polymers ... 292
3. Nanomechanical Modeling ... 298
3.1. Bottom-up Approaches ... 299
3.2. Top-down Approaches ... 309
3.3. Hybrid Modeling Methods ... 322
4. Current Issues in Nanomechanics of Nanoreinforced Polymers ... 328
5. Conclusions ... 332
Appendix ... 334
A.1 Derivation of Mori-Tanaka Method ... 334
A.2 Summary of Related Literature ... 351
References ... 352

CHAPTER 7. Modeling and Simulation of Carbon Nanotube/Polymer Composites

Jihua Gou, Kin-tak Lau

1. Introduction ... 362
2. Interfacial Bonding of Carbon Nanotube/Polymer Composites ... 363
2.1. Molecular Mechanics Simulation ... 364
2.2. Molecular Dynamics Simulation ... 368
2.3. Analytical Modeling with Classical Continuum Mechanics ... 369

3. Mechanical Behavior of Carbon Nanotube/Polymer Composites 377
3.1. Molecular Mechanics and Molecular Dynamics Simulations 378
3.2. Continuum Modeling 380
3.3. Multiscale Modeling 383
4. Experimental Investigation of Carbon Nanotube/Polymer Composites 388
4.1. Interfacial Bonding of Carbon Nanotube/Polymer Composites 388
4.2. Mechanical Properties of Carbon Nanotube/Polymer Composites 390
5. Conclusion 391
References 391

CHAPTER 8. Nano-Characterization of Materials: Silicon, Copper, Carbon Nanotubes, and Diamond Thin Films

Liangchi Zhang

1. Introduction 396
2. Molecular Dynamics Modeling of Materials 396
2.1. Initial Model 396
2.2. Simulation Temperature 397
2.3. Size of Control Volume 398
2.4. Integration Time Steps 398
2.5. Temperature Conversion 399
2.6. Stress Analysis 399
2.7. Potential Function 401
3. Nanoindentation on Monocrystalline Silicon 402
3.1. Introduction 402
3.2. Modeling 403
3.3. Phase Transformation 405
3.4. Deformation Characteristics 409
3.5. Effect of Oxygen Penetration 413
3.6. Effect of Water 415
3.7. Cyclic Indentations 417
4. Nanotribology 420
4.1. Introduction 420
4.2. Moving Control Volume 421
4.3. Diamond–Copper Sliding Systems 422
4.4. Scale Effect of Contact Size on Friction Transition 429
4.5. Diamond–Silicon Sliding Systems 433
4.6. Multiasperity Sliding 437
5. Characterization of Carbon Nanotubes 440
5.1. Introduction 440
5.2. Modeling 441
5.3. Deformation Characteristics 441
5.4. Potential 441
5.5. Number of Thermostat Atoms 443
5.6. Influence of Thermostat Schemes 444
5.7. Integral Time Step, Displacement Step, and Relaxation Step 445
5.8. Summary 446
6. Deformation of Copper Nanowhiskers 446
6.1. Introduction 446
6.2. Modeling 447
6.3. Effects of Atomic Orientation and Specimen Shape 447

6.4. Elastic Modulus and Ultimate Tensile Strength 448
6.5. Shear Banding and Necking 448
7. Buckling of Diamond Thin Films 449
7.1. Background 449
7.2. Theory and Simulation Method 450
7.3. Uniaxial Stress 452
7.4. Biaxial Stresses 452
8. Concluding Remarks 453
References 454

CHAPTER 9. Isomeric Fullerenes and Endofullerenes: Stability Computations on Promising Nanoscience Agents

Zdeněk Slanina, Kaoru Kobayashi, Shigeru Nagase

1. Introduction 458
2. Outline of Quantum-Chemical and Statistical-Mechanical Calculations 460
3. Energetics and Thermodynamics of Carbon Clusters 463
3.1. Basic Topology 463
3.2. Isolated Pentagon Rule Energetics 464
3.3. Stabilities of Clusters of Different Dimensions 465
4. Small Carbon Clusters 468
5. Topological Generations of Cages 470
6. Smaller Fullerenes 472
7. Higher Fullerenes 475
7.1. C_{72} Fullerenes 475
7.2. C_{74} Fullerenes 475
7.3. C_{76} Isolated Pentagon Rule Set 478
7.4. C_{78} Isolated Pentagon Rule Set 479
7.5. C_{80} Isolated Pentagon Rule Set 480
7.6. C_{82} Isolated Pentagon Rule Set 482
7.7. C_{84} Isolated Pentagon Rule Set 482
7.8. C_{86} Isolated Pentagon Rule Set 483
7.9. C_{88} Isolated Pentagon Rule Set 483
7.10. C_{90} Isolated Pentagon Rule Set 485
7.11. C_{92}–C_{98} Isolated Pentagon Rule Sets 487
8. Endohedral Metallofullerenes 488
9. Fullerene Derivatives 493
10. Conclusions 496
References 496

CHAPTER 10. Carbon Nanocones

Henning Heiberg-Andersen

1. Introduction 507
2. Conic Shapes in Graphite: Topology, Geometry, and Stability Considerations 510
2.1. Carbon Chemistry in Two and Three Dimensions 510
2.2. Nanocones 511
2.3. Helical Cone Growth 515
2.4. Horns, Pipettes, Zipped Structures, and Tubular Graphite Cones 518

3. Atomic and Electronic Structure: Theoretical/Computational Methods and Predictions 519
3.1. Molecular Mechanics 519
3.2. Density Functional Theory 520
3.3. The Tight-Binding Model 522
3.4. The Effective Mass Theory and Continuum Models 523
3.5. Predictions 525
4. The Nucleation Puzzle 532
References 533

CHAPTER 11. Simulation and Optimization of Composite Doped Metamaterials

Christian Hafner, Jasmin Smajic, Daniel Erni

1. Introduction 538
2. Metamaterials 540
2.1. Frequency Selective Surfaces: Gratings 540
2.2. Artificial Anisotropic Media 544
2.3. Artificial Chiral Media 546
2.4. Negative Index Materials: Left-Handed Media 548
3. Photonic Crystals 553
3.1. Perfect Photonic Crystals 555
3.2. Waveguides in Photonic Crystals 560
3.3. Waveguide Discontinuities in Photonic Crystals 564
4. Simulation of Composite Doped Metamaterials 577
4.1. Selection of Numerical Methods 578
4.2. Band Diagram Computation 582
4.3. Guided Waves 585
4.4. Waveguide Discontinuities 590
5. Optimization of Composite Doped Metamaterials 595
5.1. Probabilistic Optimizers 599
5.2. Deterministic Optimizers 603
5.3. Sensitivity Analysis 604
5.4. Towards Phenomenological Two-Dimensional Modeling of Planar Photonic Crystal Devices 605
6. Conclusions 611
References 612

CHAPTER 12. Theoretical and Computational Atomic-Scale Studies of Complex Catalytic Materials

Karl Sohlberg, Sergey N. Rashkeev

1. Introduction 615
2. Theoretical Methodologies 616
2.1. Cluster Models 616
2.2. Models Employing Periodic Boundary Conditions 617
2.3. Analysis Techniques 620
2.4. Leveraging Experiments 620
3. Typical Studies and Results 620
3.1. Nanocatalysts 620
3.2. Metal Oxide Materials 625
3.3. Catalytic Systems Based on Supported Metal Particles 630

3.4. Stability of the Support Material 634
3.5. Photocatalysis 638
4. Future Developments 640
4.1. Some Selected Outstanding Problems 640
4.2. Leveraging Infrared Spectroscopy 641
References 642

CHAPTER 13. Properties of Superconducting Nanostructures

Rosario Fazio, Fabio Taddei

1. Introduction 647
2. Hybrid Systems 648
2.1. Andreev Reflection and Proximity Effect 648
2.2. Derivation of the Bogoliubov–de Gennes Equation 650
2.3. Solutions of the Bogoliubov–de Gennes Equation 654
2.4. Scattering Theory 659
2.5. Two-Probe Differential Conductance and Conductance in More Details 665
2.6. Quasi-Classical Green's Function Approach 668
2.7. Panorama on Results in NS Systems 670
3. Coulomb Blockade and Superconductivity 671
3.1. Tunneling Rates and Parity Effects 671
3.2. NSN Transistor 674
3.3. Cooper Pair Tunneling 675
3.4. Superconducting Nanoparticles 677
3.5. Josephson Arrays 678
3.6. Quantum Computation with Superconducting Nanocircuits 681
4. Conclusions 683
References 684

CHAPTER 14. Strain Field Calculations in Embedded Quantum Dots and Wires

R. Maranganti, P. Sharma

1. Introduction 689
2. Strain Field Calculations in Quantum Dots and Wires 691
2.1. Simple Illustrative Example 692
2.2. Effect of Shape 694
2.3. Effect of Presence of a Free Surface in Near Vicinity 708
2.4. Effect of Material Anisotropy and Nonlinearity 711
2.5. Effect of Coupled Fields: Piezoelectricity 715
2.6. Effect of Size 718
2.7. Some Results from Inclusion Theory Useful for Strain Calculations in Quantum Dot Structures 726
3. Summary and Open Issues 727
References 728

CHAPTER 15. Optical Properties of Silicon Quantum Wires and Dots

Xanthippi Zianni, Androula G. Nassiopoulou

1. Introduction 733
2. Electronic Structure Calculations 735

2.1. Effective Mass Approximation 736
2.2. Atomistic Approaches 742
2.3. Quantum Confinement 746
2.4. Optical Properties 748
3. Conclusion 765
References 765

CHAPTER 16. Real-Space Electronic-Property Calculations for Nanoscale Structures

T. Torsti, V. Lindberg, I. Makkonen, E. Ogando, E. Räsänen, H. Saarikoski, M. J. Puska, R. M. Nieminen

1. Introduction 772
2. Equations of Density-Functional Theory 774
3. Real-Space Multigrid Methods 775
3.1. Previous Real-Space Approaches 775
3.2. RQMG Method 776
4. Technical Enhancements 778
4.1. Double Grid Technique 778
4.2. Traditional Mixing Schemes 780
4.3. Response-Iteration Methods 781
4.4. Higher-Order Compact Discretizations 782
4.5. RQMG with Galerkin Conditions 783
4.6. Alternative Eigenproblem Solvers 783
5. Status of the General-Purpose Real-Space Tool 784
6. Two-Dimensional Quantum Dots 785
6.1. Introduction and the Model 785
6.2. Computational Aspects 785
6.3. Zero-Field Results 786
6.4. Magnetic Fields and the Vortex Clusters 786
6.5. Impurities in Quantum Dots 788
7. Nanophysics in Axial Symmetry 788
7.1. Ultimate Jellium Model for a Breaking Nanowire 789
7.2. Adsorbed Na Quantum Dots on Cu(111) 790
8. Positron States at Nanoscale Defects 791
9. Summary and Outlook 792
References 793

CHAPTER 17. Electronic Structure of Clusters and Nanocrystals

James R. Chelikowsky, Yousef Saad

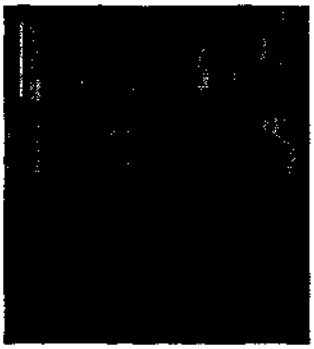
1. Introduction 798
2. Quantum Descriptions of Matter 798
2.1. The Hartree Approximation 800
2.2. The Hartree–Fock Approximation 801
3. Density-Functional Approaches 803
3.1. Free-Electron Gas 803
3.2. Hartree–Fock Exchange in a Free-Electron Gas 804
3.3. Density-Functional Theory 805
3.4. Time-Dependent Density Functional Theory 807

4. Pseudopotentials 812
5. Solving the Eigenvalue Problem 818
6. Properties of Confined Systems: Clusters 822
6.1. Structure 822
6.2. Photoemission Spectra 824
6.3. Vibrational Modes 825
6.4. Polarizabilities 828
6.5. Optical Spectra 830
7. Quantum Confinement in Nanocrystals 833
7.1. Role of Oxygen in Silicon Quantum Dots 840
7.2. Doping Quantum Dots 845
References 849

Index 853

About the Editors

Dr. Michael Rieth has been a research scientist at the Institute of Materials Research I (IMF-I) in the Forschungszentrum Karlsruhe, Germany, since 2002. He has been head of the consulting company AIFT, Karlsruhe, since 1987. He worked as a researcher at the Institute of Materials Research II (IMF-II), Forschungszentrum Karlsruhe, from 1995 to 1999 and at the Engineering Science Department of the University of Patras (Greece) from 1999 to 2000. He was product manager at AMA Systems, Pforzheim, Germany, from 2000 to 2001. He received his master of science (German Dipl. Ing.) degree in electrical engineering from the University of Karlsruhe in 1991 and his doctoral degree in physics from the University of Patras (Greece) in 2001. Dr. Rieth published 23 research articles in refereed journals, 2 book chapters, and four patents. He is the author of *Nano-Engineering in Science and Technology* (World Scientific, Singapore, 2003) and was the editor-in-chief of the *Journal of Computational and Theoretical Nanoscience* (2004–2005). His main scientific interests are in atomistic modeling of metallic nanosystems and materials development for advanced fusion reactor applications.



Prof. Dr. Wolfram Schommers is a theoretical physicist and is presently at the Research Center of Karlsruhe in Germany. He is also professor of theoretical physics, professor of physics and materials sciences, and distinguished professor in Europe, China, and the United States. He began his studies of theoretical physics at the Technical University of Munich and continued his course work at the University of Münster, receiving a diploma in physics. After a brief intermezzo in the industry, Professor Schommers joined the Research Center of Karlsruhe. He received his doctoral degree (Dr. rer. nat.) in theoretical physics from the University of Karlsruhe.

Professor Schommers concentrates his scientific activities on computational and theoretical physics. His main fields of interest include foundations of physics, liquids, solids, and gases; superionic conductors; surface science; and nanophysics as the basis for the investigation of properties of nanometer-scale atomic devices, junctions, quantum dots, and nanomachines. He has published the results of his research and thoughts in various scientific journals (214 articles and book chapters).

Some topics concerning liquids, solids, and gases concern interaction potentials, single-particle motion, diffusion, generalized phonon density of states, collective motion, and liquid-solid phase transition. Selected articles include "The Effect of van der Waals-Type Interactions in Metals: A Pseudopotential Model" (*Zeitschrift für Physik B* 121, 1976); "Liquid Argon: The Influence of Three-Body Interactions on Atomic Correlations" (*Physical Review A* 16, 327, 1977); "Theoretical Investigation of the Liquid Solid Transition. A Study for Gallium" (*Solid State Communications* 21, 65, 1977); "Pair Potentials in Disordered Many-Particle Systems: A Study for Liquid Gallium" (*Physical Review A* 28, 3599, 1983); "Many-Body Polarization and Overlap Effects in the Dynamic Structure Factor of Dense Krypton Gas" (with P. A. Egelstaff, J. J. Salacuse, and J. Ram; *Physical Review A* 34, 1516, 1986); and "Comment on 'Pair Interaction from Structural Data for Dense Classical Liquids'" (*Physical Review Letters* 58, 427, 1987).

Topics in connection with superionic conductors involve structure and dynamics, correlated motions, and collective behavior. Selected articles include "Correlations in the Motions of Particles in AgI: A Molecular-Dynamics Study" (*Physical Review Letters* 38, 1536, 1977);

"Current–Current Correlations in AgI" (*Physical Review B* 16, 327, 1977); "Structure and Dynamics of Superionic Conductors" (*Physical Review B* 21, 847, 1979); "Triplet Correlations in Solid Electrolytes" (*Solid State Ionics* 1, 473, 1980).

Topics concerning surface physics touch on temperature effects, structure, dynamics, and interaction. Selected works are as follows: "Structural and Dynamical Behaviour of Noble-Gas Surfaces" (*Physical Review A* 32, 6845, 1985); "Statistical Mechanics of the Liquid Surface and the Effect of Premelting," in *Structure and Dynamics of Surfaces II* (Springer-Verlag, Heidelberg, 1987); "The Effect of Non-Linear Interactions at the Surface of Solids" (*Surface Science* 269/270, 180, 1992); and "Steps, Point Defects and Thermal Expansion at the Au(100) Surface" (with H. Zimmermann, M. Nold, U. Romahn, and P. von Blanckenhagen; *Surface Science* 287/288, 76, 1993).

Some details regarding the work of Professor Schommers on nanophysics include study of nanoclusters, nanostructures, and nanomachines; temperature effects; and electronic states. Selected works include "Phonons and Structure in Nano-Clusters: A Molecular Dynamics Study for Al" (*Nanostructured Materials* 9, 621, 1997); "Excited Nano-Clusters" (*Applied Physics A* 68, 187, 1999); "Thermal Stability and Specific Properties of Nanosystems" (with S. Baskoutas and M. Rieth; *Modern Physics Letters B* 14, 621, 2000); "Computational Atomic Nanodesign," in *Encyclopedia of Nanoscience and Nanotechnology* (with M. Rieth; American Scientific Publishers, Stevenson Ranch, CA, 2004); "Computational Engineering of Metallic Nanostructures and Nanomachines" (with M. Rieth; *Journal of Nanoscience and Nanotechnology* 2, 679, 2002); and "Electron in an Interaction Potential of General Shape" (with M. Rieth; *Journal of Computational and Theoretical Nanoscience* 2, 362, 2005).

Concerning the foundations of physics, Professor Schommers has discussed new aspects in connection with reality, and his basic ideas can be summarized as follows: Information about reality outside flows via sense organs into the body of the observer, and the brain forms a picture of reality. On the basis of many facts, Schommers concluded that the symbols in this picture of reality should have in general no similarity with the objects in the outside world; that is, the reality outside is transformed. On the one hand, we have the reality; on the other hand, we have a picture of reality. The reality is projected on space and time, and we obtain a picture of reality; the structures in the pictures are different from those in the reality outside. This conception is discussed mathematically by Professor Schommers in connection with quantum phenomena leading to new aspects in connection with relevant basic topics. Like both Whitehead and Bergson, Schommers argues for the primacy of processes and shows that space and time are closely tied to real processes. Selected work in this regard are "Inertial Frames of Reference: Mass Coupling to Space and Time" (*International Journal of Theoretical Physics* 20, 411, 1981); "Raum-Zeit, Quantentheorie und Bilder von der Wirklichkeit" (*Philosophia N ralis* 23, 238, 1986); "Being and Becoming at the Microscopic Level" (*International Journal of Modern Physics B* 3, 1, 1989); "Space-Time and Quantum Phenomena," in *Quantum Theory and Pictures of Reality* (Springer-Verlag, Heidelberg, 1989); and "Truth and Knowledge," in *What Is Life?* (World Scientific, Singapore, 2002).

Professor Schommers is author and editor of the following books: *Fundamentals of Nanometer Structuring; Structure and Dynamics of Surfaces I and II; Quantum Theory and Pictures of Reality; The Visible and the Invisible; Das Sichtbare und das Unsichtbare; Elemente des Lebens; What is Life?; Formen des Kosmos; Space and Time, Matter and Mind; Symbols, Pictures and Quantum Reality*.

Professor Schommers is the editor-in-chief of the *Journal of Computational and Theoretical Nanoscience*. He is also an editorial board member of various scientific journals, and he is principal editor-in-charge of the book series *Foundations of Natural Science and Technology*. He is an invited member of the Humboldt Academy, an invited member of the Academic Board of the Humboldt Society, and an invited member of the Advisory Board of Medical Ethics of the 21st Century. Professor Schommers is also deputy governor of the American Biographical Institute (inauguration 2000).

Professor Schommers has been honored by various awards, medals, and appointments. He has been cited in *Who's Who in the World, Who's Who in Science and Technology, Living Science, The Europe 500, The Barons 500, 2000 Outstanding Intellectuals of the 21st Century, Leading Intellectuals of the World, 500 Leaders of Influence, and International Register of Profiles* (no. 123 of 200), and elsewhere.

List of Contributors

Number in parentheses indicates the page on which the author's contribution begins.

J. A. Alonso (43)
Departamento de Física Teórica, Universidad de Valladolid, Valladolid, Spain

Anna C. Balazs (103)
Department of Chemical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

L. Cate Brinson (253)
Department of Mechanical Engineering, Northwestern University, Evanston, Illinois, USA

Gavin A. Buxton (103)
Department of Chemical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

James R. Chelikowsky (797)
Department of Chemical Engineering and Materials Science, Institute for the Theory of Advanced Materials in Information Technology, Digital Technology Center, University of Minnesota, Minneapolis, Minnesota, USA

Daniel Erni (537)
Laboratory for Electromagnetic Fields and Microwave Electronics, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland

Rosario Fazio (647)
NEST-INFM and Scuola Normale Superiore, Pisa, Italy

Frank T. Fisher (253)
Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, New Jersey, USA

Jihua Gou (361)
Department of Mechanical Engineering, University of South Alabama, Mobile, Alabama, USA

Christian Hafner (537)
Laboratory for Electromagnetic Fields and Microwave Electronics, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland

Henning Heiberg-Andersen (507)
Institute for Energy Technology, Kjeller, Norway

Douglas L. Irving (1)
Department of Materials Science and Engineering, University of Florida, Gainesville, Florida, USA

Kaoru Kobayashi (457)
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Japan

Kin-tak Lau (361)
Department of Mechanical Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

V. Lindberg (771)
Department of Physics, Växjö University, Växjö, Sweden

Marcus Müller (151)
University of Wisconsin–Madison, Madison, Wisconsin, USA and Institut für Physik, WA 331, Johannes Gutenberg Universität, Mainz, Germany

I. Makkonen (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland

R. Maranganti (689)
Department of Mechanical Engineering, University of Houston, Houston, Texas, USA

Shigeru Nagase (457)
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Japan

Androula G. Nassiopoulou (733)
Institute of Microelectronics, NCSR, "Demokritos," Athens, Greece

R. M. Nieminen (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland

E. Ogando (771)
Donostia International Physics Center, Donostia, Spain

M. J. Puska (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland

E. Räsänen (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland

Sergey N. Rashkeev (615)
Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee, USA

Yousef Saad (797)
Department of Computer Science, Institute for the Theory of Advanced Materials in Information Technology, Digital Technology Center, University of Minnesota, Minneapolis, Minnesota, USA

H. Saarikoski (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland

Ulrike Salzner (203)
Department of Chemistry, Bilkent University, Ankara, Turkey

P. Sharma (689)
Department of Mechanical Engineering, University of Houston, Houston, Texas, USA

Susan B. Sinnott (1)
Department of Materials Science and Engineering, University of Florida, Gainesville, Florida, USA

Zdeněk Slanina (457)
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Japan, and Institute of Chemistry, Academia Sinica, Nankang, Taiwan, Republic of China

Jasmin Smajic (537)
Laboratory for Electromagnetic Fields and Microwave Electronics, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland

Karl Sohlberg (615)
Department of Chemistry, Drexel University, Philadelphia, Pennsylvania, USA

M. J. Stott (43)
Department of Physics, Queens University, Kingston, Ontario, Canada

Fabio Taddei (647)
NEST-INFM and Scuola Normale Superiore, Pisa, Italy

T. Torsti (771)
Laboratory of Physics, Helsinki University of Technology, Espoo, Finland and CSC—Scientific Computing Ltd, Espoo, Finland

Liangchi Zhang (395)
School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, New South Wales, Australia

Xanthippi Zianni (733)
Technological Educational Institution of Chalkida, Chalkida, Greece

Handbook of Theoretical and Computational Nanotechnology

Edited by

Michael Rieth and Wolfram Schommers

Volume 1. BASIC CONCEPTS, NANOMACHINES, AND MEDICAL NANODEVICES

Chapter 1. Toward Integrated Nanosystems: Fundamental Issues in Design and Modeling
K. Eric Drexler

Chapter 2. Atomic Nanodesign
Michael Rieth, Wolfram Schommers

Chapter 3. Foundations of Quantum Technology
G. J. Milburn

Chapter 4. Remarks on Basic Physical Concepts
Wolfram Schommers

Chapter 5. Foundation of Computational Nanoelectronics
Felix A. Buot

Chapter 6. Basic Theory of Electron Tunneling and Ballistic Transport in Nanostructures
An-Ban Chen

Chapter 7. Fundamentals of Nano-Thermodynamics
Michael Hartmann, Günter Mahler, Ortwin Hess

Chapter 8. Relativistic Quantum Chemistry: From Quantum Electrodynamics to Quasi-Relativistic Methods
Markus Reiher, Alexander Wolf, Bernd Artur Hess

Chapter 9. Computational Aspects of Virus Structure Determination at High Resolution
Dan C. Marinescu, Yongchang Ji, Vivek Singh, Gabriela M. Marinescu

Chapter 10. Molecular Devices, Nanotechnology, and Surfaces
Bidisa Das, K. L. Sebastian

Chapter 11. Spatio-Temporal Dynamics of Biomolecular Motors: Mesoscopic Theory and Computer Simulation
Edeltraud Gehrig, Ortwin Hess

Chapter 12. Actomyosin Complex in Skeletal Muscle: A Paradigm of a Non-Processive Molecular Motor
Jose Luis Marin, Xochitl Trujillo, Miguel Huerta, Alejandro Elizalde, Jesus Muñiz

Volume 2. ATOMISTIC SIMULATIONS—ALGORITHMS AND METHODS

Chapter 1. Time Stepping Algorithms for Classical Molecular Dynamics
Colin John Cotter, Sebastian Reich

Chapter 2. Meshfree Methods
G. E. Fasshauer

Chapter 3. The Density Matrix Renormalization Method: A Numerical Technique for Low-Dimensional and Nanoscopic Systems
Karen Hallberg

Chapter 4. Bridging Scale Methods for Computational Nanotechnology
Wing Kam Liu, Harold S. Park

Chapter 5. Finite Element Method: From Discrete Atoms to Continuum Solids
B. Liu, H. Jiang, Y. Huang, S. Qu, M.-F. Yu, K. C. Hwang

Chapter 6. Maxwell Solvers for Optics
Christian Hafner, Jasmin Smajic

Chapter 7. Virtual Reality and Haptics in Nano- and Bionanotechnology
Gaurav Sharma, Constantinos Mavroidis, Antoine Ferreira

Chapter 8. Embedded Atom Method: Theory, Development, and Applications
Kunio Takahashi

Chapter 9. Tight-Binding Molecular Dynamics Method for Nanostructure Simulations
C. Z. Wang, K. M. Ho

Chapter 10. Ultra-Large Scale Simulations of Dynamic Materials Failure
Markus J. Buehler, Huajian Gao

Chapter 11. Computational Modeling of Flow and Mass Transport Processes in Nanotechnology
D. Drikakis, M. Kalweit

Chapter 12. Potential-Based Simulation and Molecular Modeling
Patra Volarath, Robert W. Harrison

Chapter 13. Atomistic Modeling of Strain Effects in Heterostructures
Antti Kuronen, Marco Patriarca

Chapter 14. Statistical Mechanical Modeling and Its Application to Nanosystems
Keivan Esfarjani, G. Ali Mansoori

Chapter 15. Linear-Scaling Quantum Mechanical Methods for Nanoscopic Structures
ChiYung Yam, Xiao Zheng, GuanHua Chen

Chapter 16. Simulation of Nanoscale Molecular Systems
Umberto Ravaioli, Trudy A. Van der Straaten

Volume 3. QUANTUM AND MOLECULAR COMPUTING, QUANTUM SIMULATIONS

Chapter 1. Computational Methods for Simulating Quantum Computers
H. De Raedt, K. Michielsen

Chapter 2. Computational Modeling of Donor-Based Quantum Computer Architectures in Silicon
L. C. L. Hollenberg, C. J. Wellard, A. D. Greentree

Chapter 3. Evolving Quantum Programs and Protocols
Susan Stepney, John A. Clark

Chapter 4. Theory of Solid-State Quantum Information Processing
Guido Burkard

Chapter 5. Superconducting Quantum Circuits, Qubits, and Computing
G. Wendin, V. S. Shumeiko

Chapter 6. Quantum Transport and Circuit Theory
Yuli V. Nazarov

Chapter 7. *Ab Initio* Methods for Spin-Transport at the Nanoscale Level
Stefano Sanvito

Chapter 8. Theory and Simulation of Semiconductor Quantum Devices at the Nanoscale
Rita Claudia Iotti, Remo Proietti Zaccaria, Fausto Rossi

Chapter 9. Quantum Waveguide Theory
J. B. Wang

Chapter 10. Quantum Monte Carlo Methods in the Study of Nanostructures
J. Shumway, D. M. Ceperley

Chapter 11. Quantum Monte Carlo: Theory and Application to Atomic, Molecular, and Nano-Systems
Alán Aspuru-Guzik, Alexander C. Kollias, Romelia Salomón-Ferrer, William A. Lester, Jr.

Volume 4. NANOMECHANICS AND MULTISCALE MODELING

Chapter 1. Overview of Multiscale Simulations of Materials
Gang Lu, Efthimios Kaxiras

Chapter 2. Hierarchial Models of Nanomechanics and Micromechanics
Nasr M. Ghoniem, Nicholas Kioussis

Chapter 3. Computational Nanomechanics of Materials
Wing Kam Liu, Sukky Jun, Dong Qian

Chapter 4. Computational Modeling of Tribological, Adhesion, Indentation, and Fracture Processes in Nanoscale Systems
H. Rafii-Tabar

Chapter 5. Equivalent-Continuum Modeling of Nanostructured Materials
Gregory M. Odegard

Chapter 6. Continuum Mechanics for Small Systems and Fine Resolutions
C. Goldenberg, I. Goldhirsch

Chapter 7. Nanoindentation: Recent Development and Applications
Kaiyang Zeng

Chapter 8. Molecule-Based Coarse-Graining for Polymer Simulation
Cameron F. Abrams

Chapter 9. Modeling and Simulation of Nanostructure Formation in Metals and Alloys Subjected to Extensive Plastic Deformation
Alan C. Lund, Christopher A. Schuh

Chapter 10. Modeling and Simulation of Strain-Mediated Nanostructure Formation on Surface
Feng Liu

Volume 5. TRANSPORT PHENOMENA AND NANOSCALE PROCESSES

Chapter 1. Shuttle Transport in Nanostructures
R. I. Shekhter, L. Y. Gorelik, M. Jonson, Y. M. Galperin, V. M. Vinokur

Chapter 2. Dynamics of Condensed Phase Proton and Electron Transfer Processes
Raymond Kapral, Alessandro Sergi

Chapter 3. Computer Simulation of Nanofiltration Membranes and Processes
Horst Chmiel, Xavier Lefebvre, Valko Mavrov, Mohan Noronha, John Palmeri

Chapter 4. Constrained Grain Boundary Diffusion in Thin Copper Films
Markus J. Buehler, T. John Balk, Eduard Arzt, Huajian Gao

Chapter 5. Chemical Reaction and Flow Modeling in Fullerene and Nanotube Production
Carl D. Scott, Samir Farhat, Robert B. Greendyke

Chapter 6. Modeling Gas Adsorption in Amorphous Nanoporous Materials
M. B. Sweatman, N. Quirke

Chapter 7. Sliding Friction at the Atomic Scale
Annalisa Fasolino

Chapter 8. Diffusion in Elastically Strained Solids
Vladimir A. Borodin, Maria G. Ganchenkova

Chapter 9. Adsorption at Nanostructured Surfaces
Axel Groß

Chapter 10. Computational Methods for Atomistic Modeling of Nanoporous Materials and Their Properties
Muhammad Sahimi, Theodore T. Tsotsis

Chapter 11. Modeling of Electrocatalytic Surface Reactions
S. J. Mitchell, M. T. M. Koper

Chapter 12. Radiation-Induced Modifications in Nanomaterials
M. Chipara, D. Hui

Volume 6. BIOINFORMATICS, NANOMEDICINE, AND DRUG DESIGN

Chapter 1. Computational Biology
Dimiter S. Dimitrov, Igor A. Sidorov, Nikola Kasabov

Chapter 2. Computational Studies of Protein Folding
Michiel van Lun, David van der Spoel

Chapter 3. Biomolecular Machines
Artur Baumgaertner

Chapter 4. Knots, Bubbles, Unwinding, and Breathing: Probing the Topology of DNA and Other Biomolecules
Ralf Metzler, Andreas Hanke

Chapter 5. Receptor Flexibility in Ligand Docking
Claudio N. Cavasotto, Andrew J. W. Orry, Ruben A. Abagyan

Chapter 6. Enzyme Reactivity Studied by Computer Simulations
Giorgio Colombo, Massimiliano Meli, Giacomo Carrea

Chapter 7. Simulating Enzyme-Catalyzed Reactions
Anna Bowman, Adrian Mulholland

Chapter 8. Modeling of Biologically Motivated Soft Matter Systems
Ilpo Vattulainen, Mikko Karttunen

Chapter 9. Theoretical and Computational Treatments of DNA and RNA Molecules
Haijun Zhou, Yang Zhang, Zhong-Can Ou-Yang

Chapter 10. Self-Consistent P3M Simulation of Ion Channels
Campbell Millar, Asen Asenov

Chapter 11. Nanomagnetics in Biotechnology
Ching Jen Chen, Yousef Haik, Jhunu Chatterjee

Chapter 12. Computational and Theoretical Approaches to Unraveling the Permeation Dynamics in Biological Nanotubes
Shin-Ho Chung, D. Peter Tieleman

Chapter 13. Progress in Nanomedicine and Medical Nanorobotics
Robert A. Freitas, Jr.

Chapter 14. Molecular Engineering in Nanotechnology: Engineered Drug Delivery
István Majoros, Thommey Thomas, James R. Baker, Jr.

Chapter 15. Design of Protein and Enzyme Mimetics
Garland R. Marshall, Dennis P. Riley

Chapter 16. Quantum Information Processing in Nanostructures
Alexandra Olaya-Castro, Neil F. Johnson

Chapter 17. Theoretical and Computational Models for Neuro, Genetic, and Neuro–Genetic Information Processing
Nikola Kasabov, Lubica Benuskova

Chapter 18. Molecular Computation Using Hairpins and Secondary Structures of DNA
Masami Hagiya

Volume 7. MAGNETIC NANOSTRUCTURES AND NANO-OPTICS

Chapter 1. Atom Nano-optics
V. I. Balykin, V. V. Klimov, V. S. Letokhov

Chapter 2. Theoretical Investigation of Optical Properties of Single-Walled Carbon Nanotubes
Yang Zhao, XiuJun Wang, Chi-Chiu Ma, GuanHua Chen

Chapter 3. Nonlinear Optical Properties of Carbon Nanostructures
Rui-Hua Xie, Tapas Kar, Zhigang Li

Chapter 4. Models for Optical Properties of Clusters and Nanostructures
Julio A. Alonso, Angel Rubio

Chapter 5. Modeling of Photonic Crystals
Wounjhang Park

Chapter 6. Decoherence, Quantum Information, and Quantum-State Measurement in Quantum Optics
Luiz Davidovich

Chapter 7. Optical Properties of Semiconductor Nanostructures: Decoherence versus Quantum Control
Ulrich Hohenester

Chapter 8. Nanometer-Scale Electromagnetic Field Fluctuations
C. Henkel

Chapter 9. Molecular Nanomagnets
Jens Kortus, Andrei V. Postnikov

Chapter 10. Computer Simulation of Magnetic Nanolayer Systems
Willi Schepper

Chapter 11. Computational Micromagnetics
Josef Fidler, Thomas Schrefl, Werner Scholz

Chapter 12. Quantum Theory of Spintronics in Magnetic Nanostructures
J. Mathon, A. Umerski

Volume 8. FUNCTIONAL NANOMATERIALS, NANOPARTICLES, AND POLYMER DESIGN

Chapter 1. Computational Studies of Nanomaterials: A Historical Perspective
Douglas L. Irving, Susan B. Sinnott

Chapter 2. Density Functional Calculations of Clusters and Cluster Assembly
J. A. Alonso, M. J. Stott

Chapter 3. Modeling the Structural Evolution, Equilibrium Morphology, and Macroscopic Behavior of Polymer/Nanoparticle Composites
Anna C. Balazs, Gavin A. Buxton

Chapter 4. Monte Carlo Simulations and Self-Consistent Field Theory for Thin Polymer Films
Marcus Müller

Chapter 5. Conjugated Organic Polymers: From Bulk to Molecular Wire
Ulrike Salzner

Chapter 6. Nanomechanics of Nanoreinforced Polymers
Frank T. Fisher, L. Cate Brinson

Chapter 7. Modeling and Simulation of Carbon Nanotube/Polymer Composites
Jihua Gou, Kin-tak Lau

Chapter 8. Nano-Characterization of Materials: Silicon, Copper, Carbon Nanotubes, and Diamond Thin Films
Liangchi Zhang

Chapter 9. Isomeric Fullerenes and Endofullerenes: Stability Computations on Promising Nanoscience Agents
Zdeněk Slanina, Kaoru Kobayashi, Shigeru Nagase

Chapter 10. Carbon Nanocones
Henning Heiberg-Andersen

Chapter 11. Simulation and Optimization of Composite Doped Metamaterials
Christian Hafner, Jasmin Smajic, Daniel Erni

Chapter 12. Theoretical and Computational Atomic-Scale Studies of Complex Catalytic Materials
Karl Sohlberg, Sergey N. Rashkeev

Chapter 13. Properties of Superconducting Nanostructures
Rosario Fazio, Fabio Taddei

Chapter 14. Strain Field Calculations in Embedded Quantum Dots and Wires
R. Maranganti, P. Sharma

Chapter 15. Optical Properties of Silicon Quantum Wires and Dots
Xanthippi Zianni, Androula G. Nassiopoulou

Chapter 16. Real-Space Electronic-Property Calculations for Nanoscale Structures
T. Torsti, V. Lindberg, I. Makkonen, E. Ogando, E. Räsänen, H. Saarikoski, M. J. Puska, R. M. Nieminen

Chapter 17. Electronic Structure of Clusters and Nanocrystals
James R. Chelikowsky, Yousef Saad

Volume 9. NANOCOMPOSITES, NANO-ASSEMBLIES, AND NANOSURFACES

Chapter 1. Self-Organizing Nanophases: Model and Simulation
Wei Lu

Chapter 2. Computer Simulation of Surfaces
Walter Langel

Chapter 3. Molecular Organization of Gases and Liquids at Solid Surfaces
Ivan Brovchenko, Alla Oleinikova

Chapter 4. Thermodynamics of Surfaces and Adsorption
J. P. Hajra, S. Acharya

Chapter 5. Evolution of Surface-Based Nanostructures: Formation and Decay
E. G. Wang

Chapter 6. Computational Methods for the Study of Thin Layers on Semiconductor Surfaces
Laura Nurminen, Kimmo Kaski

Chapter 7. Computer Simulation of Energetic Cluster Surface Interactions
Roger Webb

Chapter 8. Molecular Dynamics Simulations of the Mechanical Properties of Polyethylene-Carbon Nanotube Composites
Michael Griebel, Jan Hamaekers

Chapter 9. Mechanics of Nanocomposite Structures of Biological Materials
Baohua Ji, Huajian Gao

Chapter 10. Electronic Properties and Reactivity of the Doped and Defected Single-Walled Carbon Nanotubes
Wei Quan Tian, Lei Vincent Liu, Yan Alexander Wang

Chapter 11. Phase Field Theory of Nucleation and Polycrystalline Pattern Formation
László Gránásy, Tamás Pusztai, Tamás Börzsönyi

Chapter 12. Modeling of Stability and Phase Transformations in Zero- and One-Dimensional Nanocarbon Systems
A. S. Barnard, S. P. Russo, I. K. Snook

Chapter 13. First-Principles Modeling of Ferroelectric Oxide Nanostructures
Philippe Ghosez, Javier Junquera

Chapter 14. Shadowing Growth and Physical Self-Assembly of 3D Columnar Structures
Tansel Karabacak, Toh-Ming Lu

Volume 10. NANODEVICE MODELING AND NANOELECTRONICS

Chapter 1. Computational Nanoelectronics
Dragica Vasileska, David K. Ferry, Stephen M. Goodnick

Chapter 2. Process Simulation for Silicon Nanoelectronic Devices
Wolfgang Windl

Chapter 3. Electron Transport in Nanostructured Systems—*Ab Initio* Study
Yoshiyuki Kawazoe, Hiroshi Mizuseki, Rodion Belosludov, and Amir Farajian

Chapter 4. Single-Electron Functional Devices and Circuits
Takashi Morie, Yoshihito Amemiya

Chapter 5. Modeling of Single-Electron Transistors for Efficient Circuit Simulation and Design
YunSeop Yu, SungWoo Hwang, Doyeol Ahn

Chapter 6. Electric Properties of Nanostructures
K. Palotás, B. Lazarovits, P. Weinberger, L. Szunyogh

Chapter 7. Transport Theory for Interacting Electrons Connected to Reservoirs
Akira Oguri

Chapter 8. Computational Nanotechnology: Computational Design and Analysis of Nanosize Electronic Components and Circuits
Jerry A. Darsey, Dan A. Buzatu

Chapter 9. Tunneling Models for Semiconductor Device Simulation
Andreas Gehring, Siegfried Selberherr

Chapter 10. Electronic Structure of Quantum Dots
J. B. Wang, C. Hines, R. D. Muhandiramge

Chapter 11. Spatiotemporal Dynamics of Quantum-Dot Lasers
Edeltraud Gehrig, Ortwin Hess

Chapter 12. Theoretical Investigations of Silicon Quantum Dots
Lin-Wang Wang

Chapter 13. Nanoscale Device Modeling
Massimo Macucci, Luca Bonci

Chapter 14. Wigner Function Based Device Modeling
Hans Kosina, Mihail Nedjalkov

Chapter 15. Logic Design of Nanodevices
Svetlana N. Yanushkevich

Chapter 16. Nanoelectromechanical Systems and Modeling
Changhong Ke, Horacio D. Espinosa

CHAPTER 1

Computational Studies of Nanomaterials: A Historical Perspective

Douglas L. Irving, Susan B. Sinnott
Department of Materials Science and Engineering, University of Florida, Gainesville, Florida, USA

CONTENTS

1. Introduction 1
2. Introduction to Computational Methods 2
   2.1. Calculating Interaction Energies and Forces 2
   2.2. Structure Optimization and Simulation 6
3. Application to Nanometer-Scale Systems 10
   3.1. Electronic Properties of Single-Walled Carbon Nanotubes 10
   3.2. Structure and Properties of Nanotube Junctions 14
   3.3. Mechanical Properties of Carbon Nanotubes 19
   3.4. Chemical Modification of Carbon Nanotubes 22
   3.5. Carbon Nanotube Gas Sensors 26
   3.6. Filling Carbon Nanotubes 29
   3.7. Nanoindentation 34
4. Summary and Future Outlooks 36
   References 37

1. INTRODUCTION

Nanomaterials have been in use for centuries (see, e.g., Ref. [1]), and have been the subject of much attention from scientists and engineers over about the last 20 years. This interest has accelerated recently as research funding agencies across the world have allocated significant funds for research in nanoscience and nanotechnology.

The last two decades have seen impressive advances in the experimental tools and techniques that allow for significant manipulation and analysis of materials at the nanometer scale. Great advances in the application of computational atomistic and molecular methods

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (1–42)

to the study of nanomaterials have occurred over this same time period. This growth has been facilitated by the development and optimization of sophisticated scientific software, substantial increases in available computing power, and a large drop in the price of computing cycles. Thus, it is increasingly common for computational methods and experimental techniques to be applied simultaneously to provide a complete picture and understanding of nanometer-scale systems. However, the area of nanomaterials is somewhat unique in that computational studies have historically had a leading or otherwise indispensable role in some crucial discoveries on which many current efforts build.

In this chapter, we review some historical seminal papers in which computational methods were used to make important predictions about, or provide crucial insight into, the behavior and properties of nanomaterials. We also review related experimental papers and discuss the relationships between the experimental and computational findings.

2. INTRODUCTION TO COMPUTATIONAL METHODS

A wide range of computational methods exists for the study of nanomaterials, from atomic-scale methods to methods that model the properties of parts created from these unique materials. In this chapter, we focus on atomic-scale methods, which range from quantum mechanical (QM) approaches with few approximations to empirical methods that include many estimates. In general, the accuracy of each technique is tied to the approximation or approximations made in its derivation. Therefore, in addition to introducing and reviewing the most commonly used computational methods to model nanomaterials, we discuss their accuracy and the conditions under which they are best applied.

2.1. Calculating Interaction Energies and Forces

Atomic-scale methods can be broken into several categories based on the manner in which they are formulated. The most theoretically rigorous methods are those that are classified as *ab initio*, or first principles. These approaches are derived from QM principles and are generally both the most accurate and the most computationally intensive. They are consequently limited to a small number of atoms (about 50–100). In contrast, empirical methods are functions containing parameters that are derived by fitting to experimental data or the results of *ab initio* calculations. These techniques can usually be relied on to correctly describe qualitative trends and are often the only choice available for modeling systems containing tens of thousands to millions of atoms. Last, semiempirical methods include elements of both empirical methods and *ab initio* methods and require both QM information in the form of, for example, atomic orbital basis sets and some fitting to experimental data.

2.1.1. Ab Initio Methods

There are two main classes of *ab initio* methods: QM wave function based methods and density functional theory (DFT), which depends on electron densities. The derivation of both these methods relies on the Born–Oppenheimer approximation [2], which states that, because of the small mass of electrons in comparison to the mass of the atomic nucleus, it is reasonable to assume that the electrons will instantaneously respond to any change of the nuclear position. This allows for the separation of the total energy into nuclear components and electronic components. In other words, the total energy of the system can be written as

$$E_{\text{total}} = E_{n-n} + E_e \tag{1}$$

where E_{total} is the total energy, E_{n-n} is the contribution to the energy from nuclear–nuclear repulsion, and E_e is the electronic contribution to the energy. The nuclear–nuclear repulsion energy is then calculated using classical methods as follows

$$E_{n-n} = \sum_{\alpha}^{\prime} \sum_{\beta} \frac{Z_\alpha Z_\beta}{|\vec{R}_\alpha - \vec{R}_\beta|} \tag{2}$$

where Z_α and R_α are the atomic number and position in space, respectively, of nucleus α. The electronic contribution to the total energy E_e, referred to only as E from now on, is then solved for by the QM or DFT methods.

QM is one of the most accurate theories of modern physics. One of the tenets of QM is the Schrödinger wave equation [3], which has been shown to be very powerful in describing the behavior of electrons [4, 5]. The time-independent form of the Schrödinger equation is

$$H\Psi = E\Psi \tag{3}$$

where H is the Hamiltonian, which contains the kinetic energy and potential energy of the system; Ψ is the wavefunction of the electron; and E is the energy of the electron. The Hamiltonian acts on the wavefunction, and the energy E is calculated.

Although this equation is both rigorously derived and very accurate, its analytic solutions are limited to a small number of simple cases, which include the H atom, the simple harmonic oscillator, and the square-well potentials [4, 5]. It is, therefore, necessary to find another, nonanalytic approach to solve for the wavefunction and energy of more complicated systems. Fortunately, the variational theorem [4] provides a conceptually simple means of finding better trial wavefunctions for more complicated Hamiltonians.

The variational theorem states that given a known Hamiltonian, any normalizable trial wavefunction, Ψ_T, can be used to approximate the true wavefunction of the system. The energy calculated with Ψ_T will always be greater than or equal to the true ground-state energy, E_0. The equality in the theorem is only satisfied when Ψ_T is the true ground-state wavefunction, Ψ_0. In other words, the only time a trial wavefunction generates the ground-state energy is when the trial wavefunction is truly the ground-state wavefunction. This is an extremely powerful theorem, and it lends itself very well to iterative computational techniques. In particular, initial wavefunctions are constructed from orbital or plane wave basis sets, and a variation of Eq. (3) is solved by numerical iteration until Ψ and E cease to change within some predetermined value. This iterative approach is termed "self-consistency" and is a viable approach for finding an approximate electronic wavefunction of complicated multielectron systems.

Different types of QM approaches exist, and these will be discussed briefly here. For a more in-depth discussion, the reader is referred to Refs. [6, 7]. In the Hartree–Fock (HF) approximation, the initial, trial wavefunction of many interacting electrons is simplified to a Slater determinant of independent, one-electron wavefunctions. With the application of the linear combination of atomic orbitals (LCAO) to express each molecular orbital (MO), Roothaan [8] showed that the HF problem could be expressed as a secular matrix equation. Finding the ground-state wavefunction now becomes a problem of varying the expansion coefficients of the LCAOs until the minimum energy is reached. The Hamiltonian is divided into multiple components that are assumed to be independent: a one-electron part that contains the kinetic energy and the electron–nuclear interactions, and a part that contains the electron–electron interactions. Thus, Eq. (3) can be rewritten as a series of one-electron equations that provide the energies of these individual electrons. The sum of these individual electron energies is then the electronic energy E of the system. These approximations, combined with the use of a Lagrange multiplier to guarantee that the one-electron wavefunctions remain normalized with respect to each other, constitute the HF approach.

In practice, the LCAOs used to expand the MOs are usually Slater-type functions or Gaussian-type functions, where Slater and Gaussian refer to the type of functions used to expand the molecular orbital and constitute the basis set of the problem. Each basis function has its advantages and disadvantages, and the reader is referred to Ref. [7] for a more thorough discussion. As noted above, once the basis set is applied to the system, the HF equations are self consistently solved until the lowest energy is found. The numerical method that is used to solve the one-electron wavefunctions in terms of the above approximations is often referred to as the LCAO or MO approach.

A known limitation of the HF approach is that it often slightly underestimates bond energies, and better bond energies can often be obtained by making additional improvements. For instance, the HF approach does not contain a term for the correlation energy of

electrons, but there are several ways to incorporate this contribution into the wavefunction approach. One common approach is configuration interaction (CI), which uses ground- and excited-state Slater determinants to better approximate the true ground-state wavefunction.

Another common approach is the application of perturbation theory to the problem. In particular, the difference between the HF operators and the operators in the true Hamiltonian is treated as a perturbation, and equations of perturbation theory are applied to correct for the differences between the two terms. This approach is termed Hartree–Fock Moller–Plesset (HF-MP) theory [6]. The HF-MP approach improves the ground-state energies and bond lengths. However, if the difference between the HF ground state and the true ground state is large, then the application of MP theory can lead to numerical divergence. The CI approach is computationally expensive and is most often used in describing excited states.

Most QM implementations do not allow for periodicity in the system under consideration [6], although relatively recent developments have yielded an implementation that can model periodic systems [9–12]. When periodicity cannot be included in the calculations, bulk systems are treated as clusters, with hydrogen atoms terminating any dangling bonds.

DFT solves for the ground-state electronic energy in a different fashion. Instead of solving for all ground state properties in terms of the ground-state wavefunction, DFT describes the ground state in terms of the electron density [13]. This is done by iteratively solving a different equation, called the Kohn-Sham equation [14], which can be shown to be related to the Schrödinger equation [15] and is written as follows:

$$E[\rho] = T_s[\rho] + V_{ee}[\rho] + V_{en}[\rho] + V_{xc}[\rho] \tag{4}$$

where $E[\rho]$, the total electronic energy, and the remaining terms are functionals (functions of functions) of the electron density, ρ. Specifically, T_s is the contribution to the energy from the kinetic energy of the electrons (assuming they are noninteracting), V_{ee} is the contribution from classical electron–electron repulsion, V_{en} is the contribution from electron–nuclear attraction, and V_{xc} is the contribution from electron exchange and correlation. The electron density ρ is related to the electron wavefunction in Eq. (3) through the following relation

$$\rho = |\Psi|^2 \tag{5}$$

DFT can be either all-electron, where it explicitly characterizes all the electrons in a system, or it can explicitly model only the valence electrons and use fitted functions, called pseudopotentials [16–21], to model the core electrons. This is a valid approximation because most chemical properties of solids are determined by the valence electrons. Pseudopotential DFT calculations are commonly used for larger systems, defined in this case as consisting of on the order of 100 atoms, because of the computational efficiency gained in not having to treat core electrons self consistently. However, although the accuracy of pseudopotential DFT calculations is generally good, it does depend on the specific pseudopotential implementation being used [6]. Furthermore, it is more difficult to deal with the properties of excited states with DFT because DFT is a ground-state theory. However, there have been several successful developments in trying to treat the excited states with time-dependent DFT (TDDFT) [6, 7].

Most flavors of DFT codes are implemented for periodic systems. Therefore, free surfaces are modeled by placing either layers of vacuum (empty space) above and below the surface slab that is replicated in three dimensions, or some implementations only periodically repeat in the two dimensions of the surface plane and have a pseudoinfinite vacuum spacing.

Within DFT and the Kohn–Sham equations, all the terms needed to calculate the energy are known exactly except for the exchange-correlation energy V_{xc}, a term that characterizes the nonclassical interactions among the electrons. Two common approximations of this energy are the local density approximation (LDA) and the generalized gradient approximation (GGA). The LDA approximates V_{xc} in the real system via V_{xc} in finite volumes of uniform electron density, where it can be calculated exactly. The GGA builds on this approximation by including information about the gradients between the volumes. The LDA generally overestimates bond energies, whereas the GGA predicts more accurate values [6].

In DFT, errors can arise as a result of electron self-exchange and self-correlation related to the approximations used to calculate exchange and correlation energies [15, 22].

Both HF and DFT approaches are self-consistent, which means that either the Schrödinger equation or the Kohn–Sham equation is solved in an iterative manner until the energies and either the wavefunctions or the electron densities stop changing. The reader is referred to Ref. [23] for an in-depth discussion of these methods and to Ref. [22] for some direct comparisons of the results of HF- and DFT-based methods.

2.1.2. Semiempirical Methods

Several methods fall into the category of semiempirical, defined as containing some QM parts and some empirically fit parts. The most important of these parts for the purposes of this article is the tight-binding (TB) method. In general, the total energy for a system of N atoms at positions R can be written as [6]

$$E = 2\Sigma\eta_\lambda\varepsilon_\lambda + E_{\text{repulsive}}(R_1, \ldots R_N) \tag{6}$$

where the first term on the right-hand side is the band term, which is generally attractive and derived from QM methods, and the second term on the right-hand side is the repulsive term, which is empirically fit to *ab initio* or experimental data. In the first term, the sum is over all occupation numbers λ. The terms η_λ and ε_λ represent the occupation number and the one-electron energy, respectively; the factor of two accounts for spin degeneracy.

The TB method thus includes information about the atomic orbitals and whether they are filled or not. As a consequence, TB methods can be used to calculate such quantities as the electron density of states. A strength of this approach is that it is fast and able to characterize orders of magnitude more atoms than true *ab initio* methods (several 1000s vs. about 100) with some quantum mechanical character. A disadvantage is that the fitting limits the transferability of the approach. In addition, the approach shown in Eq. (6) is not able to model charge and charge-transfer.

To include information about charge, a self-consistent form of the TB method is needed. In this form, Eq. (6) is modified by including an empirically fit term on the right-hand side that is dependent on the change in the electron occupation [6]. The total energy is then minimized with respect to the electron occupations at each step.

2.1.3. Empirical Methods

Empirical methods simplify the modeling of materials by treating the atoms as spheres that interact with each other via repulsive and attractive terms that can be either pairwise or many-body in nature. Thus, the electrons are not treated explicitly, although it is understood that the interactions are ultimately dependent on them. As discussed below, some empirical methods explicitly include charge through classical electrostatic interactions. However, most methods assume charge-neutral systems. The repulsive and attractive functional forms generally depend on interatomic distances or angles and contain adjustable parameters that are fit to *ab initio* results or experimental data.

The main strength of empirical methods is their computational speed. Recent simulations with these approaches have modeled several million atoms [24, 25], something that is not possible with *ab initio* or semiempirical approaches at this time. Their main weakness is their lack of quantitative accuracy, especially if they are poorly formulated or applied to systems that are too far removed from the fitting database used in their construction. In addition, because of the differences in the nature of chemical bonding in various materials (e.g., covalent bounding in carbon versus metallic bonding in nickel), empirical methods have been historically derived for particular classes of materials. Thus, they are generally nontransferable, although some methods have been shown to be theoretically equivalent [26], and there has been progress toward the development of empirical methods that can model heterogeneous material systems [27–32].

The first method to be discussed is the bond-order method to model covalently bound materials, which was first formulated by Abell [33] and developed by Tersoff for Si and

Ge [34, 35]. It was subsequently parameterized by Brenner and coworkers for hydrocarbons [36–38], Dyson and Smith for C—Si—H [39], Sinnott and coworkers for C—O—H [40], and Graves and coworkers for fluorocarbons [41].

The bond-order potential has the following functional form

$$E = \sum_i \sum_{j(>i)} [V_R(r_{ij}) - b_{ij} V_A(r_{ij})] \tag{7}$$

where $V_R(r)$ and $V_A(r)$ are pair-additive interactions that model the interatomic repulsion and electron–nuclear attraction, respectively. The quantity r_{ij} is the distance between pairs of nearest-neighbor atoms i and j, and b_{ij} is a bond-order term that takes into account the many-body interactions between atoms i and j, including those resulting from nearest neighbors and angle effects. Overall, the potential is short-ranged and only considers nearest neighbor bonds. To model long-range, van der Waals interactions, the bond-order potential is combined with pairwise potentials [42, 43].

The second technique to be discussed is the embedded atom method (EAM) approach [44, 45] and related methods [46], which were developed for modeling metals and metal alloys. The functional form is as follows

$$E = \sum_i F(\rho_i) + \sum_{i>j} \Phi(r_{ij}) \tag{8}$$

where F is called the embedding energy and is the energy required to embed an atom into a uniform electron gas with a uniform compensating positive background (jellium) of density ρ_i that is equal to the actual electron density of the system. The term $\Phi(r_{ij})$ is a pairwise functional form that corrects for the fact that actual materials have point-charge nuclei. Several parameterizations of the EAM exist (see, e.g., Refs. [44, 45, 47–49]), and it has recently been extended to model nonmetallic systems. For example, the modified EAM (MEAM) approach [27, 28] was developed so that EAM could be applied to metal oxides [29].

The third method to be discussed is the general class of Coulomb, or multipole interaction, potentials used to model charged ionic materials or molecules [6]. In this formalism, the energy can be written in the following general formalism:

$$E = \sum_i \sum_{j(>i)} [(q(r_i)q(r_j)/r_{ij})] \tag{9}$$

where $q(r_i)$ is the charge on atom i and r_{ij} is the distance between atoms i and j. More complex formalisms that take into account, for example, the Madelung constant in the case of ionic crystals are used in practice. In general, the charges are held fixed, but methods to allow charge to vary in a realistic manner have been developed in combination with other methods [30, 50].

Last, long-range van der Waals or related forces are typically modeled with pairwise additive potentials. A widely used method is the Lennard-Jones (LJ) potential [51], which has the following functional form

$$E = 4\varepsilon \sum_i \sum_{j(>i)} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{10}$$

Here ε and σ are parameters and r_{ij} is the distance between atoms i and j.

2.2. Structure Optimization and Simulation

The proceeding section discussed the various ways in which the total energy of a system can be calculated. Being able to calculate the energy of various systems is quite important, as it allows for the determination of their relative stabilities. However, knowledge of the forces allows for the optimization or relaxation of a given structure. The forces acting on individual atoms can be analytically or numerically calculated by taking the negative of the first derivative of the energy with respect to position.

The rest of this section will discuss two important ways in which the combination of information about the energies and forces in a given system can be used to optimize material structures.

2.2.1. Force Minimization

Force minimization is an approach by which the forces acting on individual atoms in a system are calculated and the atom positions are adjusted to minimize these forces. When the forces reach a minimum value, the system structure is considered to be optimized and the procedure stops. This is, therefore, a static approach that is focused on obtaining the optimal, lowest-energy structure. A complicating factor is that most potential energy surfaces contain multiple local minima in addition to the lowest-energy global minimum. An additional complication is that for systems that consist of a large number of atoms, the potential energy surface can consist of large, flat regions in which relatively large changes in the structure and configuration of the system do not result in a large change in the energy or forces. To be assured that the global minimum has been reached, it is important to calculate the atomic frequencies (if none are imaginary, than one is assured of having reached a global minimum versus a saddle point). A viable, though tedious, strategy for finding the global minimum is to restart the optimization from a variety of different starting configurations and see whether they transform into the same optimized structure.

Force minimization methods have been used with all of the methods discussed in Section 2.1, but they are most important for *ab initio* or first principles approaches. Many different types of force minimization methods exist. The most important of these are the method of steepest descent, the method of conjugate gradients, and simulated annealing.

The steepest descent method is a way of finding the local minimum value of a function by calculating its gradient. In particular, one starts at point $\mathbf{P}_0$, at position $\mathbf{r}_0$, and moves from $\mathbf{P}_i$ to $\mathbf{P}_{i+1}$ by moving downhill from point $\mathbf{P}_i$ along the descending gradient—$\nabla f(\mathbf{P}_i)$ [52]. As a practical matter, the method of steepest descent is implemented by iterating the following equation (shown for a one-dimensional system) [53]

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \varepsilon f'(\mathbf{x}_{i-1}) \tag{11}$$

where $\mathbf{x}_0$ is the initial starting position and ε is a constant that is much less than zero.

This approach is thus very good at finding the local minimum value of the function under consideration that is nearest the starting point $\mathbf{P}_0$. However, a big disadvantage of this approach is that it does not necessarily find the global minimum (it will always move only downhill). In addition, it will take a large number of small steps to "walk" downhill, even if the minimum well is deep and narrow. Thus, it is not very efficient.

The method of conjugate gradients is another way of finding the local minimum value of a function by calculating its gradient, which improves on the method of steepest descent by using the conjugate gradients of a function rather than just the local gradient to locate the minimum. In the implementation for a linear, one-dimensional function, $f(x)$, the positions are evaluated using the following iterative approach [53]

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \alpha_i \mathbf{p}_i \tag{12}$$

where $\mathbf{p}_i$ is the search direction vector and α_i is a multiple of this vector. The residuals, $\mathbf{r}_i$, which are defined as

$$\mathbf{r}_i = b - A\mathbf{x}_i \tag{13}$$

are also solved iteratively using the following expression:

$$\mathbf{r}_i = \mathbf{r}_{i-1} - \alpha_i \mathbf{q}_i \tag{14}$$

where $\mathbf{q}_i$ is related to $\mathbf{p}_i$ as follows:

$$\mathbf{q}_i = A\mathbf{p}_i \tag{15}$$

The multiplier, α_i, is chosen to minimize $\mathbf{r}_i^T A^{-1}\mathbf{r}_i$ using the following expression:

$$\alpha_i = \frac{(\mathbf{r}_{i-1}^T \mathbf{r}_{i-1})}{(\mathbf{p}_i^T A \mathbf{p}_i)} \tag{16}$$

Last, the search directions $\mathbf{p}_i$ must be orthogonal to each other to ensure that the gradients are conjugates. This is done using

$$\mathbf{p}_i = \mathbf{r}_i + \beta_{i-1}\mathbf{p}_{i-1} \tag{17}$$

where

$$\beta_i = \frac{(\mathbf{r}_i^T \mathbf{r}_i)}{(\mathbf{r}_{i-1}^T \mathbf{r}_{i-1})} \tag{18}$$

The conjugate gradient is more computationally efficient than the method of steepest descent. However, it suffers from the same drawback of not being able to guarantee location of the global minimum of the function under consideration.

Simulated annealing is a fundamentally different approach to finding the minimum of a function. In this approach, a random search is used that can accept changes that decrease or increase the value of the function f. All downhill movements are accepted outright, and uphill movements are accepted with a probability proportional to $\exp(-\Delta f/T)$, where Δf is the increase in the function and T is a control parameter that is analogous to the system temperature. The value of T is reduced slowly as the simulated annealing proceeds so that, eventually, the bottom of the minimum well of the function is reached. Thus, simulated annealing is analogous to the way in which a crystalline material cools from the liquid state at high temperature into a global minimum energy crystal structure at low temperatures. It is well known that this occurs only if the system is cooled slowly. Rapid quenching of liquids traps local minimum structures that may not correspond to the global minimum energy crystal structure. The same is true of the simulated annealing algorithm—the decrease in the control parameter T must be sufficiently slow to achieve the true global minimum of the function. A major advantage of the simulated annealing method over the methods of steepest descent and conjugate gradients is that it generally finds the global minimum and avoids being trapped in local minima.

2.2.2. Molecular Dynamics Simulations

In classical mechanics, Newton's second law states that to make a body of mass m undergo an acceleration $\mathbf{a}$, it is necessary to apply a force $\mathbf{F}$ to the object. The acceleration will then be proportional to the applied force according to the following equality

$$\mathbf{F} = m\mathbf{a} \tag{19}$$

Newton's second law can also be expressed in terms of the position vector $\mathbf{r}$ of the body as

$$\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2} \tag{20}$$

This is the basis of MD. Knowing the force $\mathbf{F}$, based on Eq. (20), we can thus study the trajectory of each particle in space and investigate the time-dependent properties. The problem is how to accurately and efficently calculate the force between atoms, which can be quite complex. From the principle of conservation of energy, we know that the kinetic energy ($1/2mv^2$) and the potential energy (U) of the body can vary, but their sum (E) is a constant.

$$\frac{1}{2}mv^2 + U = E \tag{21}$$

In terms of $\mathbf{r}$, Eq. (21) can be expressed as

$$\frac{1}{2}m\left(\frac{d\mathbf{r}}{dt}\right)^2 + U = E \tag{22}$$

Differentiating both sides of Eq. (22) with respect to time, we find

$$\frac{d}{dt}\left[\frac{1}{2}m\left(\frac{d\mathbf{r}}{dt}\right)^2 + U\right] = 0 \tag{23}$$

and so

$$m\frac{d\mathbf{r}}{dt}\cdot\frac{d^2\mathbf{r}}{dt^2} + \frac{dU}{dt} = 0 \tag{24}$$

This can be rewritten as follows because potential energy is a function of the position $U(\mathbf{r})$:

$$m\frac{d\mathbf{r}}{dt}\cdot\frac{d^2\mathbf{r}}{dt^2} + \nabla U\cdot\frac{d\mathbf{r}}{dt} = 0 \tag{25}$$

Therefore, we see

$$m\frac{d^2\mathbf{r}}{dt^2} = -\nabla U \tag{26}$$

Using Newton's second law and conservation of energy, it is finally found that the force can be calculated from the potential energy

$$\mathbf{F} = -\nabla U \tag{27}$$

The potential energy can be obtained using any of the methods discussed in Section 2.1. In MD simulations, the calculation of the potential energy and force are the most time-consuming aspects. Once the force is obtained, Newton's equation of motion [Eq. (20)] can be integrated to follow the time evolution of the atoms in response to the applied forces. For this reason, empirical methods have been most commonly used in MD simulations. TB approaches can be used in a straightforward manner on smaller systems. For many years, QM and DFT approaches were too prohibitively slow to be used in MD simulations. However, in 1985 Car and Parrinello [54] developed an approximation to allow DFT-MD simulations to be carried out in an approach that bears their name. This method uses a Lagrangian and Lagrange multipliers to efficiently find the ground-state electron density before the time evolution of the atomic system. This method was truly groundbreaking and allowed for an efficient means of performing DFT-MD. Detailed descriptions of the Carr–Parinello technique are provided in Refs. [54, 55].

In practice, numerical integration of Eq. (20), as opposed to analyitic solutions, is necessary to solve Newton's second law. There are several numerical methods for integrating Newton's equations, including the Verlet algorithm, the leapfrog algorithm, and the predictor–corrector algorithm [51].

In MD simulations, short time-steps, Δt, are required to yield reliable results. There are at least two reasons for this. One is because of the quick motion of the atoms (e.g., the timescale of atomic vibrations is typically 10^{-13} s [56]). To capture atomic motions accurately, as MD simulation desires to do, the time-step must be much smaller than the frequency of the atomic motions. The second reason is that, from the integration point of view, to achieve the predictions calculated in Eq. (19–27) as accurately as possible, a small Δt is necessary. Usually, time-steps on the order of a femtosecond (10^{-15}s) are used. Therefore, modeling of processes that occur on timescales larger than a few nanoseconds is not within the reach of conventional MD simulations on present-day computers.

MD simulations can generate information at the microscopic level, including atomic positions and velocities. Via statistical mechanics, this microscopic information can be related to macroscopic quantities such as pressure, temperature, heat capacities, and so forth. Therefore, MD simulations can be used to study both thermodynamic properties and time-dependent (kinetic) phenomena. Another advantage of MD simulations is the inherent ability to follow atomic scale mechanisms in real time. An important disadvantage is that only short timescales of up to tens of nanoseconds can be accessed. There has therefore been considerable effort spent on finding ways to accelerate MD simulations [57–60].

2.2.3. Monte Carlo Simulations

In this approach, the forces are not calculated. Instead, the energy of the starting system configuration is determined using one of the approaches discussed in Section 2.1. The position of one or more atoms is then randomly varied according to a set of predetermined conditions. If the energy of the new configuration is lower than the energy of the initial configuration, the new configuration is accepted and the process is repeated. If the energy of the new configuration, E_{new}, is higher than the energy of the initial configuration, $E_{initial}$, the new configuration is accepted if

$$e^{-(E_{new}-E_{initial})/kT} > \text{random\#} \tag{28}$$

where k is the Boltzmann constant, T is temperature, and random# is a random number generated numerically. The rationale behind accepting higher energy configurations is to move the system out of locally stable configurations into a global minimum. It is clear from Eq. (28) that at higher temperatures, more configurations that have higher energies are accepted than is the case at lower temperatures, and the system has a better chance of settling into a global minimum. Monte Carlo simulations are thus similar to the method of simulating annealing discussed in Section 2.2.1.

The advantage of Monte Carlo simulations is the ability to rapidly transform a starting configuration into a more stable, global minimum structure. The disadvantage is that no time information is contained in the simulation results. There has therefore been some work to combine MD and Monte Carlo simulation methods (see, e.g., Ref. [61]) to use the strengths of each.

3. APPLICATION TO NANOMETER-SCALE SYSTEMS

The ability of scientists to realistically probe the nanoscale world through computational experiments has been facilitated by faster computational hardware and more efficient scientific algorithms. This has led to many computational experiments that are on the cutting edge of modern science. Often, new predictions of physical properties have been made by performing computational experiments only. The added realism of new computational techniques has also played a role in understanding and interpreting experimental data, leading to the identification of new phenomena.

The following sections give examples of a variety of computational experiments that either make new predictions or assist in the understanding of new experimental findings. Some of the examples have been validated by experimental findings, while in other cases experimental evidence is not yet available. The goal is to provide a historical perspective on the significant contributions that computational modeling of materials has made to the field of nanomaterials.

3.1. Electronic Properties of Single-Walled Carbon Nanotubes

Modern technology has almost universally sought to make devices ever smaller. One example is radios, which were initially dependent on vacuum tubes to function and were consequently quite large. When Shockley, Bardeen, and Brattain invented the solid-state transistor, for which they were awarded the Nobel Prize in physics in 1956, portability was introduced, broadening the use of radios in modern life. At present, major computer chip makers are seeking to shrink the size of solid-state transistors to create faster and more functional computer processors.

When carbon nanotubes (CNTs) were discovered in 1991 by Iijima [62], there was almost an instantaneous interest in the electronic properties of these nanometer-scale materials. However, the measurement of the electronic properties of these novel structures was very difficult at the time of discovery. This was because the generation of substantial amounts of CNTs had not yet been mastered, and it was not possible to generate a clean sample of homogeneous CNTs. This left the exploration of the electronic properties of the CNT to computational methods. Mintmire et al. [63] were the first to model the electronic properties

of single-walled CNTs, followed closely by computational studies by Saito et al. [64] and Hamada et al. [65]. These studies and the subsequent experimental confirmation of their results will be the focus of this section.

Mintmire et al. [63] were faced with the challenge of modeling a new material even before it had been experimentally verified as existing [66]. Because they were studying the system computationally, they did not have the same issues of purity and structure control that concerned the experimentalists. However, they did have to determine what a reasonable geometric structure for CNTs should be. Their starting point was the C_{60} fullerene, also known as a buckministerfullerene or buckyball, which is a spherical molecule made up of hexagons and pentagons of carbon [67]. As shown in Fig. 1a, the atomic arrangement of the atoms in the buckyball is reminiscent of a soccer ball. The C_{70} molecule, shown in Fig. 1b, is formed by aligning the C_{60} fullerene to one of its C_5 axes, which has fivefold symmetry, and adding a concentric ring of 10 carbon atoms with D_{5h} symmetry. An extended tubular structure can be constructed by adding a series of these 10 atom carbon rings, as shown in Fig. 1c. Each newly added carbon ring that is added to the structure is rotated from the position of the previously added ring by half the distance necessary to rotate the structure to identical lattice positions. Adding ring after ring of carbon transforms the spherical C_{60} into an extended capsule of variable length. The original C_{60} molecules can be thought of as caps of the tubular center section. Most carbon nanotubes are produced with capped ends [67, 68].

The calculations of Mintmire et al. [63] were performed on an infinitely long cylindrical tube with all C–C distances set to 0.142 nm and a CNT diameter of about 0.7 nm. The base unit of the structure was a set of 10 carbon rings of D_{5h} symmetry, and an extended structure was generated in the code by use of a screw operator. This corresponds to a (5, 5) nanotube structure, where the indices (n, m) refer to the manner in which the "rolling up" of a graphene sheet produces nanotubes [67]. Changes to the indices produce nanotubes of various helical structures, as shown in Fig. 2. Specifically, the integer lattice vectors **n** and **m** are added together, and the tail and head of the resulting vector are placed on top of each other in the final nanotube structure. Zigzag (also called sawtooth [69]) nanotubes have $(n, 0)$ lattice vector values, whereas armchair (also called serpentine [69]) nanotubes have (n, n) lattice vector values [67]. These are the two possible achiral confirmations; all other (n, m) lattice vector values yield chiral nanotubes [67, 70]. Nanotubes may exist individually as single-walled CNTs or as multiwalled CNTs that have a nested structure. In particular, multiwalled CNTs consist of smaller-diameter CNTs within larger-diameter CNTs, which are within even larger-diameter CNTs, and so forth. In multiwalled CNTs, each nanotube is separated from its neighboring nanotubes by the van der Waals distance of about 0.35 nm.

The electronic structure calculation of Mintmire et al. [63] was performed with an all-electron *ab initio* method, which was originally used for one-dimensional band structures of polymer systems but was adapted for structures of helical symmetry. The calculated band

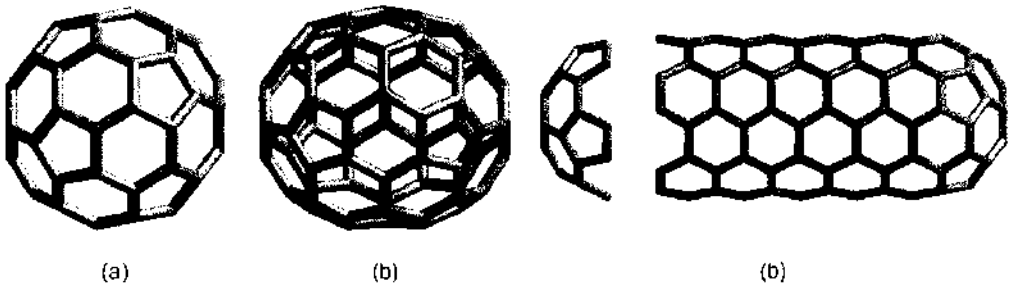


Figure 1. C_{60} is displayed in part (a). The C_2 axis of the C_{60} is aligned within the page in the horizontal direction. Part (b) displays the C_{70} fullerene with its C_5 axis within the page and pointing in the horizontal direction. As noted in the text, the C_{70} fullerene can be constructed from the C_{60} by adding 10 carbon atoms with D_{5h} symmetry. The orange atoms within the C_{70} molecule are those that came from the C_{60} and the blue atoms represent the D_{5h} ring. Part (c) shows the extended structure that can be created by adding sequential D_{5h} rings with the proper rotation. The C_5 axis is also aligned within the page and in the horizontal direction. Again, the parts of the extended structure made with the C_{60} are orange in color and the added D_{5h} rings are blue.

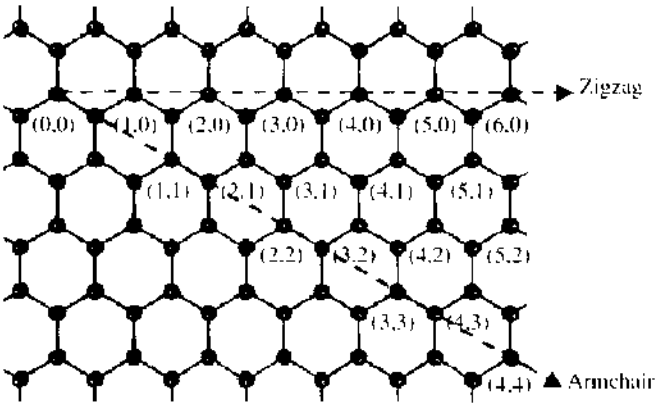


Figure 2. Image of the (n, m) indices used to describe the construction of a CNT from a graphene sheet. After Ref. [64].

structure is shown in Fig. 3. It can be seen that two bands, a_1 and a_2, crossed at the Fermi level, making this structure a one-dimensional metal. The a_1 and a_2 bands had a p-type character near the Fermi level and were equivalent to energy levels one would obtain for a graphite sheet (graphene) calculation in the same direction. Graphite is normally not metallic because it has low conductivity within the planes from low carrier concentration; graphite as a whole is insulating because of the van der Waals bonding between the sheets, across which charge does not travel.

One of the most interesting aspects of this calculation was that Peierls distortions [71] did not produce a gap at the Fermi level to destroy the metallic character of the nanotube, as had occurred in other one-dimensional materials, such as polyacetylene [63, 72, 73].

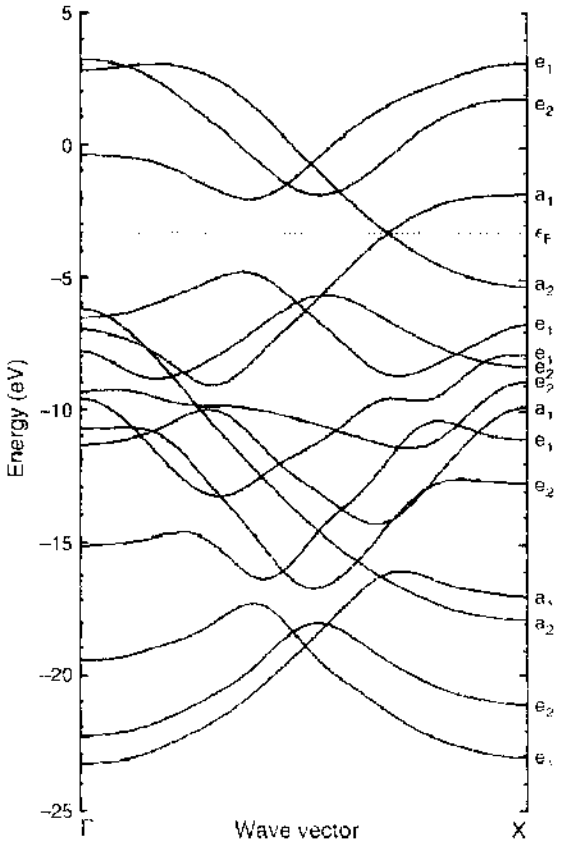


Figure 3. Calculated band structure of a (6, 6) single-walled CNT. Γ to X runs along the CNT axis from 0 to π in reciprocal space. The dotted line represents the Fermi energy. Figure reprinted with permission from [63]. J. W. Mintmire, et al., *Phys. Rev. Lett.* 68, 631 (1992). © 1992, American Physical Society.

Mintmire et al. [63] explained this by showing that the increase in electron overlap/atom that would drive a Peierls distortion of the nanotube was much less than in polyacetylene (about one-fifth as large), whereas the elastic energy/atom resisting the distortion was much larger (1.5 times as large) at temperatures less than or equal to room temperature. In addition, they estimated that the carrier concentration was on the order of more typical metal conductors (about 10^{22} cm^{-3}).

Saito et al. [64] and Hamada et al. [65] simultaneously considered, for the first time, the effects of chirality and helical structure of single-walled CNTs on their electronic properties. The electronic structure calculations of Saito et al. showed that CNTs with the same zigzag structure but different diameters had very different electronic densities of states. Specifically, the authors showed that (10, 0) nanotubes were semiconducting, whereas (9, 0) nanotubes were metallic. The researchers further derived simple, predictive rules for determining the conductivity of single-walled CNTs, as follows

$$2\mathbf{n} + \mathbf{m} = 3q \tag{27}$$

where **n** and **m** are the integral lattice vectors described above and q is an integer. It is soon obvious that this rule predicts that all armchair single-walled CNTs and all zigzag single-walled CNTs where **n** is a multiple of three are metallic. All other single-walled CNT chiralities have a band gap in their electronic structure, making them semiconducting.

The TB calculations of Hamada et al. yielded similar results, and they further showed how single-walled CNTs with (n_1, n_2) integral lattice vectors, where $n_1 \geq 2n_2 \geq 0$, which obeyed the following expression

$$\mathbf{n}_1 - 2\mathbf{n}_2 = 0 \tag{28}$$

were metals, tubes with (n_1, n_2) integral lattice vectors that obeyed the following expression

$$\mathbf{n}_1 - 2\mathbf{n}_2 = 3q \tag{29}$$

were narrow-gap semiconductors, and all other single-walled CNTS were wide-gap semiconductors (or insulators). Note that the nomenclature used by Hamada et al. was not the same as that of Saito et al. (the differences arise from the definition of the primitive lattice vectors of graphite). Hamada et al. used primitive lattice vectors separated by an angle of 120 degrees, and Saito et al. used primitive lattice vectors separated by 60 degrees. The nomenclature of Saito et al. is now the prevalent nomenclature used in the literature.

Experimental verification that single-walled CNTs are metallic conductors came in 1997 in a paper by Tans et al. [74]. The measurement was done on a nanotube draped across Pt electrodes on a Si/SiO_2 surface, as shown in Fig. 4. The experiments further showed that the nanotube conducted as a coherent quantum wire, as illustrated in Fig. 5, which plots current versus bias voltage. In other words, the current flowed through distinct and discrete electron states that are quantum-mechanically coherent from one contact point to the other. A subsequent experimental paper by Wildöer et al. [75] and one by Odom et al. [76] confirmed the theoretical prediction that the electronic properties of single-walled CNTs, as measured by scanning tunneling microscopy, vary significantly with changes in tube diameter and chirality. This finding is summarized in Fig. 6 (reprinted from Ref. [75]).

The extreme relevance of these studies is highlighted by more recent efforts to build complex electronic devices from single-walled CNTs [77–79]. In this respect, computational work also plays an important role. For example, Heyd et al. [80] used TB calculations to show that the band gap of zigzag single-walled CNTs could be varied as the CNTs were stressed in the elastic response region, and metal-to-semiconductor transitions were possible. In contrast, armchair CNTs remained metallic, even when stressed. This finding indicates that nanometer-scale strain gauges and vibration sensors could be produced from CNTs. Brenner et al. [81] used TB calculations to further explore this property of CNTs and predicted that the band gap of a (17, 0) CNT closed at a compressive strain of about 11% and at a tensile strain of about 8% and reopened at higher strains. This study further predicted that the number of charge carriers in the CNT could be doubled by straining the CNT in tension by about 5%.

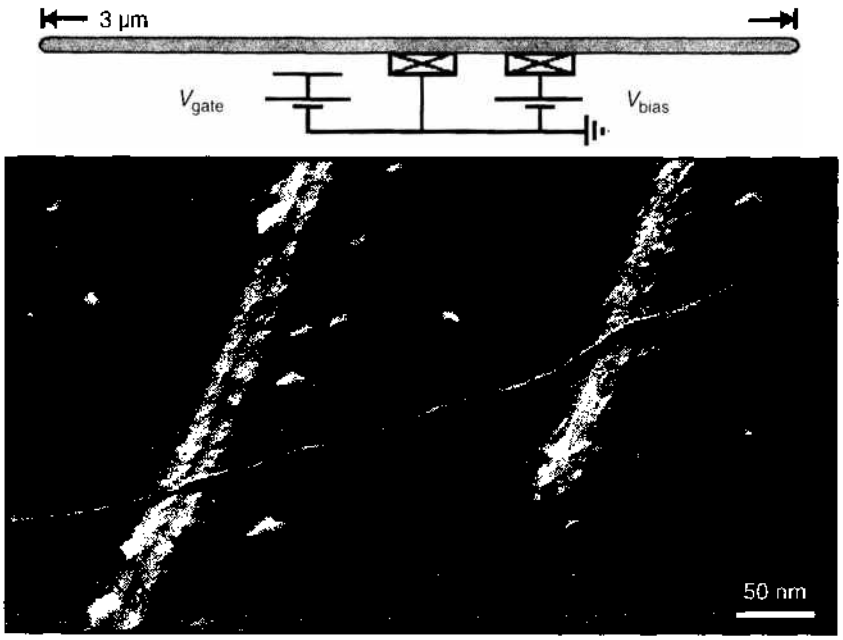


Figure 4. Experimental setup used to measure the electronic properties of the single-walled CNT. Reprinted with permission from [74], S. J. Tans et al., *Nature* 386, 474 (1997). © 1997, Macmillan Publishers Ltd.

At present, much effort is being focused on experimentally producing nanotubes of one conducting type (i.e., metallic, semiconducting, or insulating). Very recently, there has been some success in separating metallic and semiconducting CNTs from a heterogeneous mixture of tubes [82, 83].

3.2. Structure and Properties of Nanotube Junctions

The ultimate goal of researchers working to build nanometer-scale electronic devices is to be able to use as building blocks not individual CNTs, but a series of either metallic or semiconducting CNTs that are connected to one another at seamless junctions. In this way, for example, nanometer-scale transistors could be constructed by connecting a metallic nanotube to a semiconducting tube that, on its other end, is connected to another metallic

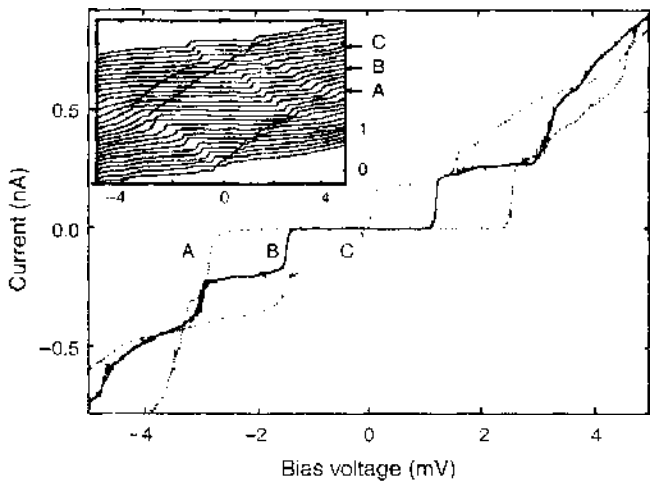


Figure 5. Experimentally measured current versus voltage plot of the single-walled CNT shown in Fig. 4 at different trace voltages. Trace voltages A, B, and C are 88.2, 104.1, and 120 meV, respectively. The inset shows current-voltage (bias) curves for different trace values of V_{gate}. The graphs are vertically offset for clarity and range from $V_{gate} = 50$ meV for the top curve, to $V_{gate} = 136$ meV for the bottom curve. Reprinted with permission from [74], S. J. Tans et al., *Nature* 386, 474 (1997). © 1997, Macmillan Publishers Ltd.

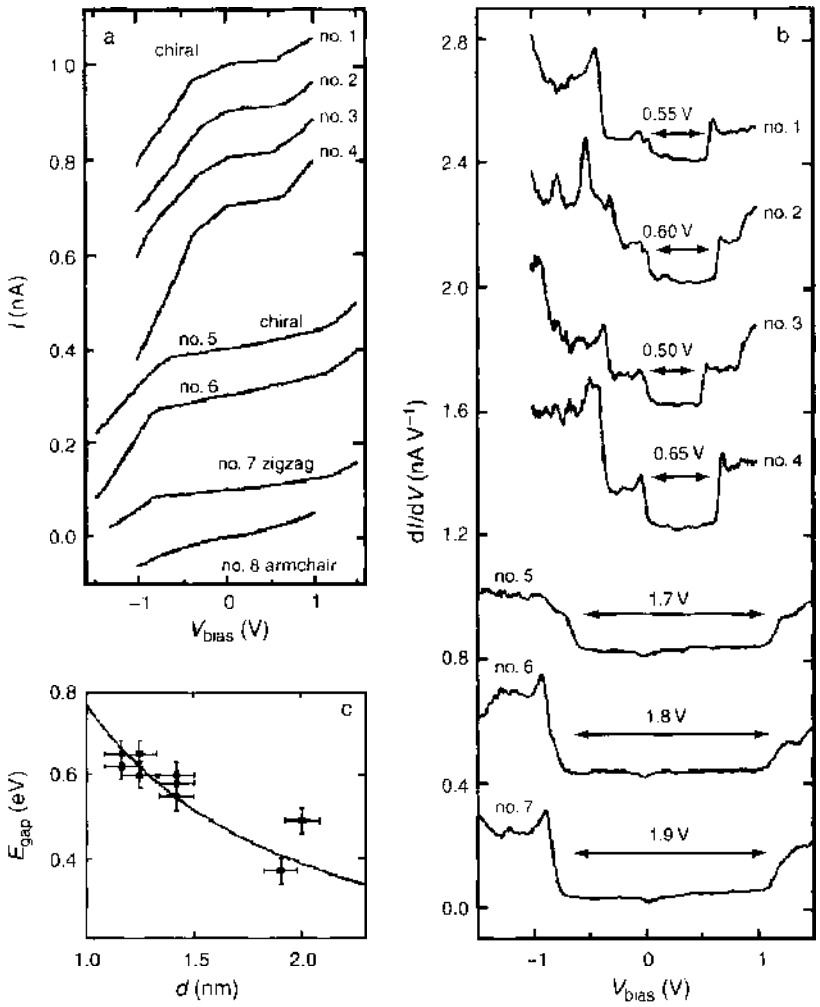


Figure 6. Experimental measurements of the electrical conductivity of a wide range of CNTs. Nanotubes 1–6 are chiral, nanotube 7 is zigzag, and nanotube 8 is armchair. Part (a) shows the I–V curves of the CNTs. Many of the chiral tubes have several kinks, but the armchair tube has none between −1 V and +1 V. Part (b) shows the dI/dV and estimated band gaps. Part (c) shows the band gaps of the nanotubes and their agreement with the theoretical predictions. Reprinted with permission from [75], J. W. G. Wildöer, et al., *Nature* 391, 59 (1998). © 1998, Macmillan Publishers Ltd.

nanotube. The study of such structures is again much easier to achieve with theoretical methods than in experiements, as discussed below.

Quite soon after the discovery of CNTs, the effect of wall defects on nanotube curvature, such as pentagons and heptagons instead of the usual hexagons, was discussed in the literature [84, 85]. In particular, pentagons cause inward curvature, whereas heptagons cause outward curvature. Such imperfections were soon found to be necessary for forming a seamless junction between two single-walled CNTs of differing diameters or chiralities [85–88]. Specifically, the presence of pentagon/heptagon pairs was predicted to allow the formation of the junctions such that there was no net curvature in the system. The electronic properties of these two-nanotube junctions were readily investigated with TB calculations by several groups [86–88]. The calculations found that the junctions possessed resonant tunneling conduction properties and effectively functioned as nanometer-scale electronic devices.

An example junction system is shown in Fig. 7 between a semiconducting (8, 0) CNT and a metallic (7, 1) CNT from Ref. [87]. The atoms forming the heptagons or pentagons at the junction are indicated by the large, gray spheres. Figure 8a and 8b show the calculated density of states for the system starting at the junction and moving away down the (8, 0) CNT and (7, 1) CNT, respectively. It is clear from these plots that the junction had a much more significant effect on the properties of the semiconducting tube, with electronic states appearing the gap region of the pristine (8, 0) CNT. In the case of the metallic tube, the

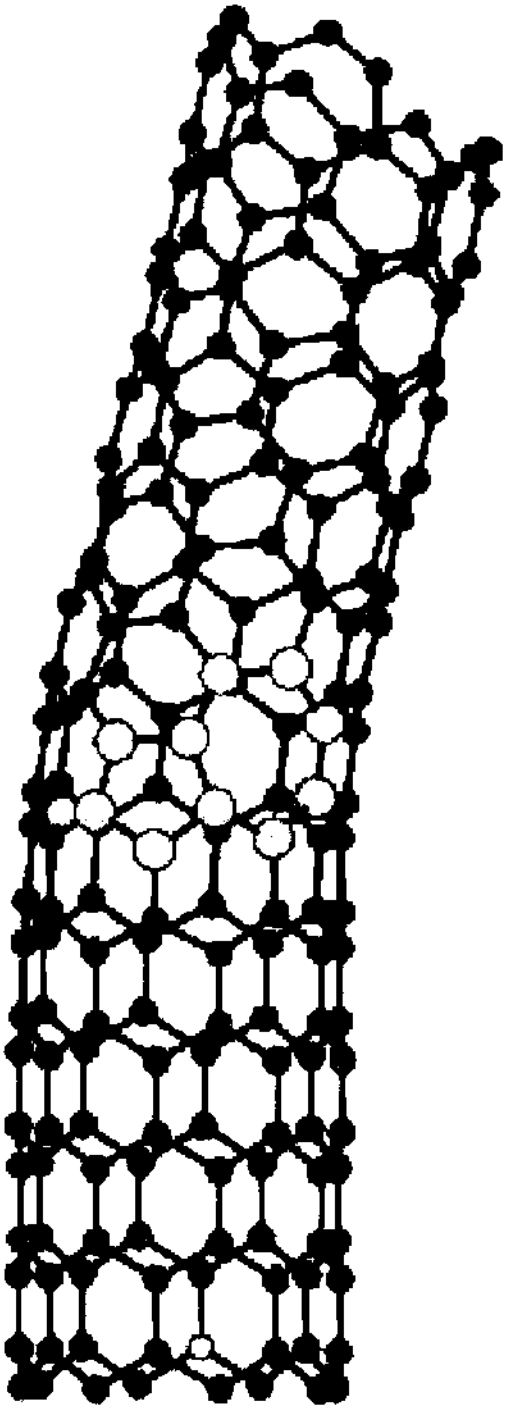


Figure 7. Computationally constructed junction between an (8, 0) single-walled CNT and a (7, 1) single-walled CNT. The large grey balls accentuate the pentagon-heptagon pair formed in joining these tubes. Figure reprinted with permission from [87], L. Chico et al., *Phys. Rev. Lett.* 76, 971 (1996). © 1996, American Physical Society.

only effect of the junction was the smearing out of some Van Hove singularities that were present in the ideal CNT [87]. In both cases, the effects of the junction were quickly lost, and the electronic properties of the pristine nanotubes recovered, as one moved away from the junction in either direction.

Subsequent work by Menon and Srivastava [89, 90] considered the electronic properties of more complex Y- and T-junctions that were not necessarily constructed from pentagon/heptagon defect pairs and could even contain octagons. Examples of these structures are shown in Fig. 9. TB calculation results for the local density of states of the T-junction presented in Fig. 9 are shown in Fig. 10. These calculations predicted that the presence of the structural defects (heptagons in this case) at the neck region caused the appearance of

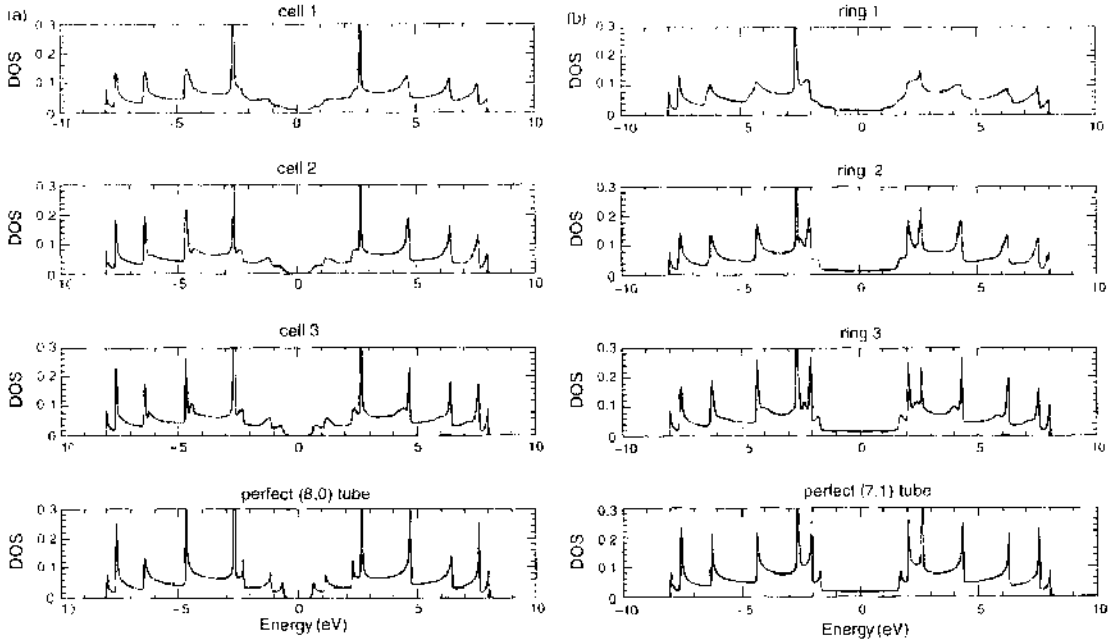


Figure 8. Calculated local density-of-states (DOS) plots for the CNTs shown in the junction in Fig. 7. Part (a) shows the local DOS of the (8, 0) CNT as one moves away from the junction and for a pristine (8, 0) CNT. Part (b) shows the local DOS of the (7, 1) CNT as one moves away from the junction and for a pristine (7, 1) CNT. In both cases, cell 1 is closest to the junction. Reprinted with permission from [87], L. Chico et al., *Phys. Rev. Lett.* 76, 971 (1996). © 1996, American Physical Society.

localized states in the gap of the semiconducting (10, 0) CNT that may act to pin the system Fermi level. Similar results were found for the Y-junctions considered by these authors.

These calculations were followed by calculations of the rectifying behavior of a Y-junction composed of a (14, 0) CNT that branched into two (7, 0) CNTs [91]. The ends of all three nanotubes were assumed to be in contact with Ni electrodes. The calculated currents as a function of applied voltage are shown in Fig. 11, where currents I_2 and I_3 flowed from the two (7, 0) CNTs to the (14, 0) CNT, and current I_1 flowed through the (14, 0) CNT away from the junction branch and toward the Ni electrode. Figure 11 indicates that this Y-junction acted as a rectifying device, because current flowed when the applied voltage was negative, but current did not flow for positive applied voltages.

Experimentally, there is unequivocal evidence for CNT Y-junctions, and some strategies have been developed for producing them. The first experimental report of nanotube L-, T-, and Y-junctions [92] considered junctions that were produced along with regular nanotubes

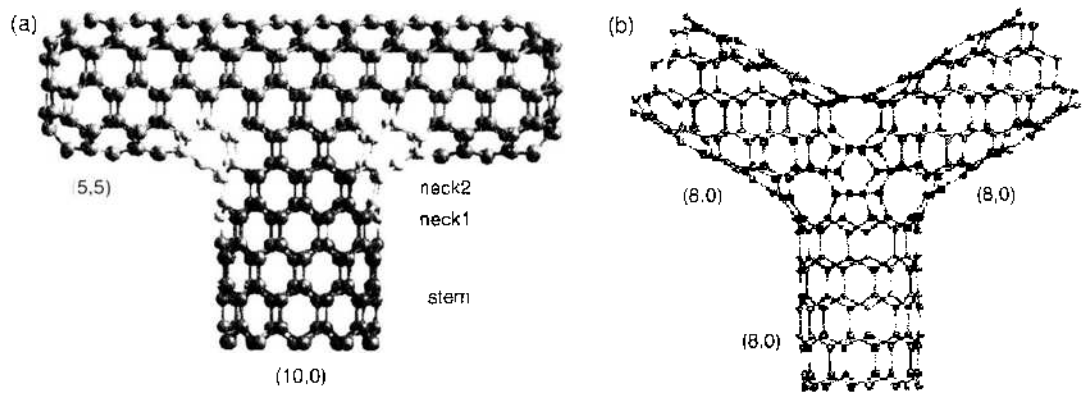


Figure 9. Computationally constructed CNT junctions. Part (a) shows a fully relaxed T-junction made up of (5, 5)-(10, 0)-(5, 5) CNTs. Part (b) shows a fully relaxed Y-junction made up of three (8, 0) CNTs. In both figures the lighter atoms denote atoms making up the heptagon–pentagon pairs. Reprinted with permission from [90], Madhu Menon and Deepak Srivastava, *J. Mater. Res.* 13, 2357 (1998). © 1998, Materials Research Society.

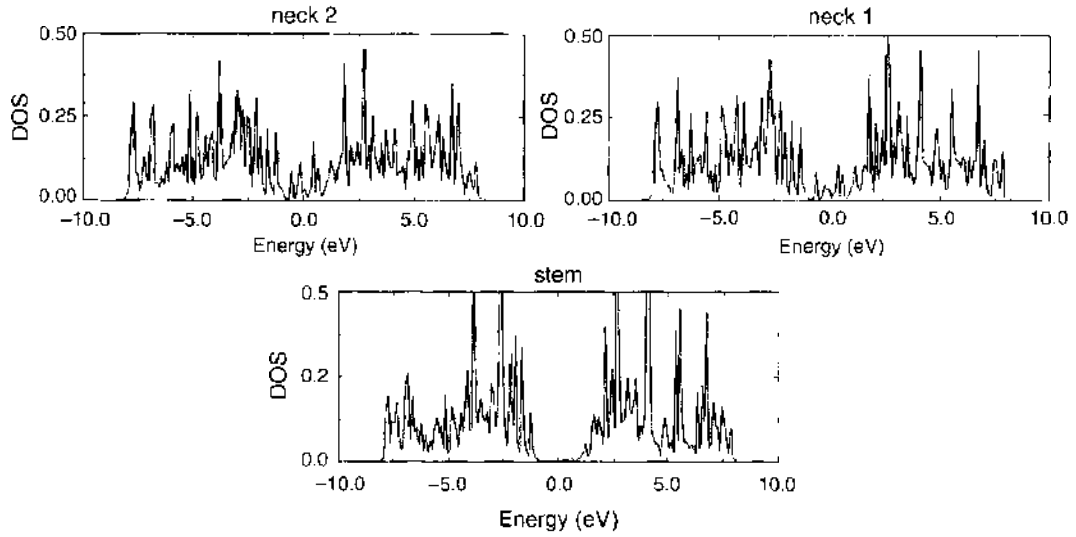


Figure 10. The calculated local DOS plots at different points on the T-junction shown in Fig. 9a. Defect induced gap states can be seen moving from stem to neck. Figure reprinted with permission from [89], M. Menon and D. Srivastava, *Phys. Rev. Lett.* 79, 4453 (1997). © 1997, American Physical Society.

in an arc-chamber. However, subsequent experimental papers showed that Y-junctions could be reproducibly formed in nanoporous templates [93], by decomposition of fullerenes in the presence of metallic powders [94], by pyrolysis of specific precursors [95], or through welding by electron irradiation [96, 97].

Experimental measurement of room-temperature electrical conduction through Y-junctions of single-walled CNTs produced in nanoporous templates [98] or through pyrolysis [95] showed the same rectifying behavior calculated in Ref. [91]. In this case, the experiments were carried out before the calculations were performed, but the good agreement between the two studies indicates the ability of the available theoretical models to characterize the behavior of these nanostructures.

MD simulations using TB methods [96] and bond-order potentials [99] revealed the likely mechanisms by which electron irradiation welds together nanotubes to form junctions. In particular, defects were created in the CNT walls that caused the damaged tubes to weld together. TB calculations of the electronic properties of X-shaped nanotube junctions produced through electron irradiation in these simulations [99, 100] indicated that the electronic

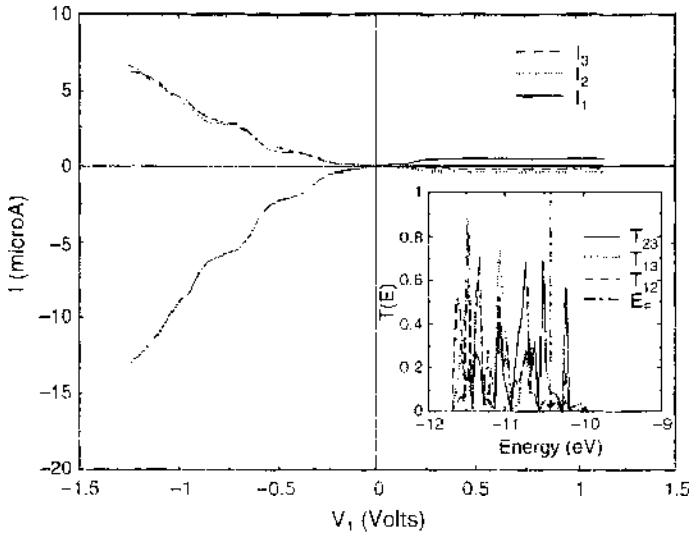


Figure 11. The predicted rectifying behavior of a Y-junction consisting of a (14, 0) CNT branching into two (7, 0) CNTs. Figure reprinted with permission from [91], A. N. Andriotis et al., *Phys. Rev. Lett.* 87, 066802-1 (2001). © 2001, American Physical Society.

properties of junctions produced through irradiation were significantly different from the properties of ideal junctions. Figure 12a shows that the irradiated junctions were much more disordered than the ideal junctions shown in Figs. 7 and 9. This disorder led to localization of electronic states, shown in Fig. 12b, effectively making irradiated CNT junctions into quantum dots.

3.3. Mechanical Properties of Carbon Nanotubes

It has been known for some time that graphite has a high Young's modulus within the plane. As carbon nanotubes are essentially graphene cylinders, it was recognized quite soon after their discovery that CNTs might have large Young's moduli as well. As was the case in the last two sections, computational and theoretical approaches led the way in exploring the mechanical properties of the nanotubes.

One of the first papers to address the mechanical properties of CNTs was by Robertson et al. [69], where both early versions of the empirical bond-order potential and DFT-LDA calculations were used to examine the elastic properties and strain energy of single-walled CNTs with diameters up to 1.8 nm. Both the computational methods used indicated that the strain energy per atom varied as the inverse radius squared relative to the energy of a perfect graphene sheet, as shown in Fig. 13. The results were found to depend only on the radius of the CNT and not on its chirality. Robertson et al. also found that the force constants related to stretching in the direction of the tube axis decreased as the nanotube diameter decreased. This behavior was found to depend on the CNT chirality, with the zigzag nanotubes showing the lowest stiffness and the armchair nanotubes showing the greatest stiffness for the same nanotube diameter. In the limit of infinite CNT diameters, these theoretical methods reproduced the elastic constants of unstrained graphene sheets.

This study was followed by many other theoretical studies of the mechanical and elastic properties of single-walled CNTs, some of which are now discussed. Calculations of Young's modulus using TB methods predicted a range of values close to 1 TPa that depended slightly on the details of the model. For example, in Ref. [101], an average value of about 1.2 TPa was predicted for single-walled CNTs of various diameters and chiralities; in Ref. [102] an average value of 0.97 TPa was predicted for single-walled CNTs that did not vary with helical structure or diameter; and in Ref. [103], values of 1.0–1.7 TPa were predicted. In the latter case, the elastic properties were found to be independent of helical structure, but the onset of plastic deformation depended significantly on helical structure, with zigzag tubes having a larger elastic limit than armchair tubes. The effects of curvature were not included in Ref. [102], which could explain the independence of the results on CNT diameter.

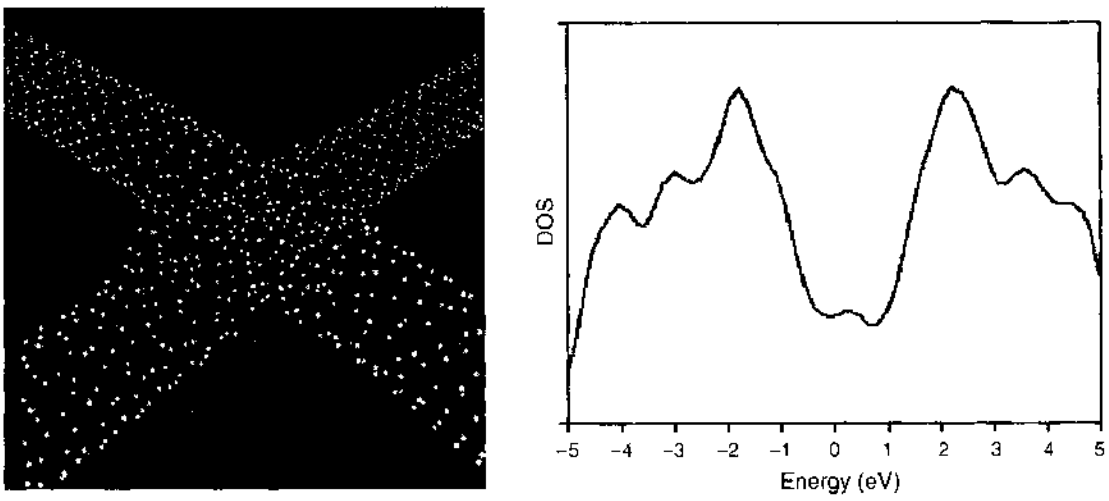


Figure 12. Part (a) shows two (5, 5) single-walled CNTs that have been welded together by electron-beam irradiation in MD simulations. Part (b) displays the calculated DOS of the welded CNT system. Courtesy of Inkook Jang.

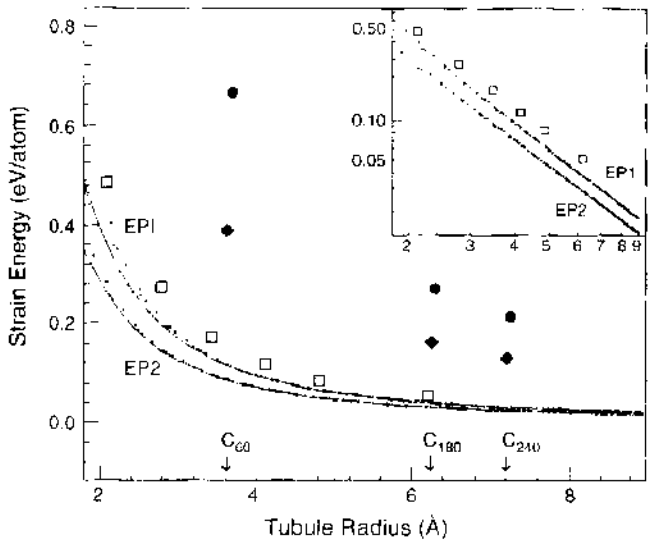


Figure 13. Minimized values for the strain energy per carbon atom as calculated with two early versions of the bond-order potential for hydrocarbons, termed EP1 and EP2. The corresponding LDA DFT results are plotted with open squares. The circles and diamonds represent strain per carbon atom for unoptimized C_{60}, C_{180}, and C_{240}. These values are placed at their respective radii. The EP1 and EP2 potentials are both zeroed with respect to −7.3995 and −7.3756 eV per atom, respectively. This is the equilibrium energy of carbon in graphite for the given potentials. Figure reprinted with permission from [69], D. H. Robertson et al., *Phys. Rev. B* 45, 12592 (1992). © 1992, American Physical Society.

Other theoretical methods were also applied to the determination of the Young's modulus of single-walled CNTs. When potentials were used that included torsional energy contributions, the calculated Young's moduli of single-walled CNTs decreased as the diameter increased and as the degree of helicity increased, yielding values of around 1.0–1.1 TPa [104]. All-electron HF-based *ab initio* calculations predicted a Young's modulus of about 1 TPa for single-walled CNTs of various chiralities and about the same radius [105]. A study by Ozaki et al. [106] examined the responses of single-walled CNTs with differing helical structures but approximately the same radius to large strain at 0 K. The strain energy per unit length did not show any helical dependence, and the resulting Young's moduli of about 0.980 TPa also showed little variation with changes in the helical symmetry, under either tension or compression. However, under large strain conditions (defined as 15% or greater), the zigzag tubes considered were significantly stiffer under tension, and the armchair tubes considered were significantly stiffer under compression.

MD simulations were also applied to study the response of nanotubes to strain and determine the Young's modulus. Importantly, MD simulations established quite early on that CNTs could be deformed easily by bending, twisting, or otherwise deforming them [107–109]. In all cases, when the CNT was returned to its original position, it was found that there was no plastic deformation as a result of the bending. Figure 14 (reprinted from the paper by Yakobson et al. [107]) shows snapshots from simulations in which single-walled CNTs were compressed and shows the effect of this deformation on the system energy. This is a remarkably elastic phenomena that would not take place in other graphitic structures or solid nanowires.

When DFT calculations by Maiti et al. were used to study the bending of single-walled CNTs under mechanical loads at the ends [110], the results showed that the tubes bent mechanically but retained their bonding structure to relatively high-bending angles. However, if the single-walled CNT was constrained from bending (e.g., held in place), then plastic failure was predicted to occur in TB calculations by Srivastava et al. [111].

Ab initio calculations and MD simulations using the bond-order potential considered the tensile loading of CNTs and agreed that although the behavior was elastic under normal conditions, plastic behavior occurred at high strain and at low temperatures [112, 113]. The brittle versus ductile behavior was found to be dependent on the diameter and helical symmetry of the CNTs under consideration. MD simulations by Yakobson et al. revealed

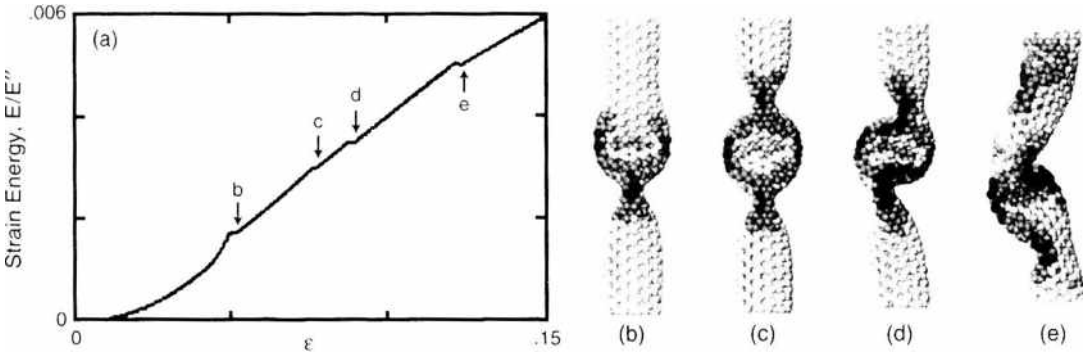


Figure 14. Calculated strain energy versus axial compression is plotted in part (a). Singularities can be seen in part (a), and these correspond to the morphologies in (b)-(e). The morphologies occur at the following strains: (b) $\varepsilon = 0.05$, (c) $\varepsilon = 0.076$, (d) $\varepsilon = 0.09$, and (e) $\varepsilon = 0.13$. The evolution of the symmetry of the CNT can also be seen in (b)-(e). Figure reprinted with permission from [107], B. I. Yakobson et al., *Phys. Rev. Lett.* 76, 2511 (1996). © 1996, American Physical Society.

information about the failure mechanisms [114] of single-walled CNTs when they were subjected to extreme tensile forces. Plastic flow was predicted at high temperatures, and brittle failure was predicted at low temperatures. The mechanism in both cases started off with the creation of a dislocation dipole of 5/7 defect structures that moved away from each other along the CNT axis in a helical path. The simulations further predicted that CNTs could be strained in tension to about 30–40% of their original length before failure occurred, with the results showing some dependence on the nanotube diameter, rate of strain, and system temperature.

There have been some important disagreements in the literature about the mechanical properties of CNTs. For instance, calculations by Yakobson et al. [107] that determined the Young's modulus from the flexural rigidity and in-place stiffness of the single-walled CNT shown in Fig. 14 predicted a value of 5.5 TPa, assuming a wall thickness of 0.066 nm, which is the radius of a carbon atom. Electronic band structure calculations, which also took into account the wall thickness of the single-walled CNTs [115], found a Young's modulus of 5.1 TPa for CNTs that was independent of radius and helical structure.

The explanation for this large difference in calculated Young's moduli of single-walled CNTs can be attributed to the value used for the wall thickness of the CNT [115]. While single-walled CNTs are one carbon atom thick, they have a van der Waals radius of about 0.35 nm. If the wall thickness was taken to be the van der Waals radius, continuum elasticity theory predicted a Young's modulus of 1 TPa but also predicted a bending stiffness that was about 25 times lower than that predicted by classical mechanics. The correct classical bending stiffness could be obtained if a wall thickness of 0.066 nm was used, but this value of wall thickness also yielded a Young's modulus of about 5 TPa. If the Young's moduli were determined in a very different manner, such as by scaling the modulus of graphene with density when it was placed in hexagonal, close-packed arrangements with hollow interior volumes, much lower values of about 600 GPa were obtained for the Young's moduli of single-walled CNTs [116].

Experimental verification of the high Young's moduli of CNTs first came in 1996 by monitoring the freestanding vibrations of nanotubes with a transmission electron microscope [117]. This approach was first applied to multiwalled CNTs [117] and was subsequently applied to single-walled CNTs [118]. The average measured value of Young's modulus was about 1.8 TPa for a range of multiwalled CNTs of differing diameters and containing a variable numbers of shells, and 1.25 ± 0.45 TPa for single-walled CNTs with diameters of 1.0–1.5 nm.

The mechanical properties of single-walled CNTs were also investigated by compressing CNTs with a proximal probe tip [119]. These studies showed that it was possible to greatly deform the nanotube without inducing plasticity [120] and that the compressibility of single-walled CNTs was determined to be reversible up to 4 GPa, after which the lattice was destroyed [121]. When proximal probe tips were used to investigate the responses of pinned

multiwalled CNTs and SiC nanorods (solid nanometer-scale cylinders), the results showed that the CNTs were twice as stiff as the SiC and that at high loads, the SiC nanorods fractured, whereas the CNTs deformed elastically [122]. This study by the Lieber group yielded a Young's modulus of 1.28 ± 0.59 TPa for multiwalled CNTs. More recently, Yu et al. used an experimental tension set-up to measure the Young's modulus of a single-wall CNT [123]. The researchers found a Young's modulus of 0.32–1.47 TPa with a mean modulus of 1.0 TPa.

CNTs were also embedded in epoxy matrix and cooled to compress the CNTs; the system was then characterized with micro-Raman spectroscopy [124]. The calculated values of Young's modulus were about 5 TPa for single-walled CNTs and about 1.8 TPa for multiwalled CNTs. It should be noted that the analysis of the experimental data in this case made use of the concentric cylinder model and the band shifts for the CNTs, which included approximations for the CNT wall thickness.

A complicating factor that should be kept in mind when comparing experimental and computational results for CNT mechanical properties is the common presence of point defects in experimental CNT walls, especially in multiwalled CNTs, which can reduce the measured modulus by a factor of 10 [125].

3.4. Chemical Modification of Carbon Nanotubes

Unlike the results discussed in the last three sections, the area of chemical modification of nanotubes is one that has been led primarily by experimental efforts. Nevertheless, computational studies have played an important role. In particular, computational studies have been used to predict the effects of chemical modification of nanotubes on their mechanical and electrical properties and to explore new methods for achieving chemical modification of CNTs.

There are numerous factors motivating the search for methods to chemically modify the structure of CNTs. For example, chemical modification of nanotube walls might be a way to tune the electrical properties of the tube. In addition, CNTs are being explored for use as fibers for polymer-matrix composites [126–138] because of their electrical properties, high Young's modulus in the direction of the nanotube axis, and high resistance to brittle failure [67, 68, 139–142].

The production of composite materials with nanometer-scale components such as CNTs is complicated by the small size of the nanotubes and their tendency to agglomerate into thick ropes or bundles [67]. In addition, CNTs, like graphite, have an inherent insolubility in most solvents. Therefore, there is incentive to disperse CNTs in solution to enhance the production of these composites. Last, CNTs and polymer chains interact only weakly through van der Waals bonds. There is, therefore, interest in increasing the strength of the bonding between the nanotubes and the polymer matrix so that the overall strength and toughness of the composite material can be increased.

The chemical functionalization of nanotube ends was first carried out by the Smalley group as a result of efforts to purify single-walled CNTs that were mixed with amorphous carbon and metal catalyst [143, 144]. Purification was achieved by refluxing the as-produced material in HNO_3 and then suspending the nanotubes in a mixture of water and surfactants. Sonication of purified CNTs in a mixture of concentrated sulfuric and nitric acid further cut them into open "pipes" that were much shorter (by a few hundred nanometers) than their original length (several micrometers) and that had carboxylic acid (–COOH) groups covalently attached to the open ends [143]. This finding was exploited by the Lieber group to make chemically sensitive proximal probe tips [145, 146]. These nanotube pipes formed stable colloidal suspensions in water in the presence of surfactants [143]. However, true dissolution of shortened single-walled CNTs was achieved by reacting the carboxylic acid groups at the nanotube ends with thionychloride ($SOCl_2$), followed by amidation reaction with octadecylamine ($CH_3(CH_2)_{17}NH_2$) (ODA) [147] or alkyl-aryl amine 4-cococyl-aniline ($4\text{-}CH_3(CH_2)_{13}C_6H_4NH_2$) [148].

Direct covalent functionalization of the CNT walls with organic groups was first achieved by the Haddon group through reaction with dichlorocarbene [147, 149]. However, the walls of single-walled CNTs were earlier fluorinated in a nondestructive manner [150–152].

Ab initio DFT calculations subsequently showed that F atoms preferred to add to neighboring sites on the walls of the single-walled CNTs [153]. Experimental studies showed that the attached fluorine could be reacted with precursors to yield nanotubes with organic groups attached to the walls [154].

Nanotubes' walls were more recently functionalized through noncovalent interactions by the irreversible adsorption of a bifunctional molecule onto the surface of a single-walled CNT [155]. This occurred through the interaction of aromatic pyrenyl group with the delocalized bonding in the nanotube walls. There were also several reports of the dissolution of single-walled CNTs in various solvents by wrapping conjugated polymers around the nanotubes or nanotube bundles [126, 156–160].

Another way in which CNTs can be chemically modified is through electron irradiation, which causes their collapse in an anisotropic manner [161, 162]. Simulations [161] and calculations [162] showed that the threshold for atomic displacement by knock-out from the electrons was strongly anisotropic and that the most damage occurred at the front and back of the nanotube. In addition, electron irradiation caused nanotubes to shrink in diameter by an order of magnitude [163] or merge through a zipper-like mechanism with other CNTs having the same chirality [164]. As discussed in Section 3.2, electron irradiation was also used to "weld" carbon nanotubes to form epitaxial junctions [96, 99, 165].

Similar findings were seen in simulations and experiments of the irradiation of carbon nanotubes with ions. Specifically, simulations by Ni et al. predicted that the deposition of ions, such as CH_3^+ and C^+, at low energies of 3–80 eV can covalently chemical modify single-walled CNTs [166] and multiwalled CNTs [167] arranged in bundles, as shown in Fig. 15. The ions were also predicted to produce a multitude of defects in the nanotube walls. Mass-selected, CF_3^+ ion beam deposition experiments at 45 eV confirmed that ion-beam deposition was a viable approach to chemically functionalize nanotube walls [168], as shown in Fig. 16. The small functional groups attached in this matter may be reacted with longer polymer chains to yield nanotubes functionalized with long polymer groups. This finding was important because it indicated that nanotubes may be chemically functionalized without first exposing them to strong acidic or otherwise harsh chemical environments. Ion deposition was also predicted by Ni et al. to induce cross-links between nanotubes in the bundle and between

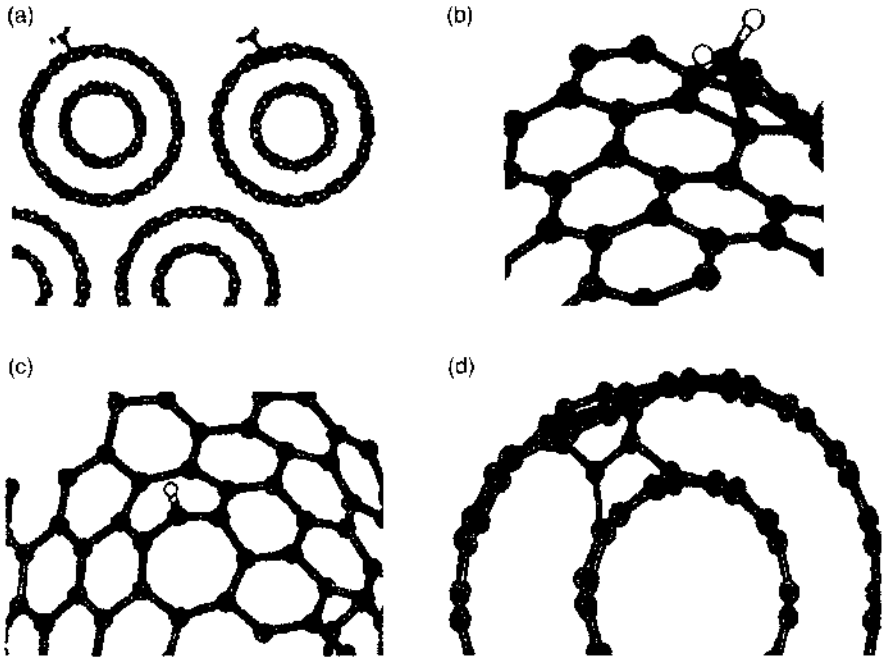


Figure 15. Snapshots from MD simulations. Part (a) shows the functionalization of a double-walled CNT after CH_3^+ ion deposition at 10 eV. Part (b) shows the functionalization of a single-walled CNT after CH_3^+ deposition at 45 eV. Examples of the heptagon defects that can form when CH_3^+ is deposited at 45 eV are shown in part (c). Part (d) shows intrashell cross-links that formed during the 45-eV ion deposition process. Reprinted with permission from [168], B. Ni et al., *J. Phys. Chem. B* 105, 12719 (2001). © 2001, American Chemical Society.

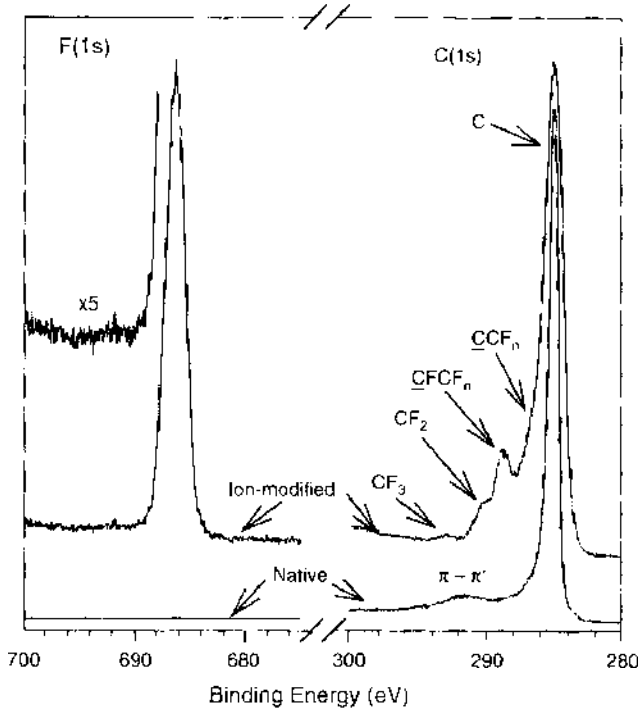


Figure 16. Core level x-ray photoelectron spectroscopy spectra in the C(1s) and F(1s) region for the unmodified and for a multiwalled CNT bundle modified by CF_3^+ ion beam deposition at incident energies of 45 eV. Structures that form during deposition are noted in the figure. Reprinted with permission from [168], B. Ni et al., *J. Phys. Chem. B* 105, 12719 (2001). © 2001, American Chemical Society.

shells in multiwalled CNTs [166–168], as shown in Fig. 15d. This was significant because the inner shells of multiwalled CNTs have been pulled out of the outer shells using a loading stage [169] with either a stick-slip or smooth pullout motion. MD simulations with classical potentials showed that sliding of the inner shells against the outer shells depended somewhat on the helicity, and hence relative corrugation, of the nanotube shells, but confirmed that little energy was needed to effect the pull-out [170]. The formation of cross links between the inner and outer shells could thus not only stabilize CNT bundles with respect to shear but anchor inner shells inside multiwalled CNTs. This could lead to efficient load-transfer between the outer and inner shells when the tube deformed.

An important combined computational and experimental study showed that introducing conformational strain caused the nanotubes to form "kink" sites that were more reactive than normal nanotube walls [171]. In the study, nanotubes were bent, compressed or twisted, and exposed to hydrogen gas in the simulations, and to nitric acid vapor in the experiments. The simulations indicated that the absorption energy of hydrogen atoms at the areas of high strain in the kink sites increased by about 1.6 eV relative to absorption on an unstrained nanotube. The experimental results were consistent with these findings. This approach for the chemical functionalization of CNTs was important for two reasons: it demonstrated that the reactivity of nanotube walls could be increased by introducing conformational strain, and it provided a mechanism by which nanotubes could be selectively functionalized along their length.

The attachment of covalently bound groups to the walls of CNT might be expected to have some effect on the mechanical properties of the tubes through the introduction of sp^3-defects in the normal sp^2-hybridized nanotube structure. Classical MD simulations with the bond order potential by Garg and Sinnott showed that the buckling forces of single-walled CNTs that were highly functionalized with $H_2C{=}C$ groups were only about 15% less than the buckling force of the pristine tubes [172]. This finding indicated that although the mechanical properties of functionalized single-walled CNTs were somewhat degraded relative to pristine tubes, the degradation was not severe. MD simulations were also used to determine the relative strain of a functionalized (10, 10) single-walled CNT relative to a bare (10, 10) CNT as a function of percentage functionalization with CH_3 groups [173].

The results showed that the nanotube contracted slightly as the amount of functionalization increased from 0 to 15%. Furthermore, the energy as a function of strain for pure (10, 10) and (17, 0) nanotubes with 15% CH_3 functionalization was determined [173]. There was little difference between the functionalized and bare nanotube curves for the (10, 10) nanotube. However, functionalization reduced the modulus of the (17, 0) nanotube relative to the pure (17, 0) nanotube by a slight amount. These predictions have not yet been quantitatively confirmed in experimental studies.

MD simulations with many-body empirical potentials by Frankland et al. were used to examine the mechanical properties of nanotube–polymer matrix composites with and without functionalization of the CNTs [174, 175]. No permanent stress transfer was observed in a nanotube/polymer composite held together by van der Waals bonds. This prediction agreed with experimental findings that CNT/polymer composites that used unmodified CNTs failed by nanotube pull out [133, 136, 137]. When the same system was considered with a chemically functionalized nanotube in the simulations, the material as a whole was found to have improved shear yield strength [175]. The simulations also indicated that shear strengths and critical lengths required for load transfer could be significantly enhanced or decreased through chemical functionalization of the CNT walls [174]. For instance, over an order of magnitude difference was predicted for cross links that involved less than 1% of the atoms in the CNT walls. Calculations further showed that the helical conformation of the polymer was as important as the binding energy between the nanotubes and the polymer [176].

MD simulations with bond order and LJ potentials were recently used by Hu and Sinnott to explore new ways of producing cross links between CNTs and polymer chains *in situ* through ion beam deposition [177, 178]. This computational study demonstrated that when reinforcing CNTs were embedded close to the surface in a CNT/polystyrene composite, polyatomic ion beam deposition of $C_3F_5^+$ ions at incident energies of 50–80 eV/ion induced the formation of cross-links between the otherwise pristine nanotubes and the polymer matrix without inducing severe damage to the nanotube structure, as shown in Fig. 17. However,

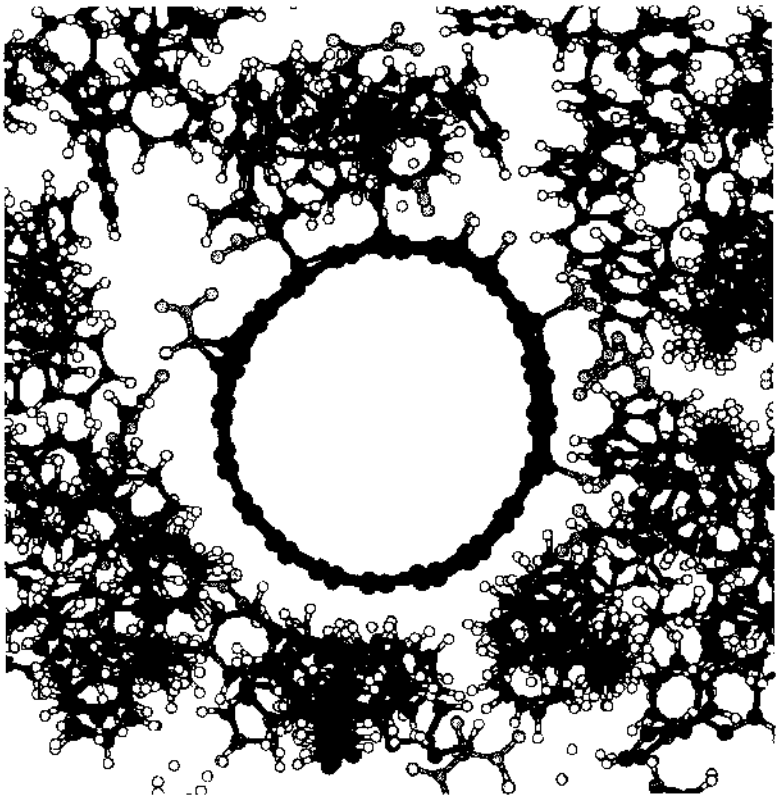


Figure 17. Snapshot from MD simulations of cross links that form between a CNT and a polymer matrix in a CNT/polystyrene composite during the deposition of a beam of $C_3F_5^+$ ions at incident energies of 80 eV. Courtesy of Yanhong Hu.

because of rapid energy decay within the substrate, this technique was limited to modification of the composite structure near the surface. This is in agreement with the conclusions of other studies on ion beam treatment of polymers that found that low-energy ion deposition (<100 eV) only modified the shallow regions of the polymer surfaces [179–181]. These studies also showed that if the ion energy is high (several hundred keV to a few MeV), the surface modification can reach much deeper regions (several micrometers), and the surface hardness of the treated polymer can be several times larger than that of steel as a result of extensive deposition-induced cross linking [182].

Chemical modification would also be expected to influence the electronic properties of the CNTs. TB calculations by Brenner et al. predicted that covalent chemical functionalization of the walls of (6, 0) CNTs with H_2CCH_2 units changed the electrical properties of the CNTs [81, 173]. For instance, at certain arrangements and concentrations of the H_2CCH_2 functional groups, the functionalized CNT behaved as a metal–semiconductor junction with a Schottky barrier of 0.44 eV [173]. In particular, the addition of H_2CCH_2 functional groups opened a band gap in the density of states of the CNT [173]. *Ab initio* calculations by Seifert et al. also examined the electronic properties of (10, 0) CNTs with various concentrations of attached fluorine atoms [183]. Three isoenergetic but very different structures with energy differences of about 0.1 eV/atom, were found for the same amount of absorbed fluorine. The F was found to make ethylene-like or polyacetylene-like units along the walls and the cross-section of the nanotube change from round to pentagonal for these different arrangements. Depending on the structure, the fluorinated (10, 0) was either semiconducting, metallic or insulating [183]. Additional calculations by Kudin et al. of fluorinated zigzag and armchair single-walled CNTs with C_2F stoichiometry found that armchair fluorinated CNTS were metallic and that fluorinated zigzag tubes were less stable than the fluorinated armchair tubes [184]. These results taken together indicate that a single CNT could have different arrangements of fluorine atoms, and hence different electronic properties, along the length of the nanotube.

These computational findings were recently followed by experiments that produced a quantum dot about 10 nm long that operated as a single electron transistor at room temperature and was created out of chemically modified single-walled CNT bundles [185]. In addition, nanometer-scale p-n [186] and p-n-p [187] junctions were produced from individual semiconducting single-walled CNTs by doping with polymer groups or K atoms. Chemical functionalization was also recently used to connect nanotubes to molecular switching and memory devices that might be the first step toward generating more complex nanometer-scale devices [188].

3.5. Carbon Nanotube Gas Sensors

In an interesting extension of the findings discussed in the last section, recent experiments and calculations indicated that the interaction of some molecules to the nanotube walls can have a significant effect on their electrical conductivity. For example, the conductivity of single-walled CNTs was originally predicted to change when exposed to NH_3 and NO_2 [189, 190], making it possible to use the CNTs as chemical sensors. Figure 18 shows the behavior of a single semiconducting single-walled CNT when it was exposed to the different gases. Part A shows an atomic force microscope image of the CNT between two metal contacts. This CNT was normally p-type and, when exposed to ammonia, the conductivity was depleted, as shown in part B of Fig. 18. In contrast, when this CNT was under a gate voltage of $V_g = +4$ V, which would normally make the CNT into an insulator, and exposed to NO_2 gas, a drastic increase in the conductivity was observed. This was believed to be caused by an excess of p-type carriers in the CNT. In later experiments, it was also shown by Collins et al. that a semiconducting single-walled CNT was sensitive to detection of O_2 gas [191] as well. In this study, it was shown that a normally semiconducting CNT was transformed into a metallic conductor when exposed to oxygen gas.

Since these original findings, there has been an intense debate about the fundamental mechanism for how the conductivity of a CNT changes when exposed to gases such as oxygen. This debate has ranged from whether a physisorbed molecule can drastically alter

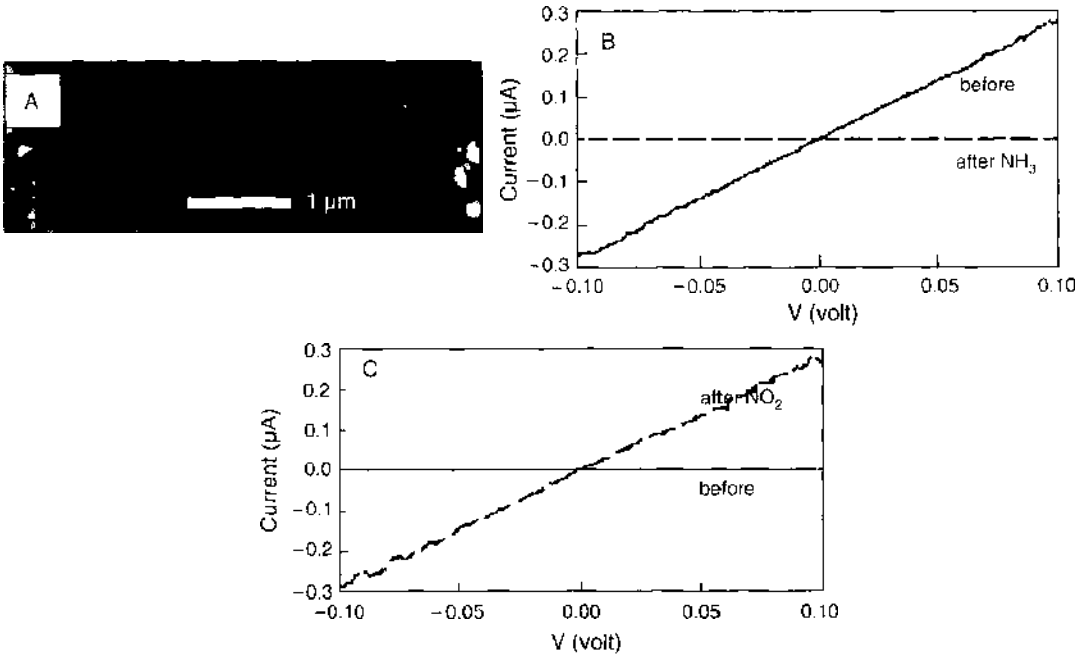


Figure 18. Experimental data. Part (a) shows an atomic force microscope image of single-walled CNT that is between two metal contacts. Part (b) shows that the normally p-type single-walled CNT looses carriers when exposed to NH_3. Under a gate voltage of +4 V, this CNT is normally an insulator. However, when exposed to NO_2, the conductivity drastically changes, as shown in part (c). Reprinted with permission from [189], J. Kong et al., *Science* 287, 622 (2000). © 2000, AAAS.

CNT conductivity to the effects of defects and other gases on the conductive properties. This section will limit itself to a discussion of the use of single-walled CNTs as chemical sensors for NO_2, NH_3, and O_2. There has been extensive research into the effects of defects on gas adsorption, and some of these are discussed below. There has also been extensive research into the detection of other gases and biological substances, and on the effect of chemical functionalization of the CNTs in increasing their chemical detection properties. To fully discuss all the current, ongoing research on the use of CNTs as chemical sensors is beyond the scope of this section.

Over the last 4 years, there has been a very active debate that has centered on what causes the conductivity change shown in Fig. 18. The initial claim was that the conductance change of the semiconducting single-walled CNTs studied in Refs. [189–191] was through direct interaction of the O_2 molecule with pristine or defected nanotubes. This was argued both theoretically in *ab initio* calculations [192–194] and experimentally [191] over the course of the last few years. This claim was disputed in a variety of theoretical [195–200] and experimental [197, 201–206] studies that found that direct contact between the O_2 molecule with the CNT did not cause the previously predicted conductivity change.

Experiments by Avorous and coworkers [205, 207] studying nanotube field effect transistors made from single-walled CNTs found that when the transistors were exposed to oxygen, the conductivity changed as previously predicted. However, when the metal/CNT contacts within the transistor were protected from the O_2 gas, no measurable change in conductance was observed. This showed that direct contact between O_2 gas molecules and the CNT walls did not play a significant role in changing the conductivity of the CNT. Rather, it was hypothesized that the conductivity change was a result of barrier height changes at the metal/CNT contact. In another experimental study, Goldoni et al. [206] showed that previous measurements of conductance changes in single-walled CNT bundles on exposure to O_2 were a result of contaminants, such as Na and Ni, left over from the growth and cleaning process. Once the bundle went through annealing cycles that removed the Na contaminants, there was no measurable conductance change of the CNT bundle when it was exposed to oxygen molecules. It should be noted that charge transfer from O_2 molecules to the CNTs in the presence of a metal contaminant was noted as a possible mechanism by Collins et al. [191].

However, Goldoni et al. did not find contaminants to be responsible for the CNT conductivity changes caused by NH_3 and NO_2 [206]. In fact, Goldini et al. saw a notable change of the C 1s peak when the annealed CNT bundle was exposed to NO_2 and NH_3 gas. In contrast, Bradley et al. reported that single-walled CNTs that were annealed and left under vacuum did not change conductivity when NH_3 was reintroduced [208]. They also studied the detection of ammonia in water and found that there was a strong signal when ammonia was introduced in an aqueous or a humid environment. They then concluded that water vapor under ambient conditions must assist in the conductivity change of the single-walled CNTs exposed to NH_3. This finding is consistent with the original predictions of Kong et al. [189], who found that the CNT conductivity change caused by NH_3 was most likely an indirect process because strong binding between the two was not found in DFT calculations. A later study by Bradley et al. [209] examined passivation of the metal/CNT contacts, as was done in the O_2 experiments in Refs. [207, 205]. In contrast to the findings for oxygen, the authors found that the conductivity only changed by a few percent and concluded that the interaction causing the conductivity change was in the CNT region rather than the contact region.

Finally, all experiments and theoretical studies appear to agree that NO_2 changes the conductivity of the single-walled CNTs. The NO_2 molecule has the strongest interaction with single-walled CNTs, and it is believed that it changes the conductivity through charge transferring from the CNT to the NO_2 molecule. There have been some assertions that the conductivity change is caused by chemisorption or other mechanisms [210], but the ideal NO_2 physisorbtion studies appear to give the clearest evidence that physisorption changes the conductivity of the CNT in this case.

It is not surprising that there is so much more disagreement between different theoretical and different experimental studies of gas interactions with CNTs. Experimentally, it is nontrivial to rule out contaminating effects. Computationally, modeling systems where dispersion interactions are important is quite complicated. To truly account for dispersion interactions, which are long-range attractive forces, it is often necessary to use very expensive wavefunction-based methods that include correlation effects. The HF method alone is not sufficient, and more computationally expensive methods such as higher-order HF-MP are required [211]. Exchange-correlation functionals in DFT are not parameterized to handle long-range dispersion interactions. LDA tends to overbind the molecule to the substrate, and each GGA parameterization responds differently, depending on the molecule used and on the methodology details. For example, one flavor of GGA tends to be slightly underbinding, and the calculated bond distances are too long [211], whereas another flavor calculates accurate bond distances and energies [212]. Although DFT does not capture all aspects of dispersion, it still can be and has been used with care to model systems in which dispersion is important.

The complete details of the computational methods used in Ref. [189] were presented in subsequent papers [190, 213]. The authors used DFT with the LDA exchange correlation potential to model the interaction of NH_3 and NO_2 with the (10, 0) semiconducting single-walled CNT. A strong interaction was found between NO_2 and the CNT, as is expected, but no binding affinity was found between NH_3 and the CNT. NH_3 was found to interact very weakly with the CNT and, as mentioned above, Kong et al. proposed that this gas molecule must change the conductivity of the CNT through an indirect mechanism that may not even be related to NH_3/CNT interactions. Subsequent studies by other authors who were also using DFT methods reported a weak interaction of NH_3 with semiconducting CNTs [193, 214].

The binding energy of NO_2 to a semiconducting CNT was found to be strong with 0.1 electron charge transfer between the CNT and molecule [189]. This strong interaction was hypothesized to be a means of doping the semiconducting CNTs. The calculations also showed that the physisorbed NO_2 changed the conductivity of the tube by taking electrons away from the CNT and adding a carrier state near the top of the CNT valence band [189, 190]. Peng et al. [190, 213] noted this argument by comparing the position of the molecule's lowest unoccupied molecular orbital to the valence band of the CNT. Later studies

clearly showed that there was a new state near the Fermi energy in the density of states plot [193, 214–216].

The fundamental issues of how CNT conductivity changes on exposure to gas molecules are by no means settled. However, it is clear that computational methods will play a major role in determining and understanding them.

3.6. Filling Carbon Nanotubes

About two decades ago, the computational study of molecular interactions in nanomaterials such as zeolites began to emerge. Zeolites are nanoporous metal oxide materials that can form a variety of shapes [217]. They are important molecular sieves and catalysts that can separate molecular mixtures through such processes as selective molecular transport and ion exchange [217, 218]. Computational studies were instrumental in identifying the mechanisms by which molecules were transported through zeolites, which included well-known mechanisms such as normal-mode diffusion, where molecules are able to pass each other [219]; single-file diffusion, where molecules are unable to pass each other [220, 221]; and a mode that is between these two that is called transition-mode diffusion [222].

The first study of molecular intercalation into small, opened, single-walled CNTs was a DFT calculation of the interaction of a HF molecule with the interior of a (6, 6) nanotube [223]. This calculation predicted that there was significant attraction between the molecule and the nanotube that was a strong driving force for intercalation. Another early theoretical study also found that it was energetically favorable for noble gas atoms to intercalate into opened CNTs [224]. Since then, several computational studies have also considered the intercalation of molecules and noble gas atoms, such as hydrogen, nitrogen, and helium, into opened single-walled CNTs or the interstitial sites between the nanotubes in a bundle [225–233]. Molecules, such as methane, and noble gas atoms, such as xenon and neon, were predicted to enter into opened nanotubes [233–235].

Computational studies were also used to study molecular transport through CNTs once the molecules entered through an opened end. Classical MD simulations with empirical, two-body potentials [236] and bond-order and LJ potentials [237] predicted that dynamical atomic and molecular flow was not likely to occur through opened nanotubes, and that even if dynamical motion started, it would rapidly stop because of the same strong attractions between the molecules and the nanotube interiors that were responsible for molecule intercalation into the CNTs. Molecular transport was rather, predicted to occur through molecular diffusion, as was the case in zeolites.

In one study by Lee and Sinnott, MD simulations using bond order and LJ potentials for methane and Coulombic and LJ potentials for oxygen, were used to compare the molecular transport of methane and oxygen molecules through opened, hydrogen-terminated CNTs [238]. A snapshot from these simulations is shown in Fig. 19. The authors predicted that these molecules showed some similarities and differences in the manner in which they flow through small-diameter CNTs. For both molecules, the transport mode followed by the molecules changed with time and could be ultimately divided into three different stages.

In the first stage, both molecules followed a transport mechanism that was intermediate between ballistic and normal-mode diffusion. However, when the molecules had flowed all the way down the CNT interior and were approaching the open exit, the significant attraction between the molecules and the CNT interiors was responsible for molecules reversing direction, which induced chaotic motion [239]. Thus, the second stage was characterized by chaotic flow that involves molecules moving forward from the gas reservoir and moving backward from the open end. In the third, near-steady-state stage, near-normal mode diffusion was followed by both sets of molecules. The differences between the methane and oxygen systems included the time needed for the molecular density change to level off within the CNT, which was lower for the oxygen systems than for the methane systems. In addition, the movement of the oxygen molecules during the second stage was more chaotic than the movement of the methane molecules. These differences were attributed to the dissimilar shapes of the molecules and the varying long-range forces that characterize their interaction with the nanotube wall.

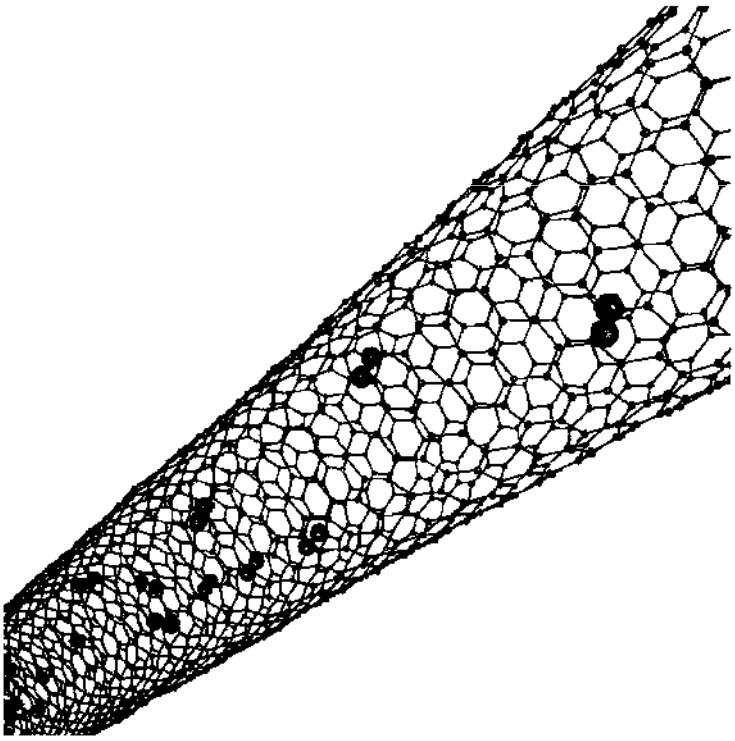


Figure 19. Snapshot from MD simulations of the molecular flow of O_2 through a (10, 10) single-walled CNT. Courtesy of Ki-Ho Lee.

Similar MD simulations by Mao and Sinnott [237] were further able to show how the CNT diameter affected the results. As the diameter increased, the diffusion coefficients of the methane molecules decreased because of increased interaction of the molecules with other gas molecules and decreased interaction of the molecules with the CNT walls. Transport that was neither normal-mode nor ballistic was very recently predicted to occur in the smallest-diameter CNTs (diameters were about 0.6 nm) by Lee and Sinnott [239]. This was predicted to occur because either the transport was transitional between normal-mode diffusion and the ballistic convection of continuum theory over the nanosecond timescales of the MD simulations, or molecular interactions were such that more scattering occurred in smaller-diameter CNTs than in larger-diameter CNTs.

The fact that at near steady-state conditions, spherical molecules such as methane flowed from a reservoir region through opened CNTs with diameters of about 1 nm via normal-mode diffusion [237, 239] agrees with what was predicted for the transport of methane through zeolites [219]. Classical MD simulations with bond order and LJ potentials by Mao et al. further predicted that nonspherical molecules, such as ethane and ethylene, followed transition-mode diffusion through most single-walled CNTs (with diameters of about 1 nm) [237, 240], which also agrees with predicted behavior of these molecules through zeolites [222]. However, the self-diffusivities and transport diffusivities of Ar and Ne in carbon nanotubes were recently predicted by Sholl and coworkers to be larger by, at most, three orders of magnitude relative to flow in zeolites [241, 242]. These equilibrium MD simulations using LJ potentials showed that the exceptionally high-flux light gas atoms and molecules through CNTs were a result of the smoothness of the CNT walls, which were very different from the corrugated structure of zeolite walls. When artificial scattering was included in the simulations, the diffusivities decreased significantly.

MD simulations by Zhang also examined the behavior of nonspherical molecules, such as linear decanes, inside CNTs [243]. The results indicated that the molecules preferred to form a single adsorption layer on the wall at most densities, and most of the transport took place in the center of the nanotube, where the molecules were not "trapped" by the wall [243]. Equilibrium MD simulations by Bhide et al. found that normal-pentane and iso-pentane molecules followed anomalous transport behavior midway between normal-mode diffusion

and ballistic motion when they flowed through opened (10, 10) CNTs [244]. Several arguments were proposed to explain this behavior, including changes in levitation, conformation, and potential energy. Supple and Quirke [245] also predicted ballistic motion for decane molecular flow through a rigid (13, 13) nanotube. The ballistic transport mode was attributed to the high driving force of convection. The fact that the walls were rigid may also have contributed to this behavior, however.

The transport of water molecules, which form strong intermolecular (hydrogen) bonds, in a small, (6, 6) CNT was recently studied in MD simulations with empirical potentials by Hummer et al. [246, 247]. In agreement with the other studies discussed earlier, the simulations predicted that water spontaneously entered the opened CNTs. However, some previously unreported behavior was predicted to occur. In particular, the water molecules formed a continuous, one-dimensional chain of individual molecules within the CNT that was transported through a pulse-like motion caused by the hydrogen bonding between water molecules. A continuous-time, random-walk model was developed that characterized the motion of the water [248]. The MD simulations were further able to characterize the kinetics of water motion into and out of the CNT [249]. Specifically, water molecules were added sequentially at one open end and removed sequentially from the other opened end. The one-dimensional chain broke when the water molecule exiting the CNT donated its hydrogen bond to the bulk water surrounding the CNT rather than to the chain inside the CNT.

MD simulations with LJ and Coulombic empirical potentials by Mashl et al. of water confinement to single-walled CNTs ranging in diameter from about 0.3 to about 1.8 nm predicted that the behavior of water changed significantly with nanotube diameter [250]. In agreement with the findings of Hummer et al. [246, 247], one-dimensional hydrogen-bonded chains of water were formed in the smallest-diameter CNTs. In the larger-diameter CNTs, more disordered arrangements of water were formed. However, water confined to CNTs with diameters of 0.86 nm underwent a transition to an icelike state with hydrogen-bonding behavior similar to that of bulk liquid water. In contrast, Noon et al. predicted the formation of helically ordered ice "sheets" inside (6, 6), (7, 7), (8, 8), (9, 9) and (10, 10) CNTs in MD simulations with nonreactive empirical potentials developed from *ab initio* methods that did not include polarization effects [251]. The water molecules were predicted to form hydrogen-bonded networks around the inner CNT walls. This finding could have been caused by all the atoms being constrained to be neutral in the simulations.

Unlike the finding of fast transport of light noble gases in CNTs [241, 242], water in CNTs was predicted to flow more slowly through the CNTs than in bulk water, most likely to because of the stabilization of hydrogen-bonded structures within the CNT [250]. Subsequent DFT-MD simulations by Mann et al. [252] and Dellago et al. [253] found that excess protons along the water chain were stabilized relative to bulk water, and that proton mobilities along the chain were thus much greater than in bulk water.

Experimental X-ray diffraction spectra and transmission electron microscopy images of water confined to the interiors of carbon nanotubes have recently been obtained [254, 255] that qualitatively agree with the predictions of the MD simulations. However, the nanotubes considered in the experiments had considerably larger diameters than the CNTs modeled in the simulations, so water meniscuses [255] or nanotubes of ice [254] were observed, rather than one-dimensional chains of water that were one molecule wide.

Computational studies have also considered the interaction of charged particles with the interiors of CNTs. For example, DFT calculations predicted that K atoms that intercalated into single-walled CNTs transferred charge to the nanotube walls leading to ionic cohesion [256]. Unlike the unusual transport behaviors discussed above for methane, light noble gas atoms, and water, the calculations predicted that the diffusive motion through the CNTs were similar to that of diffusion on a graphite surface. Shortly thereafter, experimental evidence showed that K and $FeCl_3$ could be intercalated into CNTs [257].

A combined computational and experimental study using DFT calculations and scanning transmission electron microscopy experiments showed that charged iodine molecules that were introduced in molten form formed a double-helix helical structure inside single-walled CNTs [258]. The calculations explained this behavior by the preference for the iodine to align with the carbon–carbon bonds in the CNT walls, which was responsible for the

helical shape. Subsequent MD simulations and DFT calculations predicted spiral transport for nonspherical organic (ethane and ethylene) molecules through (10, 10) CNTs at low molecule densities for similar reasons [259].

MD simulations with pairwise additive empirical potentials by Wilson and Madden [260] predict that KI crystal "fingers" could be preferentially grown into opened CNTs from the melt state. In particular, ions at the CNT openings arranged themselves into stable crystalline arrangements that then rapidly extended into the CNT interior and solidified, even while the KI ions surrounding the CNT exterior remained in the molten stage. Experimental studies indicated that nanotubes were easily filled with a variety of elements and compounds but only as long as the surface tensions of the filling material were less than about 100–200 mN/m [261].

Monte Carlo simulations by Ayappa were used to study the behavior of binary gas mixtures in CNTs [262]. Molecules with larger radii were able to adsorb to the walls more easily than smaller-diameter molecules because they were able to get closer to the CNT walls. Thus, molecular mixtures that contained molecules of differing sizes could be separated on exposure to carbon nanotubes [262]. However, for molecular mixtures in which the molecules were of similar sizes, such as methane/ethane mixtures, separation was not achieved [262]. This result was also seen in classical MD simulations by Mao and Sinnott that used combined bond-order and LJ potentials [263], which predicted significant separation of methane/butane mixtures but little separation of methane/ethane mixtures.

Aluru and coworkers have nonequilibrium MD simulations with LJ and Coulombic potentials to study the transport of ions mixed with water through CNTs that were either charged [264] or contained charged groups at the CNT opening [265]. When an opened CNT was placed in an aqueous NaCl solution and the walls were charged in the presence of a counterion, an applied external electric field caused electroosmotic flow through the CNT [264]. The adsorption behavior of the Na^+ or Cl^- counterions was predicted to be quite different from each other and to depend on several factors. One of these was the ion size, as the smaller Na^+ ions were able to get closer to and remain immobilized on the CNT surface more readily than the larger Cl^- ions. Other important factors were the strength of the electrostatic interactions between the water molecules and the ions and between the ions and the charged CNT surface, and the strength of the electric field.

In another set of MD simulations, a larger-diameter single-walled CNT that was not charged but had charged NH_3^+ groups attached to one open end and charged COO^- groups attached to the other opened end was placed in a solution of KCl in the presence of an external electric field [265]. The MD simulations predicted that the flow of Cl^- through the CNT was greater than the flow of K^+, which preferred to electrostatically attach to the COO^- groups. Hence, the functionalized system was able to function as a modulator of ionic conductivity.

A membrane of CNTs aligned within a polymer film has very recently been produced experimentally by Hinds et al. [266]. The CNTs had multiple shells with inner diameters of about 4 nm. The ends were opened and terminated with carboxyl groups. The transport of N_2 through the membrane was consistent with Knudson diffusion. In addition, the transport of $Ru(NH_3)_6^{3+}$ in aqueous solution through the membrane was substantial. This transport was significantly decreased by the presence of biotin coordinated with streptavidin, rather than carboxyl groups, at the opened ends of the CNTs. Thus, molecular transport across the membrane could be readily controlled through modifications to the chemical groups at the opened ends of the CNTs.

The dynamics of He and C_{60} motion in CNTs were studied theoretically using classical MD simulations and pairwide potentials by Tuzun et al. [267]. The results showed that flowing He can impart energy to the C_{60}. The resulting $He/C_{60}@$CNT system was proposed as a nanoscale machine [267]. The experimental synthesis of C_{60}-filled CNTs was subsequently achieved by annealing the CNTs and C_{60} molecules together in sealed containers [268–270]. The resulting structures were termed "peapods" for their distinctive structure, as shown in Fig. 20. The electronic structure of $C_{60}@$(10, 10) CNTs were calculated, and the structure was predicted to be a metallic conductor [271]. Indeed, it is the potential

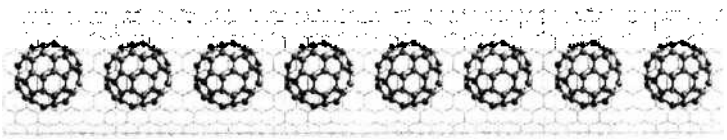


Figure 20. A schematic of a "peapod" structure. The C_{60} molecules are shown in orange, and the (10, 10) CNT is shown in blue.

unique electronic properties of these peapod-in-nanotube structures that are sparking the most intense interest [272].

These peapod structures were been proposed as unique, nanometer-scale memory devices [273]. Unlike regular peopods, which were formed in micrometer-scale CNTs, the buckyshuttle device proposed by Tomanek and Iijima [273] would only be a few nanometers long, as shown in Fig. 21. The C_{60} would further contain a K atom so that the "pea" in the pod would contain an overall charge. The entire system would operate as a memory device by movement of the $K@C_{60}$ molecule from one end of the capped CNT to the other, as illustrated in Fig. 21. The calculations used to explore the feasibility of this system used LCAO in combination with long-range van der Waals interactions.

Filling CNTs with molecules or C_{60} was also shown to affect the mechanical properties of the CNT [274–276]. In particular, classical MD simulations using bond order and LJ potentials were used by Sinnott and coworkers to study the effect of filling 10-nm-long, (10, 10) single-walled CNTs with CH_4, Ne, and C_{60} on their compressibility [275]. After a critical filling density was reached, filling the CNT increased the force at which the CNTs buckled, whereas the slope of the force versus strain plots were constant, indicating that the stiffness or Young's modulus was not affected by filling. However, filling the nanotubes significantly

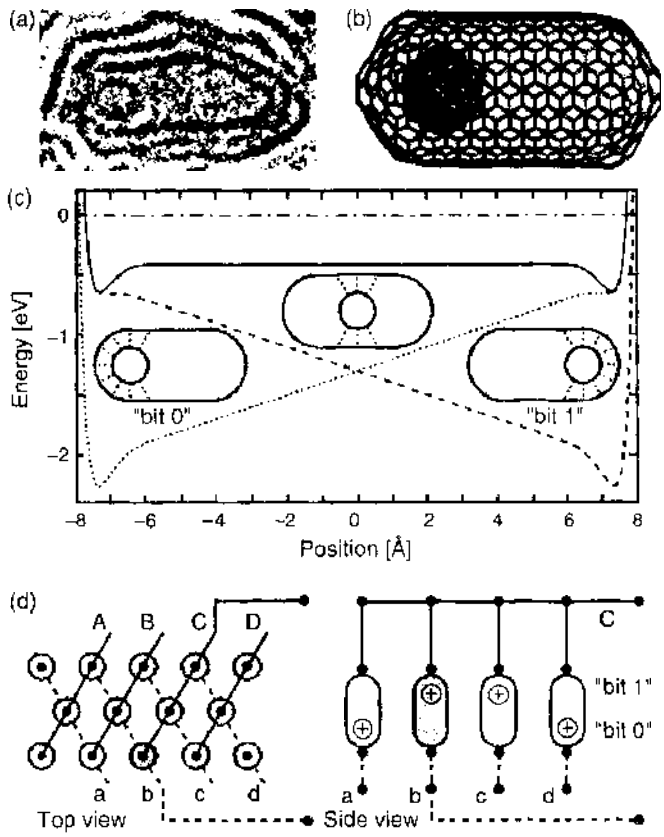


Figure 21. A small peapod "bucky-shuttle" system is shown in a transmission electron microscope image in part (a). A computational representation of the system seen in (a) is displayed in part (b). Part (c) illustrates the possible bit configurations and the energetics of the C_{60} molecule in the CNT. Also shown in part (c) are the energetics of the structure under bias. Part (d) shows a possible configuration for high-density memory. Figure reprinted with permission from [273]. Y. Kwon et al., *Phys. Rev. Lett.* 82, 1470 (1999). © 1999, American Physical Society.

decreased the effect of temperature on the buckling force, which decreased appreciably at high temperatures when the CNTs were empty but changed only a small amount or were unchanged when the CNTs were empty. A similar MD simulation study by Farajian and Mikami of the deformation of empty CNTs and peapods through compression, tension, and torsion [274] found peapods to be slightly softer, or more easily deformable, than the empty CNTs, possibly because the fullerenes in the peapods were separated by 1.0–1.5 nm, rather than by the 0.35-nm separation used in Ref. [275]. MD simulations by Sinnott and coworkers of CNTs filled with flexible molecules, such as n-butane [276] predicted that these systems would show a greater degree of yielding than CNTs filled with molecules that were constrained from yielding, such as fullerenes.

There have also been several recent studies of the mechanical properties of carbon and other nanowires that, unlike carbon nanotubes, are solid rather than hollow. Shenderova et al. [277] predicted that diamond nanorods were as energetically favorable as CNTs and that some had brittle fracture forces and zero strain stiffnesses that exceeded those of CNTs. Barnard and Snook predicted that CNTs are the most energetically favorable one-dimensional form of carbon, but that diamond nanowires could preferentially be formed for certain configurations [278]. Thermodynamic calculations also indicate that carbon nanowires may be preferentially formed inside carbon nanotubes [279], thus motivating the study of the mechanical properties of these systems.

Danailov et al. [280] used classical MD simulations with bond order and EAM potentials coupled through pairwise potentials to study the bending of CNTs containing metal nanowires. They found that the metal nanowire suppressed the buckling instability of the CNTs and damped oscillations in a manner not predicted for multiwalled CNTs. In contrast, Sinnott and coworkers [276] found no significant differences in the mechanical responses of CNTs filled with carbon nanowires versus nanotubes (in double-walled CNTs) in MD simulations. Rather, they found that CNTs filled with solid nanowires or hollow carbon nanotubes both have similar higher buckling forces than empty nanotubes or nanotubes filled with molecules. However, CNTs filled with stronger, more brittle diamond nanorods such as those discussed in Ref. [277] may display different responses.

3.7. Nanoindentation

Proximal probe microscopes, such as scanning tunneling microscopes and atomic force microscopes, have achieved wide utility for imaging and exploring surface properties since their discovery in the early 1980s [281, 282]. In these microscopes, a rigid cantilever with a sharp tip on the end is in electrical or physical contact with the surface. As the tip is indented or dragged across the surface, the cantilever is deflected as a result of changes in the surface structure, and images or other information about the surface structure is obtained.

When a tip touches a sample, the first interaction between the two is usually attractive. As the tip presses closer, the tip and the surface will usually experience a repulsive interaction. However, if the tip continues to press against the sample, the two might adhere to each other, through, for example, the formation of covalent or metallic bonds between the tip and the surface. Under ambient conditions, the presence of water or other molecules on the surface can influence tip–surface interactions [283, 284].

Because of their high Young's modulus, nanometer-scale diameters, and ability to conduct electrically, CNTs were proposed by Smalley and coworkers as ideal proximal probe microscope tips [285]. CNT proximal probe tips were formed either by gluing the CNTs to the cantilever [285] or by growing them there directly through a CVD process [286]. Such tips could be used to image fine details on surfaces [287–289] and, when a current was running between the tip and sample, to etch the sample [290]. One of the problems with these systems was thermal vibrations that degraded the performance of the tips. This was recently overcome by coating the cantilever and CNT tips with a polymer coating and then wearing away the coating to expose a short segment of the CNT at the very end [291].

Computational studies were used to determine the atomic scale mechanisms that occur when the CNT comes into contact with a surface. When a single-walled CNT indented a hydrogen-terminated diamond surface in MD simulations with a bond-order potential,

the first response was the deformation of the tube cap from a convex to a concave structure [292, 293], and then the CNT buckled [292, 294]. If the CNT was further pushed against the surface, the end of the CNT slipped on the surface [294]. The MD simulations showed that these deformations occurred elastically, regardless of the rigidity of the surface being indented in a manner analogous to what was predicted by CNT deformation [107]. Similar atomic scale mechanisms were predicted to occur during indentation with a double-walled CNT and a bundle of single-walled CNTs [295]. CNT proximal probe tips were further shown to readily etch or penetrate the Si(001) surface during indentation even without the presence of a current between the tip and the surface [296]. When CNT tips were used to indent Langmuir–Blodget films [297], the simulations predicted that the CNT would be able to deform and push aside the hydrocarbon chains.

The first computational study of metal tips indenting metal surfaces by Landman et al. [298] elucidated the nature of the adhesive interactions between metal surfaces and single metal asperities. This study used MD simulations with EAM potentials to characterize both the tip atoms and the atoms at the surface. Because clean metal surfaces have very high surface energies, they were found to be very attracted to the metal tip indenting the sample. In fact, the attraction was predicted to be so strong, surface atoms "jumped" upward to wet the tip when the tip got close enough for the surface to interact with it [298], as shown in Fig. 22a. This phenomena was termed jump-to-contact (JC) and was confirmed experimentally by several studies [298–302], using the atomic force microscopes.

If the indenting tip was pressed against the surface, the simulations [298] showed strong adhesion between the two, and the tip was flattened. This increased the contact area between the two such that when the tip was withdrawn, it resisted and a "neck" was formed between the tip and the surface, as illustrated in Figure 22b.

These phenomena were seen in several other early MD simulations using similar potentials [303–305]. Although agreeing with the initial predictions of Landman et al., these simulations also provided new insights. For example, the simulations showed that if the temperature of the system was increased enough to melt the topmost surface layers [304], the JC phenomena was enhanced and occurred at larger tip–surface separations. In addition, when the tip was significantly more stiff than the surface, the indentation caused a pile-up of surface atoms around the tip to relieve the stresses caused by the indentation [298, 303]. MD simulations by Belak et al. using perfectly rigid tips [306, 307] (i.e., tips that were not allowed to evolve in time in the simulations) showed how the surface yielded plastically after its elastic threshold was exceeded. Rigid indenter simulations are analogous to experiments that use surface passivation to prevent JC between the tip and the surface (see, e.g., Ref. [308]). The atomic scale mechanism by which the surface yielded included "popping" atoms out onto the surface under the indenting tip to produce a pileup of atoms. This prediction was subsequently verified by experimental images of "piles" near indentation sites [309].

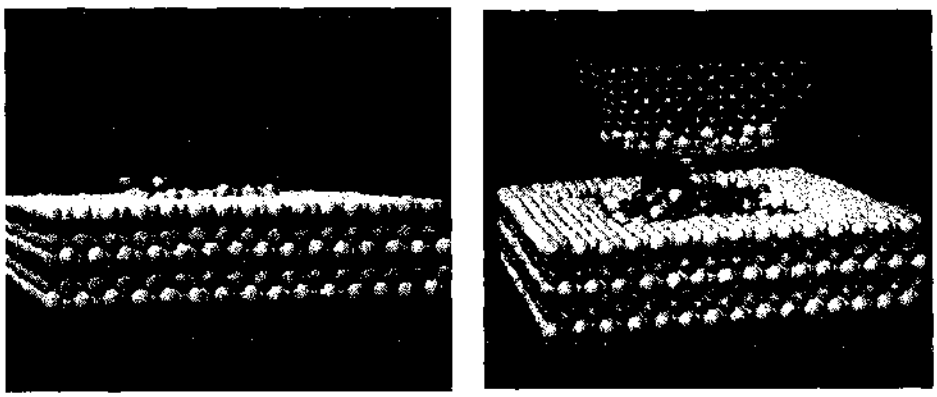


Figure 22. Snapshots from MD simulations of the indention of Ni tip into a Au substrate. In part (a) it can be seen that the Au substrate swells and jumps to contact the Ni tip. Part (b) shows the extensive neck that forms after the Ni tip has been embedded into the Au substrate. Reprinted with permission from [298], U. Landman et al., *Science* 248, 454 (1990). © 1990, AAAS.

Variations in the speed of indentation from 1 m/s to 100 m/s [307] showed that point defects created as a result of the indentation relaxed by moving through the surface at a slower speed. At the higher indentation speed, however, there was no time for the point defects to relax and move away from the indentation area. There was, therefore, interest in modeling the indentation of surfaces in a manner that was independent of indentation speed. In one such study [310], rather than use MD simulations, the indenter was pushed against the substrate, and then the system was allowed to relax using standard force minimization methods. When the system had fully relaxed, the tip was moved closer and the process was repeated. The indentation created several dislocations in the surface. If the tip was pulled back after indenting less than a critical value, the atoms in the dislocations slid back to their original positions and the surface was healed. However, if the tip was indented past the critical depth, more dislocations were created that interfered with the healing process, resulting in a surface crater.

Proximal probe tip indentation of covalently or ionically bound ceramic material surfaces would be expected to eventually cause brittle fracture. MD simulations with bond-order potentials of the indentation of Si surfaces by small Si indenters [311], revealed the rate-dependent and temperature-dependent yield strength of the substrate. The simulations predicted that crystalline Si could be converted to an amorphous phase beneath the tip under certain conditions. These results agree with the outcomes of scratching experiments [312] that showed that amorphous silicon emerged from room-temperature scratching of pure silicon. Atomistic simulations also predicted the atomic-scale mechanisms by which ceramic tips fracture following strong adhesion in ionic (CaF_2) and covalent (Si) systems [313, 305]. In contrast, the atomic-scale mechanical hysteresis experienced by a tip indenting a Si(100) was determined from DFT calculations [314], which predicted that at low speeds it should be possible to cycle repeatedly between two buckled configurations of the surface without adhesion.

MD simulations by Harrison and coworkers with bond-order potentials of diamond tips indenting hydrogen-terminated diamond surfaces [315, 316] or amorphous carbon films [317] revealed much about the nature of these interactions. The adhesion between hydrogen-terminated diamond tips and diamond surfaces was shown to be load dependent, with elastic indents occurring at low loads and adhesion occurring at loads above a certain cutoff point. Specifically, the tip goes through shear and twist deformations at low loads that changed to plastic deformation and adhesion with the surface at high loads. However, interestingly, when a diamond surface was compressed against another diamond surface, even if it is unpassivated with hydrogen, no adhesion was predicted to occur. When the diamond surface was covered with a thin amorphous carbon film [317], the tip easily penetrated the film, which "healed" easily when the tip was retracted so that no crater or other evidence of the indentation was left behind.

Thus, although nanoindentation was first discovered and explored experimentally, computational studies have made significant contributions to our understanding of the physics behind nanoindentation processes.

4. SUMMARY AND FUTURE OUTLOOKS

The proceeding sections have shown the pivotal role of computational modeling in the predictive study of nanometer-scale systems and materials. Although not every initial computational prediction was necessarily completely correct, each led the way for subsequent experiments and greatly contributed to the understanding of nanomaterials.

The future of nanomaterials contains many challenges and opportunities. The challenges include the need to improve knowledge and understanding of strengths and weaknesses of computational methods so that they can be more efficiently applied. An additional challenge is the improvement of computational methods and the expansion of existing methods to larger systems and longer time scales. Opportunities include the fact that in the nanorealm, experimental and computational results are readily and directly comparable. More direct coupling of computational and experimental methods promises to provide the most complete picture of any nanomaterials system.

ACKNOWLEDGMENTS

We are grateful to the numerous authors and publishers who granted us permission to reproduce their work here. We also thank the Army Research Office, the National Science Foundation (CHE-0200838), and the National Science Foundation supported Network for Computational Nanotechnology (EEC-0228390) for support. Yanhong Hu, Ki-Ho Lee, and Inkook Jang are thanked for their assistance in preparing this article.

REFERENCES

1. Nanotechnology: The Exhibition (Miami of Ohio University). Accessed October 27, 2004. http://www.lna.muohio.edu/amu/nano/bowls.html.
2. M. Born and J. R. Oppenheimer, *Ann. Phys.* 84, 457 (1927).
3. E. Schrödinger, *Ann. Phys.* 79, 361 (1926).
4. J. J. Sakurai, "Modern Quantum Mechanics," Addison Wesley, New York, 1995.
5. D. J. Griffiths, "Introduction to Quantum Mechanics," Prentice Hall, 1994.
6. K. Ohno, K. Esfarjani, and Y. Kawazoe, "Computational materials science," Springer, Berlin, 1999.
7. A. Szabo and N. S. Ostlund, "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory," Dover, Mineola, NY, 1996.
8. C. C. J. Roothaan, *Rev. Modern Phys.* 23, 69 (1951).
9. C. Pisani and R. Dovesi, *Int. J. Quantum. Chem.* 17, 501 (1980).
10. V. R. Sanders, *Faraday Symp. Chem. Soc.* 19, 79 (1984).
11. C. Pisani, R. Dovesi, and C. Roetti, "Hartree-Fock *ab-initio* of Crystalline Systems," Springer-Verlag, Heidelberg, 1988.
12. C. Pisani, "Quantum-Mechanical *ab-initio* Calculation of the Properties of Crystalline Materials," Springer-Verlag, Heidelberg, 1996.
13. P. Hohenberg and W. Kohn, *Phys. Rev. B* 136, B864 (1964).
14. W. Kohn and L. J. Sham, *Phys. Rev.* 140, 1133 (1965).
15. R. G. Parr and W. Yang, "Density Functional Theory of Atoms and Molecules," Oxford Univ. Press, New York, 1989.
16. D. C. Allan and M. P. Teter, *Phys. Rev. Lett.* 59, 1136 (1987).
17. P. J. H. Denteneer and W. Vanhaeringen, *J. Phys. C-Solid State Phys.* 18, 4127 (1985).
18. J. Ihm, A. Zunger, and M. L. Cohen, *J. Phys. C-Solid State Phys.* 12, 4409 (1979).
19. J. S. Lin, A. Qteish, M. C. Payne, and V. Heine, *Phys. Rev. B* 47, 4174 (1993).
20. N. Troullier and J. L. Martins, *Phys. Rev. B* 46, 1754 (1992).
21. D. Vanderbilt, *Phys. Rev. B* 41, 7892 (1990).
22. W. Koch and M. C. Holthausen, "A Chemist's Guide to Density Functional Theory," Wiley-VCH, Weinheim, 2000.
23. M. W. Finnis, *J. Phys. Condens. Matter* 8, 5811 (1996).
24. P. Walsh, R. K. Kalia, A. Nakano, P. Vashishta, and S. Saini, *Appl. Phys. Lett.* 77, 4332 (2000).
25. M. E. Bachlechner, A. Omeltchenko, A. Nakano, R. K. Kalia, P. Vashishta, I. Ebbsjö, and A. Madhukar, *Phys. Rev. Lett.* 84, 322 (2000).
26. D. W. Brenner, *Phys. Rev. Lett.* 63, 1022 (1989).
27. M. I. Baskes, J. S. Nelson, and A. F. Wright, *Phys. Rev. B* 40, 6085 (1989).
28. B.-J. Lee and M. I. Baskes, *Phys. Rev. B* 62, 8564 (2000).
29. T. Ohira, Y. Inoue, K. Murata, and J. Murayama, *Appl. Surf. Sci.* 171, 175 (2001).
30. F. H. Streitz and J. W. Mintmire, *Phys. Rev. B* 50, 11996 (1994).
31. A. Yasukawa, JSME *Intl. J. A* 39, 313 (1996).
32. T. Iwasaki and H. Miura, *J. Mater. Res.* 16, 1789 (2001).
33. G. C. Abell, *Phys. Rev. B* 31, 6184 (1985).
34. J. Tersoff, *Phys. Rev. B* 37, 6991 (1988).
35. J. Tersoff, *Phys. Rev. B* 39, 5566 (1989).
36. D. W. Brenner, *Phys. Rev. B* 42, 9458 (1990).
37. D. W. Brenner, *Phys. Status Solidi B* 217, 23 (2000).
38. D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, *J. Phys. Condensed Matter* 14, 783 (2002).
39. A. J. Dyson and P. V. Smith, *Surf. Sci.* 355, 140 (1996).
40. B. Ni, K.-H. Lee, and S. B. Sinnott, *J. Phys. C: Condensed Matter* 16, 7261 (2004).
41. J. Tanaka, C. F. Abrams, and D. B. Graves, *Nucl. Inst. Meth. B* 18, 938 (2000).
42. S. B. Sinnott, O. A. Shenderova, C. T. White, and D. W. Brenner, *Carbon* 36, 1 (1998).
43. S. J. Stuart, A. B. Tutein, and J. A. Harrison, *J. Chem. Phys.* 112, 6472 (2000).
44. S. M. Foiles, *Phys. Rev. B* 32, 3409 (1985).
45. M. S. Daw, *Phys. Rev. B* 39, 7441 (1989).
46. T. J. Raeker and A. E. Depristo, *Inter. Rev. Phys. Chem.* 10, 1 (1991).
47. R. W. Smith and G. S. Was, *Phys. Rev. B* 40, 10322 (1989).
48. R. Pasianot, D. Farkas, and E. J. Savino, *Phys. Rev. B* 43, 6952 (1991).

49. R. Pasianot and E. J. Savino, *Phys. Rev. B* 45, 12704 (1992).
50. A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.* 95, 3358 (1991).
51. D. Frenkel and B. Smit, "Understanding Molecular Simulation: From Algorithms to Applications." Academic Press, San Diego, CA, 1996.
52. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipies in fortran," 2nd edn. Cambridge Univ. Press, Cambridge, 1992.
53. Mathworld, http://mathworld.wolfram.com/. Accessed October 27, 2004.
54. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
55. M. Payne, R. Teter, D. Allan, T. Arias, and J. Joannopulos, *Rev. Mod. Phys.* 64, 1045 (1992).
56. J. B. Adams, Z. Y. Wang, and Y. H. Li, *Thin Solid Films* 365, 201 (2000).
57. F. Tassone, F. Mauri, and R. Car, *Phys. Rev. B* 50, 10561 (1994).
58. A. F. Voter, *Phys. Rev. Lett.* 78, 3908 (1997).
59. J.-C. Wang, S. Pal, and K. A. Fichthorn, *Phys. Rev. B* 63, 085403 (2001).
60. F. Montalenti, M. R. Sørensen, and A. F. Voter, *Phys. Rev. Lett.* 87, 126101 (2001).
61. J. Jacobsen, B. H. Cooper, and J. P. Sethna, *Phys. Rev. B* 58, 15847 (1998).
62. S. Iijima, *Nature* 354, 56 (1991).
63. J. W. Mintmire, B. I. Dunlap, and C. T. White, *Phys. Rev. Lett.* 68, 631 (1992).
64. R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Appl. Phys. Lett.* 60, 2204 (1992).
65. N. Hamada, S.-I. Sawada, and A. Oshiyama, *Phys. Rev. Lett.* 68, 1579 (1992).
66. J. W. Mintmire, B. I. Dunlap, and C. T. White, personal communication.
67. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic Press, San Diego, CA, 1996.
68. S. B. Sinnott and R. Andrews, *Crit. Rev. Solid State Mater. Sci.* 26, 145 (2001).
69. D. H. Robertson, D. W. Brenner, and J. W. Mintmire, *Phys. Rev. B* 45, 12592 (1992).
70. C. T. White, D. H. Robertson, and J. W. Mintmire, *Phys. Rev. B* 47, 5485 (1993).
71. R. E. Peierls, "Quantum Theory of Solids." Oxford University Press, New York, 1955, pg. 108.
72. W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. B* 22, 2099 (1980).
73. W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* 42, 1698 (1979).
74. S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, and C. Dekker, *Nature* 386, 474 (1997).
75. J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, *Nature* 391, 59 (1998).
76. T. W. Odom, J.-L. Huang, P. Kim, M. Ouyang, and C. M. Lieber, *J. Mater. Res.* 13, 2380 (1998).
77. A. F. Morpurgo, J. Kong, C. M. Marcus, and H. Dai, *Science* 286, 263 (1999).
78. S. Rosenblatt, Y. Yaish, J. Park, J. Gore, V. Sazonova, and P. L. McEuen, *Nano Lett.* 2, 869 (2002).
79. M. Radosavljevi, J. Appenzeller, P. Avouris, and J. Knoch, *Appl. Phys. Lett.* 84, 3693 (2004).
80. R. Heyd, A. Charlier, and E. McRae, *Phys. Rev. B* 55, 6820 (1997).
81. D. W. Brenner, J. D. Schall, J. P. Mewkill, O. A. Shenderova, and S. B. Sinnott, *J. Bi. Interpl. Soc.* 51, 137 (1998).
82. M. S. Strano, C. A. Dyke, M. L. Usrey, P. W. Barone, M. J. Allen, H. Shan, C. Kittrell, R. H. Hauge, J. M. Tour, and R. E. Smalley, *Science* 301, 1519 (2003).
83. Z. Chen, X. Du, M.-H. Du, C. D. Rancken, H.-P. Cheng, and A. G. Rinzler, *Nano Lett.* 3, 1245 (2003).
84. S. Iijima, T. Ichihashi, and Y. Ando, *Nature* 356, 776 (1992).
85. B. I. Dunlap, *Phys. Rev. B* 49, 5643 (1994).
86. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 53, 2044 (1996).
87. L. Chico, V. H. Crespi, L. X. Benedict, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* 76, 971 (1996).
88. J.-C. Charlier, T. W. Ebbesen, and P. Lambin, *Phys. Rev. B* 53, 11108 (1996).
89. M. Menon and D. Srivastava, *Phys. Rev. Lett.* 79, 4453 (1997).
90. M. Menon and D. Srivastava, *J. Mater. Res.* 13, 2357 (1998).
91. A. N. Andriotis, M. Menon, D. Srivastava, and L. Chernozatonskii, *Phys. Rev. Lett.* 87, 066802 (2001).
92. D. Zhou and S. Seraphin, *Chem. Phys. Lett.* 238, 286 (1995).
93. J. Li, C. Papadopoulos, and J. Xu, *Nature* 402, 253 (1999).
94. P. Nagy, R. Ehlich, L. P. Biro, and J. Gyulai, *Appl. Phys. A* 70, 481 (2000).
95. B. C. Satishkumar, P. J. Thomas, A. Govindaraj, and C. N. R. Rao, *Appl. Phys. Lett.* 77, 2530 (2000).
96. M. Terrones, F. Banhart, N. Grobert, J.-C. Charlier, H. Terrones, and P. M. Ajayan, *Phys. Rev. Lett.* 89, 75505 (2002).
97. D. Srivastava, M. Menon, and P. M. Ajayan, *J. Nanoparticle Res.* 5, 395 (2003).
98. C. Papadopoulos, A. Rakitin, J. Li, A. S. Vedeneev, and J. M. Xu, *Phys. Rev. Lett.* 85, 3476 (2000).
99. L. Jang, S. B. Sinnott, D. Danailov, and P. Keblinski, *Nano Lett.* 4, 109 (2004).
100. F. Cleri, P. Keblinski, L. Jang, and S. B. Sinnott, *Phys. Rev. B* 69, 121412 (2004).
101. E. Hernandez, C. Goze, P. Bernier, and A. Rubio, *Phys. Rev. Lett.* 80, 4502 (1998).
102. J. P. Lu, *Phys. Rev. Lett.* 79, 1297 (1997).
103. P. Zhang, P. E. Lammert, and V. H. Crespi, *Phys. Rev. Lett.* 81, 5346 (1998).
104. N. Yao and V. Lordi, *J. Appl. Phys.* 84, 1939 (1998).
105. G. V. Lier, C. V. Alsenoy, V. V. Doren, and P. Geerlings, *Chem. Phys. Lett.* 326, 181 (2000).
106. T. Ozaki, Y. Iwasa, and T. Mitani, *Phys. Rev. Lett.* 84, 1712 (2000).
107. B. I. Yakobson, C. J. Brabec, and J. Bernholc, *Phys. Rev. Lett.* 76, 2511 (1996).
108. S. Iijima, C. Brabec, A. Maiti, and J. Bernholc, *J. Chem. Phys.* 104, 2089 (1996).

109. C. F. Cornwell and L. T. Wille, *Solid State Comm.* 101, 555 (1997).
110. A. Maiti, *Chem. Phys. Lett.* 331, 21 (2000).
111. D. Srivastava, M. Menon, and K. Cho, *Phys. Rev. Lett.* 83, 2973 (1999).
112. M. B. Nardelli, B. I. Yakobson, and J. Bernholc, *Phys. Rev. B* 57, R4277 (1998).
113. B. I. Yakobson, *Appl. Phys. Lett.* 72, 918 (1998).
114. B. I. Yakobson, M. P. Campbell, C. J. Brabec, and J. Bernholc, *Comp. Mat. Sci.* 8, 341 (1997).
115. Z. Xin, Z. Jianjin, and O.-Y. Zhong-can, *Phys. Rev. B* 62, 13692 (2000).
116. A. G. Rinzler (private communication).
117. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996).
118. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14013 (1998).
119. W. Shen, B. Jiang, B. S. Han, and S.-S. Xie, *Phys. Rev. Lett.* 84, 3634 (2000).
120. M.-F. Yu, T. Kowalewski, and R. S. Ruoff, *Phys. Rev. Lett.* 85, 1456 (2000).
121. J. Tang, L.-C. Qin, T. Sasaki, M. Yudasaka, A. Matsushita, and S. Iijima, *Phys. Rev. Lett.* 85, 1887 (2000).
122. E. W. Wong, P. E. Sheehan, and C. M. Lieber, *Science* 277, 1971 (1997).
123. M. F. Yu, B. S. Files, S. Arepalli, and R. S. Ruoff, *Phys. Rev. Lett.* 84, 5552 (2000).
124. O. Lourie, and H. D. Wagner, *J. Mater. Res.* 13, 2418 (1998).
125. R. Gao, Z. L. Wang, Z. Bai, W. A. D. Heer, L. Dai, and M. Gao, *Phys. Rev. Lett.* 85, 622 (2000).
126. B. Z. Tang and H. Xu, *Macromolecules* 32, 2569 (1999).
127. P. M. Ajayan, O. Stephan, C. Colliex, and D. Trauth, *Science* 265, 1212 (1994).
128. S. Curran, A. P. Davey, J. Coleman, A. Dalton, B. McCarthy, S. Maier, A. Drury, D. Gray, M. Brennan, K. Ryder, M. L. D. L. Chapelle, C. Journet, P. Bernier, H. J. Byrne, D. Carroll, P. M. Ajayan, S. Lefrant, and W. Blau, *Synthetic Metals* 103, 2559 (1999).
129. C. Stephan, T. P. Nguyen, M. L. D. L. Chapelle, S. Lefrant, C. Journet, and P. Bernier, *Synthetic Metals* 108, 139 (2000).
130. M. S. P. Shaffer and A. H. Windle, *Adv. Mater.* 11, 937 (1999).
131. Z. Jin, X. Sun, G. Xu, S. H. Goh, and W. Ji, *Chem. Phys. Lett.* 318, 505 (2000).
132. L. S. Schadler, S. C. Giannaris, and P. M. Ajayan, *Appl. Phys. Lett.* 73, 3842 (1998).
133. D. Qian, E. C. Dickey, R. Andrews, and T. Randtell, *Appl. Phys. Lett.* 76, 2868 (2000).
134. H. D. Wagner, O. Lourie, Y. Feldman, and R. Tenne, *Appl. Phys. Lett.* 72, 188 (1998).
135. X. Gon, J. Liu, S. Baskaran, R. D. Voise, and J. S. Young, *Chem. Mater.* 12, 1049 (2000).
136. C. Bower, R. Rosen, L. Jin, J. Han, and O. Zhou, *Appl. Phys. Lett.* 74, 3317 (1999).
137. P. M. Ajayan, L. S. Schadler, C. Giannaris, and A. Rubio, *Adv. Mater.* 12, 750 (2000).
138. E. T. Thostenson, Z. Ren, and T.-W. Chou, *Composites Sci. Technol.* 61, 1899 (2001).
139. B. I. Yakobson and R. E. Smalley, *Am. Sci.* 85, 324 (1997).
140. P. M. Ajayan, *Chem. Rev.* 99, 1787 (1999).
141. E. V. Barrera, *JOM* 52, 38 (2000).
142. M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, "Carbon nanotubes: Synthesis, structure, properties, and applications." Springer, New York, 2000.
143. J. Liu, A. G. Rinzler, H. Dai, J. H. Hafner, R. K. Bradley, P. J. Boul, A. Lu, T. Iverson, K. Shelimov, C. B. Huffman, F. Rodriguez-Macias, Y.-S. Shon, T. R. Lee, D.T. Colbert, and R. E. Smalley, *Science* 280, 1253 (1998).
144. A. G. Rinzler, J. Liu, H. Dai, P. Nikolaev, C. B. Huffman, F. J. Rodriguez-Macias, P. J. Boul, A. H. Lu, D. Heymann, D. T. Colbert, R. S. Lee, J. E. Fisher, A. M. Rao, P. C. Eklund, and R. E. Smalley, *Appl. Phys. A* 67, 9 (1998).
145. S. S. Wong, E. Joselevich, A. T. Woolley, C. L. Cheung, and C. M. Lieber, *Nature* 394, 52 (1998).
146. S. S. Wong, A. T. Woolley, E. Joselevich, C. L. Cheung, and C. M. Lieber, *J. Am. Chem. Soc.* 120, 8557 (1998).
147. J. Chen, M. A. Hamon, H. Hu, Y. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon, *Science* 282, 95 (1998).
148. M. A. Hamon, J. Chen, H. Hu, Y. Chen, M. E. Itkis, A. M. Rao, P. C. Eklund, and R. C. Haddon, *Adv. Mater.* 11, 834 (1999).
149. Y. Chen, R. C. Haddon, S. Fang, A. M. Rao, P. C. Eklund, W. H. Lee, E. C. Dickey, E. A. Grulke, J. C. Pendergrass, A. Chavan, B. E. Haley, and R. E. Smalley, *J. Mater. Res.* 13, 2423 (1998).
150. A. Hamwi, H. Alvergnat, S. Bonnamy, and F. Beguin, *Carbon* 35, 723 (1997).
151. E. T. Mickelson, C. B. Huffman, A. G. Rinzler, R. E. Smalley, R. H. Hauge, and J. L. Margrave, *Chem. Phys. Lett.* 296, 188 (1998).
152. W. Zhao, C. Song, B. Zheng, J. Liu, and T. Viswanathan, *J. Phys. Chem. B* 106, 293 (2002).
153. C. W. Bauschlicher, *Chem. Phys. Lett.* 322, 237 (2000).
154. P. J. Boul, J. Liu, E. T. Mickelson, C. B. Huffman, L. M. Ericson, I. W. Chiang, K. A. Smith, D. T. Colbert, R. H. Hauge, J. L. Margrave, and R. E. Smalley, *Chem. Phys. Lett.* 310, 367 (1999).
155. R. J. Chen, Y. Zhang, D. Wang, and H. Dai, *J. Am. Chem. Soc.* 123, 3838 (2001).
156. S. A. Curran, P. M. Ajayan, W. J. Blau, D. L. Carroll, J. N. Coleman, A. B. Dalton, A. P. Davey, A. Drury, B. McCarthy, S. Maier, and A. Strevens, *Adv. Mater.* 10, 1091 (1998).
157. J. N. Coleman, A. B. Dalton, S. Curran, A. Rubio, A. P. Davey, A. Drury, B. McCarthy, B. Lahr, P. M. Ajayan, S. Roth, R. C. Barklie, and W. J. Blau, *Adv. Mater.* 12, 213 (2000).
158. M. J. O'Connell, P. Boul, L. M. Ericson, C. Huffman, Y. Wang, E. Haroz, C. Kuper, J. Tour, K. D. Ausman, and R. E. Smalley, *Chem. Phys. Lett.* 342, 265 (2001).
159. R. Bandyopadhyaya, E. Nativ-Roth, O. Regev, and R. Yerushalmi-Rozen, *Nano Lett.* 2, 25 (2002).

160. A. Star, J. F. Stoddart, D. Steuerman, M. Diehl, A. Boukai, E. W. Wong, X. Yang, S.-W. Chung, H. Choi, and J. R. Heath, *Angew. Chem. Int. Ed.* 40, 1721 (2001).
161. V. H. Crespi, N. G. Chopra, M. L. Cohen, A. Zettl, and S. G. Louie, *Phys. Rev. B* 54, 5927 (1996).
162. B. W. Smith and D. E. Luzzi, *AIP Conf. Proc.* 486, 360 (1999).
163. P. M. Ajayan, V. Ravikumar, and J.-C. Charlier, *Phys. Rev. Lett.* 81, 1437 (1998).
164. M. Terrones, H. Terrones, F. Banhart, J.-C. Charlier, and P. M. Ajayan, *Science* 288, 1226 (2000).
165. F. Banhart, *Nano Lett.* 1, 329 (2001).
166. B. Ni and S. B. Sinnott, *Phys. Rev. B* 61, R16343 (2000).
167. B. Ni and S. B. Sinnott, in "Applications of Accelerators in Research and Industry: Proceedings of the Sixteenth International Conference, Aip Conference Proceeding" (J. L. Duggan and I. L. Morgan, Eds.), Vol. 576, p. 959. American Institute of Physics Press, Melville, NY, 2001.
168. B. Ni, R. Andrews, D. Jacques, D. Qian, M. B. J. Wijesundara, Y. Choi, L. Hanley, and S. B. Sinnott, *J. Phys. Chem. B* 105, 12719 (2001).
169. M.-F. Yu, B. I. Yakobson, and R. S. Ruoff, *J. Phys. Chem. B* 104, 8764 (2000).
170. A. N. Kolmagorov and V. H. Crespi, *Phys. Rev. Lett.* 85, 4727 (2000).
171. D. Srivastava, D. W. Brenner, J. D. Schall, K. D. Ausman, M. Yu, and R. S. Ruoff, *J. Phys. Chem. B* 103, 4330 (1999).
172. A. Garg and S. B. Sinnott, *Chem. Phys. Lett.* 295, 273 (1998).
173. D. W. Brenner, O. A. Shenderova, D. A. Areshkin, J. D. Schall, and S.-J. V. Frankland, *CMES* 3, 643 (2002).
174. S. J. V. Frankland, A. Caglar, D. W. Brenner, and M. Griebel, *J. Phys. Chem. B* 106, 3046 (2002).
175. S. J. V. Frankland, A. Cagler, D. W. Brenner, and M. Griebel, *Mat. Res. Soc. Symp. Proc.* 633, A14.17.1 (2001).
176. V. Lordi and N. Yao, *J. Mater. Res.* 15, 2770 (2000).
177. Y. Hu, L. Jang, and S. B. Sinnott, *Composites Sci. Technol.* 63, 1663 (2003).
178. Y. Hu and S. B. Sinnott, *J. Mater. Chem.* 14, 719 (2004).
179. D. W. Brenner, O. A. Shendrova, and C. B. Parker, *Mater. Res. Soc. Symp. Proc.* 438, 491 (1997).
180. M. B. J. Wijesundara, L. Hanley, B. Ni, and S. B. Sinnott, *Proc. Natl. Acad. Sci. U.S.A.* 97, 23 (2000).
181. M. B. J. Wijesundara, Y. Ji, B. Ni, S. B. Sinnott, and L. Hanley, *J. Appl. Phys.* 88, 5004 (2000).
182. E. H. Lee, G. R. Rao, and L. K. Mansur, *Mater. Sci. Forum* 248, 135 (1997).
183. G. Seifert, T. Kohler, and T. Frauenheim, *Appl. Phys. Lett.* 77, 1313 (2000).
184. K. N. Kudin, H. F. Bettinger, and G. E. Scuseria, *Phys. Rev. B* 63, 045413 (2001).
185. J. B. Cui, M. Burghard, and K. Kern, *Nano Lett.* 2, 117 (2002).
186. C. Zhou, J. Kong, E. Yenilmez, and H. Dai, *Science* 290, 1552 (2000).
187. J. Kong, J. Cao, H. Dai, and E. Anderson, *Appl. Phys. Lett.* 80, 73 (2002).
188. J. L. Bahr, J. Yang, D. V. Kosynkin, M. J. Bronikowski, R. E. Smalley, and J. M. Tour, *J. Am. Chem. Soc.* 123, 6536 (2001).
189. J. Kong, N. R. Franklin, C. W. Zhou, M. G. Chapline, S. Peng, K. J. Cho, and H. J. Dai, *Science* 287, 622 (2000).
190. S. Peng and K. J. Cho, *Nanotechnology* 11, 57 (2000).
191. P. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, *Science* 287, 1801 (2000).
192. M. Grujicic, G. Cao, and R. Singh, *Appl. Surf. Sci.* 211, 166 (2003).
193. J. J. Zhao, A. Buldum, J. Han, and J. P. Lu, *Nanotechnology* 13, 195 (2002).
194. S. H. Jhi, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* 85, 1710 (2000).
195. S. Dag, O. Gulseren, T. Yildirim, and S. Ciraci, *Phys. Rev. B* 67, (2003).
196. P. Giannozzi, R. Car, and G. Scoles, *J. Chem. Phys.* 118, 1003 (2003).
197. H. Ulbricht, G. Moos, and T. Hertel, *Phys. Rev. B* 66, (2002).
198. D. C. Sorescu, K. D. Jordan, and P. Avouris, *J. Phys. Chem. B* 105, 11227 (2001).
199. S. Dag, O. Gulseren, and S. Ciraci, *Chem. Phys. Lett.* 380, 1 (2003).
200. A. Ricca, C. W. Bauschlicher, and A. Maiti, *Phys. Rev. B* 68, (2003).
201. R. Larciprete, A. Goldoni, and S. Lizzit, *Nucl. Instr. Meth. Phys. Res. B* 200, 5 (2003).
202. H. Ulbricht, G. Moos, and T. Hertel, *Surf. Sci.* 532, 852 (2003).
203. M. S. Strano, C. B. Huffman, V. C. Moore, M. J. O'Connell, E. H. Haroz, J. Hubbard, M. Miller, K. Rialon, C. Kittrell, S. Ramesh, R. H. Hauge, and R. E. Smalley, *J. Phys. Chem. B* 107, 6979 (2003).
204. S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and P. Avouris, *Phys. Rev. Lett.* 89, 106801 (2002).
205. V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, *Appl. Phys. Lett.* 80, 2773 (2002).
206. A. Goldoni, R. Larciprete, L. Petaccia, and S. Lizzit, *J. Am. Chem. Soc.* 125, 11329 (2003).
207. P. Avouris, *Chem. Phys.* 281, 429 (2002).
208. K. Bradley, J. C. P. Gabriel, M. Briman, A. Star, and G. Gruner, *Phys. Rev. Lett.* 91, (2003).
209. K. Bradley, J. C. P. Gabriel, A. Star, and G. Gruner, *Appl. Phys. Lett.* 83, 3821 (2003).
210. L. Valentini, F. Mercuri, I. Armentano, C. Cantalini, S. Picozzi, L. Lozzi, S. Santucci, A. Sgamellotti, and J. M. Kenny, *Chem. Phys. Lett.* 387, 356 (2004).
211. S. Tsuzuki and H. P. Luthi, *J. Chem. Phys.* 114, 3949 (2001).
212. D. C. Patton and M. R. Pederson, *Phys. Rev. A* 56, R2495 (1997).
213. S. Peng and K. J. Cho, *Nano Lett.* 3, 513 (2003).
214. H. Chang, J. D. Lee, S. M. Lee, and Y. H. Lee, *Appl. Phys. Lett.* 79, 3863 (2001).
215. S. Santucci, S. Picozzi, F. Di Gregorio, L. Lozzi, C. Cantalini, L. Valentini, J. M. Kenny, and B. Delley, *J. Chem. Phys.* 119, 10904 (2003).
216. W. L. Yim, X. G. Gong, and Z. F. Liu, *J. Phys. Chem. B* 107, 9363 (2003).

217. D. W. Breck, "Zeolite Molecular Sieves: Structure, Chemistry, and Use," Wiley, New York, 1973.
218. S. Bhatia, "Zeolite Catalysis: Principles and Applications," CRC Press, Boca Raton, FL, 1990.
219. P. H. Nelson and S. M. Auerbach, *J. Chem. Phys.* 110, 9235 (1999).
220. D. G. Levitt, *Phys. Rev. A* 8, 3050 (1973).
221. D. S. Sholl and K. A. Fichthorn, *J. Chem. Phys.* 107, 4384 (1997).
222. D. S. Sholl, *Chem. Eng. J.* 74, 25 (1999).
223. M. R. Pederson and J. Q. Broughton, *Phys. Rev. Lett.* 69, 2689 (1992).
224. J. Breton, J. Gonzalez-Platas, and C. Girardet, *J. Chem. Phys.* 101, 3334 (1994).
225. G. Stan, M. J. Bojan, S. Curtarolo, S. M. Gatica, and M. W. Cole, *Phys. Rev. B* 62, 2173 (2000).
226. G. Stan, J. M. Hartman, V. H. Crespi, S. M. Gatica, and M. W. Cole, *Phys. Rev. B* 61, 7288 (2000).
227. M. W. Cole, V. H. Crespi, G. Stan, C. Ebner, J. M. Hartman, S. Moroni, and M. Boninsegni, *Phys. Rev. Lett.* 84, 3883 (2000).
228. V. V. Simonyan, P. Diep, and J. K. Johnson, *J. Chem. Phys.* 111, 9778 (1999).
229. S. M. Gatica, M. W. Cole, G. Stan, J. M. Hartman, and V. H. Crespi, *Phys. Rev. B* 62, 9989 (2000).
230. Y. F. Yin, T. Mays, and B. McEnaney, *Langmuir* 15, 8714 (1999).
231. A. C. Dillon, K. M. Jones, T. A. Bekkedahl, C. H. Kiang, D. S. Bethune, and M. J. Heben, *Nature* 386, 377 (1997).
232. Y. Ye, C. C. Ahn, C. Witham, B. Fultz, J. Liu, A. G. Rinzler, D. Colbert, K. A. Smith, and R. E. Smalley, *Appl. Phys. Lett.* 74, 2307 (1999).
233. A. I. Skoulidas, D. M. Ackerman, J. K. Johnson and D. S. Sholl, *Phys. Rev. Lett.* 89, 185901 (2002).
234. S. Talapatra, A. Z. Zambano, S. E. Weber, and A. D. Migone, *Phys. Rev. Lett.* 85, 138 (2000).
235. M. Muris, N. Dufau, M. Bienfait, N. Dupont-Pavlovsky, Y. Grillet, and J. P. Palmari, *Langmuir* 16, 7019 (2000).
236. R. E. Tuzun, D. W. Noid, B. G. Sumpter, and R. C. Merkle, *Nanotechnology* 7, 241 (1996).
237. Z. Mao and S. B. Sinnott, *J. Phys. Chem. B* 104, 4618 (2000).
238. K.-H. Lee and S. B. Sinnott, in preparation.
239. K.-H. Lee and S. B. Sinnott, *J. Phys. Chem. B* 108, 9861 (2004).
240. Z. Mao, A. Garg, and S. B. Sinnott, *Nanotechnology* 10, 273 (1999).
241. A. I. Skoulidas, D. M. Ackerman, J. K. Johnson, and D. S. Sholl, *Phys. Rev. Lett.* 89, 185901 (2002).
242. D. M. Ackerman, A. I. Skoulidas, D. S. Sholl, and J. Karl Johnson, *Mol. Sim.* 29, 677 (2003).
243. F. Zhang, *J. Chem. Phys.* 111, 9082 (1999).
244. S. Y. Bhide and S. Yashonath, *J. Am. Chem. Soc.* 125, 7425 (2003).
245. S. Supple and N. Quirke, *Phys. Rev. Lett.* 90, (2003).
246. G. Hummer, J. C. Rasaiah, and J. P. Noworyta, *Nature* 414, 188 (2001).
247. A. Kalra, S. Garde, and G. Hummer, *Proc. Natl. Acad. Sci. USA* 100, 10175 (2003).
248. A. Berezhkovskii and G. Hummer, *Phys. Rev. Lett.* 89, 064503 (2002).
249. A. Waghe, J. C. Rasaiah, and G. Hummer, *J. Chem. Phys.* 117, 10789 (2002).
250. R. J. Mashl, S. Joseph, N. R. Aluru, and E. Jakobsson, *Nano Lett.* 3, 589 (2003).
251. W. H. Noon, K. D. Ausman, R. E. Smalley, and J. Ma, *Chem. Phys. Lett.* 355, 445 (2002).
252. D. J. Mann and M. D. Halls, *Phys. Rev. Lett.* 90, 195503 (2003).
253. C. Dellago, M. M. Naor, and G. Hummer, *Phys. Rev. Lett.* 90, 105902 (2003).
254. Y. Maniwa, H. Kataura, M. Abe, S. Suzuki, Y. Achiba, H. Kira, and K. Matsuda, *J. Phys. Soc. Jpn.* 71, 2863 (2002).
255. Y. Gogotsi, N. Naguib, and J. A. Libera, *Chem. Phys. Lett.* 365, 354 (2002).
256. Y. Miyamoto, A. Rubio, X. Blase, M. L. Cohen and S. G. Louie, *Phys. Rev. Lett.* 74, 2993 (1995).
257. V. Z. Mordkovich, M. Baxendale, S. Yoshimura, and R. P. H. Chang, *Carbon* 34, 1301 (1996).
258. X. Fan, E. C. Dickey, P. C. Eklund, K. A. Williams, L. Grigorian, R. Buczko, S. T. Pantelides, and S. J. Pennycook, *Phys. Rev. Lett.* 84, 4621 (2000).
259. Z. Mao and S. B. Sinnott, *Phys. Rev. Lett.* 89, 278301 (2002).
260. M. Wilson and P. A. Madden, *J. Am. Chem. Soc.* 123, 2101 (2001).
261. T. W. Ebbesen, *J. Phys. Chem. Solids* 57, 9510955 (1996).
262. K. G. Ayappa, *Langmuir* 14, 880 (1998).
263. Z. Mao and S. B. Sinnott, *J. Phys. Chem. B* 105, 6916 (2001).
264. R. Qiao and N. R. Aluru, *Nano Lett.* 3, 1013 (2003).
265. S. Joseph, R. J. Mashl, E. Jakobsson, and N. R. Aluru, *Nano Lett.* 3, 1399 (2003).
266. B. J. Hinds, N. Chopra, T. Rantell, R. Andrews, V. Gavalas, and L. G. Bachas, *Science* 303, 62 (2004).
267. R. E. Tuzun, D. W. Noid, B. G. Sumpter, and R. C. Merkle, *Nanotechnology* 8, 112 (1997).
268. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Chem. Phys. Lett.* 315, 31 (1999).
269. B. Burteaux, A. Claye, B. W. Smith, M. Monthioux, D. E. Luzzi, and J. E. Fischer, *Chem. Phys. Lett.* 310, 21 (1999).
270. B. W. Smith and D. E. Luzzi, *Chem. Phys. Lett.* 321, 169 (2000).
271. S. Okada, S. Saito, and A. Oshiyama, *Phys. Rev. Lett.* 86, 3835 (2001).
272. R. F. Service, *Science* 292, 45 (2001).
273. Y.-K. Kwon, D. Tomanek, and S. Iijima, *Phys. Rev. Lett.* 82, 1470 (1999).
274. A. A. Farajian and M. Mikami, *J. Phys. Cond. Matter* 13, 8049 (2001).
275. B. Ni, S. B. Sinnott, P. T. Mikulski, and J. A. Harrison, *Phys. Rev. Lett.* 88, 205505 (2002).
276. H. Trotter, R. Phillips, Y. Hu, B. Ni, S. B. Sinnott, P. Mikulski, and J. A. Harrison, *J. Nanosci. Nanotech.* (in press).

277. O. Shenderova, D. Brenner, and R. S. Ruoff, *Nano Lett.* 3, 805 (2003).
278. A. S. Barnard and I. K. Snook, *J. Chem. Phys.* 120, 3817 (2004).
279. Q. X. Liu, C. X. Wang, S. W. Li, J. X. Zhang, and G. W. Yang, *Carbon* 42, 629 (2004).
280. D. Danailov, P. Keblinski, S. Nayak, and P. M. Ajayan, *J. Nanosci. Nanotech.* 2, 503 (2002).
281. G. Binning, H. Rohrer, C. Gerber, and E. Weibel, *Phys. Rev. Lett.* 49, 57 (1982).
282. J. M. Soler, A. M. Baro, and N. García, *Phys. Rev. Lett.* 57, 444 (1986).
283. J. A. Harrison, S. J. Stuart, and D. W. Brenner, in "Handbook of Micro/Nanotribology" (B. Bhushan, Ed.), p. 525, CRC Press, Boca Raton, 1999.
284. S. B. Sinnott, in "Handbook of Nanostructured Materials and Nanotechnology" (H. Nalwa, Ed.), Vol. 2, p. 571, Press, San Diego, CA, 2000.
285. H. Dai, J. H. Hafner, A. G. Rinzler, D. T. Colbert, and R. E. Smalley, *Nature* 384, 147 (1996).
286. C. L. Cheung, J. H. Hafner, T. W. Odom, K. Kim, and C. M. Lieber, *Appl. Phys. Lett.* 76, 3136 (2000).
287. S. S. Wong, A. T. Woolley, T. W. Odom, J.-L. Huang, P. Kim, D. V. Vezenov, and C. M. Lieber, *Appl. Phys. Lett.* 73, 3465 (1998).
288. G. Nagy, M. Levy, R. Scarmozzino, J. R. M. Osgood, H. Dai, R. E. Smalley, C. A. Michaels, G. W. Flynn, and G. F. McLane, *Appl. Phys. Lett.* 73, 529 (1998).
289. G. Nagy, M. Levy, R. Scarmozzino, J. R. M. Osgood, H. Dai, R. E. Smalley, C. A. Michaels, E. T. Sevy, G. W. Flynn, and G. F. McLane, *Appl. Phys. Lett.* 73, 1448 (1998).
290. H. Dai, N. Fraklin, and J. Han, *Appl. Phys. Lett.* 73, 1508 (1998).
291. A. Patil, J. Sippel, G. W. Martin, and A. G. Rinzler, *Nano Lett.* 4, 303 (2004).
292. J. A. Harrison, S. J. Stuart, D. H. Robertson, and C. T. White, *J. Phys. Chem. B* 101, 9682 (1997).
293. N. Yao and V. Lordi, *Phys. Rev. B* 58, 12649 (1998).
294. A. Garg, J. Han, and S. B. Sinnott, *Phys. Rev. Lett.* 81, 2260 (1998).
295. A. Garg and S. B. Sinnott, *Phys. Rev. B* 60, 13786 (1999).
296. E. N. Dzegilenko, D. Srivastava, and S. Saini, *Nanotechnology* 10, 253 (1999).
297. A. B. Tutein, S. J. Stuart, and J. A. Harrison, *J. Phys. Chem. B* 103, 11357 (1999).
298. U. Landman, W. D. Luedtke, N. A. Burnham, and R. J. Colton, *Science* 248, 454 (1990).
299. N. A. Burnham, R. J. Colton, and H. M. Pollock, *Nanotechnology* 4, 64 (1993).
300. N. Ohmae, *Phil. Mag. A* 74, 1319 (1996).
301. N. A. Burnham and R. J. Colton, *J. Vac. Sci. Technol. A* 7, 2906 (1996).
302. N. Agraït, G. Rubio, and S. Vieria, *Langmuir* 12, 4505 (1996).
303. H. Raffi-Tabar and A. P. Sutton, *Phil. Mag. Lett.* 63, 217 (1991).
304. O. Tomagnini, F. Ercolessi, and E. Tosatti, *Surf. Sci.* 287/288, 1041 (1991).
305. U. Landman, W. D. Luedtke, and E. M. Ringer, *Wear* 153, 3 (1992).
306. J. Belak and I. F. Stowers, in "Proceedings of the American Society of Precision Engineering," 1990, p. 76.
307. J. Belak and I. F. Stowers, in "Fundamentals of Friction: Macroscopic and Microscopic Properties" (I. L. Singer and H. M. Pollock, Eds.), Kluwer Academic, Dordrecht, 1992.
308. M. Fournel, E. Lacaze, and M. Schott, *Euro. Phys. Lett.* 34, 489 (1996).
309. T. Yakohata and K. Kato, *Wear* 168, 109 (1993).
310. C. L. Kelchner, S. J. Plimpton, and J. C. Hamilton, *Phys. Rev. Lett.* 58, 11085 (1998).
311. J. S. Kallman, W. G. Hoover, C. G. Hoover, A. J. DeGroot, S. M. Lee, and F. Wooten, *Phys. Rev. B* 47, 7705 (1993).
312. K. Minowa and K. Sumino, *Phys. Rev. Lett.* 69, 320 (1992).
313. U. Landman, W. D. Luedtke, and M. W. Ribarsky, *J. Vac. Sci. Technol. A* 7, 2829 (1989).
314. K. Cho and J. D. Joannopoulos, *Surf. Sci.* 328, 320 (1995).
315. J. A. Harrison, D. W. Brenner, C. T. White, and R. J. Colton, *Thin Solid Films* 206, 213 (1991).
316. J. A. Harrison, C. T. White, R. J. Colton, and D. W. Brenner, *Surf. Sci.* 271, 57 (1992).
317. S. B. Sinnott, R. J. Colton, C. T. White, O. A. Shenderova, D. W. Brenner, and J. A. Harrison, *J. Vac. Sci. Technol. A* 15, 936 (1997).

CHAPTER 2

Density Functional Calculations of Clusters and Cluster Assembly

J. A. Alonso
Departamento de Física Teórica, Universidad de Valladolid, Valladolid, Spain

M. J. Stott
Department of Physics, Queens University, Kingston, Ontario, Canada

CONTENTS

1. Introduction 43
2. Theory 47
   2.1. Density Functional Theory 47
   2.2. Pseudopotentials 57
   2.3. Synthesis 60
3. Applications to Clusters and Cluster-Assembled Solids 63
   3.1. Electronic Shells 63
   3.2. Perturbed Clusters 70
   3.3. Clusters with Impurities 77
   3.4. Cluster Melting 87
   3.5. Cluster Assembly 92
4. Conclusions 97
   References 98

1. INTRODUCTION

Although isolated atomic clusters have been the subject of much experimental and theoretical study for some time, the self-assembling of clusters opens the possibility of building materials with novel properties. A prime example is the fullerite crystal, formed by the self-assembling of C_{60} molecules [1]. For the assembling to be successful, the clusters must retain at least some of their character, which happens in the fullerite crystal, but this requirement is demanding for metallic clusters because the interatomic interactions are more delocalized, and the clusters have a tendency to react and coalesce. This chapter reviews theoretical work that addresses the relationship between isolated clusters and cluster-assembled bulk

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (43–101)

materials. The factors that influence the stability of isolated clusters will be discussed, along with the requirements for clusters to retain their identities as they are brought together to build a bulk material.

Some simple models have been used to account qualitatively for the relative stability of certain types of clusters. In one type, the stability appears to be geometrical in nature and associated with the closing of atomic shells. These clusters, roughly spherical in shape to minimize the cost in energy due to the surface, can be viewed as a collection of close-packed spheres representing the atoms. Such compact icosohedral or cuboctahedral arrangements consist of concentric shells of atoms, and clusters with filled atomic shells containing 13, 55, 147,... atoms tend to be more stable than those with a partially filled outer atomic shell. However, the stability of clusters in which the electrons spread throughout the cluster tends to be associated with the electronic shell structure. Clusters of simple metal atoms such as the alkali atoms, Mg, Al,... fall into this group. The valence electrons in these atoms have s- and p-like character, are relatively weakly bound to the ion, and in the metallic state tend not to form directional bonds but are fairly uniformly spread. The electrons in clusters of these atoms occupy cluster states with orbitals that extend over the whole cluster. The complete filling of one group of these states closes an electronic shell, and if the total number of valence electrons in the cluster is such that all electronic shells are closed in the ground state, the cluster is stable relative to those with different numbers of electrons. These are the "magic numbers" for electronic shell closing, in close analogy with the magic numbers for shell closing in nuclei. The often used jellium model smears out the positive charge of the ions uniformly over the assumed spherical cluster so that the electrons move in a roughly spherical square well potential. The electronic shell closing magic numbers for this model are 2, 8, 18, 20, 40,..., which accords well with the valence electron count for simple metal clusters that exhibit particular stability. These simple models have undoubted value as they provide qualitative explanation for the behavior of groups of clusters and some understanding of trends. However, a detailed understanding of the properties of particular clusters, and possibly the prediction of other properties, requires treatment of the electronic structure and the atomic arrangement that goes beyond these simple models. Fortunately, calculational schemes have been developed that can simulate the behavior of a cluster and, indeed, the bulk material. These are the *ab initio* methods based on density functional theory.

The existence of atomic clusters in bulk materials is well established, but most of the examples are not the focus of this chapter. For example, the solids consisting of van der Waals bonded small stable molecules have within them well defined molecular units. As a case in point, N_2 molecules in solid nitrogen are clearly displayed. The bonding that takes place to form the solid does not disrupt the molecular unit or link them together in any significant way. On the other hand, there are many systems where it is convenient to identify small clusters in the bulk material, but this is usually only a device for classifying structures of related materials, and the clusters are strongly linked and have no separate existence. For example, the large group of silicate materials are viewed [2] as being built of tetrahedral SiO_4^{4-} units, but only in the orthosilicates are there discrete SiO_4^{4-} ions with the cations occupying interstices in which they are surrounded by O^{2-} ions. Otherwise, the SiO_4^{4-} clusters are merely a convenience for representing the atomic arrangements, as in these materials the clusters are linked through one or more shared vertex oxygens to form chains, layers, or elaborate three-dimensional networks. Indeed, the free SiO_4^{4-} ion in vacuum is not known to exist. Rather than these systems, our focus in this chapter is on stable atomic clusters, a nanometer or so in size, with which a bulk system with desirable properties can be built, a so-called cluster-assembled solid. In general, a cluster solid will be metastable (like the fullerite crystal, which is less stable that pure graphite), trapped in a deep minimum of the potential energy surface protected by sizable energy barriers.

The methods of quantum chemistry, such as configuration interaction, have been used for some time to calculate the structural properties of small molecules with sufficient accuracy to make confident predictions. Unfortunately, the computational effort involved in these calculations increases very rapidly with the number of electrons, and their application to systems comprising more than a few atoms is not feasible. Such systems are more the domain

of condensed matter physics, and for decades condensed matter theory held out the promise of approaches that would allow the simulation of larger systems such as clusters or even bulk materials with enough precision to make predictions that could guide the development of new materials. However, the achievement of this promise was always over the horizon. Very ingenious methods had been devised for treating the structural properties of materials, including those based on the "nearly free" electron approximation, "tight" binding methods, and semiempirical interatomic potentials such as the embedded atom method. Applications of these approaches have given understanding of experimental data, particularly on trends in structural properties of groups of materials. But, in many cases, these approaches involved parameters fitted to experiment and all involved uncontrolled approximations so that the accuracy of results could not be checked, and although predictions could be made, there was no way to assess the confidence to be placed in them. There was promise, but it was unfulfilled.

Clusters of atoms and bulk materials consist of electrons and nuclei, every particle interacting with every other one. The electrons must be treated quantum mechanically, but the much heavier nuclei may be treated as classical particles to a very good approximation in most circumstances. If our interest is in structural properties of the system such as the different stable arrangements of the atoms, the relative energies of these structures, the energies associated with deformations and defects, and the vibrational properties, then we require the energy of the system as a function of the positions of the nuclei. A further key assumption is that the electrons respond very rapidly to a change in position of the nuclei and remain in their ground state for the instantaneous arrangement of the nuclei. This is the celebrated Born-Oppenheimer approximation. Our task, then, is to calculate the ground state energy of the system of many interacting electrons in the potential of the nuclei for an arbitrary arrangement of the nuclei. Furthermore, the calculation must be (i) extremely efficient if many different arrangements of the nuclei are to be explored, or the trajectories of the nuclei followed as they move according to classical mechanics, and (ii) rather accurate if the energy difference between isomers or different crystal phases are to be distinguished, or the dynamics of the nuclei studied using the forces acting on them.

This is a daunting task, and a number of difficulties confront us. First and foremost, the electronic energy requires solution of the quantum mechanical many-body problem for the system of many electrons interacting strongly with one another via the Coulomb potential. A direct attack on this problem by representing in some way the many-electron wavefunction, which depends on the coordinates of all the electrons, is possible for a small number of electrons, but the methods used cannot be extended to treat many electrons. It is this difficulty that limits the scope of traditional quantum chemistry methods to small molecules. The second main difficulty is the complexity of determining the behavior (i.e., the energy and density distribution) for even a single electron moving in the potential due to an arbitrary arrangement of atoms. This one-body problem is greatly simplified if the geometry of the atomic arrangement has high symmetry, but we must consider arbitrary arrangements with no symmetry if competing cluster isomers or the molecular dynamics of the system are to be studied. Some fundamental developments in condensed matter physics, some advances in theory, and technological developments came together in the 1980s to overcome these difficulties and make possible the simulation of materials from first principles with a proper treatment of the electrons and the ions.

The main development was density functional theory, which provided a way to treat the many-electron problem. Hohenberg and Kohn [3] showed that the properties of a system of interacting electrons can be obtained from the ground state electron density distribution. In particular, the electronic ground state energy can be expressed as a functional of the ground state electron density. This is an enormous simplification. Although all information on the ground state is contained in the hideously complicated many-electron wavefunction, it is also present in principle in the ground state electron density, a real, positive definite function of a single spatial variable much easier to visualize and construct than the wavefunction, or other amplitudes. Kohn and Sham, the following year [4], showed how to extract the ground state energy from the density functional theory by casting the interacting many-electron problem

into the solution of a system of noninteracting electrons—essentially, a one-electron problem. This so-called Kohn-Sham form of density functional theory is the principal approach used today for handling the electron–electron interactions in clusters of atoms or bulk materials.

Other developments led to methods for efficient solution of the one-electron problem that density functional theory delivered. First principles, transferable pseudopotentials allowed tightly bound electrons in the atomic cores to be removed from the electrons to be considered, and their effect incorporated in an effective electron–ion interaction—a pseudopotential. This had several advantages. The core electrons are tightly bound to the nucleus, move rigidly with the atoms, and are little affected by the electronic surroundings, but large energies are associated with the core and these would obscure molecular and condensed matter effects. Removing the core electrons removes the large energies associated with them and leaves for close attention the outer, valence electrons whose distribution determines the structural properties of molecules, clusters, and bulk material. The focus is on the electrons that are sensitive to the atomic arrangement. The idea of a pseudopotential to describe the interaction between valence electrons and the ion was introduced much earlier, but usually these contained parameters that were fitted to calculated or experimentally determined band structure features or to other experimental data [5]. These pseudopotentials were particularly successful in studies of simple metal properties where they were taken to be weakly scattering and used with second-order perturbation theory. But, two issues limited the use of these pseudopotentials. The first was transferability. There was no guarantee that a pseudopotential constructed to describe the atom in one situation, for example immersed in a metal, would give an adequate description when transfered to a different system, for instance, when the atom is chemisorbed at a surface. Second, these pseudopotentials were not designed to give the correct distribution of valence electrons between the core region of the atom and outside, which led to substantial errors in the electrostatic potential due to the electrons, as well as other problems. These two limitations were remedied in the norm-conserving pseudopotential constructed from a Kohn-Sham density functional calculation of the free atom [6, 7].

The introduction of pseudopotentials for describing the valence electrons has an added advantage. Removing the core electrons themselves also removed their effects on the valence electron wavefunctions. The valence wavefunctions of the all-electron atoms are orthogonal to the core electron wavefunction, consequently they must oscillate in the core region. These "orthogonality wiggles" in the valence electron wavefunctions are removed along with the core as part of the pseudizing procedure. The resulting electron pseudowavefunctions are smooth in the core region and can be represented much more easily than the original rapidly varying valence wavefunctions. The solution of the one-electron problem requires a representation for the electron wavefunction—a way to represent the wavefunction must be chosen. This could be as wavefunction values at a set of grid points in real space or as a linear combination of basis functions. Whatever the choice of basis, the number of basis functions required to describe adequately a smooth pseudopotential is much less than is required for the original wavefunction, which obviously permits the study of systems containing a much larger number of atoms.

The final ingredient in the synthesis that allowed the simulation of systems of many atoms with a quantum mechanical treatment of the electrons was the availability of large memory, high-speed computers. These elements were first brought together by Yin and Cohen [8] in their study of the structural properties of Si and Ge, including crystal stability and pressure-induced phase transitions. In what has become known as an *ab initio* study, they used density functional theory for treating the electron–electron interactions, normconserving pseudopotentials for describing the interaction of valence electrons with the ions, and large plane wave basis sets for representing electron wavefunctions. A further development due to Car and Parrinello [9] introduced a scheme for following the dynamics of the atoms that treated the electrons and the ions on a similar footing. Since the introduction of the *ab initio* methods, there have been advances in density functional theory giving improved functionals accounting for the effects of electron–electron interactions, advances in pseudopotentials allowing

description of a wider range of atoms so that most of the periodic table can be treated, technical advances in the representation of electron wavefunctions, such as real-space methods, and order of magnitude increases in the computing power that can be brought to bear on materials simulation due, for example, to the availability of multiprocessor machines.

In this chapter, we shall review these theoretical ingredients of the *ab initio* calculations: the application of density functional theory for treating the electron–electron interactions, pseudopotentials for describing the electron–ion interaction, and the miscellany of technical items that allows the realistic simulation of clusters and cluster-assembled solids to be performed. Next, we describe applications of the theory to clusters and cluster-assembled materials, beginning with the observation of shell effects in clusters of simple metals, and their explanation in terms of simple theoretical models within the framework of density functional theory [3, 4]. A number of examples are presented indicating how the shell effects have a strong influence in many properties of the clusters. Simulations of the melting of clusters is discussed including the application of a scheme free of electron orbitals, based entirely on the electron density. The next section of this chapter is devoted to a discussion of computer simulations of the assembling of metallic clusters. The simulations again highlight the important idea that closed shell clusters are the best candidates for self-assembling, but other favorable factors are also revealed. Further insight into the self-assembling phenomenon is also provided by an analysis of the chemical bonding in some ordered crystalline alloys known to display clustering.

2. THEORY

2.1. Density Functional Theory

The origins of the density functional theory of interacting systems of electrons can be traced to the statistical Thomas-Fermi model, and the main features of the theory can be illustrated by reviewing the Thomas-Fermi ideas [10]. Consider N electrons interacting in the normal way and in the presence of an external potential, $V_{ext}(\mathbf{r})$. In practice, $V_{ext}(\mathbf{r})$ is the potential due to the ions or nuclei at fixed positions in the system that contribute a density of positive charge, $n_+(\mathbf{r})$. An electron at r will have potential energy $V_{eff}(\mathbf{r})$ with contributions from the external field and the electron distribution. The kinetic energy[1] of the fastest electron $\frac{1}{2}k_F(\mathbf{r})^2$ will vary from point to point so that the total energy, kinetic plus potential, is constant throughout the system and equal to the chemical potential μ, giving

$$\frac{1}{2}k_F(\mathbf{r})^2 + V_{eff}(\mathbf{r}) = \mu \tag{1}$$

A reasonable choice for the effective potential would be the electrostatic potential due to the total charge distribution with a contribution from the electron distribution as well as $V_{ext}(\mathbf{r})$,

$$V_{eff}(\mathbf{r}) = V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2}$$

where $n(\mathbf{r})$ is the electron density distribution. Alternatively, Poisson's equation gives

$$\nabla^2 V_{eff}(\mathbf{r}) = -4\pi[n(\mathbf{r}) - n_+(\mathbf{r})] \tag{3}$$

For a uniform system of electrons in its ground state, the kinetic energy of the fastest electron is related to the constant electron density, $\bar{n}$, through $\bar{n} = k_F^3/3\pi^2$. If the electron density varies sufficiently slowly, we assume that in the vicinity of $\mathbf{r}$ the electronic system behaves as though it were uniform, with density $n(\mathbf{r})$ and k_F adjusts so that $n(\mathbf{r}) = k_F(\mathbf{r})^3/3\pi^2$. Using this slowly varying approximation in Eq. (1) gives the Thomas-Fermi equation:

$$\frac{1}{2}(3\pi^2 n(\mathbf{r}))^{2/3} + V_{eff}(\mathbf{r}) = \mu \tag{4}$$

[1] Atomic units with $\hbar = e = m_e = 1$ are used throughout except where otherwise stated.

or, employing Poisson's equation we obtain the differential equation for the effective potential

$$\nabla^2 V_{eff}(\mathbf{r}) = -4\pi\left(\frac{1}{(3\pi^2)^{2/3}}\{2[\mu - V_{eff}(\mathbf{r})]\}^{3/2} - n_+(\mathbf{r})\right) \tag{5}$$

Solution of Eq. (5) for a given external charge distribution yields estimates for V_{eff} and hence for $n(\mathbf{r})$, where μ is adjusted to give the required number of electrons.

In an alternative approach, we attempt to construct an expression for the ground state energy, E, of the N electrons moving in the external potential $V_{ext}(\mathbf{r})$. Two terms in the potential energy must be the electrostatic self-energy of the electron distribution

$$E_{ee} = \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$$

and the potential energy of the electrons due to $V_{ext}(\mathbf{r})$

$$E_{ext} = \int d\mathbf{r}V_{ext}(\mathbf{r})n(\mathbf{r})$$

both of which are expressed in terms of the electron density. To complete the total energy, an expression for the electron kinetic energy is required. Noting that the kinetic energy per electron in a uniform gas of noninteracting electrons is $3E_F/5 = 3k_F^2/10$, and assuming that the system is slowly varying, the kinetic energy of the $n(\mathbf{r})d\mathbf{r}$ electrons in volume element $d\mathbf{r}$ is $(3k_F(\mathbf{r})^2/10)(n(\mathbf{r})d\mathbf{r})$, where $n(\mathbf{r}) = k_F(\mathbf{r})^3/3\pi^2$. The total electron kinetic energy becomes

$$T = \int d\mathbf{r}\left(\frac{3}{10}(3\pi^2 n(\mathbf{r}))^{2/3}\right)n(\mathbf{r})$$

Assembling the pieces of the total energy, we have

$$E[n] = c_0\int d\mathbf{r}\, n(\mathbf{r})^{5/3} + \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \int d\mathbf{r}\, V_{ext}(\mathbf{r})n(\mathbf{r}) \tag{6}$$

where $c_0 = 3(3\pi^2)^{2/3}/10$. The notation $E[n]$ indicates that E is a *functional* of $n(\mathbf{r})$, acknowledging that the state of the electronic system is described in terms of the density distribution $n(\mathbf{r})$.

If the ground state electron density for the system is known, an estimate of the ground state energy can be obtained by substituting into the energy functional $E[n]$, Eq. (6). But, we shall see that Eq. (6) may be used to estimate the electron density. For a fixed total number of electrons $N = \int d\mathbf{r}\, n(\mathbf{r})$, the value of the $E[n]$ depends on the way the electrons are distributed; the kinetic energy is a minimum when the electrons are uniformly distributed, whereas the potential energy coming from the last two terms in Eq. (6) is lowered when electrons heap up in the regions of attractive potential. The ground state electron density must be a compromise between these two extremes, which minimizes the total energy. This suggests a variational principle for the ground state energy in which $E[n]$ is minimized with respect to variations in $n(\mathbf{r})$ subject to the constraint of a fixed total number of electrons. The variational principle is compactly expressed

$$\delta\left\{E[n] - \mu\int d\mathbf{r}\, n(\mathbf{r})\right\} = 0 \tag{7}$$

where δ denotes variations with respect to the density, and the constraint is incorporated using a Lagrange undetermined multiplier μ, chosen so that there are the required number of electrons in the system. In terms of the functional derivative $\delta/\delta n(\mathbf{r})$, we have

$$\frac{\delta E[n]}{\delta n(\mathbf{r})} = \mu \tag{8}$$

which gives the Euler-Lagrange equation corresponding to the variational principle, and μ is clearly identified as the electron chemical potential. Taking the functional derivative of our approximate energy functional Eq. (6), we obtain the Euler-Lagrange equation

$$\frac{5}{3}c_0 n(\mathbf{r})^{2/3} + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{ext}(\mathbf{r}) = \mu \tag{9}$$

which is precisely the Thomas-Fermi equation, Eq. (4). In summary, we can view the Thomas-Fermi approach as:

- an attempt to express the ground state energy of an electronic system as a *functional* of the electron density distribution $n(\mathbf{r})$, and
- *a variational principle* with the actual ground state density minimizing the value of the functional subject to fixed particle number, the minimum value being the ground state energy.

Density functional theory shows that this approach is in principle exact, and the Thomas-Fermi method is an approximate special case of the exact theory. With this view of Thomas-Fermi as a variational principle, it is easy to see where improvements are needed. The quantity $c_0 n(\mathbf{r})^{2/3}$ is the energy per particle for a uniform noninteracting gas so no account is taken in $E[n]$ given, by Eq. (6), of the exchange and correlation effects of the electron–electron interactions. Furthermore, the expression for the electron kinetic energy, $c_0 \int d\mathbf{r}\, n(\mathbf{r})^{5/3}$, is correct in the limit of a uniform gas but misses the direct effects of inhomogeneieties in the electron density. We now show how density functional theory remedies these defects.

The Hamiltonian for our N electron system with external potential $V_{ext}(\mathbf{r})$ is

$$\hat{H} = \hat{T} + \hat{U} + \int d\mathbf{r}\, V_{ext}(\mathbf{r})\hat{n}(\mathbf{r}) \tag{10}$$

where $\hat{T} = \sum_i \mathbf{p}_i^2/2$ is the electron kinetic energy, $\hat{U} = 1/2 \sum_i \sum_{j \neq i} v(r_{ij})$ is the interaction energy, and $\hat{n}(\mathbf{r})$ is the electron density operator. If it is understood that the interacting particles are electrons (i.e., fermions with $v(r) = 1/r$ and the electron mass), then the system is characterized by $V_{ext}(\mathbf{r})$. For example, it is $V_{ext}(\mathbf{r})$ that distinguishes the different 2-electron systems: He, H_2, Li^+, Be^{2+}, and so forth. Given $V_{ext}(\mathbf{r})$, the N-electron Schrödinger equation can in principle be solved to obtain the electronic properties of the system, for example, the ground state density $n_o(\mathbf{r})$, and energy E_o. We would say that $n_o(\mathbf{r})$ and E_o are functionals of V_{ext} with the functional relationship involving the solution of the Schrödinger equation. Hohenberg and Kohn [3] showed that the inverse is also true; they showed that there is a one-to-one relationship between $n_o(\mathbf{r})$ and the V_{ext}, apart from an additive constant. Therefore, given $n_o(\mathbf{r})$, V_{ext} can in principle be found along with other properties of the system (e.g., E_o), and the external potential and ground state energy are functionals of the ground state density, that is, $V_{ext}[n_o]$, and $E_o[n_o]$. This establishes the possibility of our first goal, of describing the electronic system in terms of the electron density rather than the many-body wavefunction or similar construction.

Further progress in obtaining the ground state energy and other electronic properties using a density functional approach requires more than just knowledge that there is a functional, and this is provided by the variational principle obtained by Hohenberg and Kohn [3]. We begin by identifying terms in the ground state energy that we know explicitly as functionals of n_o and extracting these from the functional. Two such terms are the electrostatic self-energy of the electronic charge distribution and the external potential energy. Extracting these gives

$$E_o[n_o] = G[n_o] + \frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}' \frac{n_o(\mathbf{r})n_o(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \int d\mathbf{r}\, V_{ext}(\mathbf{r})n_o(\mathbf{r}) \tag{11}$$

The residual functional $G[n_o]$ contains all kinetic and remaining potential energy contributions and is universal for all electron systems. Consider now another system of N electrons with a different external potential, V'_{ext}. The ground state $|\psi'_o\rangle$ of this system has

electron density $n(\mathbf{r}) = \langle\psi_a'|\hat{n}(\mathbf{r})|\psi_a'\rangle$. Because G is universal, $\langle\psi_a'|\widehat{H}|\psi_a'\rangle = E_a[n]$ and the Rayleigh-Ritz variational principle gives

$$\langle\psi_a'|\widehat{H}|\psi_a'\rangle = E_a[n] > E_a$$

where E_a is the ground state energy of our system of interest, namely $E_a[n_a]$. Finally, we see

$$E_a[n] \geq E_a[n_a] \tag{12}$$

with the equality holding only for $n = n_a$, and the ground state energy functional is a minimum for $n = n_a$, subject to $\int d\mathbf{r}\, n(\mathbf{r}) = N$. This variational principle can be expressed in terms of a functional derivative:

$$\frac{\delta E[n]}{\delta n(\mathbf{r})} = \mu \tag{13}$$

where the Lagrange undetermined multiplier μ is chosen to give the correct number of particles in the system. This is the variational principle Eq. (7) we applied earlier to the Thomas-Fermi energy functional Eq. (6). This variational principle is the route we seek to obtain the ground state electron density and the energy. Although on the face of it, the inequality Eq. (12) applies only to densities n that are ground state densities for some external potential, much work has been devoted to showing that the variational principle, in fact, holds for functions $n(\mathbf{r})$ that could be physically reasonable densities. A formulation of the energy functional that illuminates this generalization was proposed by Levy [11].

Progress in direct use of the variational principle to calculate the electronic structure is hampered by lack of knowledge of the nature of the universal functional $G[n]$, which contains all the kinetic energy contributions. The problem is illustrated by comparing for an atom the electronic kinetic energy calculated within the Hartree-Fock approximation, which is very close to the exact value, with that obtained from the Thomas-Fermi kinetic energy functional $c_0 \int d\mathbf{r}\, n(\mathbf{r})^{5/3}$, where n is taken to be the Hartree-Fock density, n_{HF}. For the sodium atom, the Hartree-Fock kinetic energy is 324 a.u.; the corresponding Thomas-Fermi value is 10% less. Although this may seem to be encouraging agreement, the discrepancy is about 700 eV, a massive error compared with the 1 eV per atom that would begin to be of interest in structural chemistry or condensed matter physics. Nevertheless, there would be considerable merit in a practical scheme based entirely on the density, and some progress is being made in this area to which we shall return later. Kohn and Sham [4] showed how to circumvent this obstacle and capture the greater part of the electron kinetic energy. The Kohn-Sham scheme that they introduced maps the interacting many-electron system onto an exactly soluble system of particles, and in doing so they provided a firm theoretical foundation for the self-consistent field approach to electronic systems—atoms, molecules, and bulk systems—which has been applied in various forms since the pioneering work of Hartree.

2.1.1. Kohn-Sham Scheme

A system of N noninteracting fermions moving in an effective external potential $V_{eff}(\mathbf{r})$ has Hamiltonian $\widehat{H}_s = \sum_i \hat{h}_i$, where

$$\hat{h}_i = \frac{\mathbf{p}_i^2}{2} + V_{eff}(\mathbf{r}_i)$$

The ground state wavefunction for this independent particle system is a Slater determinant Ψ_s of single particle orbitals $\phi_i(\mathbf{r})$ and the ground state energy, E_s, is the sum of the N lowest single particle energy eigenvalues ϵ_i. The solution of the single particle Schrödinger equation

$$\left[-\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}), \quad i = 1, 2, \ldots, N \tag{14}$$

gives us the ground state for this system. Other useful quantities can also be obtained simply; for example, the particle density is

$$n_o(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{15}$$

and the kinetic energy is

$$\begin{aligned} T_s &= \sum_i \int d\mathbf{r}\, \phi_i(\mathbf{r})^* \left(-\frac{1}{2}\nabla^2\right)\phi_i(\mathbf{r}) \\ &= \sum_i \epsilon_i - \int d\mathbf{r}\, V_{eff}(\mathbf{r}) n_o(\mathbf{r}) \end{aligned}$$

But, the Hohenberg-Kohn theorem applies equally well to this noninteracting system as to our interacting system of interest. Consequently, there is a ground state energy functional

$$E_s[n] = T_s[n] + \int d\mathbf{r}\, V_{eff}(\mathbf{r}) n(\mathbf{r})$$

whose minimum value is the ground state energy $E_s[n_o] = \sum_i \epsilon_i$, where n_o is the ground state density, which itself can be obtained from

$$\left.\frac{\delta T_s[n]}{\delta n(\mathbf{r})}\right|_{n=n_o} + V_{eff}(\mathbf{r}) = \mu_{ind} \tag{16}$$

The kinetic energy for this system in its ground state is

$$T_s[n_o] = \sum_i \epsilon_i - \int d\mathbf{r}\, V_{eff}(\mathbf{r}) n_o(\mathbf{r}) \tag{17}$$

Returning now to the interacting electron system, we separate out of the universal functional $G[n]$ the kinetic energy functional of a set of independent particles with density n(r) and write $G[n] = T_s[n] + E_{xc}[n]$. The ground state energy functional becomes

$$E_o[n] = T_s[n] + \frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n] + \int d\mathbf{r}\, V_{ext}(\mathbf{r}) n(\mathbf{r}) \tag{18}$$

$E_{xc}[n]$ is a catch-all for all the contributions to the energy that are not included elsewhere in Eq. (18). It is the so-called exchange-correlation energy. Applying the variational principle to $E_o[n]$, the ground state density of the interacting system is the solution of

$$\frac{\delta T_s[n]}{\delta n(\mathbf{r})} + \left\{ \int d\mathbf{r}\, \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} + V_{ext}(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \right\} = \mu \tag{19}$$

We see that Eqs. (16) and (19) have the same form and the quantity in $\{\ldots\}$ plays the role of an effective potential acting on a system of noninteracting fermions. This is the Kohn-Sham potential

$$V_{KS}[n, \mathbf{r}] = V_{ext}(\mathbf{r}) + \int d\mathbf{r}'\, \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + V_{xc} \tag{20}$$

where

$$V_{xc}[n, \mathbf{r}] = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \tag{21}$$

is the exchange-correlation potential. The Kohn-Sham approach amounts to solving the set of single particle Schrödinger equations for an auxiliary system of independent particles moving in an effective external potential, V_{KS}, and with the same ground state density n_o as

the interacting system

$$\left(-\frac{1}{2}\nabla^2 + V_{KS}[n_o, \mathbf{r}]\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}). \qquad i = 1, 2, \ldots, N \tag{22}$$

$$n_o(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{23}$$

These are the Kohn-Sham equations, and the ϵ_i, the Kohn-Sham eigenvalues. The solution of the Kohn-Sham equations must be made self-consistent because V_{KS} depends on the electron density. The ground state density can now be substituted into the energy functional Eq. (18), and although we do not know the independent particle kinetic energy functional, $T_s[n]$, we do its value for the density $n = n_o$; it is given by Eq. (17). The ground state energy in terms of the density n_o and the Kohn-Sham eigenvalues is

$$E_o = \sum_i \epsilon_i + \frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \int d\mathbf{r}\, n_o(\mathbf{r})\,[V_{ext}(\mathbf{r}) - V_{KS}(\mathbf{r})] + E_{xc}[n_o] \tag{24}$$

Extensions of the Kohn-Sham form of density functional theory have been made to treat spin polarized systems described in terms of spin-up and spin-down electron densities, $n_\uparrow$ and $n_\downarrow$, respectively [12, 13]. A system of independent particles with the same densities as the interacting electron system is introduced along with the exchange-correlation functional of the up- and down-spin densities. The ground state is obtained through solution of the Kohn-Sham equations with separate potentials for up- and down-spin electrons. The Kohn-Sham scheme summarized by Eqs. (22) and (24) is exact, provided an auxiliary independent particle system exists with the same density as the interacting electron system, and this does not seems to be a serious restriction, but knowledge of the exchange-correlation functional is required if the scheme is to be implemented.

2.1.2. Exchange-Correlation

The functional E_{xc} and its derivative V_{xc} contain all the complex many-body effects. Exchange effects are those captured by the Hartree-Fock approximation and act between parallel-spin electrons, and correlation effects are all those beyond the Hartree-Fock level. Most of E_{xc} is interaction potential energy due to the exchange-correlation hole in the electron distribution around a given one, but these spatial correlations imply additional variations in the electron wavefunction that lead to a kinetic contribution to E_{xc} of a few percent of the total. The total exchange-correlation energy is a small part of the total energy for atomic systems. For Be, it amounts to about 18% of the total, but the fraction diminishes with increasing atomic number so that for Ne and Ar it is 10% and 6%, respectively. Most of E_{xc} is due to exchange, correlation amounting to 3% or less for the atoms noted above. Also of interest is the exchange and correlation energy for a uniform electron gas, often taken as a model for simple metal systems. Kinetic energy dominates the energy in the limit of very high electron density, such as would be encountered in the inner regions of an atom, but in the range of mean valence electron densities of the simple metals such as the alkalis the negative exchange energy becomes the main contribution. At lower densities such as those found in the outer regions of atoms, exchange and correlation dominate the total. In assessing the magnitude of E_{xc}, we must bear in mind that it is not so much the total energy of clusters and bulk materials that are of interest, but differences in total energy between different molecular or cluster isomers or between different crystal structures, and these may be only 1 eV per atom or much less, and the treatment of exchange and correlation could be the determining factor. An important feature of the Kohn-Sham scheme is that apart from numerical uncertainties in the solution of the single particle Schrödinger equation, approximations are all concentrated in the functional E_{xc}. It has been clear from its first introduction where to focus theoretical effort to improve the accuracy of the scheme.

The first approximation for E_{xc}, which has seen and continues to see a great deal of use in simulations, is the local density approximation (LDA) and, if there is spin polarization, the LSDA. This is the same sort of approximation made for the electron kinetic energy in

the Thomas-Fermi model. If the system is spatially slowly varying so that near $\mathbf{r}$ it may be replaced by a uniform electron gas of mean densities $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$ for up- and down-spin electrons, a corresponding approximation for the exchange-correlation energy would be

$$E_{xc}[n_\uparrow, n_\downarrow] = \int d\mathbf{r}\, n(\mathbf{r})\, \epsilon_{xc}^{u}[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})] \tag{25}$$

where $\epsilon_{xc}^{u}(n_\uparrow, n_\downarrow)$ is the exchange-correlation energy per electron of a uniform electron gas of spin densities $n_\uparrow$ and $n_\downarrow$, and $n = n_\uparrow + n_\downarrow$ is the total density. The corresponding potential for up-spin particles is

$$V_{xc\uparrow}(\mathbf{r}) = \frac{\partial}{\partial n_\uparrow}[n\epsilon_{xc}^{u}(n_\uparrow, n_\downarrow)]\Big|_{\substack{n_\uparrow = n_\uparrow(\mathbf{r}) \\ n_\downarrow = n_\downarrow(\mathbf{r})}} \tag{26}$$

and similarly for down-spin. Practical application of the LSDA for exchange and correlation requires a convenient parameterized form of electron gas data of which there are many. The most common ones in use today are due to Vosko, Wilk, and Nusair [14] and to Perdew and Zunger [15] and are different fittings to the Monte Carlo electron gas results of Ceperley [16] and Ceperley and Alder [17]. This simple approximation works very well in density functional calculations of atoms, molecules, and more extended systems. A summary of the successes compiled by von Barth [18] includes binding energies often better than 1 eV, equilibrium bond lengths generally accurate to within 0.1 Å, electron densities usually better than 2%, and physical trends are usually reproduced. Furthermore, errors are usually systematic (e.g., molecular binding energies and solid cohesive energies are usually overestimated and bond lengths and lattice parameters are underestimated). This is not to say the LSDA has no serious difficulties. The tail of the Kohn-Sham potential outside localized systems—atoms, clusters, a solid surface—falls off too rapidly with distance. This is due to residual electron self-interaction effects that do not cancel completely in the LSDA as they should when an electron is far from the rest of the system. Self-interaction corrections have been proposed that improve matters in a quantitative sense, but which lead to an undesirable orbital-dependent potential [15]. Although the LSDA has clear deficiencies, the strength of its success in atomic and molecular systems is surprising in view of the rapid variation of the electron density in the atomic core regions, and it has been the subject of much analysis.

Corrections to the LSDA must involve some measure of variations in the electron density from place to place in the system, and a common way of accomplishing this is through density gradients so that the integrand in E_{xc} depends on $\nabla n(\mathbf{r})$ and possibly higher order gradients, as well as the density at $\mathbf{r}$.

$$E_{xc} = \int d\mathbf{r}\, n(\mathbf{r})\, \epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}), \ldots) \tag{27}$$

with the obvious generalization for spin polarized systems. The first few terms of an expansion of ϵ_{xc} in ascending order of density gradients have been evaluated, but gradient corrections for atomic-like systems do not improve on the LSDA. Indeed, Langreth and Mehl [19] have shown that the criterion for the LSDA to work is less restrictive than that for the validity of the gradient corrections. In response to this setback, various functional forms for $\epsilon_{xc}(n, \nabla n, \ldots)$ have been developed giving what are known as generalized gradient approximations (GGA) for exchange and correlation. These are of two types. The first uses known properties of the exact E_{xc} to construct a functional, whereas the second semiempirical type of functional involves parameters that are fitted to particular electronic systems. One of the first type in common use in simulations is PW91 due to Perdew and Wang [20], but more sophisticated functionals have since been developed by Perdew and coworkers [21]. A popular semiempirical functional uses the formula for exchange due to Becke [22] and the correlation functional of Lee, Yang, and Parr [23]. Currently available GGAs achieve impressive accuracy for features of atomic, molecular, and solid-state systems. For example, a mean absolute error of 5 or so kcal/mol for the atomization energies of the G2 set of molecules is obtained, and agreement for lattice constants of a group of solids averages a few hundredths of Å [21, 24]. However, average measures of errors over a wide range of

systems are not particularly useful when a DFT simulation of a new system is performed on which basis predictions of properties are made. In such a situation, much experimentation is needed on electronically similar systems in order to estimate the degree of confidence in the results.

The exact and universal E_{xc} is a worthy but unrealistic goal, but recent efforts suggest that steady progress can be expected toward an approximate functional achieving 1 kcal/mol accuracy for binding energies. Developments may include methods based on Hartree-Fock or its best local equivalent, the optimized effective potential [25–27], to capture the exchange energy that is the bulk of E_{xc}, with the remaining correlation part handled with an approximate functional [27]. Particularly important developments for the study of clusters and cluster-assembled solids are the introduction of nonlocal effects into E_{xc}. Functionals of the type (27) are a spatial average involving the density and density gradients at a point so that it is only the density in the vicinity of $\mathbf{r}$ that enters the corresponding $V_{xc}(\mathbf{r})$, and the approximation for exchange and correlation is still a local approximation. Consequently, the long-range van der Waals interaction due to the correlated motion of electrons in distant even nonoverlapping systems is not accounted at this level of DFT. The van der Waals energy is likely to be an important contribution in the interaction energy of stable neutral clusters, those very clusters that are candidates for cluster-assembled materials. A number of approaches for introducing nonlocality into E_{xc} have been proposed and tested on separated atoms and molecules, atoms and molecules outside surfaces, and interacting slabs, with encouraging results [28].

The density functional theory presented here is a theory for the ground state. The results that are on a firm footing are the electron density and the ground state energy and quantities that can be obtained from these including equilibrium structures, deformation energies, and molecular dynamics within the adiabatic approximation. Properties that involve excited electronic states are not accessible through DFT and require the time-dependent generalization of DFT established by Runge and Gross [29], but applications to realistic systems are not yet commonplace [30]. In addition to the ground state density and energy, the Kohn-Sham scheme also yields the Kohn-Sham energy eigenvalues, and it is tempting to use the differences in these eigenvalues as estimates of one-electron excitation energies, in the same way that Hartree-Fock energy eigenvalues are used through Koopman's theorem. However, the Kohn-Sham eigenvalues appear to be artifacts of the theory with no real physical interpretation except for the highest occupied energy, which has been shown to equal the ionization potential of a localized system such as a molecule or cluster or the work function of an extended system [31]. For localized systems, the eigenvalues of lower occupied states are found to be above the one-electron excitation energies to the vacuum, and those that are unoccupied are below, and this is not a consequence of using an approximate E_{xc} although the LDA does accentuate the differences. Corresponding discrepancies occur for extended systems. In particular, a Kohn-Sham band structure calculation for a semiconductor underestimates the band gap. Nevertheless, the Kohn-Sham eigenvalues give useful information on the electronic structure. The sum of the Kohn-Sham energy eigenvalues is a major ingredient in the ground state energy (24), and the spectrum of eigenvalues of a cluster or the density of states for a solid often gives insights into the reason for stability of one structure over another, and similar questions.

2.1.3. Orbital-Free Approach

The Kohn-Sham approach circumvents the need of an explicit density functional for the electron kinetic energy, or at least $T_s[n]$ which is the major part of it, by introducing the kinetic energy of a set of noninteracting fermions. This allows the value of $T_s[n]$ to be determined precisely, for the particular density of interest, but at the cost of handling not just the density distribution but a set of N mutually orthogonal single particle orbitals, a task that becomes expensive in computational time and memory for large systems. An alternative scheme more in the spirit of the Thomas-Fermi approach outlined earlier uses an explicit but approximate functional for $T_s[n]$ and offers considerable savings. The system is characterized by $n(\mathbf{r})$ alone, and the method is free of the set of Kohn-Sham of orbitals. This is the orbital-free method that is seeing a number of applications [32].

The orbital-free method uses an approximate form for $T_s[n]$, as well as for $E_{xc}[n]$. $T_s[n]$ has seen much less study than the exchange-correlation functional, and approximations are still rather primitive; in addition, in most systems of interest the value of T_s is more than an order of magnitude greater than E_{xc} and so approximations to it are likely to have more serious consequences. Nevertheless, the savings in computational effort compensate for this in carefully chosen systems, and useful results are being obtained in cases where the full Kohn-Sham procedure would not be feasible at present, for example, for systems of many particles and when long molecular dynamics runs are required.

Approximate kinetic energy functionals are usually designed to give known limiting forms of the exact kinetic energy of a noninteracting electron gas. The kinetic energy in the limit of slowly varying density is known in the form of an expansion in gradients of the density [12],

$$T_s[n] = T^{(0)} + T^{(2)} + T^{(4)} + \cdots \tag{28}$$

where the leading term

$$T^{(0)} = T_{TF} = c_0 \int d\mathbf{r}\, n^{5/3} \tag{29}$$

is the Thomas-Fermi result,

$$T^{(2)} = \frac{1}{72} \int d\mathbf{r}\, \frac{(\nabla n)^2}{n} \tag{30}$$

and subsequent terms are higher order in density gradients, but interestingly, $T^{(4)}$ does not involve a fourth derivative of the density and can be evaluated with knowledge of only n, $|\nabla n|^2$, and $\nabla^2 n$. The von Weizsäcker functional

$$T_{vW}[n] = 9\,T^{(2)} \tag{31}$$

is exact for one and two-electron systems and is generally acknowledged to be appropriate in circumstances where the density is rapidly varying, such as in the exponentially decaying tail outside a localized system. Also known is the form of the kinetic energy functional for a weakly perturbed uniform electron gas in terms of density response functions

$$T_s[n] = T^{(0)}[\bar{n}] + \frac{1}{2\Omega} \sum_{\mathbf{q}\neq 0} \tilde{K}^{(2)}(\mathbf{q})\tilde{n}(\mathbf{q})\tilde{n}(-\mathbf{q}) + \cdots \tag{32}$$

The first term in (32) is the kinetic energy of the uniform gas of density $\bar{n}$ and volume Ω, and the next term is written in terms of the Fourier coefficients of the density, $\tilde{n}(\mathbf{q})$, and $\tilde{K}^{(2)} = \chi_0^{-1}(q)$, where χ_0 is the familiar Lindhard screening function. Higher order terms in (32) involve increasingly complicated response functions.

Approximate functionals based on these limiting cases have been used with some success in variational density functional treatments of simple metal systems where the electron density varies rather slowly. Various linear combinations of T_{TF} and T_{vW} [33, 34], as well as the first two terms in the gradient expansion [35] have been tried, but these functionals fail to give any shell structure in the electron density of an atom or Friedel oscillations in the screening cloud of an atomic impurity in an electron gas. Linear response theory does lead to Friedel oscillations around an impurity, and functionals incorporating this limit as well as the slowly and rapidly varying limits discussed above have been proposed. Two of these that use different *ansatz* [36, 37] have been used by Smargiassi and Madden [38] in studies of Na and Al, and for a variety of static and dynamic properties of the solid metals, the two types of kinetic energy functional give satisfactory results, although, as expected, the agreement with experiment for Al is the poorer because of the larger valence leading to stronger variations in the electron density. The functionals we have discussed so far can be used very efficiently; a calculation of the total energy requires order-N operations for a system of N atoms. But, there is a difficulty with the application of these functionals beyond bulk simple metal systems for which conventional second-order perturbation theory using pseudopotentials works fairly well regardless of density functional methods. This is the

need for a uniform "reference" electron density for the system in order to meet the linear response limit. In the case of bulk simple metal systems, the mean electron density will suffice for the reference, but for systems with directional bonds the bonding is associated with variations in the electron density, and for surfaces or clusters of even simple metals the electron density varies from bulk values inside to zero away from the surface, and there is no appropriate uniform reference density. A type of functional that overcomes this difficulty somewhat was introduced in 1985 by Chacon, Alvarellos, and Tarazona [39] (CAT) and has seen development by this group since then [40].

The new idea in the CAT functional is to use in an admixture of the Thomas-Fermi and von Weizsäcker functionals not only $n(\mathbf{r})$ but also a functional of the density

$$\bar{n}(\mathbf{r}) = \int d\mathbf{r}' \, n(\mathbf{r}') \, \omega[|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})] \tag{33}$$

which can be viewed as an average density in the vicinity of $\mathbf{r}$. The weighting or sampling function $\omega(r, n)$ is chosen so that the linear response limit is met in the case of a weakly perturbed gas. This procedure introduces nonlocality into the kinetic energy functional and deserves more attention than it has thus far received, but this comes at the cost of significantly greater computational cost for its evaluation than the simpler functionals. Whereas the evaluation of the local functionals requires order-N operations, the nonlocal one requires order-N^2, because for every $\mathbf{r}$-point, the $\bar{n}(\mathbf{r})$ must be evaluated, which requires an integral of the weighted density over all space. An approximate form of this nonlocal kinetic energy functional, which has been applied with some success to liquid metals, preserves the main CAT idea of using an average density $\bar{n}(\mathbf{r})$ but with the weighting function

$$\omega[|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})] \approx \omega(|\mathbf{r} - \mathbf{r}'|, \bar{n}) \tag{34}$$

where $\bar{n}$ is again some constant reference density appropriate for the system. The reversion to a uniform reference density restricts the application of the functional to systems with fairly uniform density, but the $\bar{n}(\mathbf{r})$ can now be obtained efficiently by, for example, a single fast Fourier transform and the method scales as order-N.

The functionals described above are designed to yield a total kinetic energy when a given electron density is substituted. However, to be useful in *ab initio* total energy calculations, the functional must also be used to obtain the ground state electron density either by direct minimization of the total energy functional or by solving the Euler-Lagrange equation. Consequently, a useful $T_s[n]$ should incorporate known limiting forms but have, in addition, accurate functional derivatives. Wang et al. [41] built this consideration into their study of a generalized gradient approximation for $T_s[n]$. They adopted a form proposed earlier by DePristro and Kress [42]

$$T_s[n] \approx c_0 \int d\mathbf{r} n(\mathbf{r})^{5/3} P(X, Y, \ldots) \tag{35}$$

where

$$X = \frac{1}{72 c_0} \left| \frac{\nabla n}{n^{4/3}} \right|^2 \tag{36}$$

is a dimensionless measure of the density gradient, and Y correspondingly for the Laplacian of the density, and so on for higher derivatives. DePristro and Kress [42] used a Padé approximant form for $P(X)$ with parameters chosen to satisfy some limiting cases and to obtain agreement with the total kinetic energy for some atomic all-electron systems. The objective of kinetic energies for all-electron systems to an accuracy useful in cluster and condensed matter physics of 1 eV or less is too ambitious for this primitive functional, and instead Wang et al. [41] designed a similar functional for use with relatively smooth densities such as the valence electron density of pseudoatoms. They used model systems and adjusted the functional in order to obtain agreement for $T_s[n]$ and its functional derivative, which apart from an additive constant is the Kohn-Sham potential for the system. Their functional worked reasonably well for group IV atoms, and there were indications that a generalized

gradient approximation like Eq. (35), which reduced to the gradient expansion through fourth-order for slowly varying densities and to $T_{vW}[n]$ in the density tail outside localized systems, would be a promising development.

The future of the orbital-free method depends largely on the development of improved kinetic energy functionals. Its merit is the efficiency of calculation, in speed, allowing molecular dynamics simulations for long times, and memory requirements allowing simulation of large samples, but the main drawback is the error in T_s, which is such a large part of the total energy. Realistic simulations can only be expected at present for simple metal systems. Most approximate functionals incorporate $T_{TF}[n]$ and give good results for slowly varying densities as would be found in bulk simple metal systems when the electron–ion interaction is described by a pseudopotential. Pseudopotentials developed from first principles give an accurate description of the valence electronic structure of the whole of the periodic table, but these are nonlocal in the sense that different potentials act on orbitals with different angular momentum. This complication is necessary if the details of the energy spectrum and orbitals for different angular momenta are to be reproduced, as we shall see in the next section. Because the density in the orbital-free approach is not, as the name suggests, broken down into contributions from individual orbitals, the electron potential energy due to the interaction with the ions is $\int d\mathbf{r} V_{ext}(\mathbf{r}) n(\mathbf{r})$ and the external potential must be local, or at least, the angular momentum dependent nonlocal pseudopotentials cannot be used. This is not to say that a local pseudopotential that gives an accurate description of the valence electronic structure cannot be developed, but at present the restriction to local pseudopotentials is a limitation of the orbital-free approach.

2.2. Pseudopotentials

2.2.1. Model Pseudopotentials

The development of accurate pseudopotentials greatly simplified the application of density functional methods to *ab initio* materials simulation, and it is to these pseudopotentials that we now turn. The idea of pseudopotentials developed in the early days of condensed matter physics from questions such as the following. In view of the very strong attractive potential that electrons experience in the core region of an atom, why are there simple metals such as Na and the other alkalis, Mg, Al, which behave very like a Sommerfeld model in which the potential inside the system is constant? Why also is the valence-conduction band gap in Si and Ge so small, ≈ 1 eV, suggesting a weakly perturbing periodic potential? The answer to these questions was discovered in the 1960s and simple empirical and semiempirical pseudopotentials were developed [5]. Accurate, transferable, first principles pseudopotentials of the sort that are used now in *ab initio* simulations were first developed in the early 1980s [6, 7].

The valence electron orbitals must be orthogonal to the core electron orbitals of the same angular momentum, and consequently they must oscillate as functions of r in the core region. This leads to valence electrons having a large expectation value of the kinetic energy. But, the potential experienced by the valence electrons is strongly attractive in the core region and so the expectation value of the potential is correspondingly large but negative. At a few eV, the relatively small binding energy of the valence electrons $\epsilon_{val} = \langle T \rangle + \langle V \rangle$ is the result of a large degree of cancellation between the kinetic and potential contributions. This cancellation of large energies can be removed by introducing a potential, the pseudopotential, that is much weaker than the full Kohn-Sham effective potential for the atom and gives the valence state at the right energy, which means that the corresponding valence pseudo-orbital will therefore have the correct shape in the outer region of the atom. The pseudopotential is not unique, different ones will give different shapes for the pseudo-orbital in the core region, and it is useful to choose a potential that will give a smooth pseudo-orbital in the core with no orthogonality "wiggles." This potential will have the valence state as its lowest state for the given angular momentum, and the maximum amount of cancellation between kinetic and potential energies will have been removed. A spherical potential well provides a simple illustration of this procedure. A well of radius 1 a.u. and 32 a.u. deep has bound s-states at -28.1 a.u., and -16.8 a.u., these are the "core" levels, and a "valence" state at

$\epsilon_{val} = -0.34$ a.u. A much shallower well 12.3 a.u. deep with the same radius has a valence state again at -0.34 a.u. and a single deeper s-state at -8.9 a.u. Continuing, a well only 2.3 a.u. deep has only the valence state bound at $\epsilon_{val} = -0.34$ a.u. The valence orbitals for these three potentials are shown in Fig. 1. The valence orbital for the original deep well being orthogonal to two deeper s-states has two nodes, that for the intermediate well has one node, and the one for the shallow well is nodeless, but all behave in the same way outside the well, and all correspond to an energy of -0.34 a.u.

An example of an early pseudopotential for describing the electron–ion interaction is the so-called empty core potential of Ashcroft [43]

$$\begin{aligned} v(r) &= 0, \qquad r < r_c \\ &= -\frac{Z_v}{r}, \qquad r > r_c \end{aligned}$$

where Z_v is the valence. The potential is local acting in the same way on all angular momentum components and contains a single parameter, the core radius, r_c, which was fitted so that the potential gave good agreement for selected properties. The potential was then used to investigate other properties. A further example is the model potential of Heine and Abarenkov [44]. They recognized that the pseudoion potential may have to be different for different partial waves if the spectrum of valence energy levels is to be correct for all angular momenta. If $V_l(r)$ is the potential for lth partial wave, the total pseudoion potential can be expressed in terms of the operator $|l, m\rangle\langle l, m|$ which projects out of the wavefunction the (l, m) angular momentum component

$$V_{ion}^{PS} = \sum_{l,m} V_l(r)|l, m\rangle\langle l, m| \tag{37}$$

The individual $V_l(r)$ were taken to be $-Z_v/r$ outside the ion core, and a constant inside chosen so that the lowest valence energy eigenvalue for angular momentum l was given correctly. Both the Ashcroft potential and the non-local Heine-Abarenkov potential were regarded as weakly scattering and used within second-order perturbation theory for calculating total energies and related quantities.

2.2.2. First-Principles Normconserving Pseudopotentials

Two problems with these early potentials were (i) their "transferability" and (ii) the division of the electron distribution between the core region and the outside region of the atom.

The first issue, (i), is the degree to which the pseudopotential, which by design describes accurately the valence states of the atom in one environment, can be transferred to give

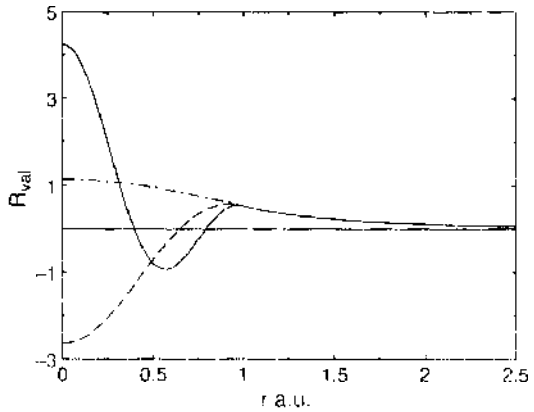


Figure 1. Valence s-orbitals each corresponding to energy $\epsilon_{val} = -0.34$ a.u. for square well potentials with radii 1 a.u. but different depths V_0. The full line for $V_0 = 37$ a.u. has two more tightly bound s-states; the dashed line for $V_0 = 12.3$ a.u. has one deeper s-state; the dash-dot line for $V_0 = 2.3$ a.u. is the lowest s-state for the potential. Because they correspond to the same energy, the three orbitals have the same shape outside "the core," that is, beyond $r = 1$ a.u.

an equally good description in another situation. For example, the pseudopotential may be chosen to give the correct energies for the valence states of the free atom, but if it is to be used in the corresponding solid to compute the conduction and valence bands, it should describe the electron–ion interaction not only at the discrete atomic bound state energies but over a range of energies covering the energy bands in the solid. In this respect, we note that the solution of the radial Schrödinger equation for energy E can be obtained beyond some radius, say r_c, apart from a normalizing factor, by integrating out, if the logarithmic derivative of the solution at r_c is known:

$$\Gamma_l(E) = \frac{d}{dr} ln[r\, R_l(E, r)]|_{r_c} \tag{38}$$

If $\Gamma_l(E)$ is known over a range of energies, the correct exterior solutions can be obtained over that range.

The second difficulty, (ii), concerns the so-called norm of the valence pseudo-orbital. A realistic description of the bonding in the region between atoms requires the correctly normalized pseudo-orbital $R_l^{ps}(\epsilon_{val}, r)$ to have not only the right shape but also the right amplitude outside the core. Furthermore, we require the pseudo-orbital to contribute the same amount of charge in the core region as the full orbital $R_l(\epsilon_{val}, r)$, otherwise the electrostatic contributions to the Kohn-Sham effective potential will be badly in error. The requirement that the full and the pseudo-orbital have the same norm

$$\int_0^{r_c} dr[rR_l^{ps}(\epsilon_{val}, r)]^2 = \int_0^{r_c} dr[rR_l(\epsilon_{val}, r)]^2 \tag{39}$$

where ϵ_{val} is the valence energy eigenvalue for angular momentum l, solves both of these difficulties at one stroke. Manipulation of the radial Schrödinger equation gives a relationship between the norm and the energy dependence of $\Gamma_l(E)$ [6]. Because the pseudopotential gives the correct valence bound state energy, ϵ_{val}, it also gives correctly $\Gamma_l(\epsilon_{val})$. If the pseudopotential also conserves the norm $\frac{d}{dE}\Gamma_l(E)|_{E=\epsilon_{val}}$ is also correct, and a Γ_l correct over a range of energy around ϵ_{val} is guaranteed.

The first principles, normconserving pseudopotentials first introduced by Hamann, Schlüter, and Chiang [6], and expanded upon by Bachelet, Hamann, and Schlüter [7], form the basis of the pseudopotentials presently applied in *ab initio* simulations. The pseudopotentials are obtained by fitting to aspects of a Kohn-Sham calculation of the free atom including all electrons, not just the valence electrons. Often, an artificial electron configuration is chosen including some valence states that would not be occupied in the ground state, so that there is some occupancy of $l = 0,\ 1,\ 2$ valence states that could well be occupied in a situation to which the pseudopotential is transferred (e.g., a cluster of atoms). For instance, the ground state configuration of the free Si atom is the neon core $+3s^2 3p^2$, but in a solid we expect some occupancy of d-like states, and the pseudopotential is more likely to describe the situation if it reproduces a free atom configuration such as the neon core $+3s^2 3p^{1.75} 3d^{0.25}$.

Next, for each orbital angular momentum l-value, the all-electron valence radial wavefunction is modified inside the core so that orthogonality wiggles are removed giving a pseudo-orbital R_l^{ps}. The corresponding potential that has this orbital as its solution for the true valence energy eigenvalue is obtained by inverting the radial Schrödinger equation. This procedure is far from unique. The flexibility in modifications that can be made inside the core and give a R_l^{ps} and an accompanying pseudopotential that still give the required valence energy can be used to advantage. If the modification leads also to the conservation of the norm (39), then R_l^{ps} and R_l when properly normalized will be identical outside the core, and the transferability of the potential will be enhanced. This gives the Kohn-Sham effective potentials, one for each l-value, that yield the corresponding pseudo-orbital. These potentials include the Hartree and exchange-correlation contributions due to the electron density of the pseudoatom, which can be subtracted off leaving the pseudopotential for the bare ion, V_{ion}^{ps}, a so-called normconserving pseudopotential. But, there still remains flexibility in the core region, and the requirements above for a normconserving pseudopotential can be met for a whole family of potentials. A simple procedure for choosing from these a normconserving potential that is smooth and is particularly well suited for use with a plane wave basis set is due to Troullier and Martins [45]. Pseudopotentials generated with their

procedure are extremely efficient in those cases for which the expansion in plane waves converges slowly, such as first row elements and $3d$ transition elements, and their procedure is widely used in *ab initio* simulations. The ion pseudopotential is usually written as

$$V_{ion}^{ps} = V^{loc}(r) + \sum_{l,m} \Delta V_l^{nonl}(r)|l, m\rangle\langle l, m| \tag{40}$$

where $V^{loc}(r)$ is a local potential that falls off outside the core as $-Z_v/r$ and acts equally on all angular momentum components, and the nonlocality is contained in the deviations $\Delta V_l^{nonl}(r)$ for just a small number of l-values.

Simulations that employ first principles normconserving pseudopotentials such as those of Troullier and Martins do not use them directly in their nonlocal form, Eq. (40), as this would be very expensive computationally. Instead, a separable form suggested by Kleinman and Bylander [46] is used in which $\Delta V_l^{nonl}(r)|l, m\rangle\langle l, m|$ in Eq. (40) is replaced by

$$\Delta V_l^{KB} = \frac{|\phi_{lm}^{ps}\Delta V_l^{nonl}\rangle\langle\Delta V_l^{nonl}\phi_{lm}^{ps}|}{\langle\phi_{lm}^{ps}\Delta V_l^{nonl}\phi_{lm}^{ps}\rangle} \tag{41}$$

where the ϕ_{lm}^{ps} are pseudoatom wavefunctions at the reference energy. With this substitution, if there are N_b basis functions, all N_b^2 matrix elements of V_{ion}^{ps} can be obtained from the N_b quantities $\langle\Delta V_l^{nonl}\phi_{lm}^{ps}|\phi_{basis}\rangle$, a great saving.

The "ultrasoft" pseudopotential scheme of Vanderbilt [47] is also becoming widely used. Here, the conservation of the norm in the core region is relaxed, which allows much smoother pseudo-orbitals in the core than would otherwise be possible. This is particularly useful for the first row elements and the $3d$ transition and rare earth elements. Taking the oxygen atom in the first row as an example, we note that the $2s$ orbital can be pseudized in the core region while preserving the norm because of the orthogonality to the $1s$ orbital, resulting in a smooth s pseudo-orbital that is the lowest s-state of the pseudopotential. The same is not possible for the $2p$ orbital, there is no lower p-state in the core, and the $2p$ orbital cannot be smoothed or pseudized in the core to any significant degree while preserving the norm. The [illegible]n is similar for the $3d$ orbitals and the $4f$ orbitals of the third row transition and rare earth elements, respectively. The smoother pseudo-orbitals of the ultrasoft potentials can be represented by a much smaller set of plane waves than if the norm were conserved. Relaxing the norm means that the charge density can no longer be given simply by the sum of squares of the orbitals, and additional terms are necessary. These are formulated in Vanderbilt's scheme.

2.2.3. First-Principles Local Pseudopotentials

In the orbital-free method described earlier, the electronic energy is expressed in terms of the electron density, $n(\mathbf{r})$, the electron–ion interaction energy has the form $\int d\mathbf{r}\, V_{ext}(\mathbf{r})n(\mathbf{r})$, and there is no decomposition of the density into orbitals. Consequently, a pseudopotential describing the interaction with the ions must be local. The high efficiency of the orbital-free scheme has stimulated some interest in the development of first-principles local pseudopotentials. Gonzalez and coworkers have constructed local pseudopotentials suitable for ions in a metallic environment by fitting to a density functional calculation of the ion positioned in an atom-sized cavity in a uniform electron gas [48]. In this procedure the pseudopotential depends not only on the type of ion but also on the mean electron density of the system to be studied, and its use is mainly restricted to simple metal systems. It has been [illegible]successfully in simulations of molten simple metals and alloys [48]. A scheme for obtaining [illegible]al equivalent of a given first-principles nonlocal pseudopotential by solving the Kohn-Sham equations inversely has been presented by Wang and Stott [49] and applied to group IV elements. The results for Si, Ge, and Sn, while certainly inferior to those of the parent nonlocal potential, are encouraging, and the potentials are adequate for use in orbital-free total energy simulations given the current deficiencies in kinetic energy functionals.

2.3. Synthesis

Within the framework of a density functional theory treatment of the electron–electron interactions, a number of elements need to be drawn together if reliable simulations of clusters

and cluster-assembled solids are to be performed. At the outset, some choices have to be made about the sort of approach to be taken. Foremost of these is the choice between including all electrons, core as well as valence electrons, in the calculation of the electronic energy, a so-called all-electron calculation, or removing the core electrons and treating the valence electrons only in a pseudopotential calculation. At first sight, the former, with all the electrons moving in the external field due to the point charge nuclei, would seem to be preferred, as this approach involves no uncertainities associated with the quality of the pseudopotentials. Many all-electron computer codes are readily available based on a linear combination of atomic orbitals (LCAO) for representing the Kohn-Sham one-electron wavefunctions, and employing Gaussian, Slater orbital, or numerical basis sets.

However, there are drawbacks to this approach. The large contributions of the core electrons that are included in such calculations to the total energy and the electron density tend to obscure the valence electronic structure that is responsible for the clustering and solid-state effects. Furthermore, the very different length scales present in the all-electron wavefunctions makes their representation in terms of a basis set prone to error. The deep core orbitals have an extent $\simeq 1/Z$ a.u., and variations on this scale occur in the valence orbitals because of orthogonality: these variations must be represented well because there are large energies associated with them. But, in addition, there are variations in the valence orbitals on the scale of the interatomic distance that incorporate the bonding features we wish to investigate. In contrast, the valence pseudowavefunctions have the rapidly varying core electron features removed and are comparatively smooth. They contain only a single length scale $\simeq 1$ a.u. and are relatively easy to represent. Also, the choice of which atomic orbitals to include in the LCAO is somewhat subjective. Finally, the LCAO basis set required for an all-electron calculation depends explicitly on the positions of the atoms; the basis set changes if the atoms are moved. This feature creates difficulties in the calculation of the forces on the atoms that are needed if the atom positions are to be relaxed to determine the equilibrium arrangement or if molecular dynamics simulations are to be performed [50]. Overall, the balance currently favors the plane wave pseudopotential approach for cluster assembly studies, and most of the studies that have been reported use this approach. For the simulation of isolated clusters the choice is more open, but we again prefer the plane wave pseudopotential approach for which checks of convergence with respect to the basis set are decisive. Even so, all-electron calculations are useful in testing the pseudopotential, provided uncertainties due to the LCAO basis set are recognized.

We now briefly review the various ingredients of the plane wave, pseudopotential, total energy simulation method. A fuller discussion can be found in the comprehensive article of Payne et al. [51].

2.3.1. The Pseudopotential

The plane wave pseudopotential approach requires a choice of pseudopotential and the testing of it. The electrons in the free atom to be treated in the calculations as valence electrons subject to the pseudopotential have to be decided. In some cases, it is not sufficient to take just the outermost electrons, but so-called semicore electrons should also be included (e.g., for Cu, to the outer 4*s* electron should be added the 10 3*d* electrons that are close by in energy and easily polarized). But, this is at the cost of a harder pseudopotential and a larger total number of electrons to be treated. When there is significant overlap between the core and valence electron densities, such as for Na, a correction [52] should be applied to account for the nonlinear dependence of the exchange-correlation potential on the electron density. Extensive testing and fine-tuning of the pseudopotential is required to determine error limits on calculated total energies. Usually, tests are conducted on free atom quantities comparing results such as excitation energies for the pseudoatom and the all-electron atom. Tests of the transferability of the pseudopotential into molecular and bulk systems for which results of other calculations or of experiment are available should also be conducted.

2.3.2. Superlattice Geometry

A set of plane waves that have the periodicity of a superlattice is used to represent the Kohn-Sham orbitals. The atoms of interest, whether they form an isolated cluster or are the atoms

in a unit cell of a crystal, are positioned in a unit cell of the superlattice, so that periodicity is forced on the system. This allows the Kohn-Sham orbitals to be expanded in a discrete set of plane waves that have the periodicity of the superlattice. In choosing the size of the supercell and the sample of atoms to be positioned in the cell, it must be borne in mind that the system for which the energy, electron density, and so forth, is being calculated is the periodic superlattice array of atoms. To simulate an atomic cluster, the atoms must be placed in a supercell with enough empty space around the cluster so that the interaction between clusters in the periodic array is negligible. A cluster positioned in a supercell and surrounding images of the cluster in the superlattice are illustrated in Fig. 2. One simple test of the strength of the interaction between clusters is the effect on the energy of changing the orientation of the cluster in the supercell. Similarly, slabs of atoms used to investigate surface properties must be separated by enough vacuum to make the interaction between adjacent slabs in the periodic array negligible. The Kohn-Sham orbitals, $\psi_{n,\mathbf{k}}(\mathbf{r})$, will be Bloch waves for the superlattice, labeled by the band n and the wavevector $\mathbf{k}$ spanning the superlattice Brillouin zone.

2.3.3. Plane Wave Basis

The supercell geometry allows the orbitals to be expanded in the discrete set of plane waves that have the periodicity of the superlattice,

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n,\mathbf{k},\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \tag{42}$$

where the G-vectors are the reciprocal lattice vectors of the superlattice. In practice, the sum over G-vectors is cut off at a G_c given by $G_c^2 = E_c$, where E_c is known as the cutoff energy. Figure 2 illustrates the reciprocal lattice associated with the real-space superlattice, which is also shown. The number of plane waves included in the sum in Eq. (42) is given by $N_{pw} = \frac{4\pi}{3} G_c^3/\Omega_B = \Omega E_c^{3/2}/6\pi^2$, where Ω is the volume of the supercell and $\Omega_B = (2\pi)^3/\Omega$ is the volume of the corresponding Brillouin zone. The energy cutoff depends on the pseudopotential and can range from a few Ryd. for a soft pseudopotential, for say Na, to a few tens of Ryd. for the Troullier-Martins pseudopotentials for first row elements C and O, and ≈70 Ryd. for Cu and Zn [45]. The number of plane waves in the basis set can be large, for example, calculating the electronic energy of a cluster involving Cu in a cubic supercell of edge 30 a.u. requires $N_{pw} \approx 10^5$. It is only because of the fast Fourier transform algorithm that this size of plane wave basis set can be manipulated.

The periodicity of the superlattice leads to energy bands with k-vectors spanning the super Brillouin zone. The sum of Kohn-Sham energy eigenvalues and the electron density involved in calculation of the total energy require an integration over the Brillouin zone as well as a

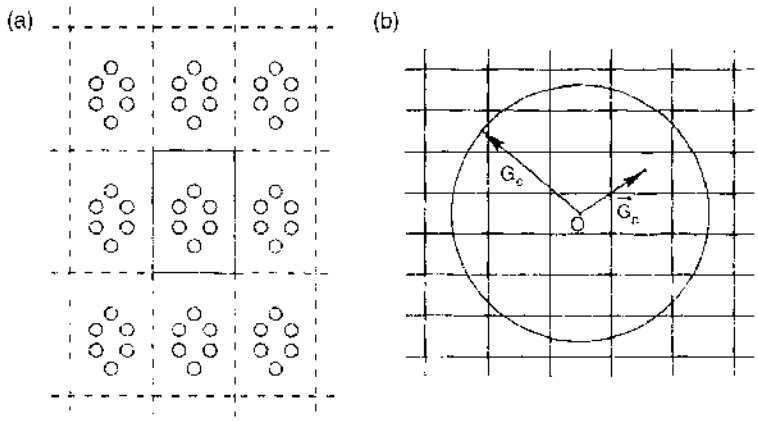


Figure 2. Illustration of real-space and reciprocal-space situations involved in the simulation of a cluster. The cluster positioned in a supercell and the images of the cluster in the surrounding superlattice are depicted in (a). The corresponding reciprocal-space is illustrated in (b). The wave-vector G_c giving the cut-off of the plane wave expansion is shown along with one of the reciprocal lattice vectors included in the sum.

sum over occupied bands. The integration is replaced by a weighted sum over special **k**-points, for each of which $\psi_{n,\mathbf{k}}$ is calculated [53, 54]. The number of **k**-points required to achieve convergence of the energy and other calculated quantities must be the subject of tests; the smaller the supercell, the larger the super Brillouin zone, the more **k**-points required to adequately sample the bands. Exceptions are when the bands are artificial results of the enforced periodicity, such as for isolated clusters. It is customary to use only the $\mathbf{k} = 0$, Γ-point for clusters.

2.3.4. Ground State Energy Calculation

The number of plane waves describing the orbitals is too large for direct diagonalization of the Hamiltonian to be feasible. Instead, the calculation of the total energy for a given configuration of atoms in the supercell is performed variationally by iterating from an initial trial state described by a set of $c_{n,\mathbf{k},G}$'s, until an energy minimum corresponding to the ground state is reached. The initial state can be found by using an approximate electron density to construct the Kohn-Sham Hamiltonian and diagonalizing this with a truncated set of plane waves. Preset tolerances on changes in the total energy and perhaps other quantities such as the forces on the ions are used to terminate the minimization. At each step in the iterations, an adjustment to the wavefunctions is made using some form of damped dynamics based on the direction of steepest descent of the energy, which can be calculated by operating with the current Hamiltonian, while maintaining the orthogonality of the Kohn-Sham wavefunctions [55]. The mass associated with the dynamics, the degree of damping, and the quality of the initial state all affect the approach to the minimum and should be the subject of experimentation.

2.3.5. Atomic Arrangement

At this point, the forces on the ions may be calculated using the Hellmann-Feynman theorem, but unlike the energy the error in the forces is first order in the difference in the electron density from the ground state density, and so care must be taken to obtain an accurate ground state. Using the calculated forces and some ficticious damped dynamics, the atoms may be moved to new locations to lower the total energy if relaxation to the equilibrium atomic arrangement is wanted. However, unless an exorbitantly expensive annealing process is simulated, the system will relax to a local energy minimum of which the system could have many. Consequently, care must be taken to explore all the atomic arrangements that are candidates for the ground state. If a molecular dynamics simulation is to be performed, the Newtonian equations of motion of the ions can be integrated over a time-step.

3. APPLICATIONS TO CLUSTERS AND CLUSTER-ASSEMBLED SOLIDS

Great strides have been made in the ability to calculate using density functional theory (DFT) the electronic and structural properties of clusters. *Ab initio* methods are able to provide in great detail the electronic density and the constituent Kohn-Sham one-electron energies and orbitals. Relaxation of the atomic positions enables the spectrum of isomers to be explored, and even the thermal properties of clusters can be studied using molecular dynamics techniques. Nevertheless, simple models, also based on DFT, are a source of understanding of many of the properties of clusters: the stability of some, the fragility of others, the influence of impurities, and the response to external perturbations.

3.1. Electronic Shells

3.1.1. Electronic Shells in Clusters of Simple Metallic Elements

A remarkable discovery by Knight and coworkers [56, 57] motivated the strong interest in clusters of the alkali metals. The abundance population in cluster beams obtained by gas aggregation techniques shows a non-monotonic variation as a function of cluster size, with maxima at some particular sizes $N = 2, 8, 20, 40, 58, 92 \ldots$, immediately followed by sharp drops in

the population. These sizes are known as *magic numbers*. The same authors found [57] that the magnitude of the ionization potential I drops substantially after the same *magic* sizes, and after $N = 18$.

It was immediately recognized that these features reflect electronic shell structure in the clusters, and that the magic sizes correspond to clusters with filled shells. Due to the weakness of the pseudopotential describing the effect of the ion core in alkali atoms, the conduction electrons (one per atom) behave as nearly-free electrons in the bulk metal, and a good first approximation is to smear out the ions uniformly. A corresponding model for the finite cluster has the ions replaced by a spherical distribution of constant positive charge density and radius R, that is

$$n_+(\mathbf{r}) = \bar{n}_+ \Theta(R - r) \tag{43}$$

where $\Theta(R - r)$ is the step function, with values 1 for $r < R$ and 0 for $r > R$. This is the so-called *spherical jellium model*. When the Kohn-Sham formulation of DFT is applied to this model [57–59], the valence electrons of the cluster move in a self-consistent potential well, which is the sum of several contributions

$$V_{KS}(\mathbf{r}) = V_{ext}(\mathbf{r}) + V_e(\mathbf{r}) + V_{xc}(\mathbf{r}) \tag{44}$$

Here, $V_{ext}(\mathbf{r})$ is the electrostatic potential due to the positive charge distribution $n_+(\mathbf{r})$ representing the ions, $V_e(\mathbf{r})$ is the electrostatic potential due to the electron density $n(\mathbf{r})$, and $V_{xc}(\mathbf{r})$ is the exchange-correlation potential discussed earlier for which the local density approximation (LDA) [4] is often used. In this self-consistent potential, the electrons group in shells that are characterized by a principal quantum number ($k = 1, 2, \ldots$) and a quantum number giving the orbital angular momentum [$l = 0(s), 1(p) \ldots$]. The degeneracy of each electronic shell, taking spin into account, is $2(2l + 1)$. For the spherical jellium model, the shells become filled in the sequence $(1s)^2\ (1p)^6\ (1d)^{10}\ (2s)^2\ (1f)^{14}\ (2p)^6\ (1g)^{18} \ldots$ where the superscript indicates the number of electrons required to fill the shell. The shells are separated by energy gaps, so clusters with the precise number of electrons required to fill shells are more stable than clusters of neighboring sizes, a behavior parallel to that of the inert gas atoms. The closed shell clusters predicted by the jellium model, $N = 2, 8, 18, 20, (34), 40, 58, 92 \ldots$ are in agreement with the observed magic numbers. The jellium model overestimates the magnitude of the gap between the $1f$ and $2p$ shells, and, therefore, also the stability of the $N = 34$ cluster. A simple modification of the model by smoothing the sharp surface of the positive background [60] reduces that gap and the stability of $N = 34$, leading to better agreement with experiment. The effective spherical potential well confining the electrons in the cluster has the form of a rounded square well, and the case of Na_{20} is shown in Fig. 3. This cluster

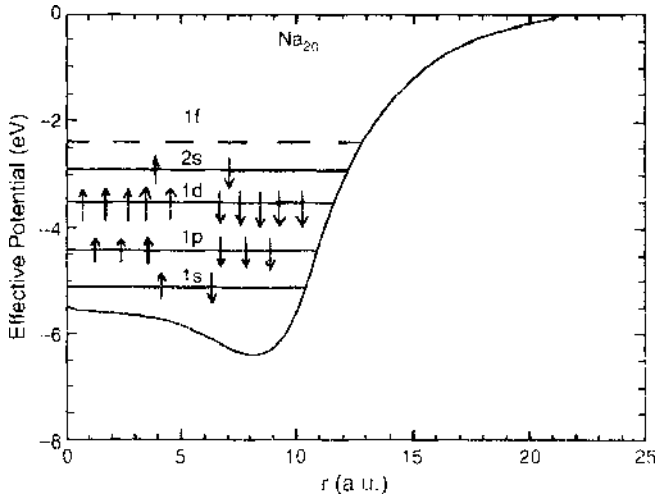


Figure 3. Self-consistent effective potential for Na_{20} in the spherical jellium model. The occupied electronic shells are represented by the continuous lines, and the lowest unoccupied shell, $1f$, by the dashed line.

has a closed-shell electronic configuration $(1s)^2\ (1p)^6\ (1d)^{10}\ (2s)^2$ and a HOMO–LUMO gap of 0.5 eV between the highest occupied molecular orbital ($2s$) and the lowest unoccupied molecular orbital ($1f$).

The cohesive energy per atom of a cluster X_N can be written in terms of the energies of the free atom, $E(atom)$, and the cluster, $E(X_N)$.

$$E_{coh}(X_N) = E(atom) - \frac{E(X_N)}{N} \tag{45}$$

When plotted as a function of cluster size, $E_{coh}(X_N)$ is a local maximum for the magic sizes. The following argument explains the larger abundance of the magic clusters. The variations of the abundance population in cluster beams arise from the evaporative cooling of warm clusters [57] and, as highly stable clusters evaporate less frequently than other clusters, their abundance grows at the expense of less stable clusters. This implies that the population of the magic clusters can be increased deliberately if evaporation of atoms is enhanced by laser heating of the flying clusters. Hansen and coworkers [61] have verified the sharpening of the abundance spectrum by laser heating of a beam of clusters of C_{60} molecules. In that case, laser heating enhanced the population of $(C_{60})_{13}$, which is a magic number in the family of $(C_{60})_N$ clusters, although in this family the magic numbers arise from atomic packing rather than electronic shell effects.

A quantity often used in the analysis of theoretical calculations that clearly displays the effect of the electronic shells on cluster stability is the second derivative of the calculated energy

$$\Delta_2(N) = E(X_{N+1}) + E(X_{N-1}) - 2E(X_N) \tag{46}$$

This quantity measures the relative stability of a cluster with N atoms in comparison with neighboring clusters of $N + 1$ and $N - 1$ atoms. Figure 4 shows sharp peaks in $\Delta_2(N)$ for the closed shell clusters. In addition, thermodynamic arguments [57] establish the following relationship between $\Delta_2(N)$ and the measured abundance populations P_N,

$$\ln\left(\frac{P_N^2}{P_{N+1}P_{N-1}}\right) = \frac{\Delta_2(N)}{k_B T} \tag{47}$$

where k_B is the Boltzmann constant and T is the nozzle tube temperature. We see that the predictions of the jellium model based on density functional theory account for the observed magic numbers of the alkali clusters.

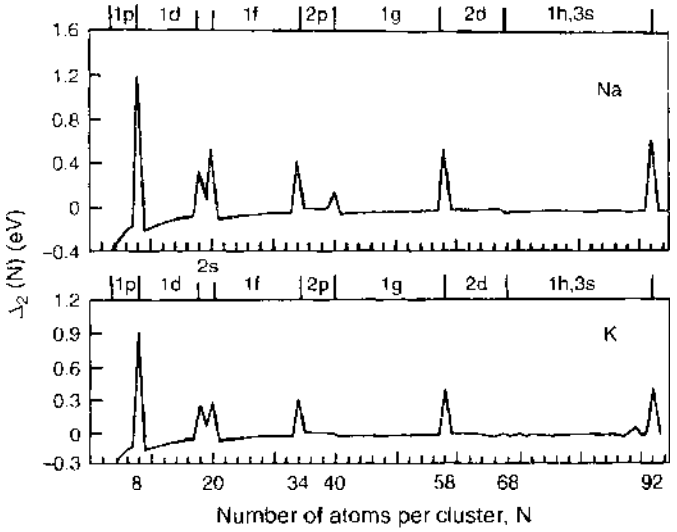


Figure 4. Relative stability $\Delta_2(N)$ [given by Eq. (46)] of sodium and potassium clusters in the jellium model, as a function of the cluster size. Peaks appear for the closed shell clusters.

Other observed features also reflect the magic numbers. For instance, drops after the magic numbers are seen in the measured ionization potential [57]

$$I = E(X_N^+) - E(X_N) \tag{48}$$

the energy required to remove one electron from the cluster. Results of jellium model calculations for the ionization potential display a local maximum for a cluster with closed shells, and a substantial drop afterwards.

The common explanation of the enhanced stability due to shell closing effects in alkali clusters is based on the results of the spherical jellium model for the ionic background, but more accurate DFT calculations taking explicit account of the granular structure of the ionic skeleton support the picture provided by the spherical jellium model [62–64]. Most of these calculations use a nonlocal pseudopotential to represent the electron–ion interaction, but studies have been restricted to small sizes, up to about $N = 20$, because the difficulties in performing a full optimization of the geometries grow quickly with N.

3.1.2. Electronic Shells in Large Clusters

When the cluster size increases, the number of valence electrons in the cluster also increases. Because the depth of the confining potential remains roughly constant, the gaps between electronic subshells become narrower [58], and for N sufficiently large, the discrete energy levels evolve into the quasicontinuous energy bands of the solid. The point at which this occurs, when electronic shell effects are no longer discernible, is of interest. Experiments indicate that shell effects remain important for clusters with a few thousand valence electrons [65–68]. When cluster abundances are plotted on a $N^{1/3}$ scale ($N^{1/3}$ gives the linear dimension of the clusters), the observed magic numbers appear at nearly equal intervals of 0.6. This periodicity in the appearance of the shell closings can be explained qualitatively [66]. An expansion of N in terms of K, the principal quantum number of the highest occupied shell, has a leading term proportional to K^3. One power of K arises from the sum over all shells up to K in order to obtain the total number of particles. A second power of K arises because the number of subshells in a shell increases approximately linearly with the shell index. Finally, the third power of K arises because the number of particles in the largest subshell also increases with K. Then

$$N_K = aK^3 \tag{49}$$

This qualitative argument is supported by theoretical calculations. When the number of electrons in the cluster increases, the number of electronic shells also increases. Even so, calculations for large clusters have shown that groups of energy levels bunch together and leave sizable energy gaps between bunches. DFT calculations reproduce this bunching effect and the $N^{1/3}$ periodicity, giving magic numbers in close agreement with experiment [69–71]. The results of calculations performed by Genzken [70] for large Na clusters in the spherical jellium model are shown in Fig. 5. A smooth interpolated function $E_{av}(N)$ that averages out the shell effects has been subtracted from $E_{jellium}(N)$, leaving the quantity

$$E_{shell}(N) = E_{jellium}(N) - E_{av}(N) \tag{50}$$

which highlights the shell effects and is plotted in Fig. 5. $E_{av}(N)$ is well represented by a liquid drop model as a sum of volume, surface, and curvature terms

$$E_{av}(N) = e_b N + a_s N^{2/3} + a_c N^{1/3} \tag{51}$$

The bulk energy per atom e_b can be obtained from calculations on the homogeneous electron gas, and a_s from the jellium model for a planar surface. But a_s and a_c can also be obtained from a direct fit to the calculated data for $E_{jellium}(N)$, as was done by Genzken [70]. The shell contribution $E_{shell}(N)$ obtained in this way oscillates with N, displaying sharp minima at the shell closing numbers [69, 70]. This is the shell effect with periodicity $\Delta N^{1/3} = 0.6$, and there are just small differences between the predicted and observed magic numbers [67, 68].

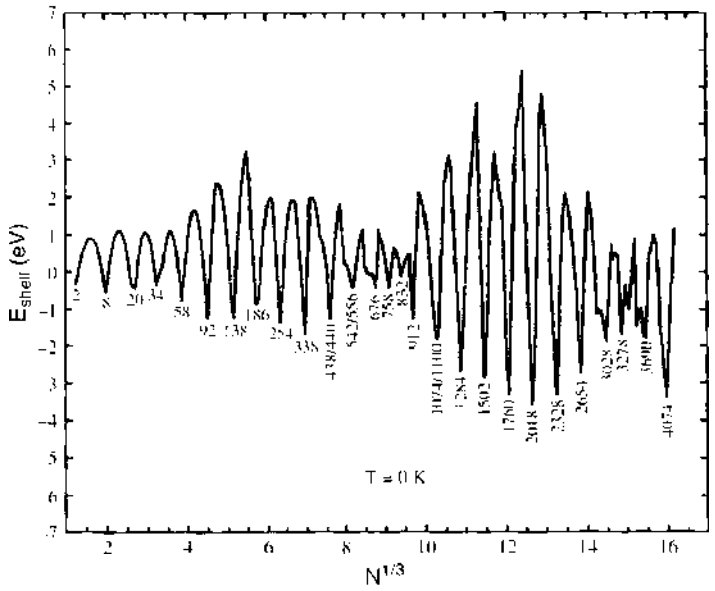


Figure 5. Periodically varying contribution of valence electrons to the binding energy of spherical sodium clusters. Magic numbers are indicated. Adapted with permission from [70]. O. Genzken, *Mod. Phys. Lett. B* 7, 197 (1991). © 1991, World Scientific.

In addition, there is a supershell effect with the amplitude of the shell oscillations varying periodically on a larger size scale $\Delta N^{1/3} = 6$. The first supershell node occurs in Fig. 5 near $N = 850$. Calculations by Nishioka et al. [69] give N around 1000. This node has been observed, although the experiments are not in complete agreement: the first node is located near $N = 1000$ in [67], and near $N = 800$ in [68]. Supercell structure was predicted by nuclear physicists to be a rather general property of a system consisting of a large number of fermions in a confining potential, although supershells have not been observed in nuclei due to an insufficient number of particles. The supershell structure of lithium clusters has also been studied [70, 71], and the agreement with experiment is even better than for sodium. The experimental and the theoretical first supershell nodes are both found near $N = 820$. Shell and supershell effects have been observed in clusters of the trivalent elements Al, Ga and In [72], but in order to explain the details it is necessary to go beyond the spherical jellium model.

The effect of temperature on the shells and supershells of sodium clusters has been studied. Calculations of the free energies of clusters were performed by treating the valence electrons as a canonical ensemble in the heat bath of the ions [73]. The spherical jellium model that was used for the ionic background is expected to be an even better approximation at finite temperature. The amplitudes of shell and supershell oscillations decrease with increasing T. This is particularly important in the region of the first supershell node, which becomes smeared out at a temperature of 600 K. However, the temperature does not shift the positions of the magic numbers.

3.1.3. Electronic Shell Effects in Noble Metal Clusters

Experiments on clusters of the noble metals Cu, Ag, and Au have also shown electronic shell structure [74]. The mass spectra of Cu_N^+, Ag_N^+, and Au_N^- clusters obtained by bombarding the metal with inert gas ions show steep drops of the cluster intensities after $N = 3, 9, 21, 35, 41, 59\ldots$. Because the outer electrons of the noble metal atoms have the configuration $d^{10}s^1$, the features in the mass spectrum are easily explained by a jellium model applied to the outer s-electrons only, provided it is recognized that in the sputtering experiments the clusters are ionized and so the number of s-type electrons corresponding to the list of masses above is $N - 1 = 2, 8, 20, 34, 40, 58\ldots$. Anionic clusters like Au_N^- display the same shell closings. For these clusters, the intensity drops after $N = 7, 19, 33, 39\ldots$, corresponding to the electronic shell closing numbers $N + 1 = 8, 20, 34, 40\ldots$. The shell effects are

confirmed by measurements of ionization potentials [75] and also from the electron affinities, obtained by electron detachment from the corresponding anionic clusters [76]. The observation and theoretical analysis of anionic clusters has been extended to multiply charged cluster anions [77]. Silver di-anionic clusters, Ag_N^{2-}, were formed in a Penning-trap. The relative abundance shows very pronounced dips at $N = 33$ and $N = 39$, reflecting shell closings for 34 and 40 electrons: Ag_{33}^{2-} and Ag_{39}^{2-} have 35 and 41 electrons, respectively, one electron in excess of a closed shell. The mass spectrum of Au_N^{3-} also showed evidence of shell closing at 58 electrons.

The HOMO–LUMO gap is another quantity that can reveal electronic shell closing. The gap for Au_{20}, measured by photodetachment spectroscopy of the Au_{20}^- anion [78], was found to be 1.77 eV, even larger than the HOMO–LUMO gap of C_{60} (1.57 eV). This large HOMO–LUMO gap has its origin in a shell closing, for DFT calculations predict a very symmetrical tetrahedral geometry formed by planar facets [78, 79], shown in Fig. 6, and a calculated HOMO–LUMO gap of 1.82 eV. Although the cluster is not spherical, 20 is a shell-closing number for clusters with tetrahedral symmetry.

The band structure of noble metals supports a picture of rather localized d-electrons, extended s-electrons, and substantial s–d mixing [80], far removed from that of the free electrons in the alkali metals. It is, therefore, intriguing how well the shell model works for clusters of the noble metals. Fujima and Yamaguchi [81] have performed DFT calculations for Cu_N clusters with sizes up to $N = 19$ and a variety of model structures. The analysis of the molecular orbitals (MO) shows that these are of two types. The first type is formed by MOs built from atomic $3d$ orbitals. These span a narrow energy range of a width comparable to that of the d-band of the solid and do not mix much with the second type of MOs, which are derived from atomic $4s$-$4p$ orbitals. The $3d$ charge is localized around the atoms, whereas the sp charge is extended over the whole cluster. Fujima and Yamaguchi related their results to the shell model. Disregarding the MOs with d-character on the atoms, the sequence of the remaining MOs can be reproduced rather well by considering a spherical model potential with a small anharmonic term. This is essentially the form of the effective potential in the spherical jellium model (however, when the cluster lacks a central atom, as is the case of the icosahedral structure of Cu_{12}, a three-dimensional Gaussian potential barrier was added to simulate the effect of the missing atom). Also important is the positioning of the d-band relative to the sp levels in the full calculation: for $3 \leq N < 8$, the d-band is located in the energy gap between the molecular levels with overall symmetries comparable to those of the $1S$ and $1P$ jellium levels (here, the angular momentum quantum number of the delocalized jellium levels are labeled with capital letters, to avoid confusion with the atomic s, p, and d orbitals), between the $1P$ and $1D$ levels for $9 \leq N < 18$, between the $1D$ and $2S$ levels for $19 \leq N < 20$, and so on. The atomic d levels are always full. The combination of these two features—the good one-to-one correspondence between the energy levels of the model potential and those of the full calculation, and the appearance of the d levels in the gaps between sp levels minimizing the sp–d mixing—explains why the simple shell model also works for Cu clusters.

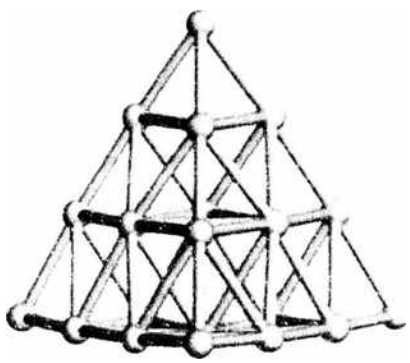


Figure 6. Calculated ground state structure of Au_{20}. Adapted with permission from [79], Wang et al., *Chem. Phys. Lett.* 380, 716 (2003) © 2003, Elsevier.

3.1.4. Clusters of the Aluminum Group

After the alkali clusters, aluminum clusters are among the most studied. Aluminum is a nearly-free electron metal, so the electronic structure of these clusters is expected to be relatively simple. The electronic configuration of the valence electrons in the free atom is $3s^2$ $3p^1$ and the s and p levels are separated by an energy gap of 4.99 eV. Aluminum behaves as a monovalent atom in very small clusters. However, experiments [82–84] and calculations [85] indicate that the gap is reduced as the cluster grows and s–p hybridization begins to develop at Al_8. Already for Al_{13} the s–p hybridization is complete and photoelectron spectroscopy [82–84] confirms the validity of the jellium model for the electronic structure of Al_N clusters for N larger than 12. The calculated structures [86] up to Al_5 are planar and coincide with those of the alkali clusters, a result consistent with the monovalent character of the Al atoms in those small clusters. The geometries become three-dimensional starting with Al_6 (see Fig. 7). The first cluster that develops a pentagonal arrangement of atoms is Al_9. Clusters with 11 atoms or more contain at least one inner atom with a bulk-like coordination. The calculated evaporation energy suggests that Al_7^+ is very stable. This is because the cluster has 20 valence electrons. The high stability is also a property of neutral Al_7, and, in fact, this has been reported as a magic cluster in the experiments of Jarrold et al. [87]. In general, the geometries of small charged clusters are similar to those of the neutrals, with small exceptions.

Special interest in Al_{13} arises because this cluster has 39 valence electrons, one electron short of the 40 needed for a closed-shell cluster. For this reason, the doping of Al_{13} has been proposed as a way to force the shell closure and give a highly stable cluster that could be a possible candidate for cluster-assembled solids. The lowest energy structure of Al_{13} is a distorted icosahedron but a distorted decahedron, in which two pentagonal caps join to form square faces, lies only 0.2 eV above in energy [86, 88, 89]. The distortions are Jahn-Teller effects due to the open-shell electronic configuration. The distortion gives the icosahedral cluster a slightly oblate shape. The 40 electron anionic Al_{13}^- has closed electronic shells, and consequently a more regular icosahedral structure [90]. In contrast, for the cation, a larger distortion compared with the neutral, and a small volume expansion, are obtained.

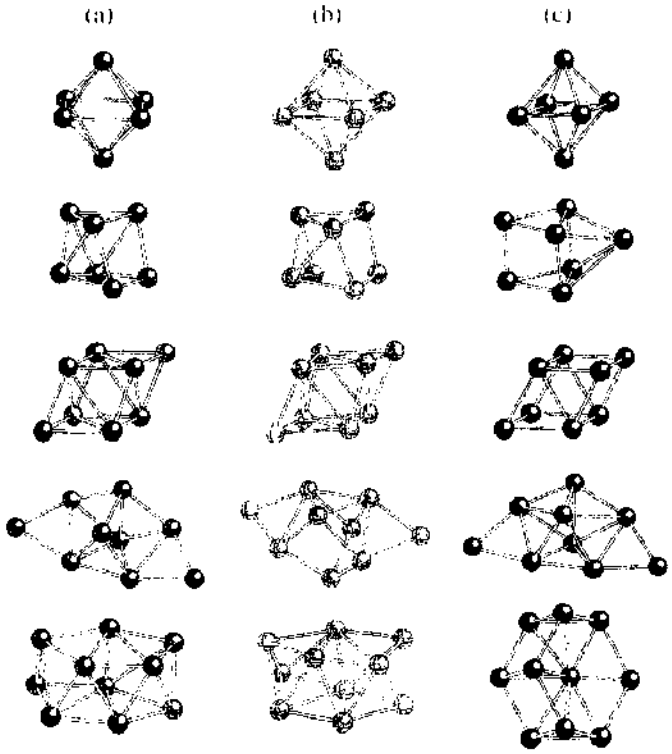


Figure 7. Calculated ground state geometries of neutral (a), cationic (b) and anionic (c) aluminum clusters with 6–10 atoms. Reprinted with permission from [86]. Rao et al., *J. Chem. Phys.* 111, 1980 (1999). © 1999, American Institute of Physics.

3.2. Perturbed Clusters

3.2.1. Electric Polarizability of Clusters

The electric dipole polarizabilities of clusters can be obtained by measuring the deviation of a beam of size-selected neutral clusters traveling through a region where an inhomogeneous electric field has been applied perpendicular to the beam direction [57]. Measurements for Na_N and K_N showed that the polarizabilities per atom, α/N, follow a downward trend toward the bulk value with some abrupt drops related to electronic shell closing, and minima of α/N at $N = 2, 8, 18$ [91]. Results for selected Na clusters are given in Table 1. The classical static polarizability of a metallic sphere of radius R, equal to R^3 [92], provides a useful reference. The table shows that the experimental results, given in units of R^3, exceed the classical values. This enhancement over the classical value is due to the spilling out of the electronic charge beyond the classical cluster radius R in the field-free system.

The polarizability can be calculated using linear response within DFT [93]. If the electric field is characterized by a potential $\delta V = \mathscr{E}_0 z$, where $\mathscr{E}_0$ is small, and considering the dipole case only with $l = 1$, the cluster develops an induced dipole moment **D** of magnitude $D = \alpha \mathscr{E}_0$ in response to the field. To first order, the response of the system is characterized by a small change in the Kohn-Sham one-electron orbitals: $\psi_i(\mathbf{r}) \rightarrow \psi_i(\mathbf{r}) + \delta\psi_i(\mathbf{r})$. Using first-order perturbation theory, a set of equations is obtained for the changes $\delta\psi_i(\mathbf{r})$:

$$\left[-\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r}) - \varepsilon_i\right]\delta\psi_i(\mathbf{r}) = -\delta V_{eff}(\mathbf{r})\psi_i(\mathbf{r}) \tag{52}$$

Here, ε_i are the one-electron energy eigenvalues associated with the unperturbed (i.e., field-free) orbitals ψ_i, and

$$\delta V_{eff}(\mathbf{r}) = \mathscr{E}_0 z + \int d\mathbf{r}' \frac{\delta n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \int d\mathbf{r}' \frac{\delta V_{xc}(\mathbf{r})}{\delta n(\mathbf{r}')}\delta n(\mathbf{r}') \tag{53}$$

is the change in the self-consistent Kohn-Sham potential. The second term on the r.h.s. of Eq. (53) represents the change in the classical electrostatic potential of the electronic cloud, and the last is the change in the exchange-correlation potential V_{xc}. The calculation of the polarizability proceeds by first solving the Kohn-Sham equations for the ground state of the field-free cluster to obtain ψ_i and ε_i. Solving self-consistently Eqs. (52) and (53) yields the $\delta\psi_i(\mathbf{r})$ from which the density change $\delta n(\mathbf{r})$ can be obtained and used to determine the static dipole polarizability

$$\alpha = \frac{1}{\mathscr{E}_0} \int d\mathbf{r}\, z\delta n(\mathbf{r}) \tag{54}$$

as the ratio between the induced dipole moment and the external field strength.

The electric dipole polarizabilities of alkali clusters have been calculated using the LDA and the spherical jellium model (SJM) [60, 94, 95]. Results reported in the column LDA-SJM of Table 1 show the expected enhancement over the classical value, but the theory still underestimates the measured polarizability by about 20%. The other columns show results of applying different corrections. The first one improves the treatment of exchange and correlation. In a

Table 1. Static dipole polarizabilities (in units of R^3, the classical Mie value of neutral Na_N clusters in the jellium model using different descriptions of exchange and correlation (LDA, WDA, and SIC).

	LDA [60]				
N	SJM	DJM	WDA [97, 98]	SIC [94]	Exp. [91]
8	1.45	1.71	1.81	1.66	1.77
18	1.33	1.53	1.59	1.55	1.71
20	1.37	1.61	1.63	1.59	1.68
34	1.27	1.46	1.47	1.47	1.61
40	1.32	1.56	1.53	1.56	1.51

R is the radius of the spherical positive background.

neutral cluster, the LDA exchange-correlation potential decays to zero exponentially, outside the cluster, whereas the exact asymptotic decay should be proportional to $-1/r$. The asymptotic behavior can be improved by using a nonlocal description of exchange and correlation effects known as the weighted density approximation (WDA) [96]. The slower decay of V_{xc} produces a more extended electron density tail and a larger number of bound unoccupied states in the single-particle spectrum. These effects lead to higher polarizabilities [97, 98], improving the agreement with experiment. Self-interaction corrections (SIC) also include nonlocal effects beyond the LDA by removing from the Hartree potential the self-interaction of one electron with itself. The polarizabilities also improve [94]. A further improvement smooths the sharp discontinuity of the positive jellium background at the cluster surface. For this purpose the abrupt step density [see Eq. (43)] has been replaced by a continuous function modeling a surface with a finite thickness [60]. This diffuse jellium model (DJM) results in a more extended electron density and an increased polarizability (column labeled LDA-DJM in Table 1). This suggests that a better description of the polarizability per atom is obtained with the formula

$$\alpha_{jellium} = \frac{(R+\delta)^3}{N} \tag{55}$$

where δ measures the spatial extent of the electronic spill-out and is nearly independent of the cluster radius R.

Polarizabilities have also been calculated using a finite-field approach [63] in which the Kohn-Sham equations are solved with and without a small applied electric field. The polarizability is then estimated from the difference in the dipole moments

$$\alpha_{ij} = \frac{\Delta D_i(\mathscr{E})}{\Delta \mathscr{E}_j} = \frac{D_i(\mathscr{E}) - D_i(\mathscr{E}=0)}{\mathscr{E}_j} \tag{56}$$

where $\mathscr{E}_j$ is the magnitude of the electric field applied along the jth axis. The average polarizability is then

$$\alpha_{av} = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{57}$$

Using this finite-field approach and the GGA [99] to treat exchange-correlation effects, Chelikowsky and coworkers [63] have performed accurate calculations of the electric dipole polarizability of sodium clusters with sizes up to $N = 20$. The results are compared with experiment in Fig. 8. The calculations take into account the structure of the ionic skeleton of the clusters

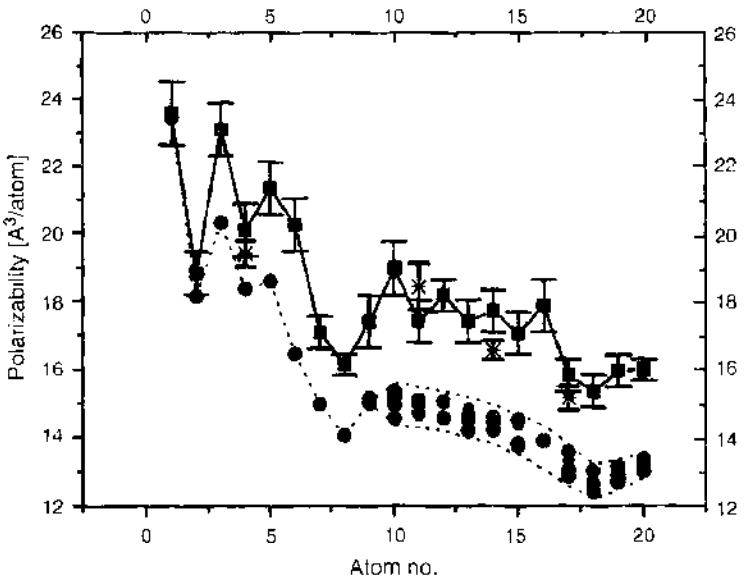


Figure 8. Polarizability of Na clusters as a function of cluster size: squares, experimental data [91]; circles, theory at $T = 0$ K; stars, theory at $T = 750$ K. Reprinted with permission from [63]. Kronik et al., *Phys. Rev. B* 62, 9992 (2000). © 2000, American Physical Society.

and reproduce well the observed trend including minima of the polarizability for Na_7, Na_8, and Na_{18}. The closed-shell clusters, like the rare gas atoms, are less polarizable than those with open shells because of an energy gap to any excited states. The *band* of calculated values between Na_{10} and Na_{20} is due to the several isomers with energies close to the ground state. The figure also shows that the calculations for $T = 0$ K underestimate the measured values of the polarizabilities. This was ascribed to the fact that the clusters in the experiment may have a temperature of several hundred degrees. To simulate this effect, Kronik and coworkers performed molecular dynamics simulations at a temperature $T = 750$ K for several different clusters ($N = 4, 11, 14, 17$) and computed the average polarizability for each cluster by sampling several cluster structures. These values, shown in the figure as stars, are much closer to the measured values. It is fair to claim that not only the main features, but also the detailed variation of the polarizabilities with size N are reasonably reproduced by DFT calculations, which include the cluster geometry [63, 64, 100, 101]. The polarizabilities also provide information on the atomic arrangement of the cluster. A comparison between measured polarizabilities and calculations for several isomeric structures has been used to identify the most stable geometries [64, 102].

3.2.2. Reactivity

The association of high stability of a cluster with an electronic closed-shell configuration is very strong. Closed-shell clusters have a high ionization potential I and a low electron affinity A and are expected to be much less reactive than clusters with unfilled electronic shells. The change ΔE in the energy of an atom, a molecule or a cluster due to a small change ΔN_e in the number of electrons, maintaining the nuclei at fixed positions, can be written as an expansion in powers of ΔN

$$\Delta E = \frac{\partial E}{\partial N_e}\Delta N_e + \frac{1}{2}\frac{\partial^2 E}{\partial N_e^2}(\Delta N_e)^2 + \cdots \tag{58}$$

where ΔN_e can be smaller than 1 when the system of interest is in contact with other species. Using DFT and applying Eq. (58) to atoms, Parr and coworkers [103] have identified the coefficient $\frac{\partial E}{\partial N_e}$ of the first-order term with the electronegativity μ, which is understood in chemistry as the power of an atom to attract or lose electrons in a molecule. The coefficient $\frac{1}{2}\frac{\partial^2 E}{\partial N_e^2}$ of the second-order term was identified with the hardness κ of the system [104] and gives the resistance to a change in the number of electrons. Evidently, $1/\kappa$ is a measure of the reactivity of the system. These ideas can be extended to a cluster, to define its electronegativity [105, 106] and hardness [106].

Using a finite difference approximation to estimate the derivatives, the energy of the cluster becomes

$$E(N_e) = E(N_0) - \frac{1}{2}(I + A)(N_e - N_0) + \frac{1}{2}(I - A)(N_e - N_0)^2 + \cdots \tag{59}$$

where N_0 is the number of electrons in the neutral cluster. The electronegativity and hardness are then given by

$$\mu = \frac{1}{2}(I + A) \tag{60}$$

and

$$\kappa = I - A \tag{61}$$

respectively. Shell effects lead to structure in $\kappa(N)$. The value of $I(N)$ drops between N_c and $N_c + 1$, where N_c indicates a shell-closing size. On the other hand, A drops between $N_c - 1$ and N_c. Consequently, a cluster of size N_c has a large I and a small A, and a large value of the hardness is expected. The results of calculations for Na clusters [106], shown in Fig. 9, predict local maxima of the hardness for the closed-shell clusters. These are the less reactive clusters.

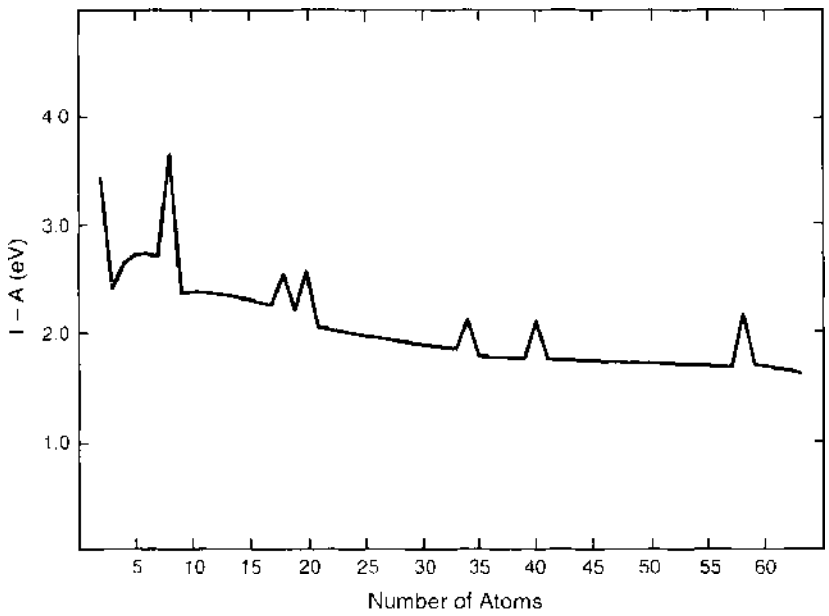


Figure 9. Effects of shell closing on the chemical hardness of sodium clusters described by the spherical jellium model. Reprinted with permission from [106], Alonso and Balbás, *Struct. Bonding* 80, 229 (1993). © 1993, Springer-Verlag.

In an exact DFT treatment

$$I = -\epsilon_F \tag{62}$$

where ϵ_F is the energy eigenvalue of the highest occupied Kohn-Sham (KS) orbital. Correspondingly, A is equal to minus the eigenvalue of the highest occupied orbital of the anionic cluster. For the usual treatments of exchange and correlation, Eq. (62) is only approximate, but a reasonable measure of the hardness is given by the difference between the energies of the lowest unoccupied molecular orbital and the highest occupied molecular orbital of the neutral cluster

$$\kappa = \epsilon_{LUMO} - \epsilon_{HOMO} \tag{63}$$

Clusters with closed shells have large HOMO-LUMO gaps, as shown in Fig. 9. These are chemically hard clusters. Their resistance to either donation or acceptance of an electron makes the closed-shell clusters relatively inert, and two closed-shell clusters are expected to interact only weakly.

Reactivity experiments verify the inert character of closed shell clusters. Leuchner et al. [107] have reacted anionic and cationic Al clusters with oxygen in a flow tube reactor. An etching reaction was observed and rate constants were reported. The most striking feature was the negligible reactivity of Al_7^-, Al_{13}, and Al_{23}. Not only were these clusters unreactive, they were also products of reactions of larger clusters. According to the jellium model, also applicable to Al clusters, the three species are closed-shell clusters with 20, 40, and 70 electrons, respectively.

3.2.3. Subshells and Distortion of the Cluster Shape

The spherical jellium model is not quite correct for cluster sizes between two shell closing numbers because a partially filled outermost shell leads to a nonspherical electron density that induces a Jahn-Teller distortion of the ionic background. This leads to a splitting of the highly degenerate spherical shells into subshells, and these effects are reflected in the mass spectrum [57]. To account for the shape deformations, the jellium model has been modified [108] by allowing the positive background to have a spheroidal shape. The deformation, prolate or oblate, of the positive background is described by a distortion parameter $\beta = 2(a-b)/(a+b)$, with a and b denoting the lengths of the semiaxes of the spheroid. The distortion is then independently optimized for each cluster size by minimizing the total energy. Clusters with closed

shells remain spherical, whereas alkali clusters with $N = 3$ and 4 become prolate and those with $N = 5 - 7$ are oblate. Following the spherical cluster with $N = 8$, clusters are prolate for $N = 9 - 13$ and oblate for $N = 14 - 19$, and so on. The stability function $\Delta_2(N)$ of Eq. 46 predicts peaks at $N = 8, 10, 14, 18, 20, 26, 30, 34, 40, 46, 50, 54, 58\ldots$ reflecting the filling of subshells as well as the usual shell-closing numbers of the spherical model. The corresponding features in the mass spectra have been observed, although these are less pronounced than the main shell effects.

3.2.4. Collective Electronic Excitations

Photoabsorption experiments also provide evidence for the shell-closing and shape deformations. By exposing a beam of size-selected clusters to laser pulses of the appropriate wavelength, photons are absorbed that excite collective oscillations of the electrons against the ionic skeleton. For example, in small Na clusters, the energy of the collective dipole excitation is about 3 eV, while the energy required to evaporate one atom is less that 1.1 eV. The collective electronic excitation decays by inducing the fragmentation of the parent cluster leading to a deviation of the daughter cluster away from the initial direction of motion of the molecular beam. The ratio between the number of clusters of a given size arriving at the detector with and without exposure to the laser gives the photoabsorption cross section $\sigma(\omega)$ [109]. Using a statistical model and assuming that the energy of the collective mode is converted into vibrations, the time required to evaporate an atom is found to be several orders of magnitude smaller than the time of flight of the molecular beam in the spectrometer, which is about 1 ms. Consequently, it can be assumed that the photoabsorption and photoevaporation cross sections are equal. The collective excitation, or surface plasmon, of clusters is similar to the giant dipole resonance in nuclei.

The dynamical response of clusters can be investigated using the time-dependent density functional formalism (TDDFT) [110–112]. An external time-dependent electric field characterized by the potential $V_{ext}(\mathbf{r}; \omega) = \varepsilon_0 r^l Y_l^0 e^{i\omega t}$ with multipolarity l ($l = 1$ for a dipolar field) induces a perturbation of the density

$$\delta n(\mathbf{r}; \omega) = \int d\mathbf{r}' \chi_0(\mathbf{r}, \mathbf{r}'; \omega) \delta V_{eff}(\mathbf{r}'; \omega) \tag{64}$$

Here, $\chi_0(\mathbf{r}, \mathbf{r}'; \omega)$ is the independent-particle dynamical susceptibility, and ω is the frequency of the electric field. This equation describes the electrons responding as independent particles to an effective Kohn-Sham potential

$$\delta V_{eff}(\mathbf{r}'; \omega) = V_{ext}(\mathbf{r}'; \omega) + \int d\mathbf{r}'' \frac{\delta n(\mathbf{r}''; \omega)}{|\mathbf{r}' - \mathbf{r}''|} + \int d\mathbf{r}'' \frac{\delta V_{xc}(\mathbf{r}')}{\delta n(\mathbf{r}'')} \delta n(\mathbf{r}''; \omega) \tag{65}$$

which is the sum of the external potential and the electrostatic and exchange-correlation potentials due to the induced density. The parallel with the Kohn-Sham equations giving the orbitals generating the ground state density is evident. Equations (64) and (65) allow for a self-consistent calculation of the induced density $\delta n(\mathbf{r}; \omega)$ from which the dynamical electric dipole polarizability can be obtained

$$\alpha(\omega) = \frac{1}{\varepsilon_0} \int d\mathbf{r}\, z \delta n(\mathbf{r}; \omega) \tag{66}$$

Using now the *Golden rule*, the photoabsorption cross section of the cluster becomes [110, 111]

$$\sigma(\omega) = \frac{4\pi\omega}{c} Im\, \alpha(\omega) \tag{67}$$

where c is the velocity of light and $Im\alpha(\omega)$ represents the imaginary part of the dynamical polarizability. In practice, the V_{xc} strictly appropriate only for the electronic ground state is used in Eq. (65).

The classical theory for the dynamic polarizability dating back to Mie predicts a dipole frequency

$$\omega_{Mie} = \left(\frac{Ze^2}{mR^3}\right)^{1/2} \tag{68}$$

where Z is the number of electrons taking part in the collective oscillation equal to the number of atoms in monovalent clusters, and R is the radius of the cluster. This gives for ω_{Mie} a value equal to $1/\sqrt{3}$ of the bulk plasma frequency. The experimental dipole resonance frequency of small clusters is *red shifted*, by 10–20% with respect to ω_{Mie} and TDDFT calculations using the jellium model and the LDA indicate that this shift is due to the spill-out of electrons beyond the jellium edge [112]. Often, the collective resonance is broadened by the temperature, and it can be fragmented due to interaction with particle-hole excitations. Fedrigo and coworkers [113] have measured the optical spectrum of silver clusters with sizes up to $N = 39$ trapped in an argon matrix. From the spectra they calculated the mean absorption frequency

$$\omega_{mean} = \frac{\int \omega\sigma(\omega)d\omega}{\int \sigma(\omega)d\omega} \tag{69}$$

The mean absorption energy $\hbar\omega_{mean}$ shows oscillations as a function of N, with a tendency to higher values for closed-shell clusters. In particular, maxima were found for $N = 2$, 8, 18, and 34, the sizes at which the static electric dipole polarizabilities show minima. The optical response of Ag clusters has been simulated by a smooth polarizable background, with the dielectric function ϵ_d of the bulk metal, acting on the valence (5*s*-type) electrons [114]. The 4*d* core electrons of Ag have only an indirect effect on the optical response of the clusters at low energies. Intrinsic effects of the *d* electrons occur at higher energy [115].

The static deformation of the cluster shape affects the photoabsorption spectrum. The splitting of the collective dipole resonance in two or three peaks that has been observed in alkali [116, 117] and noble metal clusters [118] is thought to be due to collective electronic oscillations in the directions of the principal axes of a triaxially deformed cluster. This has motivated the further extension of the deformed jellium model to fully triaxial shapes [119]. The results are consistent with experiment and also support the findings of the spheroidal model, with prolate shapes predicted after the magic numbers $N = 2$ and $N = 8$, and oblate shapes before $N = 8$ and $N = 20$. A transition region of triaxial shapes is found separating prolate and oblate clusters. The only triaxial cluster in the region of the 1*p* shell is Na_5, but there are several triaxial clusters in the region of the 1*d* shell: Na_{11}, Na_{13}, Na_{15}, and Na_{17}. The triaxial character was found to be strong for Na_5, but very weak for the other clusters, and thermal effects could easily wash out the triaxial signature in the dipole resonance spectrum.

3.2.5. Shell Effects in Cluster Fragmentation

A striking manifestation of the shell closing effects occurs in the fragmentation of multiply charged clusters. Soon after the discovery of the shell effects in clusters, Alonso and coworkers [120] predicted that the most favorable fragmentation channel of a doubly-charged cationic alkali cluster

$$X_N^{2+} \longrightarrow X_{N-p}^{+} + X_p^{+} \tag{70}$$

corresponds to the production of a singly charged magic fragment with 2 or 8 electrons, that is $p = 3$ or $p = 9$. This prediction resulted from the analysis of the heat of fragmentation

$$\Delta H_f = E(X_{N-p}^{+}) + E(X_p^{+}) - E(X_N^{2+}) \tag{71}$$

calculated from the total energies of the parent and fragment clusters. Experiments for alkali [121] and noble metal clusters [122] confirmed this prediction. The dynamical process of fragmentation is, however, more complex. A doubly charged cluster is, in general, metastable, and an energy barrier exists preventing fragmentation. The barrier arises when the shape of the cluster deforms from an initial compact shape to an elongated shape on its way along the dissociation path.

Shape deformations are a common feature of open shell clusters and have been discussed above. This tendency is enhanced in multiply charged clusters because of the additional electrostatic repulsive energy. A striking interplay between deformation and shell effects has been found in a number of calculations [123–125]. In some favorable cases, the *magic* fragments are preformed in the ground state of the charged X_N^{z+} cluster and the energy required to further elongate the cluster and overcome the barrier to dissociation following the path leading to a magic fragment is smaller than for other channels. As an example, the electron density distribution for the ground state of Na_{24}^{2+}, obtained in a calculation with a deformed jellium model, is given in Fig. 10 [126]. The charged cluster can be interpreted as a supermolecule built from two pieces, Na_{21}^{+} and Na_3^{+}, both magic clusters with 20 and 2 electrons, respectively. The most favorable dissociation channel is evidently

$$Na_{24}^{2+} \longrightarrow Na_{21}^{+} + Na_3^{+} \tag{72}$$

and the dissociation barrier is only 0.14 eV. The lower panel of the figure shows a snapshot of the electron density of the fissioning cluster after overcoming the fission barrier.

3.2.6. Ultimate Jellium Model

In an extension of these ideas, Manninen and coworkers [127] have introduced the ultimate jellium model, in which the positive background charge is allowed to be completely deformable, both in shape and density distribution. In order to minimize the cluster energy the background adapts itself perfectly to the electronic profile and $n_+(\mathbf{r}) - n(\mathbf{r})$ vanishes everywhere. Consequently, the electrostatic Coulomb energy of the cluster vanishes due to

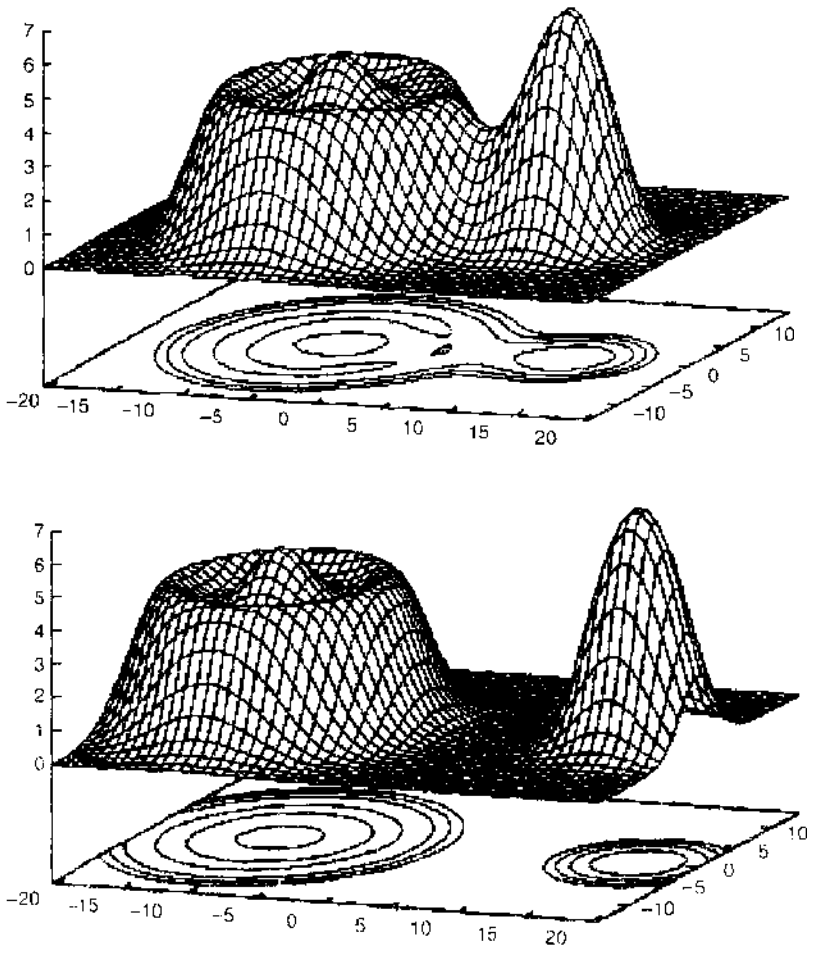


Figure 10. Electron density in the ground state of Na_{24}^{2+}. The Na_{21}^{+} and Na_3^{+} fragments are already preformed. The lower panel corresponds to a configuration when the fissioning cluster has surpassed the maximum of the barrier. Adapted with permission from [126], Garcias et al., *Heavy Ion Phys.* 1, 227 (1995), © 1995, Akademiai Kiado.

the local charge neutrality and the electrons just move in their own exchange-correlation potential $V_{xc}(\mathbf{r})$. Qualitatively, the clusters follow the pattern of shapes described above. The shapes up to $N = 8$ can be simply understood in terms of the symmetries of the s, p_x, p_y, p_z orbitals, and a similar explanation in terms of the filling of d and s orbitals is possible between $N = 8$ and $N = 20$. An interesting outcome of the ultimate jellium model is the appearance of some clusters as supermolecules built from magic clusters weakly bound to each other. A few examples are shown in Fig. 11. Na_4 is a supermolecule built from two Na dimers with a weak separation energy of only 0.10 eV. Na_{10} can be viewed as a dimer attached to Na_8, with a separation energy of 0.15 eV, and Na_{12} has the appearance of two dimers attached to Na_8.

3.3. Clusters with Impurities

3.3.1. Monovalent and Divalent Impurities in Alkali Metal Clusters

The shell model provides a framework to interpret some experiments on mixed metallic clusters. Clusters with formula BX_m have been produced by supersonic expansion of mixed atomic vapors [128]. Here, X indicates an alkali element (Na or K) acting as the matrix and B is either an alkali (Li, Na) or a divalent (Mg, Ca, Sr, Ba, Zn, Hg, Yb, Eu) impurity. For some impurity-matrix combinations, the usual magic number $N_e = 8$ disappeared and was replaced by $N_e = 10$, and the magic number $N_e = 18$ also disappeared. Baladrón and Alonso [129] noticed that these changes occur for $X - B$ combinations where the difference

$$\Delta\bar{n}_+ = \bar{n}_+(B) - \bar{n}_+(X) \tag{73}$$

between the jellium background densities of the B and X metals is larger than a critical positive value, prompting an extension of the spherical jellium model. In this generalization, a BX_m cluster is modeled as a spherical matrix, characterized by a background density $\bar{n}_+(X)$, containing an impurity, described as a small jellium sphere with background density $\bar{n}_+(B)$, inserted in a cavity at the center of the matrix cluster. DFT calculations for this jellium-on-jellium model show that as the difference $\Delta\bar{n}_+$ increases the magnitude of the energy gap between the $2s$ and $1d$ shells is reduced because the s state, being finite at the center of the cluster, is more sesitive to the larger attractive potential in that region. When $\Delta\bar{n}_+$ is large enough, a reversal occurs in the order of these two shells. In such a case, the ordering of the electronic shells becomes $1s1p2s1d$, and the shell closing numbers become 2, 8, 10, 20. The calculations predict, in full agreement with experiment, the particular impurity–host combinations for which the transition to the new magic numbers occurs. *Ab initio* DFT calculations confirm that small size impurities sit preferentially at the center of the matrix cluster [130]. In other cases, such as small $MgNa_m$ clusters, the impurity does not sit at the center, although it becomes increasingly surrounded by Na atoms as the cluster grows. Even in these cases the *jellium-on-jellium* model can be justified. Figure 12 shows the electron density in the ground state of the cluster $PbLi_4$. The Pb atom is part of the cluster surface, and one can observe that the contours of constant electron density are centered on the Pb atom [131]. Evidently, when an alkali cluster contains a single impurity atom with a pseudopotential substantially more

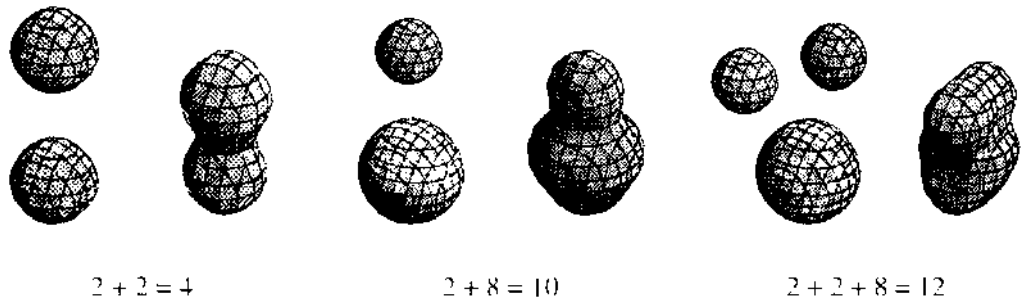


Figure 11. Molecular interpretation of some ground state cluster shapes in the ultimate jellium model. Adapted with permission from [127], Koskinen et al., *Z. Phys. D* 35, 285 (1995), © 1995, Springer-Verlag.

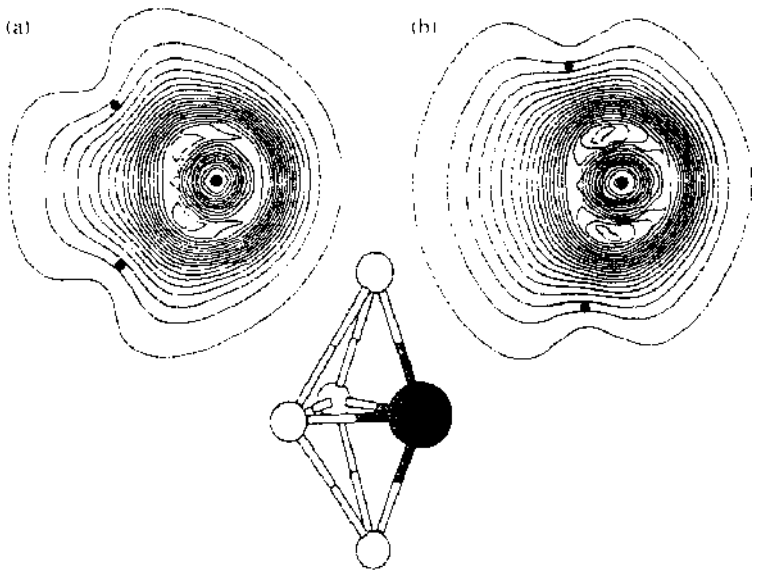


Figure 12. Structure of Li_4Pb and contours of constant electron density in two planes: (a) common basal plane of the bipyramid, (b) vertical plane containing the Pb atom and the two apical atoms. Notice than the contours are centered on the Pb atom. Reprinted with permission from [131], Alonso et al., *Chem. Phys. Lett.* 289, 451 (1998). © 1998, Elsevier.

attractive than that of the matrix atoms, the contours of constant electron density surround the impurity, and the location of the impurity defines the natural center of the electron density distribution in the cluster.

The optical response of alkali metal clusters containing an impurity has been studied using the *jellium-on-jellium* model [132]. The calculated optical spectrum of Na_8Zn is characterized by two closely spaced lines at 2.87 eV, carrying 26% of the total strength, and a stronger line at 2.57 eV carrying 42% of the strength. This is in good agreement with the experimental double peak, formed by a higher energy component at 2.97 eV, which carries less strength than the lower energy component at 2.63 eV . The fragmentation is due to the near degeneracy between the plasmon peak and the $2s \rightarrow 2p$ and $1p \rightarrow 3s$ particle-hole transitions. The calculations were performed for an impurity background density corresponding to $r_s(Zn) = 1.15$ a.u., a value that is substantially smaller than the usual value of $r_s = 2.31$ a.u. for bulk Zn (r_s is a parameter related to the average bulk electronic density n; it is the radius of a sphere which on average contains one electron). The small value of r_s indicates a strong attractive potential at the impurity site, which produces the shift of those transitions required for the degeneracy to develop with the plasmon. The spectrum of Na_8 contains a single line at 2.53 eV, so the effect of the Zn atom on the optical response is evident. The need for such a small value of r_s may be related to the application of the jellium model to an element like Zn, at the end of the transition metal group. As another example of the influence of impurities on the optical response, one can consider the three clusters $K_{19}Rb$, K_{20}, and $K_{19}Na$, all of them having 20 valence electrons [132]. The calculated optical spectrum of K_{20} shows a split plasmon due to the degeneracy between the collective plasmon and a $2s \rightarrow 3p$ particle-hole transition. In the case of $K_{19}Na$ the impurity shifts the $2s$ level downwards, so the energy of the $2s \rightarrow 3p$ transition increases. At the same time, the energy of the plasmon remains unchanged, and a single line dominates the spectrum. The opposite effect occurs for a Rb impurity. The energy of the $2s \rightarrow 3p$ transition is lowered and the plasmon splitting is more pronounced.

In order to test the reliability of the model, Yannouleas et al. [132] have performed parallel jellium-on-jellium and molecular orbital calculations for K_8Mg, Na_8Zn, and K_6Mg, comparing the electronic structures resulting from the two methods. In the molecular orbital calculations, the geometry was assumed to be a body centered cube for K_8Mg and Na_8Zn, and a centered octahedron for K_6Mg. For these two geometries, the molecular orbital electronic configurations are of the type $(1a_{1g})^2(1t_{1u})^6(2a_{1g})^0$, and $(1a_{1g})^2(1t_{1u})^6$, respectively. Comparison with

the jellium calculations is based on the correspondence $1s \rightarrow 1a_{1g}$, $1p \rightarrow 1t_{1u}$, $2s \rightarrow 2a_{1g}$, and $1d \rightarrow (1d_t + 1d_e)$, which was validated. Both calculations reproduce the downward shift of the $2s$ level below the $1d$ level.

The location of the impurity is an interesting question that has been studied by comparative calculations for Li_mBe and Li_mMg [133]. The ground state geometries of both systems are rather similar up to $m = 5$, at which point the growth patterns differ significantly. The structure of Li_6Be is an octahedron with the Be atom at the center. The Be atom is also trapped in the interior for larger clusters. In contrast, the Mg atom avoids internal positions and this affects strongly the structure of the clusters. Even for $m = 12$, the Mg atom resides on the surface of the cluster. The different behavior can be explained on the basis of the atomic size of the constituent atoms and the relative strength of the bonds between them. The ionic radius of Mg (2.57 a.u) is larger than that of Li (2.36 a.u.), while that of Be (1.68 a.u.) is substantially smaller. In addition, the Li–Be bond (0.39 eV) is stronger than the Li–Mg bond (0.26 eV), and both are weaker than the Li–Li bond. The strain associated with positioning the larger Mg atom inside the cluster and the tendency to lower the energy by maximizing the number of stronger Li–Li bonds both lead to the Mg atom residing on the surface of the cluster.

3.3.2. A Full Description of the Cluster Structure

Experiments provide only indirect information on the geometrical structure of clusters. In contrast, *ab initio* calculations based on density functional theory can predict the structures directly although there are uncertainties because of approximations associated with exchange and correlation, the pseudopotentials, and the completeness of the basis set. In addition, such calculations require a substantial computational effort, in part because the number of isomers grows rapidly with size N and the identification of the lowest energy configuration between isomers with similar binding energies becomes a hard task. For this reason, the calculations are usually restricted to small clusters, but with improvements in the methods and in computing power the scope of *ab initio* calculations is steadily advancing. For the prototypical case of sodium clusters, *ab initio* calculations have been performed for sizes up to $N = 20$ [64, 100, 134–137]. Figures 13 to 15 show the results of calculations by Solovyov et al. [64]. For N up to 8, both DFT and Moller-Plesset (MP) many-body perturbation theory were used to optimize the geometries. The particular implementation of DFT employed by Solovyov et al. was the so called B3LYP developed by Becke [22]. This hybrid method expresses the exchange-correlation energy as a parameterized functional that includes a mixture of Hartree-Fock and DFT exchange, along with DFT correlation. The electronic wavefunctions are calculated in the standard DFT way by solving Kohn-Sham equations. For clusters larger than Na_8, only the B3LYP method was used in the optimizations.

The point-symmetry group is indicated for each cluster in the figures, and several low-lying isomers are given for $N = 3$, 4, 6, 10, 11, and 20. Clusters with sizes up to $N = 5$ are planar. MP predicts the isosceles triangle, C_{2v}, symmetry, as the lowest energy structure of Na_3, although the C_{2v} structure with a broken bond lies only 0.008 eV higher in energy. Na_4 illustrates how the geometry can be influenced by the multiplicity of an electronic state: the rhombic geometry, with D_{2h} symmetry, occurs for spin multiplicity 1, while a multiplicity of 3 leads to a square; the former is more stable. Three-dimensional structures first appear for Na_6: the flat pentagonal pyramid is more stable than the planar isomer, but only by 0.02 eV. Na_3 and Na_4 are prolate, and Na_5 to Na_7 oblate, in agreement with the deformed jellium models. B3LYP leads to some differences with respect to MP. The ground state of Na_3 is the linear isomer, and the broken triangle lies 0.008 eV above. The planar isomer is the ground state of Na_6. This shows the difficulties in predicting structures by total energy calculations.

A noticeable feature of the clusters larger than Na_7 studied using B3LYP is that Na_8 and Na_{20} have higher symmetry (T_d) than others. This is consistent with the formation of closed shells in the jellium model. The ground state of Na_{10} has the C_2 symmetry, and the D_{4d} isomer lies 0.07 eV above. The binding energies of the two isomers of Na_{11} only differ by 0.03 eV. Na_{10}, Na_{11}, and Na_{12} are approximately prolate, and Na_{14} approximately oblate, again in agreement with the deformed jellium models. For other clusters the triaxiality is substantial. Nevertheless, Na_{17}, Na_{18}, and Na_{19} appear close to prolate. This is at variance with the jellium

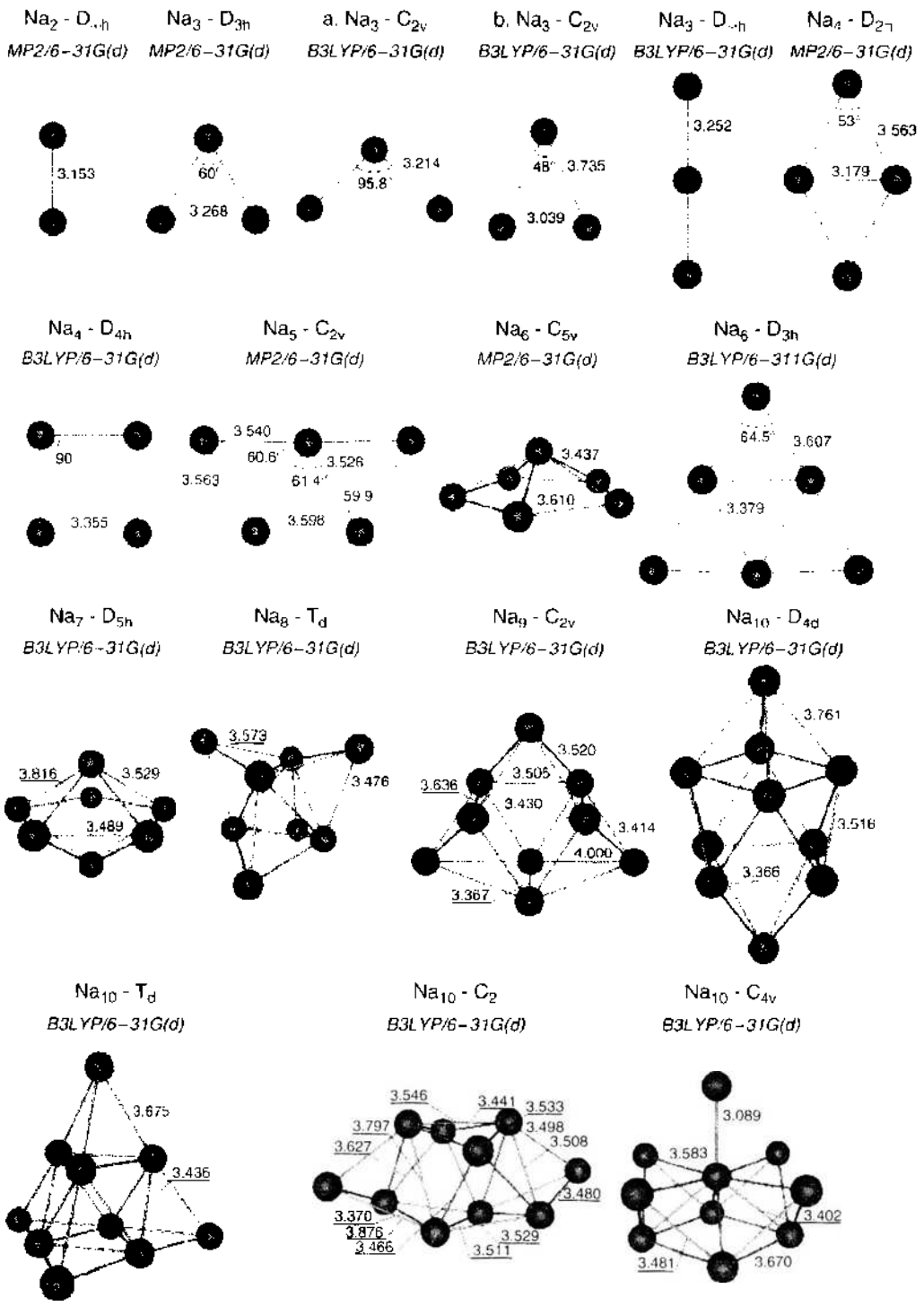


Figure 13. Optimized geometries of sodium clusters with $N = 2$ to 10. Several isomers are given for $N = 3$, 4, 6, 10. Reprinted with permission from [64], Solovyov et al., *Phys. Rev. A* 65, 53203 (2002). © 2002, American Physical Society.

models and originates in the underlying double-icosahedron structure of these three clusters, complete for Na_{16} and incomplete for Na_{17} and Na_{18}. That structure, well-known for inert gas clusters, arises from atomic packing effects, which are not captured by the deformed jellium models. These packing effects are especially important for very large clusters.

Ionization affects the geometry for the smallest clusters. The structure of Na_3^+ is an equilateral triangle. MP calculations predict that Na_4^+ is a near-equilateral triangle with the fourth atom attached to the apex and that the rhombus is close in energy (0.01 eV); the order of these isomers is reversed by B3LYP. Na_5^+ can be viewed as two oppositely oriented isosceles

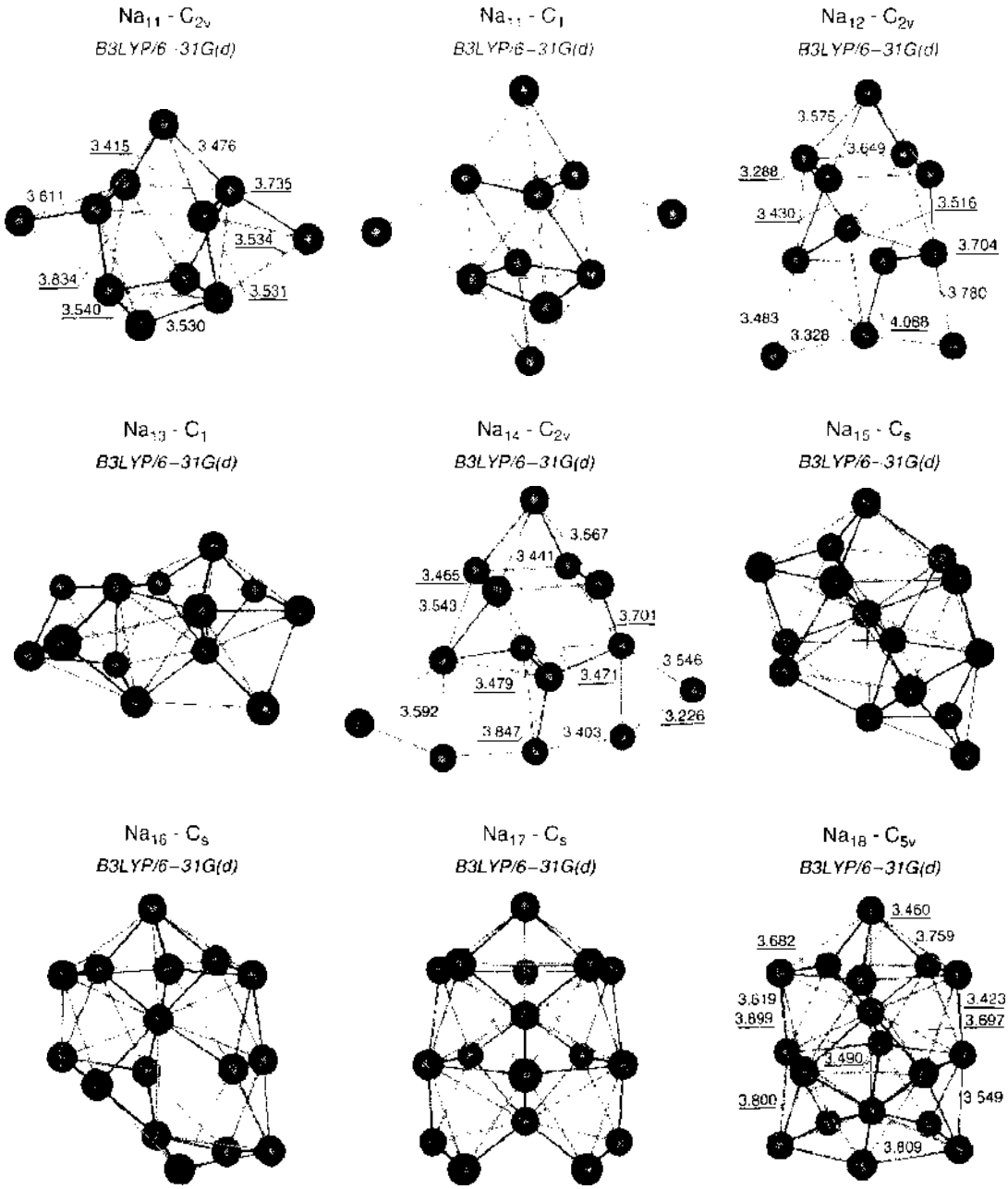


Figure 14. Optimized geometries of sodium clusters with $N = 11$ to 18. Reprinted with permission from [64], Solovyov et al., *Phys. Rev. A* 65, 53203 (2002). © 2002, American Physical Society.

triangles sharing an atom along a common axis. The structure is planar, but one triangle can easily rotate about the axis so that the planes of the two triangles become perpendicular to each other. The structures of Na_6^+ and Na_8^+ are also different from those of Na_6 and Na_8. The structural changes become less drastic for N higher than 9.

Figure 16 shows the dependence of the binding energy per atom with the cluster size. Here the B3LYP energies are compared with the results of the MP and configuration interaction (CI) calculations [135, 137]. The three methods give results in reasonable agreement. One can observe that the binding energy per atom of the clusters Na_8 and Na_{20} are a little higher than the binding energies of clusters of similar sizes. The visibility of the magic number effects would become enhanced by plotting, instead, the second derivative of the binding energy, given by Eq. (46).

Some of the clusters plotted in Figures 13 to 15 have permanent electric dipole moments because the centers of gravity of the electronic and ionic charge distributions differ. The calculations indicate that only clusters with C_1 point symmetry groups have dipole

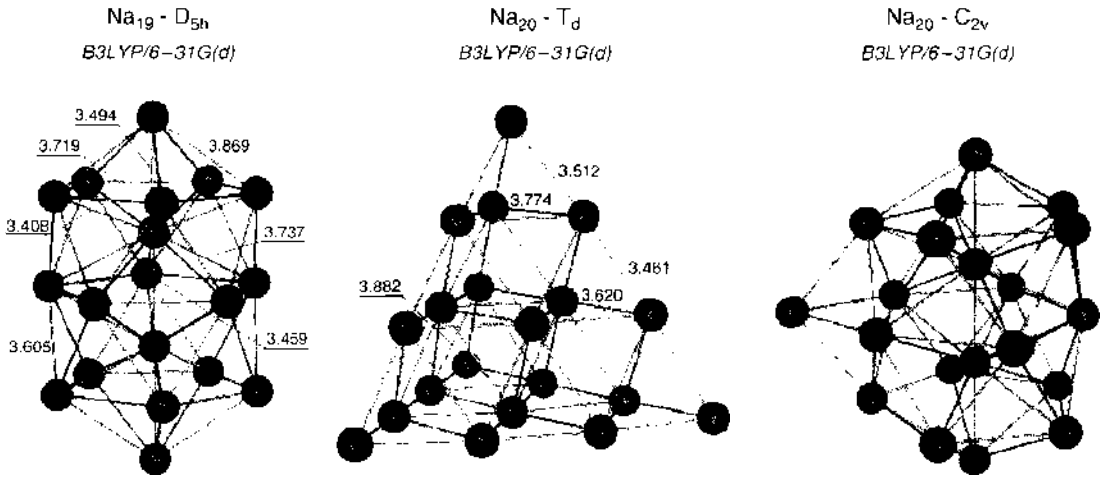


Figure 15. Optimized geometries of sodium clusters with $N = 19$ and 20. Reprinted with permission from [64], Solovyov et al., *Phys. Rev. A* 65, 53203 (2002). © 2002, American Physical Society.

moments with a magnitude up to 1 debye or a little higher [64]. For instance, the dipole moments of the two C_{2v} isomers of Na_3, labeled a and b in Fig. 13, are equal to 0.28 and 1.30 debye, respectively. The sizable differences between the dipole moments of different isomers have led Solovyov and coworkers to propose a method for isomer separation by passing a mass-selected cluster beam through an inhomogeneous electric field. For a cluster with intrinsic dipole moment **D**, the force acting on the cluster in an inhomogeneous electric field $\mathscr{E}(\mathbf{r})$ is equal to

$$\mathbf{F}(\mathbf{r}) = \nabla[\mathbf{D} \cdot \mathscr{E}(\mathbf{r})] \tag{74}$$

If the cluster spends a time t passing through the inhomogeneous electric field, the distance by which the clusters are deflected from the original direction of motion is estimated to be $\Delta = Ft^2/2M$, where M is the mass of the cluster. For typical values of the time period, $t = 10^{-3}$s, and the inhomogeneity of the electric field $\nabla E = 5000$ V/cm^2, the isomers with $N = 3$ and dipole moment difference of 1 debye become separated by $\Delta = 0.7$ mm, which

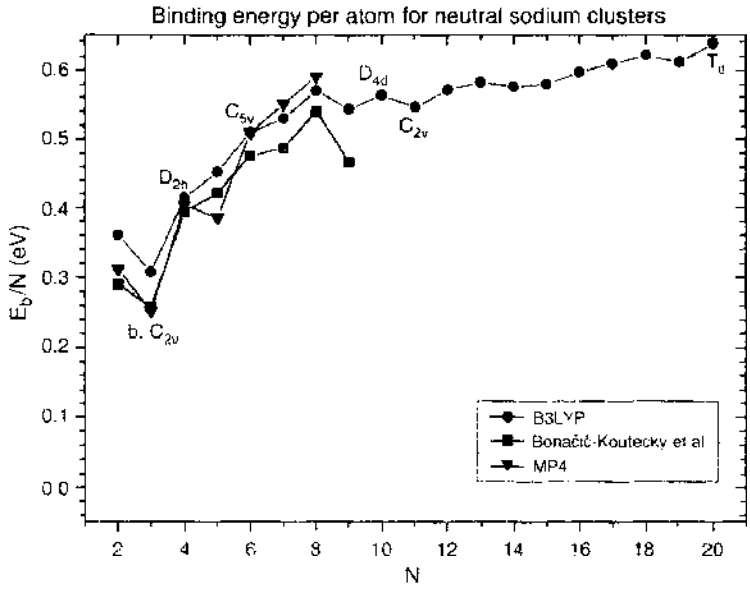


Figure 16. Binding energy per atom of neutral sodium clusters as a function of cluster size. Circles: DFT calculations using the B3LYP method. Triangles: Moller-Plesset calculations. Squares: Configuration interaction calculations [135, 137]. Some clusters in the figure have labels indicating the point symmetry group of the isomers represented. Adapted with permission from [64], Solovyov et al., *Phys. Rev. A* 65, 53203 (2002). © 2002, American Physical Society.

illustrates the capabilities of the proposed method. An analysis for clusters having quadrupole moments shows similar promise.

3.3.3. High Valence Impurities in Alkali Clusters

Impurities with a valence higher than two produce a stronger perturbation of a host alkali cluster. The calculated equilibrium structures of Li_mAl clusters [138] are three-dimensional starting at Li_5Al, which is in contrast to pure Li clusters, where two-dimensional structures prevail up to Li_6. For $m < 6$, the Al atom does not occupy an internal position, but for larger clusters the Al atom is internally located, and surrounded by the Li atoms. The binding energy of the cluster, the evaporation energy of a Li atom $E_{evap}(m) = E(Li_{m-1}Al) + E(Li) - E(Li_mAl)$, and the ionization potential I increase monotonically up to $m = 5$, and then show a drop, indicating the high stability of Li_5Al which has 8 electrons. The structure of this cluster is similar to that of Na_5Pb shown in Fig. 17, with the Al atom a little more distant from the plane of four Li atoms. The analysis of the orbitals of Li_5Al shows a doubly occupied orbital at -7.5 eV with s character about the Al site, and a manifold of three closely spaced and doubly occupied states with predominant p character at -3.5 eV. The charge density indicates that the bonding is covalent with charge accumulation in the regions connecting Al and Li ions. The electronic structure of Li_mAl clusters with $m < 6$ reflects the atomic-like nature of the orbitals, and the properties of Li_5Al are related to the closing of the Al $3p$ shell, albeit perturbed by the lithium environment. Addition of Li atoms to Li_5Al leads to a picture consistent with the delocalized shell model, with the ordering of the $1d$ and $2s$ shells reversed from that corresponding to the homogeneous jellium model, just as in the case of some divalent impurities already discussed.

The tetravalent impurities of group 14 (C, Si, Ge, Sn, Pb) form another interesting case. The supersonic expansion of lead–sodium vapor from a hot oven source led to the observation of an exceptional abundance of the Na_6Pb cluster [139]. Several DFT calculations have been performed to study the nature of this species [140–142]. The calculated ground state geometries of Na_mPb clusters are given in Fig. 17. Up to $m = 7$, each added Na atom binds directly to the Pb atom and the coordination of the Pb atom increases. Up to Na_5Pb, the Pb atom is on the surface to allow the Na atoms to come in closer contact with one another and bind, albeit weakly. The lowest energy structure of Na_6Pb is an octahedron with the Pb atom in the interior of the cluster. A seventh Na atom also binds to Pb, but at this stage the Pb atom appears to be fully coordinated and an additional Na atom, in Na_8Pb, begins to form a second shell and is not directly bound to Pb [140].

The electronic configuration of the free Pb atom is $6s^26p^2$ and the electronic structure of the occupied valence orbitals of Na_6Pb is $(1a_{1g})^2(1t_{1u})^6(2a_{1g})^2$. The $1a_{1g}$ level, which is localized on the Pb atom and has dominant s character, is substantially deeper than the other cluster levels due to the very attractive s-part of the nonlocal Pb pseudopotential. This $1a_{1g}$

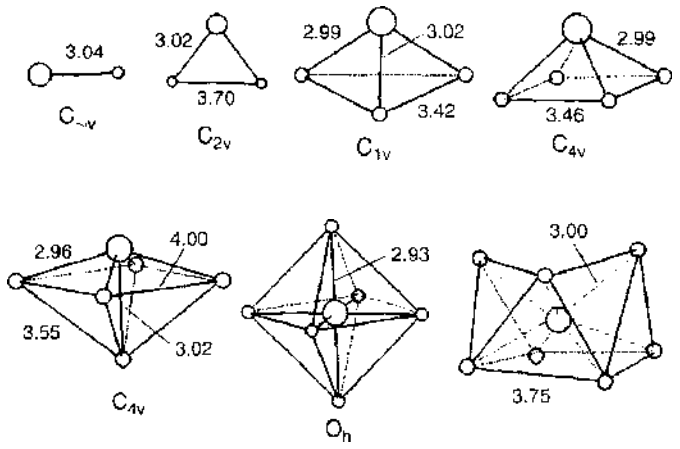


Figure 17. Calculated lowest energy structures of Na_mPb, $m = 1$ to 7. Small spheres represent Na atoms and the large one the Pb atom. Symmetries are indicated, and bond angles are given in angstroms. Reprinted with permission from [141], Balbas and Martins, *Phys. Rev. B* 54, 2937 (1996). © 1996, American Physical Society.

level is associated with the 6s level of the Pb atom. The HOMO level $2a_{1g}$ is doubly occupied, and there is a gap of 1 eV between this level and the LUMO, which has e_g symmetry. The electron density is spherically symmetric, consistent with the close-shell character of Na_6Pb [141]. The filling of the $1t_{1u}$ levels in the series Na_mPb ending with Na_6Pb can be viewed as the filling of the atomic $6p$ shell of Pb, perturbed by the effect of the surrounding Na atoms. In summary, the electronic structure of small Na_mPb clusters is dominated by the highly attractive potential of the Pb atom.

The closed-shell structure of Na_6Pb is responsible for the high abundance of this cluster, but a full explanation requires some analysis of the formation process. The energy to remove a Na atom from the cluster, the evaporation energy, is plotted in Fig. 18. There is a pronounced odd-even oscillation of the evaporation energy as a function of m and the smallest value occurs for Na_7Pb. The explanation proposed for the high abundance of Na_6Pb is based on a two-step mechanism [143]. For the experimental conditions with a Pb concentration in the mixed Pb-Na vapor of the order 10%, growth of mixed clusters containing a single Pb atom is favorable compared with other possibilities. Formation of pure Pb clusters is unlikely because of the small concentration of Pb. The gain in binding energy in the step

$$Na_{m-1}Pb + Na \longrightarrow Na_mPb \tag{75}$$

which might be called the capture energy, can be read from Fig. 18; it is equal to the evaporation energy of one Na atom from Na_mPb. This is positive so the process is exothermic and the clusters tend to grow. This capture energy is higher than the capture energy for the growth of small pure Na clusters; consequently, mainly Na_mPb clusters will form if there is a supply of Pb atoms. The competition between different Pb atoms to form Na_mPb clusters and the fact that the capture energy quickly decreases as m increases, suggest that most clusters will not grow larger than Na_7Pb or Na_8Pb, and this completes the first part of the argument.

On the other hand, during the growth process the clusters become hot and cool down by evaporating Na atoms. The evaporation energy increases by a factor of more than 2 between the two consecutive evaporation events $Na_7Pb \rightarrow Na_6Pb + Na$ and $Na_6Pb \rightarrow Na_5Pb + Na$, so the evaporation cascade is expected to stop after the first of these two reactions, and in this way the population of Na_6Pb becomes highly enriched. Evidence for enormous differences in the abundance arising from relatively small differences in the evaporation energies is well documented from experiments for $(C_{60})_N$ [61].

Li_6C and Na_6Pb are just two clusters in a broad class M_6X, where X = C, Si, Ge, Sn, and Pb, and M is an alkali element. This family has been studied by Marsden [144] and Schleyer [142] using DFT. Their calculated structure is the centered octahedron of Fig. 17. The single exception is Li_6Sn; in this case the Sn atom is on the surface but coordinated to the six Li atoms [145]. This special location appears to be due to the similar atomic radii of Sn and Li. The bonding in the entire family is similar, with considerable ionic character. Most of

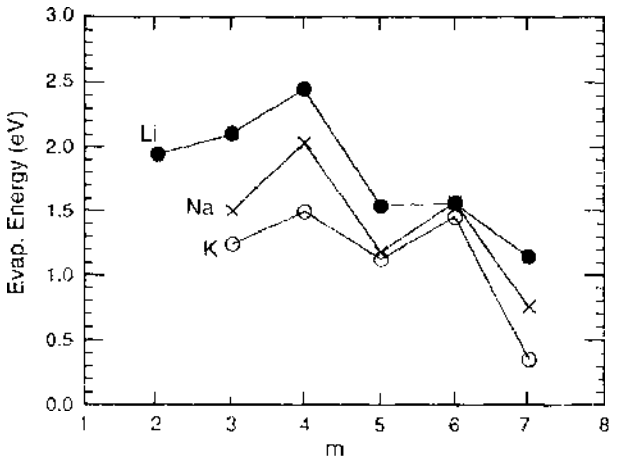


Figure 18. Energy to remove an alkali atom from Li_mPb, Na_mPb, and K_mPb. Reprinted with permission from [131], J. A. Alonso et al., *Chem. Phys. Lett.* 289, 451 (1998). © 1998, Elsevier.

the clusters are quite floppy, and the most rigid one is Li_6C, which is exceptionally stable (1108 KJ/mol, relative to separated atoms). The binding energy increases as the alkali atom changes from Cs to Li, reflecting the growth in alkali–alkali bonding. On the other hand, by changing the tetravalent atom, the binding energies decrease in the order M_6C, M_6Si, M_6Ge, M_6Sn, and M_6Pb, an order that follows the variation of the electronegativity [146] down group 14. The stability of Na_6Pb is then relatively low compared to other M_6X clusters, and a large number of clusters in this family await experimental synthesis.

The charge transfer to the impurity weakens the bonds between the alkali atoms compared with the corresponding pure alkali clusters, and this affects their thermal properties. Comparative DFT simulations of the thermal behavior of Li_6Sn and Li_7 indicate that Li_7 exhibits solid-like behavior at 100 K, showing only small atomic vibrations, whereas liquid-like behavior is predicted for Li_6Sn at that temperature [145].

The ionization potentials of Cs clusters containing oxygen impurities display pronounced drops at some particular sizes [147] that reveal the occurrence of closed electronic shells. The electronic configuration of the oxygen atom is $1s^2 2s^2 2p^4$. In the clusters, the $2p$ shell of the oxygen atom becomes filled by electrons donated by the Cs atoms, forming O^{2-} anions. The remaining valence electrons of the cluster behave according to the shell model [148], and so clusters Cs_mO_x have $m - 2x$ nearly free electrons. The drops in the ionization potential are reproduced by the DFT calculations [148].

3.3.4. Impurities in Aluminium Clusters

Al_{13} has a very symmetrical (icosahedral) structure and 39 valence electrons. One additional electron will produce a closed-shell cluster with 40 electrons, and as expected, the measured electron affinity of Al_{13} is large, 3.6 eV [84], a value similar to that of the Cl atom (3.61 eV), the highest electron affinity of all the atoms of the periodic table. Al_{13}^- is then a very stable anion. A configuration of 40 electrons is also achieved by replacing one Al atom of the cluster by a tetravalent atom C, Si, Ge, Sn, or Pb, and it has been conjectured that icosahedral $Al_{12}X$ clusters with X = C, Si, Ge, Sn, Pb could be candidates for cluster-assembled materials. Calculations for $Al_{12}C$ have predicted that the icosahedral structure is preserved after substitution, with the C atom located at the center of the icosahedron [149–151]. In addition, the electron affinity of $Al_{12}C$ is lower than for the clusters of neighboring sizes $Al_{11}C$ and $Al_{13}C$, and the energy to remove an Al atom is larger. A prediction of icosahedral structure was also made for $Al_{12}Si$ and $Al_{12}Ge$, but for $Al_{12}Sn$ the Sn atom was predicted to substitute for an Al atom on the surface of the icosahedron [151–153].

The electronic structure of this family has been probed by photoelectron spectroscopy (PES) [154]. The PES spectra of $Al_{12}X^-$, with X = C, Ge, Sn, Pb, measured at two different photon energies, 193 and 266 nm, are shown in Fig. 19, together with the spectrum of Al_{13}^-. The spectrum of Al_{13}^- is well understood. Starting with the shell model and reducing the symmetry by taking into account the icosahedral structure leads to the electronic configuration $1a_g^2\ 1t_{1u}^6\ 1h_g^{10}\ 2a_g^2\ 1t_{2u}^6\ 1g_u^8\ 2t_{1u}^6$. DFT calculations suggest that the features labeled A and B in panels (a) and (b) of this figure correspond to electron detachments from the $1t_{2u}$, $1g_u$, and $2t_{1u}$ orbitals, which are closely spaced in energy [83]. The extra electron of the species $Al_{12}X^-$ is expected to occupy the next level, $1h_u$, and a large gap should be observed in the spectra. The gap is observed in the spectrum of $Al_{12}Pb^-$, with a weak feature (X) at lower energies followed by broad features (A and B) at higher energies. The close similarity with the A and B features of Al_{13}^- indicates a rigid filling of the electronic shells for the icosahedral structure. Confirming this idea, features X, A, and B are also evident in the spectra of $Al_{12}Sn^-$ and $Al_{12}Ge^-$. The weaker features observed in the spectrum of $Al_{12}Pb^-$ taken at 266 nm, and marked as solid bars, are interpreted as corresponding to a minor isomer, which is more prominent in the spectra of $Al_{12}Sn^-$ and $Al_{12}Ge^-$. The relative intensities of the extra features in $Al_{12}Sn^-$ and $Al_{12}Ge^-$ depend strongly on the source conditions [154]. Calculations for $Al_{12}Sn$ had predicted a structure in which the Sn atom substitutes for an Al atom on the surface of the icosahedron. The shape of the PES spectrum indicates that both this isomer and the isomer with the Sn atom at the center of the cluster are probably present in the experiments.

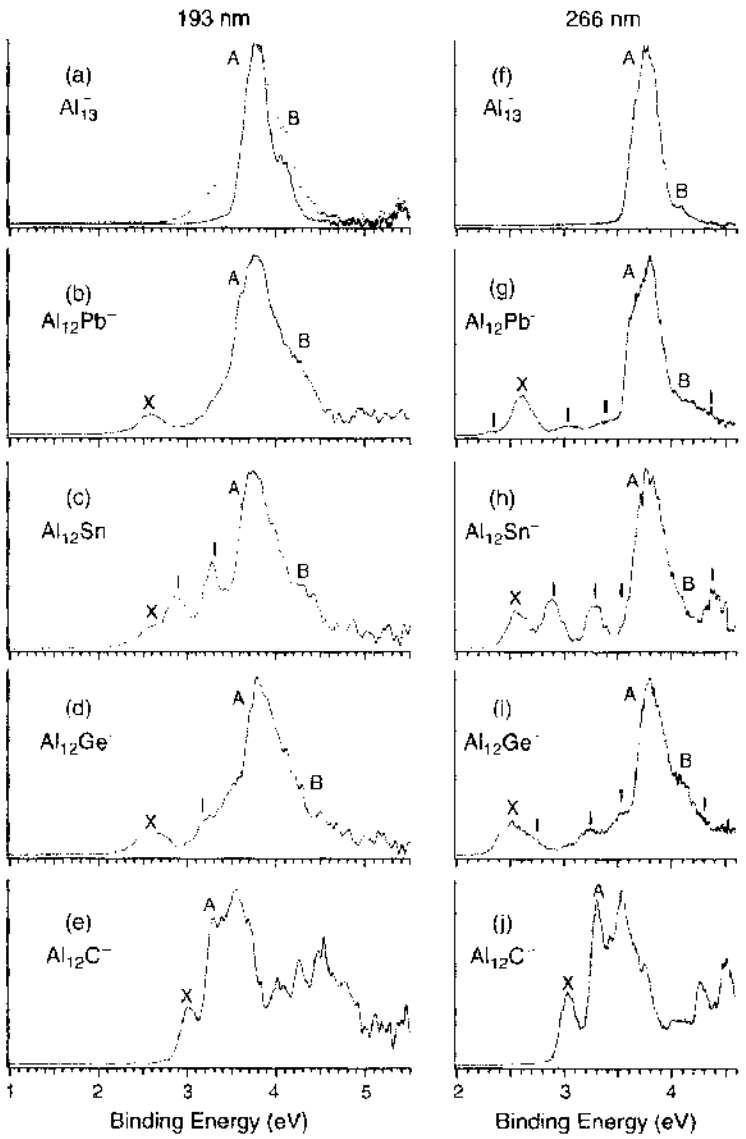


Figure 19. Photoelectron spectra of Al_{13}^- and $Al_{12}X^-$ (X = C, Ge, Sn, Pb) taken at 193 and 266 nm. The continuous and dotted spectra in panel (a) correspond to vibrational temperatures of about 260 K and 570 K, respectively. The bar symbols represent nonicosahedral features. Reprinted with permission from [154], Li and Wang, *Phys. Rev. B* 65, 153404 (2002). © 2002, American Physical Society.

The PES of $Al_{12}C^-$ is completely different. It shows higher adiabatic and vertical electron detachment energies and a smaller gap. In this case, the spectral features are not dependent on the source conditions, indicating that a single isomer is responsible for the spectrum. DFT calculations have consistently predicted the icosahedral structure for neutral $Al_{12}C$ [84, 151–153], but the anionic species appears to have a lower symmetry C_{2v} [84], accounting for the complexity of the spectrum.

Doping Al_{13} with a hydrogen atom also leads to a cluster with 40 valence electrons. $Al_{13}H$ clusters have been produced and exhibit a substantial HOMO–LUMO gap of about 1.4 eV, measured by photoelectron spectroscopy of the negatively charged species [155]. The calculated ground state structure of $Al_{13}H$ [90] preserves the icosahedral structure of Al_{13}. This icosahedron has a small oblate distortion with two triangular faces closer to the central atom than the other faces, and the H atom sits in a hollow position above the center of one of these two special faces. The distance from the H atom to the cluster center is 5.68 a.u., and the distances between the central Al atom and those in the surface range between 5.00 and 5.08 a.u. The density of electronic states for the predicted structure [90] is consistent with the measured photoelectron spectrum. The binding energy of the H atom to Al_{13} is 3.4 eV [90, 156] and the barrier for diffusion to another hollow position in a neighboring face is 0.08 eV.

The state of the H atom can be described as a negatively charged impurity screened by the surrounding electron gas [157], similar to a H impurity embedded in a vacancy in the Al metal.

3.4. Cluster Melting

3.4.1. Orbital-Free Simulations of the Melting Transition

Apart from its scientific interest, the melting of small particles may have some technological implications. In sintering processes, fine powders are compressed and heated until the particles coalesce. A linear reduction of the melting temperature as a function of the inverse particle radius $1/R$ has been observed, which is mainly due to the substantial fraction of atoms on the surface. Consequently, lower sintering temperatures are required for particles with very small radii. In addition, given the current trend toward nanoscale technologies, the extremely small size of the components will affect their mechanical stability at elevated temperatures. The case of metallic clusters presents interesting features, and the alkali metals in particular present an intriguing case.

The melting temperatures of Na clusters measured by Haberland and coworkers [158–160] are shown in Fig. 20. The experimental method is based on the analysis of the temperature dependence of the photofragmentation pattern of a size-selected cluster, which is used to determine directly the caloric curve $E = E(T)$. Melting is then associated with a change of slope in the caloric curve. Large oscillations of the melting temperature T_m as a function of cluster size are observed in the figure. The trend in most properties of clusters is a rather rapid approach to the bulk value as the cluster size increases, but this is not the case for the melting temperature (the melting temperature of bulk Na is 371 K). Even more surprising is the absence of a correlation between the maxima of T_m and the magic numbers for the filling of electronic or geometrical shells. The electronic magic numbers control the size dependence of a number of properties of sodium clusters. For the range of sizes in Fig. 20, the electronic magic numbers are $N = 59, 93, 139, 199, 255$, and 399 (notice that the clusters are singly ionized, so the number of valence electrons is $N - 1$). These are indicated by the vertical dotted lines. The magic numbers for having complete icosahedral structures, $N = 55, 147$, and 309, are indicated by the dashed lines.

Computer simulations reveal a complex picture of the melting of alkali metal clusters. Often in molecular dynamics simulations a simplifying assumption is to represent the interaction between the ions by a sum of pair potentials. This allows the simulation of large samples for

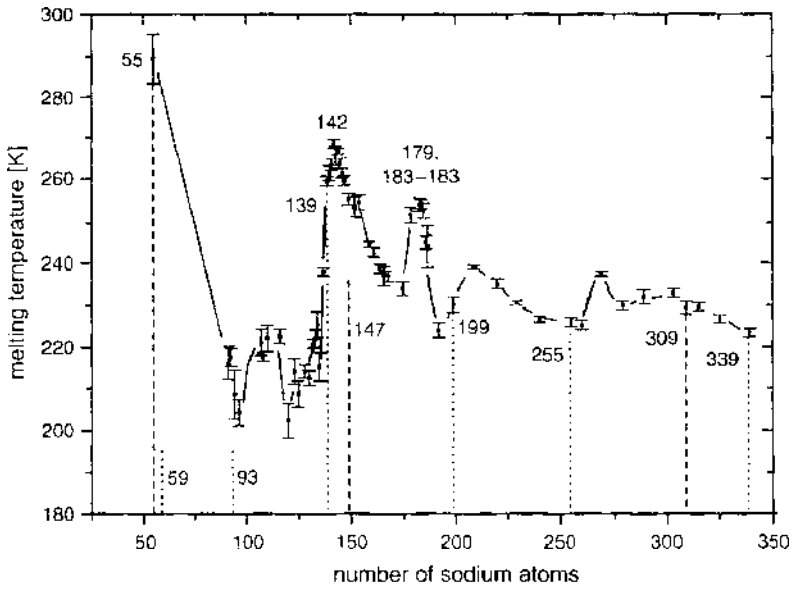


Figure 20. Melting temperature of Na_N clusters as a function of the number of atoms in the cluster. Sizes corresponding to electronic (dotted lines) and geometric (dashed lines) shell closings are indicated. Adapted with permission from [160]. Schmidt et al., "Physics and Chemistry of Clusters," p. 22 (2001). © 2001, World Scientific.

long times. An essential difficulty in the simulation of metallic clusters in general and alkali metal clusters in particular is that the interactions between the ions cannot be represented as the sum of two-body terms because of the itinerant nature of the electrons in metallic systems; this is especially so near a surface where the electron density decays from bulk values to zero. The unification of the Kohn-Sham form of DFT and molecular dynamics (MD) formulated by Car and Parrinello [9] allows for an explicit treatment of the electronic degrees of freedom, and this technique has been used for modeling the melting of some clusters (see next section). However, in order to calculate the forces on the atoms, the Kohn-Sham equations have to be solved at each step, that is, for each geometrical configuration of the atoms in a MD trajectory, and this is very demanding of computer resources. At present, full Kohn-Sham MD simulations can only be performed for small samples of ions, typically fewer than 50 and for short simulation runs, usually less then 10 ps. Large computational savings can be obtained by using an orbital-free (OF) energy functional in the OFT–MD simulations, in which the kinetic energy of the electrons appears as an explicit functional of the valence electron density $n(\mathbf{r})$ as discussed in Section 2.1. The expression

$$T_s[n] = T^{(0)} + T^{(2)} \tag{76}$$

is a simple example of such kinetic energy functionals, retaining the first two terms in the expansion, Eq. (28), of the kinetic energy in gradients of the density, appropriate for a system of slowly varying density. Adding to the kinetic energy the other contributions to the energy of the system gives the total energy as an explicit functional of the electron density $n(\mathbf{r})$. Using a large plane wave basis to represent the density, the ground state energy for a given geometrical configuration of the atoms is obtained directly by minimization of the total energy functional.

Using this functional for the kinetic energy of the valence electrons, and replacing the ion cores by local pseudopotentials [161, 162], the melting of Na clusters has been studied [35, 163] by constant energy MD simulations. The melting transition was identified in the computer simulations by an analysis of the caloric curve and the bond length fluctuations. The caloric curve represents the thermal response to an increase in energy and is given as a plot of the total energy as a function of the internal temperature T. In the MD simulations, the cluster temperature is defined by the expression

$$T = \frac{2}{(3N-6)k_B}\langle E_{\text{kin}}\rangle_t \tag{77}$$

where $\langle\ \rangle_t$ represents the time average over the entire trajectory, E_{kin} is the instantaneous kinetic energy of the cluster, N is the number of atoms, and k_B is the Boltzmann constant. The specific heat C_v is given by the slope of the caloric curve. In a MD simulation, the C_v per atom can be conveniently obtained from the expression (in units of the Boltzman constant) [164]

$$C_v = \left[N - N\left(1 - \frac{2}{3N-6}\right)\langle E_{kin}\rangle_t \langle E_{kin}^{-1}\rangle_t\right]^{-1} \tag{78}$$

A quantitative measure of the atomic motion is given by the relative root mean square bond length fluctuation

$$\delta = \frac{2}{N(N-1)}\sum_{i<j}\frac{(\langle r_{ij}^2\rangle_t - \langle r_{ij}\rangle_t^2)^{1/2}}{\langle r_{ij}\rangle_t} \tag{79}$$

where r_{ij} is the instantaneous distance between atoms i and j, and the sum runs over all the pairs of atoms in the cluster.

For small clusters, a melting transition in stages is predicted [35]. For Na_8 and Na_{20}, the melting transition is spread over a broad temperature interval of 100 K. The specific heat of Na_8, shown in Fig. 21, has a sharp peak at $T = 110$ K, coinciding with a jump in δ. This is followed by a steady increase of $\delta(T)$, until a leveling off occurs at T near 220 K. The temperature $T = 110$ K marks the onset of isomerization transitions among the permutational isomers of the low-temperature structure. The orbital-free method predicts an octahedral

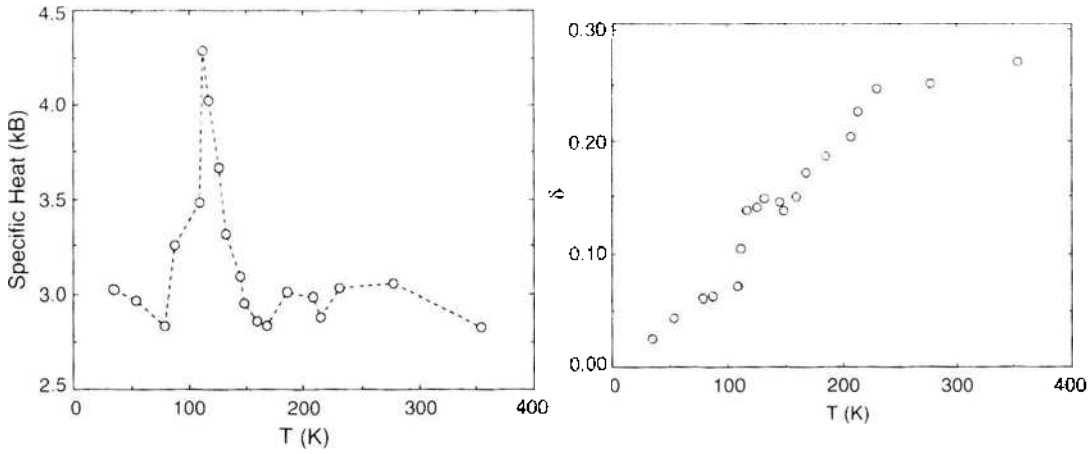


Figure 21. Specific heat and rms bond length fluctuation of Na_8 as functions of the internal temperature. Adapted with permission from [35], Aguado et al., *J. Chem. Phys.* 111, 6026 (1999). © 1999, American Institute of Physics.

low-temperature structure in agreement with the predictions of a number of Kohn-Sham calculations [62]. With increasing temperature, the isomerizations become more frequent and induce progressively stronger distortions of the structure until, at $T = 220$ K, all the atoms diffuse across the cluster volume and its shape changes continuously, indicating a fully established liquid phase. Because this effect is progressive, it does not result in a peak in the specific heat. The specific heat of Na_{20} has two peaks at 110 and 170 K, correlating with abrupt increases of δ at 110 and 160 K, respectively. Then, δ levels off at 220 K. The peak at 110 K marks the onset of isomerization transitions between permutational isomers almost preserving the low-temperature structure, a capped double icosahedron. The peak at 160 K arises from the transformation of the double icosahedron to a new structure with a single central atom. The exchanges of this atom with one of the 19 surface atoms occur at a slower rate than the interchanges between surface atoms. The surface atoms are very mobile and the cluster shape fluctuates a great deal. Finally, at $T = 220$ K the structure is very fluid.

Constant temperature MD simulations using a phenomenological interatomic potential found for Na_8 a transition at $T = 100$ K from the solid-like phase to one characterized by occasional atom interchanges while the atoms stay for relatively long periods close to their equilibrium positions [165]. Monte Carlo (MC) simulations by Calvo and Spieglemann [166] using an empirical potential, predicted melting at $T = 80$ to 100 K, while tight-binding MC simulations [167] predict $T_m = 200$ K. In view of the complex behavior revealed by the orbital-free simulations, the other results are not surprising. For Na_{20}, the tight-binding [167] and empirical potential simulations [165, 166] predict melting in two steps, the first one involving the surface atoms only. The occurrence of melting in steps has been associated with a soft repulsive, short-range interatomic interaction and this is, indeed, the case for sodium [163].

Direct calorimetric measurements of the melting temperature of such small clusters have not been performed, but the temperature dependence of the photoabsorption cross section has been reported [168, 169] for Na_N^+, $N = 4$ to 16. Although the spectra do not show evidence of a sharp melting transition, comparison between theory and experiment is encouraging. There is no appreciable change in the spectrum of Na_8 until $T = 105$ K, at which point it begins to evolve in a continuous way. It is noteworthy that this temperature is close to $T = 110$ K predicted by the orbital-free simulations for the beginning of a broad melting transition.

The complex behavior of small sodium clusters can be expected, because most of the atoms lie on the surface. Larger clusters are required for the inner atoms to experience a bulk-like environment. The melting of Na_{55}, Na_{92}, and Na_{142} has also been simulated using the orbital-free method [163]. The precise low-temperature structure of these large clusters is not known. DFT calculations predict a structure close to icosahedral for Na_{55} [170], and using empirical potentials Calvo and Spiegelmann [166] predicted icosahedral structures for Na_{55}, Na_{93}, Na_{139}, and Na_{147}. Consequently, complete or incomplete icosahedral structures were

taken as plausible low-temperature structures to start the OF simulations of the heating process for Na_{55}, Na_{92}, and Na_{142}. The specific heat of Na_{142}, shown in Fig. 22, displays a main peak at $T = 270$ K and a smaller one at $T = 240$ K. These two peaks are so close that only one slope change can be distinguished in the caloric curve. An analysis of the atomic motions suggests that the peak at $T = 240$ K can be associated with surface melting. The diffusion coefficient $D(T)$ shows a sharp increase in slope at this temperature. The peak at $T = 270$ K represents complete melting. At low temperatures, the radial atomic density distribution calculated with respect to the cluster center displays the atomic shells characteristic of the icosahedron, and these shells broaden as T increases. Above the surface melting temperature, the shells gradually vanish, and further changes in the slope of $D(T)$ are not detected. At $T = 270$ K, the atomic density distribution is nearly uniform over the volume of the cluster. This theoretically predicted melting temperature is in excellent agreement with the observed melting temperature for Na_{142} of 280 K. In addition, the latent heats of fusion, q_m, estimated from the step at T_m in the caloric curve, q_m(experimental) = 14 meV/atom and q_m(theoretical) = 15 meV/atom, are in good agreement.

Two-steps melting is also predicted for Na_{92}, with a small pre-peak in the specific heat at $T = 130$ K, and a large peak at 240 K corresponding to homogeneous melting. In this case, the two features are well separated. The position of the large peak and the latent heat, $q_m = 8$ meV/atom, are again in excellent agreement with the corresponding experimental values: $T = 250$ K and $q_m = 7$ meV/atom. Once more the diffusion coefficient is very sensitive to the appreciable diffusive motion that begins when surface melting occurs. The results of the MC simulations [166] for clusters with sizes in the range of 100 atoms are consistent with the OF simulations, namely two step melting for Na_{139} and Na_{93}, with the two features close together in Na_{139} and well separated in Na_{93}.

The orbital-free simulations for Na_{55} predict melting in a single step at $T = 190$ K. Single-step melting is also predicted by the MC simulations [166]. The experiments indicate a substantial enhancement of the melting temperature at $N = 55$ (see Fig. 20). The melting temperature of this cluster is even higher than that of Na_{142}^{+}, which is a local maximum. Neither the OF simulations discussed above nor the MC simulations reproduce this feature.

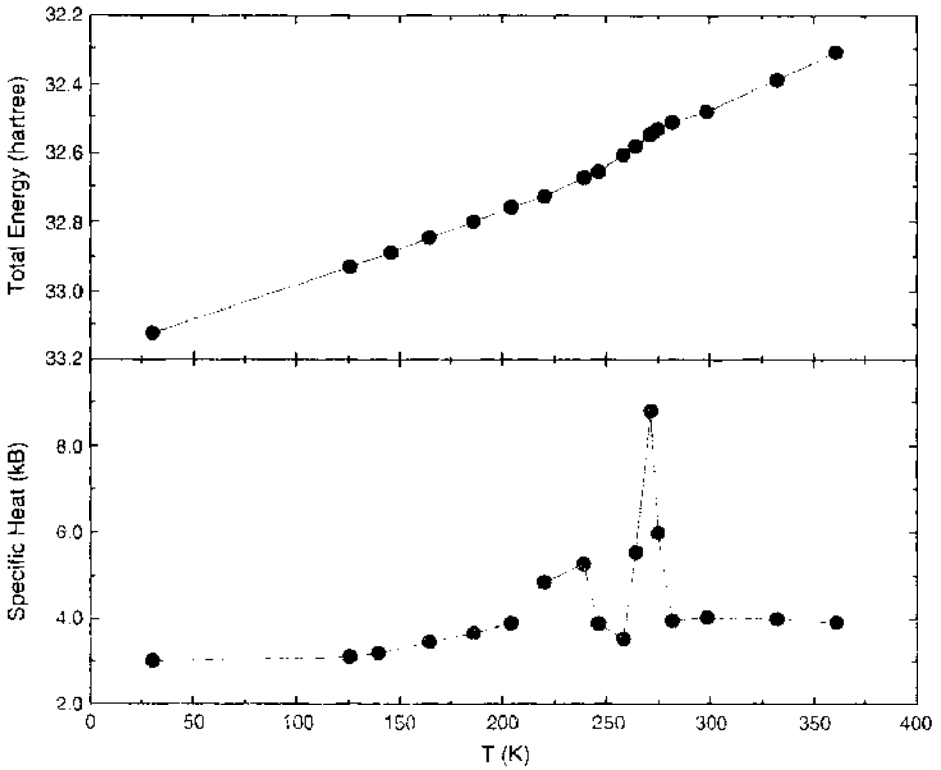


Figure 22. Caloric and specific heat curves of Na_{142}. Reprinted with permission from [163], Aguado et al., *J. Phys. Chem.* 105, 2386 (2001), © 2001, American Chemical Society.

Blaise and Blundell [171] performed OF calculations using for the $T^{(2)}$ term of Eq. (76) an enhancement coefficient $\lambda = 1.44$. The calculated melting points, shown in Fig. 23, give the melting temperature of Na_{55}^+ as a local maximum, but again this temperature is lower than those of clusters in the size range of about 140 atoms. The figure also shows that the calculated melting temperatures are substantially smaller than the measured ones and smaller than the melting temperatures obtained by the current authors [35, 163]. The main reasons for this discrepancy may be differences in the pseudopotential and the kinetic energy functional that were used in the calculations. There were also differences in the statistical mechanics prescription used to obtain T_m. Clearly, further work is required to explain fully the trends of the measured melting temperatures.

3.4.2. Ab Initio DFT Simulations of Melting

The full Kohn-Sham formulation of DFT has been used in molecular dynamics simulations of the melting of small clusters of simple metals. Blundell and coworkers [172] studied Na_8 and Na_{20} with the objective of testing ionic pseudopotentials and of comparing the results with those of the OF simulations. The results of all models are qualitatively similar, but the precise positions of the peaks in the specific heat are sensitive to the pseudopotential. Na_8, Na_{10}, Na_{14}, Na_{20}, Na_{40}, and Na_{55}^+ were studied by Rytkonen et al. [173]. The simulations for each cluster started from a low-temperature structure that was expected or known from other works. The melting transition was explored with two customary indicators: the caloric curve and the diffusion coefficient. A melting region could not be identified for clusters smaller than Na_{20}. Na_{20} was observed to melt in the range 235–275 K. Na_{40} melted around 300–350 K. Melting was found to broaden the density of electronic states of this cluster, but the HOMO–LUMO gap, approximately 0.4 eV wide, remained open until higher temperatures were reached. The Na_{55}^+ cluster was found to melt around 310–360 K. There were well defined deformations of the molten clusters: Na_{40} exhibited an octupole deformation, and liquid Na_{20} and Na_{55}^+ were prolate. The melting temperatures of this work appear to be too large compared with the measured values shown in Fig. 20. This discrepancy may be due to the fast heating rate used in the simulations which was necessitated by the time consuming Kohn-Sham procedure, but which was probably too rapid for thermal equilibrium to be achieved at a given temperature.

The clusters Al_{13} and Al_{14} behaved differently on heating [174]. There are no distinct transitions, and isomerizations between the starting icosahedral structures and other structures, mainly decahedral, were seen frequently. The isomerizations led to peaks in the potential energy as a function of temperature and occur by a mechanism in which the atoms collectively twist between the different isomers. For instance, in the case of Al_{14}, the outermost atom does not move around the cluster; instead, the atom often inserts itself into the cluster, causing a "new" outer atom to be expelled on the opposite side. Some of the isomerizations in Al_{13}^- involve more drastic structural changes, and the behavior at the highest temperatures studied

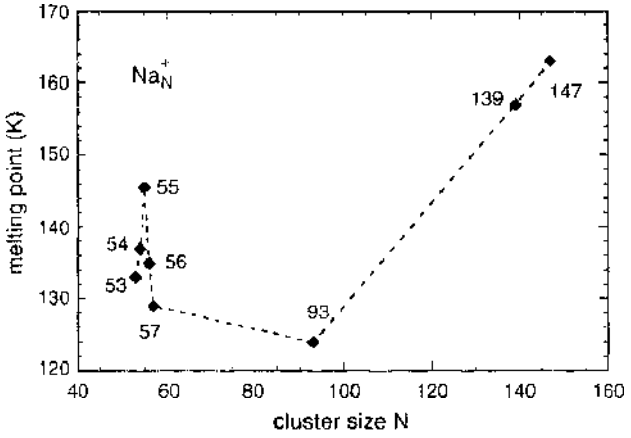


Figure 23. Melting points calculated by orbital-free simulations for singly charged sodium clusters. Reprinted with permission from [171], Blaise and Blundell, *Phys. Rev. B* 63, 235409 (2001). © 2001, American Physical Society.

(1500 K) is consistent with a molten phase. The vibrational density of states for Al_{13}^- shows a shift toward lower frequencies and a gradual smearing out as the temperature increases, supporting the interpretation of gradual melting above 1000 K. However, the melting of bulk aluminium metal occurs at 933 K, and it is likely that the difficulties in performing long simulations may account for the overestimation.

3.5. Cluster Assembly

The *ab initio* DFT calculational schemes provide the opportunity for simulating the assembling of a bulk material using stable clusters as the building blocks. The supercell geometry lends itself to positioning the clusters initially in an arrangement to be investigated in an expanded supercell, and then shrinking the superlattice in stages drawing the clusters together. At each stage, the electronic structure can be relaxed into its ground state and atomic positions relaxed to the equilibrium arrangement. As the clusters become closer, the molecular orbitals will begin to overlap, the clusters will interact, and the atomic arrangement within a cluster will respond. As the shrinking of the superlattice continues and the interaction between clusters increases, the identity of the clusters may be preserved, or atoms or larger fragments may leave the clusters and occupy positions in the interstices, or if the interaction is strong enough the clusters may distort and merge so that the cluster building blocks cannot be identified in the bulk structure. The equilibrium structure of a given type is obtained when the total energy as a function of superlattice spacing is a minimum. A few simulations of this type have been reported. They provide insight into the interaction between clusters and the conditions for cluster assembly of solids.

3.5.1. Clustering in Crystalline Alloys of Alkali Metals and Elements of the Lead Group

The equiatomic alloys of Pb with Na, K, Rb, or Cs contain tetrahedral Pb_4 clusters surrounded by the alkali atoms. The structure of the alloys is shown in Fig. 24. The same clustering occurs for the alloys of the alkali metals with Si, Ge, or Sn, which are elements in the same column as Pb in the periodic table [175] (the exceptions are the Li alloys in which Pb_4 clusters do not appear). In these alloys, each tetrahedral Pb_4 cluster is surrounded by alkali atoms capping the faces and edges of the tetrahedron. The free Pb_4X_4 clusters, where X denotes the alkali element, are the key for understanding the structure of the alloys and have been investigated by DFT calculations [131]. The calculations used the LDA, but tests performed for selected clusters with the generalized gradient approximation for exchange and correlation [176] support the conclusions obtained with the LDA. The calculated lowest energy structure of all the free Pb_4X_4 clusters is a Pb_4 tetrahedron surrounded by another, oppositely oriented alkali tetrahedron with the alkali atoms capping the faces of the Pb_4 tetrahedron. However, the ground state structure of the free Pb_4 cluster is a planar rhombus, and the structural change to the tetrahedron is a result of a transfer of electronic charge from the alkali atoms. The outer electronic configuration of the Pb atom is $6s^26p^2$, so the free Pb_4 cluster has 16 valence electrons. However, Pb_4 has also a tetrahedral isomer, less stable than the planar rhombic one, but with an interesting characteristic: the LUMO of the tetrahedral isomer is a fourfold degenerate level (including spin) and a substantial gap exists above this LUMO. A transfer of four electrons to Pb_4 would result in a total of 20 electrons, which is a shell-closing number for clusters of tetrahedral symmetry. Indeed, the higher electronegativity of Pb leads to electron transfer from the alkali atoms to the Pb_4 unit, and the transferred electrons fill the LUMO of the Pb_4 tetrahedral isomer. The four alkali atoms added to Pb_4 convert this cluster into a Pb_4^{4-} polyanion with closed electronic shells. The calculations predict large HOMO–LUMO gaps for this family of clusters: 2.54 eV for Pb_4Li_4, 1.92 eV for Pb_4Na_4, and 1.36 eV for Pb_4K_4. Even so, the presence of a large HOMO–LUMO gap in the free clusters is not by itself enough to explain the clustering in the bulk alloys, as Pb_4 clusters do not form in the alloys when the alkali partner is Li, and interestingly, the exception occurs in the case of the largest HOMO–LUMO gap.

A cluster with a large HOMO–LUMO gap is less reactive than others with small or no gaps. Chemical passivity is, nevertheless, difficult to achieve in metallic clusters because the

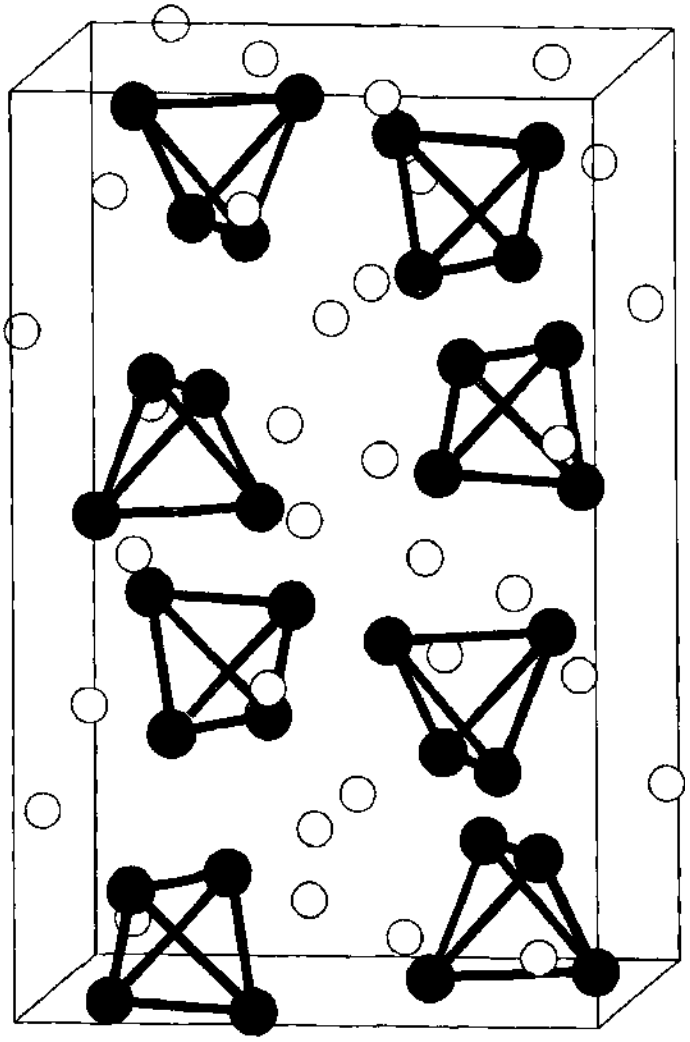


Figure 24. Unit cell of the crystalline compound $PbNa_4$. Tetrahedral Pb_4 clusters are surrounded by the Na atoms.

molecular orbitals are less localized than in covalent clusters, and overlap between orbitals of adjacent clusters often leads to substantial interaction. For the family under study, the interaction is small when the size of the cations is large enough to prevent the clusters coming close. This idea is illustrated by the map in Fig. 25, where a number of clusters have been represented in terms of two coordinates, the HOMO–LUMO gap and the diameter of the alkali cation [177]. In addition to the Pb_4X_4 clusters, other closed-shell clusters of compositions PbX_4 and PbX_6 have been included. The clusters PbX_4 are also very stable. The structure of $PbLi_4$ is shown in Fig. 17 and for other alkaline elements the structure becomes a pyramid with the Pb atom at the apex [143]. In spite of the high stability of the PbX_4 clusters, crystalline alloys with compositions near 20% Pb do not show any clustering features. The cluster $PbNa_6$ was found to be extremely abundant in gas phase experiments [139] and has been suggested as a candidate to form cluster-solids [178]. This cluster was discussed in Section 3.3. The boundary drawn in Fig. 25 separates clusters related to *stable* cluster solids from the rest. The conclusion is that both a high HOMO–LUMO gap and a large cation size are stringent requirements for the occurrence of clustering in the solid alloys. Clusters of composition PbX_6 are, evidently, bad candidates. The emerging picture is that the alkali cations are not part of the cluster *in the crystal*. The Pb_4^{4-} polyanions are the clustered units, with the cations playing a secondary but important role by providing a barrier between the anions and donating electrons to them so that the polyanions survive in the solid. Clusters with compositions PbX_4 and PbX_6 do not lead to clustering in solid alloys because in these cases the cluster core reduces to a single Pb atom.

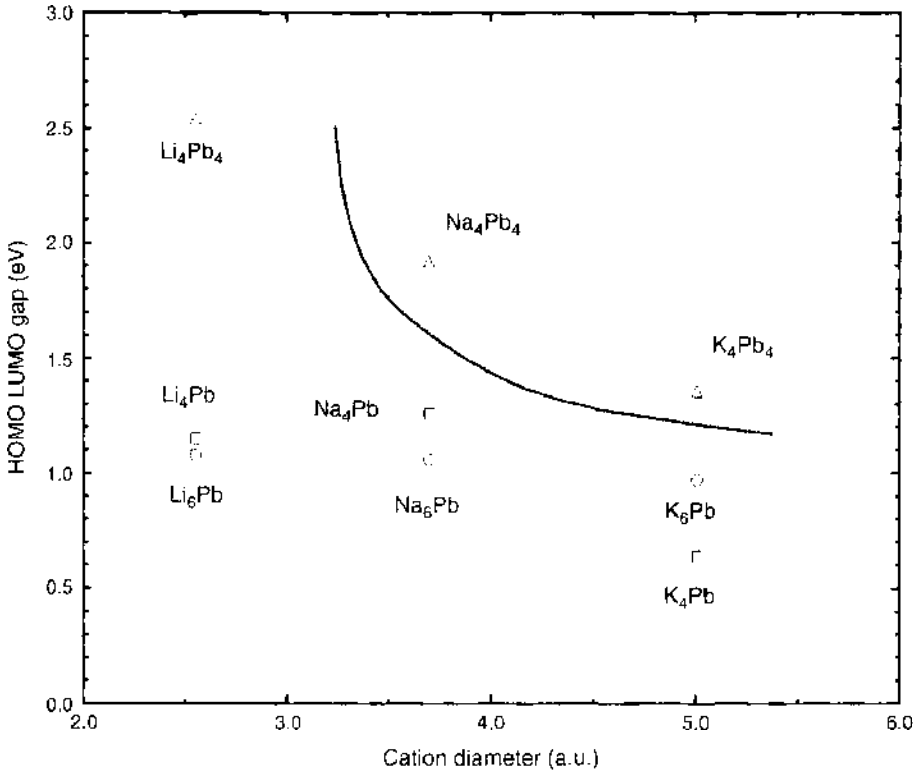


Figure 25. Map for clusters in terms of two coordinates: the HOMO-LUMO gap and the cation diameter. Solid alloys showing clustering effects are related to the clusters on the upper right corner. Reprinted with permission from [177], Molina et al., "Recent Advances in Density Functional Methods, Part III," p. 234 (2002). © 2002, World Scientific.

3.5.2. Simulation of Assembling of Alkali–Lead Clusters

These ideas have been confirmed by computer simulations of the assembling of Pb_4Na_4, Pb_4K_4, and Pb_4Li_4 clusters [179]. The calculations were performed using the *ab initio* method synthesizing the supercell geometry, plane wave basis, and nonlocal normconserving pseudopotentials described earlier. In the equiatomic PbX solid compounds, the Pb_4 clusters form a body-centered-tetragonal (b.c.t.) lattice with axis ratio $c/a = 1.676$ as illustrated in Fig. 24. The simulations started by placing the free Pb_4Na_4 clusters on a b.c.t. lattice with the same c/a ratio but with dimensions c and a enlarged by a factor $f = 2$. The factor f was then reduced in steps maintaining $c/a = 1.676$, and at each step the electronic energy and all the atomic positions were allowed to relax to minimize the energy. Initially, the clusters were far enough apart for the interaction between them to be negligible, but as the cell size was reduced, the Pb_4Na_4 units were unable to maintain independent identities, and in order to achieve an efficient packing some rearrangement of the Na atoms occurred. However, each Pb_4 tetrahedron remained surrounded by four cations capping its four faces. More precisely, each Pb_4 tetrahedron is fully covered by alkali cations capping the faces and the edges of the tetrahedron, and each cation is shared by two Pb_4 units. The structure obtained at the end of the simulation when the lattice parameters minimized the total energy is precisely the structure observed for the crystalline compound. The electron density in the regions of the crystal occupied by the cations is smaller than the average electron density in bulk Na, supporting the picture of the alloy as formed from charged polyanions surrounded by the alkali cations. The density of electronic states for the equilibrium structure has a small gap at the Fermi level, so the crystal is a narrow gap semiconductor. The binding energy gained by assembling the clusters into a solid is $\delta_{assembling} = 4.09$ eV per cluster, to be compared with the value $\delta_{clustering} = 11.19$ eV of the binding energy of a single Pb_4Na_4 cluster with respect to the separated Pb and Na atoms. $\delta_{assembling}$ is nearly one-third of $\delta_{clustering}$. The results are similar for the assembling of the PbK solid, with the polyanions even more widely separated from each other because of the larger size of the K^+ cation compared with Na^+. In this case

$\delta_{assembling} = 4.14$ eV, and $\delta_{clustering} = 10.54$ eV, so the ratio of the two quantities is 0.39. In the fullerite crystal, which is the prototypical case of a cluster-solid, the ratio $\delta_{assembling}/\delta_{clustering}$ is, however, much smaller.

The simulation of the assembling of Pb_4Li_4 clusters gives totally different results. As the cell size of the supercrystal is reduced the clusters interact strongly. Each Pb_4 tetrahedron opens up forming a bent rhombus (or butterfly), a shape midway between the tetrahedron and the planar rhombus, that reflects a lower charge transfer from the Li to the Pb atoms in comparison to the free cluster. Distances between Pb atoms in neighboring butterflies are similar to bond lengths inside the butterflies. The strong interaction between clusters leads to metallic character for the assembled solid, whose structure, based on an interconnected network of Pb_4 butterflies, is very different from that of PbNa (or PbK, PbRb, PbCs) shown in Fig. 24. The assembled solid is metastable and cannot be characterized as a solid containing clusters. Furthermore, calculations for the PbLi alloy with the experimental CsCl structure give a cohesive energy of 2.4 eV/atom which is larger (Table 2) than the cohesive energy of the metastable PbLi solid. Also in agreement with experiment, Table 2 shows that the body-centered-tetragonal clustered solid is more stable than the CsCl phase for PbNa and PbK.

3.5.3. Assembling of Al-Based Clusters

It has been suggested that icosahedral clusters $Al_{12}X$ with X = C, Si, Ge, Sn, Pb could be good candidates for cluster-assembled materials [180]. This family of clusters, which was discussed earlier, has a 40-electron closed-shell configuration, which is associated with high stability. Khanna and Jena [180] used DFT methods to calculate the binding energy of two tetrahedral Mg_4 8-electron closed-shell clusters and found weak binding. Based on this result, they speculated that similarly stable and symmetrical $Al_{12}Si$ and $Al_{12}C$ closed-shell clusters might interact weakly when assembled into a solid.

This suggestion was investigated by Seitsonen et al. [181] using *ab initio* methods to simulate the cluster assembly. As a first step, $Al_{12}Si$ rigid icosahedral clusters were assembled on a face-centered-cubic (f.c.c.) lattice and the equilibrium lattice was found by minimizing the total energy. A cohesive energy with respect to separated clusters of 6.1 eV per cluster was calculated for this minimum energy structure, but there was no energy gap in the density of states at the Fermi level and the system appeared to be metallic. In addition, the equilibrium lattice parameter was so small that the shortest intercluster Al–Al distance was less than the intracluster Al–Al bond length. Next, full relaxation of the atomic positions was allowed, and when equilibrium had been achieved the shortest Al–Al bonds had increased to 2.71 Å and open regions of the unrelaxed cluster lattice were more filled. The clusters had merged, and the short range atomic order indicated that the atoms had moved toward a close-packed arrangement. Similar conclusions were reached for $Al_{12}C$ by Seitsonen et al. [150]. Again using *ab initio* methods, these authors found that although the isolated clusters were very stable, the f.c.c. solid composed of $Al_{12}C$ clusters was unstable against melting when the atoms were allowed to relax individually.

A simulation that successfully achieved cluster-assembled $Al_{12}C$ and $Al_{12}Si$ solids was reported by Gong [151]. The electron density of $Al_{12}C$ is not spherical; there are protrusions and indentations. Taking account of these asymmetries in the electron densities of the isolated clusters, Gong designed a cubic-like solid structure with 8 clusters per unit cell in which each cluster is oriented 90° with respect to all its near neighbors. The orientation of the clusters is illustrated in Fig. 26. In this structure, the overlap of the electron densities of neighboring clusters shown in Fig. 27 is reduced over the overlap for a f.c.c. structure. The equilibrium structure was found by minimizing the total energy, and the resulting solid had a very small

Table 2. Cohesive energy (eV/atom) for solid compounds in two structures.

Alloy	Assembled from clusters	CsCl
LiPb	2.325	2.415
NaPb	1.910	1.895
KPb	1.835	1.545

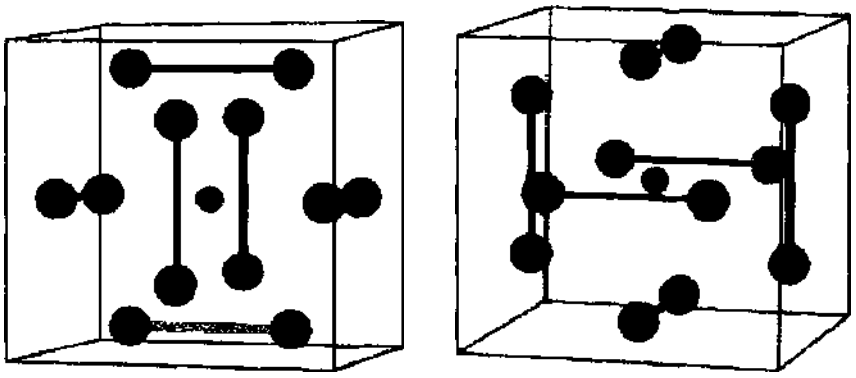


Figure 26. The atomic structure of $Al_{12}C$ cluster dimer, and the relative orientation of $Al_{12}C$ clusters in the solid. Big balls for Al atoms and the small balls for C atoms. Reprinted with permission from [151], Gong, *Phys. Rev. B* 56, 1091 (1997). © 1997, American Physical Society.

cohesive energy of ~1.1 eV per cluster implying that the $Al_{12}C(Si)$ clusters are condensed by a rather weak van der Waals force. The clusters retained their identity in the equilibrium solid with the shortest intercluster Al–Al distance being much longer than the Al–Al intracluster bond length or the nearest neighbor distance in bulk Al. A MD simulation of the solid at a temperature of ~20 K provided a further check of the stability. Apart from vibrational motion of the atoms, the icosahedral structure of the $Al_{12}C$ clusters was unchanged during the 1.1 ps MD run.

We finally discuss the assembling of $Al_{13}H$ clusters, which also highlights the importance of the relative orientation of clusters in cluster assembly. These are clusters with 40 valence electrons and a sizable HOMO–LUMO gap of 1.4 eV, and we recall from a previous section that the structure is an icosahedron with the H atom chemisorbed on the center of a triangular face. The study of the $(Al_{13}H)_2$ cluster-dimer provides insight on cluster assembling [183–185]. Two isomers characterized by different relative orientations of the two Al_{13} units were studied. In the first dimer, the two clusters had parallel faces in contact, rotated 180° relative to one another. The potential energy of the dimer was calculated, for frozen cluster geometries, as a function of the cluster–cluster distance, and the energy minimum was obtained for a separation of 15.74 a.u. All the atomic coordinates were then relaxed for that separation. The new atomic arrangement preserved the intrinsic structure of the dimer. The binding energy of the cluster–dimer with respect to the separated clusters was 1.74 eV. The relative orientation

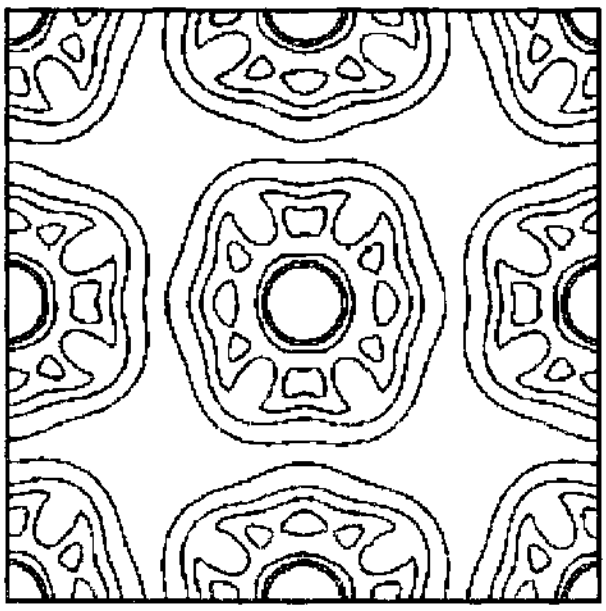


Figure 27. Contour map of the electron density in a (001) plane for $Al_{12}C$. Reprinted with permission from [151], Gong, *Phys. Rev. B* 56, 1091 (1997). © 1997, American Physical Society.

in the second dimer is such that the clusters have edges in contact, but one of the clusters is rotated 90° with respect to the other so that the contact edges are perpendicular to each other. In this configuration, the relaxed dimer structure has a binding energy of 3.03 eV, and the increase of 1.28 eV with respect to the first isomer is a consequence of the favorable relative orientation of the two clusters, a point that has obvious implications for cluster assembling. A similar structure with perpendicular contact edges was also found to be the ground state of the dimer composed of two icosahedral clusters in the pair potential calculations of Sun and Gong [186]. The stability of these dimers suggests that linear nanowires built from these clusters could be synthesized.

Simulations aiming to model the assembling of a nanostructured material from $Al_{13}H$ clusters were performed using *ab initio* methods [183–185]. The optimal relative cluster–cluster orientation in the cluster-dimer suggested a favorable configuration for an assembled solid with neighboring clusters having edges in contact and perpendicular to each other. This condition leads to a simple cubic lattice. The energy of the assembled solid was first calculated while maintaining the structures of the individual clusters as frozen. A curve with two minima was obtained. The less relevant outer minimum corresponded to a large lattice constant of 32 a.u. and a small binding energy with respect to the isolated clusters of only 0.35 eV per cluster. A similar minimum was found by Gong [151] in the earlier simulations of the assembling of $Al_{12}C$ and $Al_{12}Si$ clusters. The lattice constant of the inner minimum was 24.2 a.u. In this arrangement the Al–Al distances between atoms in neighboring clusters were comparable to the intracluster Al–Al distances. The solid was metallic but the density of states was very different from that of bulk Al. The binding energy of the assembled material was 1.5 eV per cluster, about one-third of the internal binding energy of the free $Al_{13}H$ cluster. This fraction is similar to that found for the assembling of the Pb-alkali clusters, and seems to give the characteristic order of magnitude for the assembling of clusters formed by typical metallic elements. In contrast, the ratio for the fullerite is one order of magnitude smaller. Molecular dynamical simulations at 150 K were performed to test the stability of the clusters. During the 3 ps simulation, the assembly was stable with only mild distortions. However, the temperature was high enough for the H atoms to migrate toward the open regions in the assembled lattice between the Al_{13} clusters.

We should conclude from the theoretical work on the assembling of Al_{13}-based clusters that it is no simple matter to test the stability of a cluster-assembled solid through a simulation. The internal structure of the clusters, the relative orientation of the clusters, and the structure into which they are assembled are all factors that have to be considered.

4. CONCLUSIONS

Simulations based on density functional theory have provided a very useful theoretical tool for calculating and interpreting the electronic and structural properties of clusters. The most distinctive characteristic of small clusters of the simple *sp*-metals is the existence of magic numbers in the abundance mass spectrum of clusters formed by the usual supersonic expansion techniques. This and other evidence shows that the clusters with magic numbers are particularly stable with respect to neighboring clusters. A simple droplet-like model based on DFT in which details of the atomic structure are smeared out explains the magic number effect. Electronic shells are formed, and enough electrons to completely fill a shell leads to a closed-shell structure and stability paralleling the rare gas atoms. Full *ab initio* calculations taking into account the detailed geometrical structure of the ionic skeleton of the clusters corroborate this interpretation. The existence of electronic shells has a strong influence on the variation of many electronic properties of the clusters as a function of size. Structural properties are also influenced. For example, the geometrical structure of aggregates is closely related to their electronic structure, and in extreme cases, like some doubly ionized clusters, the structure can be viewed as supermolecules composed of interacting closed-shell fragments.

The chemical reactivity of the clusters with a closed-shell electronic structure is weaker than the reactivity of open-shell clusters. This provides motivation for a search for closed-shell clusters with exceptional stability that could form the building blocks of new cluster-assembled materials. However, a closed-shell and a large HOMO–LUMO gap are necessary conditions,

but are not enough for achieving a successful cluster self-assembling. The reactivity of metallic clusters is normally large and we have from the simulations of the family of alloys formed by alkali metals and elements of the Pb group that an efficient way of passivating the clusters is by coating. For example, in the case of the PbNa alloy, the building blocks are Pb_4Na_4 clusters, but on assembling, the coatings of Na^+ ions leave the clusters, move into the interstices, and form a barrier separating the robust Pb_4 cluster cores. In fact, experimentalists have already used this tool to produce assemblies of size-selected gold clusters coated by organic molecules [187].

The density functional–based *ab initio* methods can be useful in assessing the reactivity of clusters that are candidates for assembling. The simulation of cluster dimers as a function of the separation of the clusters can be used to estimate the strength of the cluster–cluster interaction, and the relaxation of the individual atoms is a measure of the degree to which the clusters retain their integrity. In addition, there are a few reports of the simulation of the assembling of clusters into a solid [150, 151, 179, 181–183]. These provide understanding of the interactions between clusters as they are brought together to form bulk material. In particular, the work of Gong illustrates that the detailed electron distribution of the isolated cluster is an important factor in determining the sort of arrangement of clusters that can lead to successful cluster assembling.

Developments in the theory underpinning the *ab initio* simulation methods should be expected, and these will lead to a more refined calculational tool. In the density functional theory itself, steady improvements are being made to the exchange-correlation functional for the ground state, where the only approximations in the Kohn-Sham approach are made. Although the discovery of the exact functional is a vain hope, approximate functionals are good enough to yield energies to chemical accuracy, which is necessary if the energies of close-lying cluster isomers are to be ordered correctly. Theoretical effort in developing improved explicit electron kinetic functionals for use in the orbital-free approach would be very profitable, as this approach is much faster and more efficient than the full Kohn-Sham scheme. However, first principles normconserving pseudopotentials cannot be used in the orbital-free approach as it is presently formulated and more work is needed to develop accurate and transferable pseudopotentials that can be used in the approach.

ACKNOWLEDGMENTS

J. A. A. wishes to acknowledge the support of MCyT (grant MAT2002-04499-C02-01) and Junta de Castilla y León (grant CO01/102). M. J. S. acknowledges the support of the NSERC of Canada.

REFERENCES

1. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, *Nature* 318, 162 (1985).
2. A. F. Wells, "Structural Inorganic Chemistry," Oxford Science Publications, Oxford, 1984.
3. P. Hohenberg and W. Kohn, *Phys. Rev.* 136B, 864 (1964).
4. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
5. M. L. Cohen and V. Heine, *Solid State Phys.* 24, 37 (1970).
6. D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* 43, 1494 (1979).
7. G. B. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* 26, 4199 (1982).
8. M. T. Yin and M. L. Cohen, *Phys. Rev. B* 26, 5668 (1982).
9. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
10. N. H. March, *Adv. Phys.* 6, 1 (1957).
11. M. Levy, *Proc. Natl. Acad. Sci. U.S.A.* 76, 6062 (1979).
12. R. M. Dreizler and E. K. U. Gross, "Density Functional Theory: An Approach To The Quantum Many-Body Problem," Springer-Verlag, Berlin, 1990.
13. U. von Barth and L. Hedin, *J. Phys. C* 5, 1629 (1972).
14. S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* 58, 1200 (1980).
15. J. P. Perdew and A. Zunger, *Phys. Rev. B* 23, 5048 (1981).
16. D. M. Ceperley, *Phys. Rev. B* 18, 3126 (1978).
17. D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* 45, 566 (1980).
18. U. von Barth, in "Methods of Electronic Structure Calculations," Proceedings from Trieste (V. Kumar, O. K. Anderson, and A. Mookerjee, Eds.), World Scientific, Singapore, 1994, p. 21.
19. D. C. Langreth and M. J. Mehl, *Phys. Rev. B* 28, 1809 (1983).

30. J. P. Perdew and Y. Wang, *Phys. Rev. B* 45, 13244 (1992).
31. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* 91, 146401 (2003).
32. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
33. C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* 37, 785 (1988).
34. T. Tsuneda, T. Suzumura, and K. Hirao, *J. Chem. Phys.* 111, 5656 (1999).
35. R. T. Sharp and G. K. Horton, *Phys. Rev.* 30, 317 (1953).
36. J. D. Talman and W. F. Shadwick, *Phys. Rev. A* 14, 36 (1976).
37. S. Kümmel and J. P. Perdew, *Phys. Rev. Lett.* 90, 043004 (2003).
38. W. Kohn, Y. Meir, and D. E. Makarov, *Phys. Rev. Lett.* 80, 4153 (1998); E. Hult, H. Rydberg, and B. I. Lundqvist, *Phys. Rev. B* 59, 4708 (1999); J. Dobson and J. Wang, *Phys. Rev. B* 62, 10038 (2000); M. Fuchs and X. Gonze, *Phys. Rev. B* 65, 235109 (2002).
39. E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* 52, 997 (1984).
40. E. K. U. Gross, J. F. Dobson, and M. Petersilka, in "Density Functional Theory" (R. F. Nalewajski, ed.), Springer Series, Topic in Current Chemistry, Springer, New York, 1996, p. 81.
41. C.-O. Almbladh and U. von Barth, *Phys. Rev. B* 31, 3231 (1985).
42. Y. A. Wang and E. A. Carter, in "Progress in Theoretical Chemistry and Physics" (S. D. Schwartz, ed.), p. 117, Kluwer, Dordrecht, 2000.
43. M. L. Plumer and M. J. Stott, *J. Phys. C: Solid State Phys.* 18, 4143 (1985).
44. D. Nehete, V. Shah, and D. G. Kanhere, *Phys. Rev. B* 53, 2126 (1996).
45. A. Aguado, J. M. Lopez, J. A. Alonso, and M. J. Stott, *J. Chem. Phys.* 111, 6026 (1999).
46. F. Perrot, *J. Phys.: Condens. Matter* 6, 431 (1994), see also M. Pearson, E. Smargiassi, and P. A. Madden, *J. Phys.: Condens. Matter* 5, 3221 (1993).
47. L.-W. Wang and M. P. Teter, *Phys. Rev. B* 45, 13196 (1992).
48. E. Smargiassi and P. A. Madden, *Phys. Rev. B* 49, 5220 (1994).
49. E. Chacón, J. E. Alvarellos, and P. Tarazona, *Phys. Rev. B* 32 7868 (1985).
50. P. García-González, J. E. Alvarellos, and E. Chacón, *Phys. Rev. A* 57, 4192 (1998).
51. B. Wang, M. J. Stott, and U. von Barth, *Phys. Rev. A* 63, 52501 (2001).
52. A. E. DePristo and J. D. Kress, *Phys. Rev. A* 35, 438 (1987).
53. N. W. Ashcroft, *Phys. Lett.* 23, 48 (1966).
54. V. Heine and I. Abarenkov, *Philos. Mag.* 9, 451 (1964).
55. N. Troullier and J. L. Martins, *Phys. Rev. B* 43, 1993 (1991).
56. L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* 48, 1425 (1982).
57. D. Vanderbilt, *Phys. Rev. B* 41, 7892 (1990).
58. D. J. González, L. E. González, J. M. López, and M. J. Stott, *Phys. Rev. E* 65, 184201 (2002).
59. B. Wang and M. J. Stott, *Phys. Rev. B* 68, 195102 (2003).
60. P. Pulay, *Mol. Phys.* 17, 197 (1969); C. Satoko, *Phys. Rev. B* 30, 1754 (1984).
61. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* 64, 1045 (1992).
62. S. G. Louie, S. Froyen, and M. L. Cohen, *Phys. Rev. B* 26, 1738 (1982).
63. D. J. Chadi and M. L. Cohen, *Phys. Rev. B* 8, 5747 (1973).
64. H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* 13, 5188 (1976).
65. M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, *Comp. Phys. Commun.* 107, 187 (1997) and references therein.
66. W. D. Knight, K. Clemenger, W. A. de Heer, W. A. Saunders, M. L. Chou, and M. L. Cohen, *Phys. Rev. Lett.* 52, 2141 (1984).
67. W. A. de Heer, W. D. Knight, M. Y. Chou, and M. L. Cohen, *Solid State Phys.* 40, 93 (1987).
68. W. Ekardt, *Phys. Rev. B* 29, 1558 (1984).
69. D. E. Beck, *Solid State Commun.* 49, 381 (1984).
70. A. Rubio, L. C. Balbás, and J. A. Alonso, *Z. Phys. D* 19, 93 (1991).
71. K. Hansen, H. Hohmann, R. Müller, and E. E. B. Campbell, *J. Chem. Phys.* 105, 6088 (1996).
72. J. L. Martins, J. Buttet, and R. Car, *Phys. Rev. B* 31, 1804 (1985); U. Röthlisberger and W. Andreoni, *J. Chem. Phys.* 94, 8129 (1991).
73. L. Kronik, I. Vasiliev, and J. R. Chelikowsky, *Phys. Rev. B* 62, 9992 (2000).
74. I. A. Solovyov, A. V. Solovyov, and W. Greiner, *Phys. Rev. A* 65, 53203 (2002).
75. S. Bjornholm, J. Borgreen, O. Echt, K. Hansen, J. Pedersen, and H. D. Rasmussen, *Phys. Rev. Lett.* 65, 1629 (1990).
76. T. P. Martin, T. Bergmann, H. Göhlich, and T. Lange, *Z. Phys. D* 19, 25 (1991).
77. J. Pedersen, S. Bjornholm, J. Borgreen, K. Hansen, T. P. Martin, and H. D. Rasmussen, *Nature* 353, 733 (1991).
78. T. P. Martin, S. Bjornholm, J. Borgreen, C. Brechignac, Ph. Cahuzac, K. Hansen, and J. Pedersen, *Chem. Phys. Lett.* 186, 53 (1991).
79. H. Nishioka, K. Hansen, and B. R. Mottelson, *Phys. Rev. B* 42, 9377 (1990).
80. O. Genzken, *Mod. Phys. Lett. B* 7, 197 (1993).
81. O. Genzken and M. Brack, *Phys. Rev. Lett.* 67, 3286 (1991).
82. M. Pellarin, B. Baguenard, C. Bordas, M. Broyer, M. Lermé, and J. L. Vialle, *Phys. Rev. B* 48, 17645 (1993); M. Pellarin, B. Baguenard, M. Broyer, J. L. Vialle, and A. Pérez, *J. Chem. Phys.* 98, 944 (1993); B. Baguenard, M. Pellarin, C. Bordas, M. Lermé, J. L. Vialle, and M. Broyer, *Chem. Phys. Lett.* 205, 13 (1993).
83. M. Brack, O. Genzken, and K. Hansen, *Z. Phys. D* 19, 51 (1991).

74. I. Katakuse, T. Ichihara, Y. Fujita, T. Matsuo, T. Sakurai, and H. Matsuda, *Int. J. Mass Spectrom. Ion Proc.* 67, 229 (1985); I. Katakuse, T. Ichihara, Y. Fujita, T. Matsuo, T. Sakurai, and H. Matsuda, *Int. J. Mass Spectrom. Ion Proc.* 74, 33 (1986).
75. M. Knickelbein, *Chem. Phys. Lett.* 192, 129 (1992).
76. D. G. Leopold, J. Ho, and W. C. Lineberger, *J. Chem. Phys.* 86, 1715 (1987); C. L. Pettiette, S. H. Yang, M. J. Craycraft, J. Conceicao, R. T. Laaksonen, O. Cheshnovsky, and R. E. Smalley, *J. Chem. Phys.* 88, 5377 (1988); G. Gantefor, M. Gausa, K. H. Meiwes-Broer, and H. O. Lutz, *J. Chem. Soc. Faraday Trans.* 86, 2483 (1990).
77. C. Yannouleas, U. Landman, A. Herlert, and L. Schweikhard, *Phys. Rev. Lett.* 86, 2996 (2001).
78. J. Li, X. Li, H. J. Zhai, and L. S. Wang, *Science* 299, 864 (2003).
79. J. Wang, G. Wang, and J. Zhao, *Chem. Phys. Lett.* 380, 716 (2003).
80. J. A. Alonso and N. H. March, "Electrons in Metals and Alloys," Academic Press, London, 1989.
81. N. Fujima and T. Yamaguchi, *J. Phys. Soc. Jpn.* 58, 1334 (1989).
82. G. Gantefor and W. Eberhardt, *Chem. Phys. Lett.* 217, 600 (1994).
83. J. Akola, M. Manninen, H. Hakkinen, U. Landman, X. Li, and L. S. Wang, *Phys. Rev. B* 60, R11297 (1999).
84. X. Li, H. Wu, X. G. Wang, and L. S. Wang, *Phys. Rev. Lett.* 81, 1909 (1998).
85. F. Duque and A. Mañanes, *Eur. Phys. J. D* 9, 223 (1999).
86. B. K. Rao and P. Jena, *J. Chem. Phys.* 111, 1980 (1999).
87. M. F. Jarrold, J. E. Bower, and J. S. Kraus, *J. Chem. Phys.* 86, 3876 (1987).
88. H. P. Cheng, R. S. Berry, and R. L. Whetten, *Phys. Rev. B* 43, 10647 (1991).
89. J. E. Fowler and J. M. Ugalde, *Phys. Rev. A* 58, 383 (1998).
90. F. Duque, A. Mañanes, L. M. Molina, M. J. López, and J. A. Alonso, *Int. J. Quantum Chem.* 86, 226 (2002).
91. W. D. Knight, K. Clemenger, W. A. de Heer, and W. Saunders, *Phys. Rev. B* 31, 2539 (1985); K. Selby, M. Vollmer, J. Masui, V. Kresin, W. A. de Heer, and W. D. Knight, *Phys. Rev. B* 40, 5417 (1989).
92. G. Mie, *Ann. Phys. (Leipzig)* [Folge 4] 25, 377 (1908).
93. D. E. Beck, *Phys. Rev. B* 30, 6935 (1989).
94. J. M. Pacheco and W. Ekardt, *Ann. Physik* 1, 255 (1992); *Phys. Rev. B* 47, 6667 (1993).
95. P. Stampfli and K. H. Bennemann, *Phys. Rev. A* 39, 1007 (1989).
96. J. A. Alonso and L. A. Girifalco, *Phys. Rev. B* 17, 3735 (1978).
97. A. Rubio, L.C. Balbás, LI. Serra, and M. Barranco, *Phys. Rev. B* 40, 10950 (1990).
98. A. Rubio, Ph.D. Thesis, University of Valladolid, 1991.
99. J. P. Perdew, K. Burke, and M. Ernzenhof, *Phys. Rev. Lett.* 77, 3865 (1996).
100. D. Rayane, A. R. Allouche, E. Benichou, R. Antoine, M. Aubert-Frecon, P. Dugourd, M. Broyer, C. Ristori, F. Chandezon, B. A. Huber, and C. Guet, *Eur. Phys. J. D* 9, 243 (1999).
101. S. Kümmel, T. Berkus, P. G. Reinhard, and M. Brack, *Eur. Phys. J. D* 11, 239 (2000).
102. P. Calaminici, A. Jug, and A. M. Koster, *J. Chem. Phys.* 111, 4613 (1999).
103. R. G. Parr, R. A. Donelly, M. Levy, and E. W. Palke, *J. Chem. Phys.* 68, 3801 (1978).
104. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.* 105, 7512 (1983).
105. A. Rubio, L. C. Balbás, and J. A. Alonso, *Physica B* 167, 19 (1990).
106. J. A. Alonso and L. C. Balbás, *Structure and Bonding* 80, 229 (1993).
107. R. E. Leuchner, A. C. Harms, and A. W. Castleman, *J. Chem Phys.* 91, 2753 (1989); *J. Chem. Phys.* 94, 1093 (1991).
108. Z. Penzar and W. Ekardt, *Z. Phys. D* 17, 69 (1990); W. Ekardt and Z. Penzar, *Phys. Rev. B* 38, 4273 (1988).
109. W. A. de Heer, K. Selby, V. Kresin, J. Masui, M. Vollmer, A. Chatelain, and W. D. Knight, *Phys. Rev. Lett.* 59, 1805 (1987).
110. M. J. Stott and E. Zaremba, *Phys. Rev. A* 21, 12 (1980).
111. A. Zangwill and P. Soven, *Phys. Rev. A* 21, 1561 (1980).
112. M. Brack, *Rev. Mod. Phys.* 65, 677 (1993).
113. S. Fedrigo, W. Harbich, J. Belyaev, and J. Buttet, *Chem. Phys. Lett.* 211, 166 (1993).
114. LI. Serra and A. Rubio, *Phys. Rev. Lett.* 211, 1428 (1997).
115. O. Cheshnovsky, K. J. Taylor, J. Conceicao, and R. E. Smalley, *Phys. Rev. Lett.* 64, 1786 (1990).
116. C. Brechignac, P. Cahuzac, F. Carlier, M. de Frutos, and J. Leignier, *Chem. Phys. Lett.* 189, 28 (1992).
117. J. Borgreen, P. Choudury, N. Kebaili, L. Lundsberg-Nielsen, K. Lutzenkirchen, M. B. Nielsen, J. Pedersen, and H. D. Rasmussen, *Phys. Rev. B* 48, 17507 (1993).
118. J. Tiggesbäumker, L. Köller, H. O. Lutz, and K. H. Meiwes Broer, *Chem. Phys. Lett.* 190, 42 (1992).
119. C. Kohl, B. Montag, and P. G. Reinhard, *Z. Phys. D* 35, 57 (1995).
120. M. P. Iñiguez, J. A. Alonso, M. A. Aller, and L. C. Balbás, *Phys. Rev. B* 34, 2152 (1986).
121. C. Brechignac, Ph. Cahuzac, F. Carlier, J. Leygnier, and A. Sarfati, *Phys. Rev. B* 44, 11386 (1991); C. Brechignac, Ph. Cahuzac, F. Carlier, and M. de Frutos, *Phys. Rev. B* 49, 2825 (1994).
122. I. Katakuse, H. Ito, and T. Ichihara, *Z. Phys. D* 20, 101 (1991).
123. C. Brechignac, Ph. Cahuzac, F. Carlier, M. de Frutos, N. Barnett, and U. Landman, *Phys. Rev. Lett.* 72, 1636 (1994).
124. B. Montag and P. G. Reinhard, *Phys. Rev. B* 52, 16365 (1995).
125. A. Rigo, F. Garcias, J. A. Alonso, J. M. López, M. Barranco, A. Mañanes, and J. Nemeth, *Surf. Rev. Lett.* 3, 617 (1996).
126. F. Garcias, J. A. Alonso, J. M. López, M. Barranco, and J. Nemeth, *Heavy Ion Phys.* 1, 227 (1995).
127. M. Koskinen, P. O. Lipas, and M. Manninen, *Z. Phys. D* 35, 285 (1995).

128. M. M. Kappes, M. Schar, C. Yeretzian, U. Heiz, M. Vayloyan, and E. Schumacher, "Physics and Chemistry of Small Clusters" (P. Jena, B. K. Rao, and S. N. Khanna, eds.),Vol. 158, p. 263. NATO ASI Series B. Plenum, New York, 1987.
129. C. Baladrón and J. A. Alonso, *Physica B* 154, 73 (1988).
130. U. Röhlisberger and W. Andreoni, *Chem. Phys. Lett.* 198, 478 (1992).
131. J. A. Alonso, L. M. Molina, M. J. López, A. Rubio, and M. J. Stott, *Chem. Phys. Lett.* 289, 451 (1998).
132. C. Yannouleas, P. Jena, and S. N. Khanna, *Phys. Rev. B* 46, 9751 (1992).
133. M. Deshpande, A. Dhavale, R. R. Zope, S. Chacko, and D. Kanhere, *Phys. Rev. A* 62, 63202 (2000).
134. I. Boustani, W. Pewestorf, P. Fantucci, V. Bonacic-Koutecky, and J. Koutecky, *Phys. Rev. B* 35, 9437 (1987).
135. V. Bonacic-Koutecky, P. Fantucci, and J. Koutecky, *Phys. Rev. B* 37, 4369 (1988).
136. I. Boustani and J. Koutecky, *J. Chem. Phys.* 88, 5657 (1988).
137. V. Bonacic-Koutecky, I. Boustani, M. F. Guest, and J. Koutecky, *J. Chem. Phys.* 89, 4861 (1988).
138. H. P. Cheng, R. N. Barnett, and U. Landman, *Phys. Rev. B* 48, 1820 (1993).
139. C. Yeretzian, U. Röthlisberger, and E. Schumacher, *Chem. Phys. Lett.* 48, 334 (1995).
140. J. Chang, M. J. Stott, and J. A. Alonso, *J. Chem. Phys.* 104, 8043 (1996).
141. L. C. Balbás and J. L. Martins, *Phys. Rev. B* 54, 2937 (1996).
142. P. Von Schleyer and J. Kapp, *Chem. Phys. Lett.* 255, 363 (1996).
143. L. M. Molina, M. J. López, A. Rubio, J. A. Alonso, and M. J. Stott, *Int. J. Quantum Chem.* 69, 341 (1998).
144. C. J. Marsden, *Chem. Phys. Lett.* 245, 475 (1995).
145. K. Joshi and D. G. Kanhere, *Phys. Rev. A* 65, 43203 (2002).
146. A. L. Allred and E. J. Rochow, *J. Inorg. Nucl. Chem.* 5, 264 (1958).
147. T. Bergman, H. Limberger, and T. P. Martin, *Phys. Rev. Lett.* 60, 1767 (1988).
148. U. Lammers, G. Borstel, A. Mañanes, and J. A. Alonso, *Z. Phys. D* 17, 203 (1990).
149. B. K. Rao and P. Jena, *J. Chem. Phys.* 115, 778 (2001).
150. A. P. Seitsonen, K. Laaksonen, R. M. Nieminen, and M. L. Klein, *J. Chem. Phys.* 103, 8075 (1995).
151. X. G. Gong, *Phys. Rev. B* 56, 1091 (1997).
152. X. G. Gong and V. Kumar, *Phys. Rev. Lett.* 70, 2078 (1993).
153. V. Kumar, S. Bhattacharjee, and Y. Kawazoe, *Phys. Rev. B* 61, 8541 (2000).
154. X. Li and L. S. Wang, *Phys. Rev. B* 65, 153404 (2002).
155. S. Burkart, N. Blessing, B. Klipp, J. Muller, G. Ganteför, and G. Seifert, *Chem. Phys. Lett.* 301, 546 (1999).
156. S. N. Khanna and P. Jena, *Chem. Phys. Lett.* 218, 383 (1994).
157. A. Mañanes, F. Duque, F. Méndez, M. J. López, and J. A. Alonso, *J. Chem. Phys.* 119, 5128 (2003).
158. M. Schmidt, R. Kusche, B. von Issendorf, and H. Haberland, *Nature* 393, 238 (1998).
159. R. Kusche, M. Hippler, M. Schmidt, B. von Issendorf, and H. Haberland, *Eur. Phys. J. D* 9, 1 (2000).
160. M. Schmidt, M. Hippler, J. Donges, W. Kronmuller, B. von Issendorf, and H. Haberland, "The Physics and Chemistry of Clusters" (E. E. B. Campbell and M. Larsson, eds.), p. 22. World Scientific, Singapore, 2001.
161. C. Fiolhais, J. P. Perdew, S. Q. Armster, J. M. McLaren, and H. Brajczewska, *Phys. Rev. B* 51, 14001 (1995); *Phys. Rev. B*, 53, 13193 (1996).
162. F. Nogueira, C. Fiolhais, J. He, J. P. Perdew, and A. Rubio, *J. Phys.: Condens. Matter* 8, 287 (1996).
163. A. Aguado, J. M. López, J. A. Alonso, and M. J. Stott, *J. Phys. Chem.* 105, B2386 (2001).
164. S. Sugano, "Microcluster Physics," Springer-Verlag, Berlin, 1991.
165. A. Bulgac and D. Kusnezov, *Phys. Rev. B* 45, 1988 (1992); N. Ju and A. Bulgac, *Phys. Rev. B* 48, 2721 (1993).
166. F. Calvo and F. Spiegelmann, *Phys. Rev. Lett.* 82, 2270 (1999); *J. Chem. Phys.* 112, 2888 (2000).
167. R. Poteau, F. Spieglemann, and P. Labastie, *Z. Phys. D* 30, 57 (1994).
168. M. Schmidt, C. Ellert, W. Kronmuller, and H. Haberland, *Phys. Rev. B* 59, 10970 (1999).
169. H. Haberland, in "Metal Clusters" (W. Ekardt, ed.), p. 181, John Wiley & Sons, New York, 1999.
170. S. Kummel, M. Brack, and P. G. Reinhard, *Phys. Rev. B* 62, 7602 (2000).
171. P. Blaise and S. A. Blundell, *Phys. Rev. B* 63, 235409 (2001).
172. A. Vichare, D. G. Kanhere, and S. A. Blundell, *Phys. Rev. B* 64, 045408 (2001).
173. A. Rytkonen, H. Hakkinen, and M. Manninen, *Phys. Rev. Lett.* 80, 3940 (1998); *Eur. Phys. J. D* 9, 451 (1999).
174. J. Akola and M. Manninen, *Phys. Rev. B* 63, 193410 (2001).
175. W. Van der Lugt, *J. Phys.: Condens. Matter* 8, 6115 (1996).
176. L. M. Molina, M. J. López, A. Rubio, L. C. Balbás, and J. A. Alonso, *Adv. Quantum Chem.* 33, 329 (1999).
177. L. Molina, J. A. Alonso, A. Rubio, M. J. López, and M. J. Stott, in "Recent Advances in Density Functional Methods, Part III" (V. Barone, A. Bencini, and P. Fantucci, eds.), p. 234. World Scientific, Singapore, 2002.
178. Y. Jinlong, D. Kaiming, X. Chuanyun, and W. Kelin, *Phys. Rev. B* 55, 13293 (1997).
179. L. M. Molina, J. A. Alonso, and M. J. Stott, *J. Chem. Phys.* 111, 7053 (1999).
180. S. N. Khanna and P. Jena, *Phys. Rev. Lett.* 69, 1664 (1992).
181. A. P. Seitsonen, M. J. Puska, M. Alatalo, R. M. Nieminen, V. Milman, and M. C. Payne, *Phys. Rev. B* 48, 1981 (1993).
182. C. Ashman, S. N. Khanna, F. Lui, P. Jena, T. Kaplan, and M. Mostoller, *Phys. Rev. B* 55, 15868 (1997).
183. F. Duque, A. Mañanes, L. M. Molina, M. J. López, and J. A. Alonso, *Int. J. Quantum Chem.* 86, 226 (2001).
184. F. Duque, L. M. Molina, M. J. López, A. Mañanes, and J. A. Alonso, *Eur. J. Phys. D* 16, 285 (2001).
185. J. A. Alonso, M. J. López, L. M. Molina, F. Duque, and A. Mañanes, *Nanotechnology* 13, 253 (2002).
186. D. Y. Sun and X. G. Gong, *Phys. Rev. B* 54, 17051 (1996).
187. R. L. Whetten, J. T. Koury, M. Alvarez, S. Murthy, I. Vezmar, L. Wang, P. W. Stephens, C. L. Cleveland, W. D. Luedtke, and U. Landman, *Adv. Mater.* 8, 428 (1996).

CHAPTER 3

Modeling the Structural Evolution, Equilibrium Morphology, and Macroscopic Behavior of Polymer/Nanoparticle Composites

Anna C. Balazs, Gavin A. Buxton
Department of Chemical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

CONTENTS

1. Introduction 104
2. Models for Determining the Structure of Nanocomposites . . . 106
   2.1. Self-Consistent Field/Density Functional Theory Approach 106
   2.2. Cahn-Hilliard/Brownian Dynamics Model 108
3. Models for Determining Macroscopic Properties of Nanocomposites 111
   3.1. Mechanical Properties: Lattice Spring Model 111
   3.2. Electrical Properties: Finite Difference Model 113
   3.3. Optical Properties: Finite Difference Time Domain Method 114
4. Results and Discussion 116
   4.1. Properties of Mixtures of Diblock Copolymers and Spherical Nanoparticles 116
   4.2. Behavior of Nanorods in Binary Polymer Blends 126
   4.3. Properties of Mixtures of Nanorods and Diblock Copolymers 132
   4.4. Structure and Optical Properties of Nanoparticle-Filled Diblock Copolymers 142
5. Conclusions 146
   References 147

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (103–149)

1. INTRODUCTION

The combination of organic polymers and inorganic nanoparticles can lead to a nanocomposite that is more useful than either of the individual components. For example, if the nanoparticles are metals or semiconductors, the nanocomposite can exhibit the mechanical, optical, or electromagnetic properties of the inorganics and the flexibility and processability of the polymers. One of the critical challenges facing researchers in the area of nanocomposites is formulating *fundamental, nonempirical* structure–property relationships for these materials. Without the aid of such basic relationships, the fabrication of nanocomposites for specific applications remains a largely Edisonian and costly process. The stumbling block to the derivation of such fundamental correlations has been a lack of models for predicting (1) the structure of these materials and (2) the properties of those specific structures. Such predictive models are crucial for establishing how changes in the characteristics of the components affect the morphology of the mixture and, similarly, how changes in the morphology affect the macroscopic behavior of the material.

Our recent research efforts have been focused at addressing this issue. In particular, we developed computational models for the structural evolution and final equilibrium morphology of mixtures of polymers and nanoparticles. We then used the specific morphological information obtained from these models as the input to computer simulations for the mechanical, electrical, and optical properties of the composites. In this chapter, we describe a number of examples that illustrate this approach and then discuss our subsequent findings on the relationship between structure and properties in nanocomposites.

Our general methodology is outlined schematically in Fig. 1. We start by specifying the size and shape of the nanoparticles and the composition of the polymeric matrix. These characteristics serve as parameters in our models for determining the mesoscale morphology of the mixture. Specifically, we recently developed two distinct techniques for modeling the structure of copolymer/nanoparticle mixtures. To predict the equilibrium morphology, we derived a mean field theory for mixtures of soft, flexible chains and hard spheres [1–10]. Applied to mixtures of diblock copolymers and nanoparticles, the theory predicts ordered phases where the particles and diblocks self-assemble into spatially periodic structures [9]. The model integrates a self-consistent field theory (SCFT) for polymers and a density functional theory (DFT) for particles. The SCFT has been remarkably successful in describing the thermodynamics of pure polymer systems [11], whereas DFTs capture particle ordering and phase behavior in colloidal systems [12, 13]. The integrated SCF/DFT approach provides

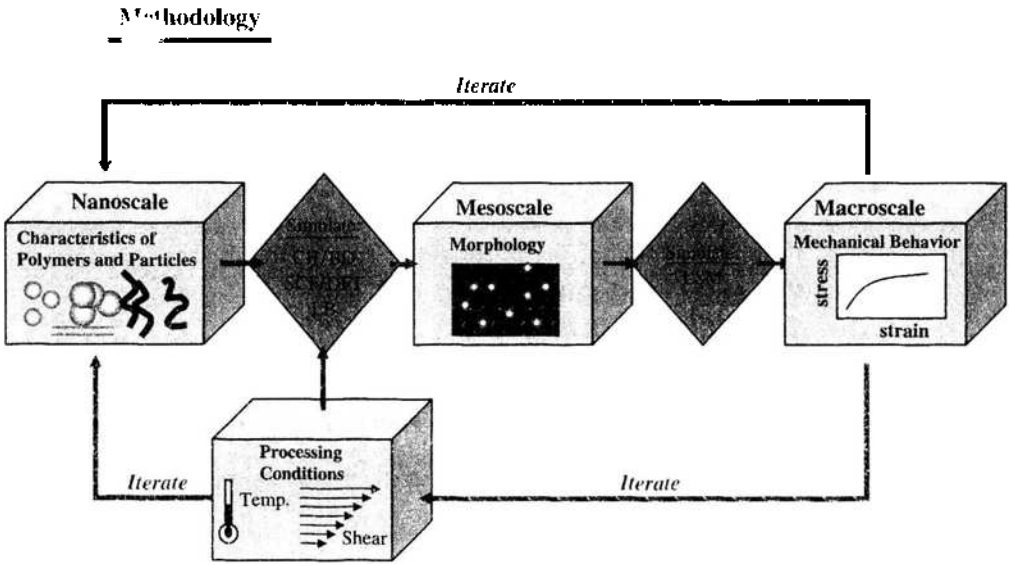


Figure 1. General methodology: Size and shape of constituents and composition of polymeric matrix are parameters for morphological studies. The resultant morphologies serve as input to performance models, such as the LSM, which predicts the mechanical properties of the system. This procedure can be iterated to optimize the constituent properties necessary for obtaining the desired physical properties of the polymer composite.

a powerful technique for determining the structure of a broad class of polymer–particle mixtures.

It is noteworthy that experimental studies [14] have recently confirmed our SCF/DFT predictions on the entropically driven size segregation of binary particle mixtures within a diblock melt [6, 7]. Experiments [15] have also validated our observation that added nanoparticles can promote transitions between the different structures of the diblock copolymers [8].

We also developed a dynamic model that allows us to capture the structural evolution of the system [16–25]. The model combines a Cahn-Hilliard (CH) theory for the phase-separation dynamics of binary blends or diblocks with the Brownian motion of the nanoparticles. We subsequently refer to this technique as the CH/BD (Brownian dynamics) model. The technique allows us to probe the interactions in a system of phase-separating polymers and mobile, wettable particles. Using this CH/BD approach, we can elucidate factors that govern the dispersion of the particles, the growth of the fluid domains, and the structure of the solid–liquid and liquid–liquid interfaces [21]. Recent experimental studies [26] have confirmed our predictions based on the CH/BD simulations [24, 25] that nanoparticle additives effectively pin the sizes of the polymer domains within a phase-separating blend.

As indicated in Fig. 1, the output from these models was then used in the lattice spring model (LSM) to obtain the response of the material to mechanical deformation [3, 16, 18]. The LSM is a network of nearest and next-nearest neighbor interactions, which are harmonic in nature [27]. These harmonic interactions, or "springs," result in linear forces between lattice sites (nodes); consequently, the simulation yields the linear elastic behavior of the composite. A distinct feature of the LSM is the ease with which various heterogeneities can be incorporated through local variations in the characteristics of the springs.

To examine electrical properties, we use an approach similar to the one outlined in Fig. 1 [16]. We now map the morphology of the mixture onto a finite difference model (FDM) for the electrical conductivity of the material. In particular, to investigate the electrical conductivity of the heterogeneous polymer/nanoparticle composites, we discretize Laplace's equation for a static potential distribution onto a square or cubic lattice [28]. This is equivalent to solving Ohm's law (potential difference is linearly related to current) and Kirchoff's law of current conservation (total current flowing into any point is zero) for a network of resistors [29]. The resistor network model offers an expedient method for discretizing the electrical behavior of a continuous medium, while allowing for the heterogeneous nature of this material.

We have also combined the SCF/DFT morphological studies with a computational model for the propagation of light [30]. In particular, the finite difference time domain (FDTD) technique is used to simulate the propagation of light through this diblock/nanoparticle composite. The FDTD method is a flexible, numerical means of analyzing interactions between waves and complex materials containing dielectric or metallic objects [31]. The technique involves approximating the integration of Maxwell's equations in real space by the use of finite differences. Specifically, the FDTD is a "time marching" algorithm used to solve the wave equation at each point on a grid [32]. The field is set to zero at the initial time step, and at the next time step a source is turned on to generate an optical signal. From this initial condition, the wave fields at all points on the grid at any later time can be calculated through a simple, iterative scheme. The FDTD technique is highly useful for examining the behavior of heterogeneous materials [33] because one can readily compute the propagation of light in materials that include particles of arbitrary shape or interparticle distance [31]. The results of this study allow us to determine how the polymer–particle interactions affect the spatial distribution of fillers within the polymer matrix and how this distribution in turn affects the optical properties of the nanocomposite.

By integrating the morphological and mechanical, electrical, or optical models, we can isolate how specific modifications in the geometry or properties of the components affect the macroscopic behavior. Thus, we can establish how choices made in the components affect the ultimate performance of the system. In addition, we can perform trade-off analyses and compare the advantages of one choice over another. We can also iteratively optimize the system parameters to establish the ideal additives that will yield the desired, specified macroscopic

properties. For example, we can ascertain the optimal polymer matrix or particle architecture for enhancing the reinforcing behavior of the nanoparticles and thus increasing the strength of the materials. We can formulate optimal mixtures of nanotubes and polymers that yield electrically conducting composites with only a very low volume fraction of nanotubes. We can also determine the optimal polymer–particle interactions for creating photonic band gap materials.

Finally, as indicated in Fig. 1, by introducing variations in temperature, pressure, or an imposed shear flow, the effects of processing on both the structure and properties of the system can be analyzed. We recently focused on a pure binary polymer blend (without nanoparticles) and used our computational models to examine how shearing this polymer melt affects the dynamic fracture mechanics of the final solid material [34]. We do not discuss this paper here because our focus is on the structure and properties of polymer/nanoparticle composites. Nonetheless, we briefly note how these calculations were carried out. In particular, the phase separation of the blend under an imposed shear is simulated through the Cahn-Hilliard method, where an advection term is added to introduce the flow field [35]. Using this model, we obtained the structural evolution and late-stage morphology of the sheared mixture. We coupled these morphological results to a dynamic LSM to simulate crack propagation through the solid blend structure. In the dynamic LSM, the behavior of the nodes is dictated by Newtonian dynamics. In addition, in order to introduce a crack and model the subsequent degradation of the material, springs are selectively removed from the system. The model allows us to simulate crack propagation through these heterogeneous structures and determine the strength and toughness of the material. Consequently, we could correlate the relative orientation of the interfacial regions to the overall mechanical behavior of the system. We also contrasted these results with our findings on the unsheared samples and thereby probed the effect of processing on the performance of polymer blends. These studies are particularly important because the variations in processing could be exploited to create novel behavior, and ultimately new products, from off-the-shelf materials.

In the next section, we begin by briefly describing each of the models that we used to obtain the structure (SCF/DFT, CH/BD) and macroscopic behavior (LSM, FDM, FDTD) of the polymeric nanocomposites. We then describe our findings on the effect of such critical factors as the particle geometry and polymer architecture on the structural organization and properties of these complex materials.

2. MODELS FOR DETERMINING THE STRUCTURE OF NANOCOMPOSITES

2.1. Self-Consistent Field/Density Functional Theory Approach

To predict the equilibrium morphology of polymeric composites, we created a mean field theory for mixtures of flexible polymers and solid nanoparticles [1–10]. The model combines a self-consistent field theory (SCFT) for polymers and a density functional theory (DFT) for the particles. We note that in the area of composites, the particles are commonly referred to as fillers and the polymers containing these fillers are referred to as filled systems. The SCF/DFT is a powerful method for determining the phase behavior of filled systems because we are not constrained to make *a priori* assumptions about the structure of the phase or the distribution of particles. Here, we outline the fundamental equations of a system consisting of an incompressible mixture of molten AB diblock copolymers and solid spherical nanoparticles. An AB diblock copolymer is composed of a block of A monomers that are chemically linked to a block of B monomers. For sufficiently low temperatures (discussed later), the A and B blocks are chemically incompatible, and this incompatibility drives the A and B blocks to "microphase separate" into an array of periodically regular morphologies. (Because they are chemically linked, the fragments cannot undergo macroscopic phase separation.)

In the systems considered here, all the particles have the same radius R. Each AB diblock consists of N segments, each of volume ρ_0^{-1}. The fraction of A segments per chain is denoted by f. The enthalpic interaction between an A segment and a B segment is described by

the dimensionless Flory-Huggins parameter, χ_{AB}, which is inversely proportional to temperature. As a function of (χ_{AB}N) and f, a pure diblock melt can form spatially periodic microstructures with lamellar, cylindrical, spherical, or more complicated phases [36].

In SCF theory, pair-wise interactions between differing segments are replaced by the interaction of each segment with the average field created by the other segments. Here, $w_A(\mathbf{r})$ is the value at a point $\mathbf{r}$ of the mean field felt by the A segments, $w_B(\mathbf{r})$ denotes the field for B segments, and $w_p(\mathbf{r})$ represents the field for particles. Using this approach, the free energy for our system is given by $F_T = F_c + F_d - F_p$. The first term, F_c, details the enthalpic interactions in the system:

$$F_c = (1/V)\int dr[\chi_{AB}N\varphi_A(\mathbf{r})\varphi_B(\mathbf{r}) + \chi_{BP}N\varphi_B(\mathbf{r})\varphi_P(\mathbf{r}) + \chi_{AP}N\varphi_A(\mathbf{r})\varphi_P(\mathbf{r})] \quad (1)$$

where V is the volume of the system and $\varphi_A(\mathbf{r})$, $\varphi_B(\mathbf{r})$, and $\varphi_P(\mathbf{r})$ are the local concentrations of A segments, B segments, and particles, respectively. The parameters χ_{AP} and χ_{BP} describe the enthalpic interactions between the particles and the respective blocks.

The diblock entropic free energy F_d is [11]:

$$F_d = (1 - \phi_P)\ln[V(1 - \phi_P)/Q_d] - (1/V)\int dr\,[w_A(\mathbf{r})\varphi_A(\mathbf{r}) + w_B(\mathbf{r})\varphi_B(\mathbf{r})] \quad (2)$$

where Q_d is the partition function of a single diblock subject to the fields $w_A(\mathbf{r})$ and $w_B(\mathbf{r})$. Here, the total volume fraction of particles in the system is given by ϕ_p and thus, the volume fraction of diblocks in the mixture is $(1 - \phi_P)$.

Finally, the particle entropic contributions to the free energy is given by

$$F_p = (\phi_P/\alpha)\ln(V\phi_P/Q_p\alpha) - (1/V)\int dr[w_p(\mathbf{r})\rho_p(\mathbf{r})] + (1/V)\int dr\,\rho_p(\mathbf{r})\Psi_{CS}[\bar{\varphi}(\mathbf{r})] \quad (3)$$

where Q_p is the partition function of a single particle subject to the field $w_p(\mathbf{r})$. The local particle volume fraction, $\varphi_p(\mathbf{r})$, is related to the dimensionless center of mass distribution, $\rho_p(\mathbf{r})$, by

$$\varphi_P(\mathbf{r}) = \left[4\alpha/(3\pi R^3)\right]\int_{r'<R} dr'\,\rho_p(\mathbf{r} + \mathbf{r}') \quad (4)$$

The parameter $\alpha = [4\pi R^3\rho_0/(3N)]$ denotes the particle-to-diblock volume ratio. Recall that the model is for an incompressible system and the mixture obeys the following incompressibility constraint: $\varphi_A(\mathbf{r}) + \varphi_B(\mathbf{r}) + \varphi_P(\mathbf{r}) = 1$. In a mean field manner, this incompressibility constraint prohibits significant overlap between the particles and monomers, as well as significant overlap between the A and B components. However, to accurately capture the behavior of the system as the polymer concentration goes to zero, we must also include a term that explicitly describes the steric interactions (the nonideal term) between the particles. The last term of F_p describes the excess (nonideal) steric free energy of the particles through the DFT derived by Tarazona [12]. In particular, the Carnahan-Starling equation of state for the excess free energy of a hard-sphere fluid, Ψ_{CS}, is now evaluated with the "weighted" (locally averaged) particle volume fraction, $\bar{\varphi}(\mathbf{r})$ [9]. This density functional contribution is included so that the model can describe not only homogeneous (liquid) but also inhomogeneous (crystalline) distributions of the hard spheres. In the case of the spherical nanoparticles, we chose the Tarazona DFT because it is a simple and physical approach specifically developed to reproduce the liquid-solid transition in hard spheres. Some DFTs more accurately describe the liquid equation of state but fail to predict the liquid-solid transition; other approaches are more computationally intensive.

This approach is valid for a finite range of particle sizes. If the particle radius is beyond the root mean-squared end-to-end distance of the chain—that is, $R \gg R_0$—it may be more appropriate to view a particle as a substrate and include boundary conditions that characterize the polymer–substrate interactions. If, on the other hand, $R \ll R_0$, where the particle is on the same size scale as a monomer, then it may be sufficient to treat the particle as a solvent molecule. In our studies [1–10], we have focused on cases where $0.1R_0 \leq R \leq 0.3R_0$,

which clearly falls between these two extreme situations. In addition, we have focused on particle volume fractions that are less than 20% to ensure that mixture is not macroscopically phase separated into a two-phase mixture.

In addition to examining the behavior of mixtures that contained spherical nanoparticle additives, we examined systems that contained rodlike nanoparticles. To introduce these rods, we modified the free energy functional F_p and adopted another form of DFT, which is called the fundamental-measure theory (FMT) [37–39]. Instead of the term $\Psi_{CS}[\bar{\varphi}(\mathbf{r})]$ in Eq. (3), we now have $\Psi_{FMT}[\varphi(\mathbf{r})]$, where Ψ_{FMT} assumes a different functional form than Ψ_{CS} [9, 37–39]. A distinct feature of FMT is that the density functionals are constructed from the geometric features of the particles. This approach has recently been used to model the behavior of a system of parallel hard cubes (PHC) [37–39]. Following this derivation, we extended the theory to parallel hard rectangles (PHR), the "rods" in our mixtures, and parallel hard platelets [3]. In these systems, the long axes of all the particles are assumed to lie parallel to each other. This is a rather stringent constraint that is, nonetheless, applicable to certain liquid-crystalline rods and clay platelets. In addition, this assumption might be appropriate for effectively high concentration of rods that are confined within the nanoscale domains of the microphase-separated diblocks.

Each solid, rectangular particle is characterized by a cross section of $2\sigma \times 2\sigma$ and a length of $2\sigma_l$. (In other words, σ is 1/2 an edge length.) The dimensions of the solids are given in terms in R_0. To simplify the calculation and make it computationally tractable, we constrain all the parallel particles to be oriented along one of three possible directions, namely, the x, y, or z directions. For each set of $(\sigma_x, \sigma_y, \sigma_z)$, we carry out two distinct simulations, allowing the particles to be oriented in or out of the plane of the mesophase. We compare the resultant free energies of the simulations and determine the equilibrium morphology as the one with the lowest free energy. Further details about the calculation can be found in Ref. [3].

In the mean-field approximation, $w_A(\mathbf{r})$, $w_B(\mathbf{r})$, and $w_p(\mathbf{r})$ are determined by locating saddle points in the free energy functional F_T subject to the incompressibility constraint. This yields a system of equations that is solved numerically and self-consistently to give possible equilibrium solutions. To obtain these solutions, we implement the "combinatorial screening" technique of Drolet and Fredrickson [40]. The advantage of this technique is that it requires no *a priori* knowledge of the equilibrium morphology; this is particularly important in our studies because polymer/nanoparticle systems can exhibit structures that are as yet unknown. We make an initial guess for the fields and calculate all the densities and the free energy at each step; the fields are then recalculated, and the entire process is repeated until changes in the diblock densities at each step become sufficiently small. In addition, we minimize our free energy with respect to the size of the simulation box, as proposed by Bohbot-Raviv and Wang [41].

2.2. Cahn-Hilliard/Brownian Dynamics Model

We also developed a hybrid model that couples the phase-separation dynamics of binary homopolymer blends or diblocks with the Brownian motion of the nanoparticles [16–25]. As noted in Section 1, we refer to this approach as the CH/BD model. This model not only reveals the steady-state structure of the system but also allows us to capture the structural evolution of the system as it moves toward this final state.

Using this CH/BD model, we considered an A/B binary blend or AB diblock copolymers that contain particles, which have an affinity for the A phase. This affinity is introduced via a polymer-particle coupling term in the free energy (as described later). Thus, the phase-separation of the blend, or the microphase separation of the diblocks, can affect the spatial distribution of the particles, and the particles can influence the size and morphology of the polymer domains.

Our description of the blend is based on the Cahn-Hilliard approach [42, 43]. We define a scalar order parameter $\psi(\mathbf{r}) = \phi_A(\mathbf{r}) - \phi_B(\mathbf{r})$, which is the difference between the local volume fraction of A and B at a position $\mathbf{r}$. Note that $\psi = \pm\psi_{eq}$ corresponds to the equilibrium order parameter for the A-rich (B-rich) phase. In this model, $\psi_{eq} = 1$ for the A phase

and $\psi_{eq} = -1$ for the B phase. In the Cahn-Hilliard model, the phase separation dynamics of a binary A/B blend is described by the following equation.

$$\partial\psi/\partial t = M\nabla^2(\delta H\{\psi\}/\delta\psi) + \xi \tag{5}$$

where M is a kinetic coefficient (taken as a constant for simplicity), $H\{\psi\}$ is a free energy functional, and ξ is a conserved zero mean Gaussian white noise. Here, ξ is set to zero. To describe the dynamics of microphase separation in a melt of AB diblocks, the equation is modified in the following way [44]:

$$\partial\psi/\partial t = M\nabla^2(\delta H\{\psi\}/\delta\psi) - \Gamma(\psi - \mathrm{F}) \tag{6}$$

The variable $\mathrm{F} = 2f - 1$ describes the asymmetry of the diblock; for a symmetric diblock, $f = 0.5$ and $\mathrm{F} = 0$. The parameter Γ determines the thickness of the domain structure and is related to the degree of polymerization, N, of the diblock [44].

The free-energy functional $H\{\psi\}$ can be written as the sum of two terms [23]: $H\{\psi\} = H_{\mathrm{GL}}\{\psi\} + H_{\mathrm{CPL}}\{\psi\}$. The first term describes the local and gradient energy contributions and is given by:

$$H_{\mathrm{GL}}\{\psi\} = \int -A\ln[\cosh(\psi)] + (1/2)\psi^2 + (D/2)(\nabla\psi)^2\,dr \tag{7}$$

where A and D are constants. Here, we set $A = 1.3$ and $D = 0.5$.

The second term, $H_{\mathrm{CPL}}\{\psi\}$, takes into account the coupling between the particles and the order parameter field and is given by [23]:

$$H_{\mathrm{CPL}}\{\psi\} = \sum_i \int V(\mathbf{r} - \mathbf{R}_i)(\psi(\mathbf{r}) - \psi_s)^2\,dr \tag{8}$$

Here, $\mathbf{R}_i$ is the position of the center of mass of the ith particle and ψ_s is the order parameter at the surface of the particle. The summation is over all particles. For short-ranged interactions, we can take

$$V(\mathbf{r}) = C\exp(-|\mathbf{r}|/r_0) \tag{9}$$

where r_0 represents a microscopic length scale and C captures the strength of the wetting interaction. A variety of alternative choices for $V(\mathbf{r})$ will yield the same qualitative features [23].

In the first set of studies presented here, the particles in the system are spherical in shape. To describe their interaction with the background fluid, we impose the following boundary condition: the order parameter at the surface of each particle is set to a constant value, such as $\psi_s = 1$. In this case, the particles are effectively "coated" by a layer of A and thus are wet by the A fluid. By varying C [Eq. (9)] or ψ_s, we can modify the strength of the wetting interaction or the chemical nature of the particle.

The particles exert an influence on the binary mixture through this preferential wetting interaction. In addition, through Eq. (8), we see that the free energy of the system is reduced when the value of ψ near a particle is similar to the value of ψ_s. Thus, the nature and extent of the wetting interaction play a role in the evolution of the fluid domains.

The particles in the system undergo Brownian dynamics, and their motion is described by the following Langevin equation (in the high friction limit, where the inertial term can be neglected) [23]:

$$d\mathbf{R}_i/dt = M_p(\mathbf{f}_i - \partial H_{\mathrm{CPL}}/\partial\mathbf{R}_i) + \eta_i \tag{10}$$

where M_p is mobility and η_i represents a Gaussian white noise. The second term is the force from the polymer matrix that modifies the motion of the particles. In effect, $(-\partial H_{\mathrm{CPL}}/\partial\mathbf{R}_i)$ represents a thermodynamic force that drives the particles into the compatible, wetting phase. The term $\mathbf{f}_i$ is the force acting on the ith particle due to all the other particles; for example, such a force arises when a Lennard-Jones interaction or other potential is introduced between the particles. Here, we set $\mathbf{f}_i$ equal to zero.

In this version of the model, the particles in the system are "soft" or penetrable to the fluid because we neglect excluded volume interactions between the particles and the fluid. (Excluded volume interactions between the particles and fluid can be explicitly included in the model [24, 25]; for the corresponding systems, both the "hard" and "soft" particles yield similar results for the particle distributions.) On the other hand, the excluded volume interactions are included for particle–particle interactions. In particular, if a move results in overlap between particles, the move is rejected.

In addition to examining A-like, spherical particles with this model, we considered A-like nanorods [16, 22]. The free energy $H\{\psi\}$ in Eq. (5) now contains an additional term H_{rr}, which represents the rod–rod interaction term. This term is taken to be repulsive and is dependent upon the distance and angle between two rods:

$$\begin{aligned} H_{rr} &= \chi \sum_i \sum_j (L - |r_i - r_j|)^2 \left[\frac{4}{3} - \cos^2(\theta_i - \theta_j)\right] \qquad \text{for } |\mathbf{r}_i - \mathbf{r}_j| < L \\ &= 0 \qquad \text{for } |\mathbf{r}_i - \mathbf{r}_j| \geq L \end{aligned} \tag{11}$$

where the constant χ characterizes the strength of the interaction. L is the rod length, and $\mathbf{r}_i$ and θ_i are the respective position of the ith rod and its angle relative to a fixed direction. One way to simultaneously achieve the preferential wetting interaction described previously and the rod–rod repulsion is to anchor A chains onto the surface of the rods. The compatibility between the anchored chains and the chemically identical A phase will yield the desired wetting behavior, and the steric interactions between the grafted chains will lead to an effective repulsion between the rods. In the absence of the binary blend, H_{rr} leads to an isotropic-nematic ordering for the pure rod system [22].

The rigid rods are described as discrete entities [16, 22], each of which has a center-of-mass position $\mathbf{r}_i$ and an orientation angle θ_i, measured from a fixed direction. The position $\mathbf{r}_i$ and angle θ_i of the ith rod is updated with the following Langevin equations

$$\begin{aligned} \partial r_i/\partial t &= -M_r(\partial H/\partial r_i) + \xi_r \\ \partial \theta_i/\partial t &= -M_\theta(\partial H/\partial \theta_i) + \xi_\theta \end{aligned} \tag{12}$$

where M_r and M_θ are the respective translational and rotational mobility constants, and ξ_r and ξ_θ represent thermal fluctuations that satisfy the fluctuation–dissipation relations.

The preceding differential equations for the evolution of the system are discretized and solved numerically. The model allows us to calculate how the characteristics of the polymer matrix (for example, blend versus diblock) and particles (for example, size, shape, wettability) affect the kinetics of structure formation. We can also include the effects of an applied flow field (such as simple shear) [45] and thus determine how processing affects the morphology of the system.

The arrays that hold the spatial information in the morphological models serve as the input to the LSM. We illustrate this point through the schematic in Fig. 2. Consider a blend of two immiscible homopolymers, A and B; the image on the left represents the output from simulations on the morphology of the mixture. The locations of these different polymer domains are superimposed onto the LSM network of springs, which is represented by the black mesh in the middle image. The springs that lie within the A domains (in red) are assigned one set of force constants, and in a similar manner the springs that lie in the B phase (in blue) are assigned separate characteristic values. The green circle in the central image encompasses springs at the interface between the A and B domains; one set of these springs is enlarged and shown on the right (where the dashed black line is the interface).

In mixtures involving nanoparticles, the particle positions are also mapped onto the LSM, and the particle regions are assigned distinct spring constants. This procedure is used to initiate the calculations described later. We note that in these simulations, each filler particle is represented by several nodes of the LSM. Thus, the nanoparticles are sufficiently large that they can be assigned and characterized by an appropriate spring constant.

To examine electrical properties, we use a similar approach of mapping the morphology of the films onto a model for the electrical conductivity of the material. Here again, each

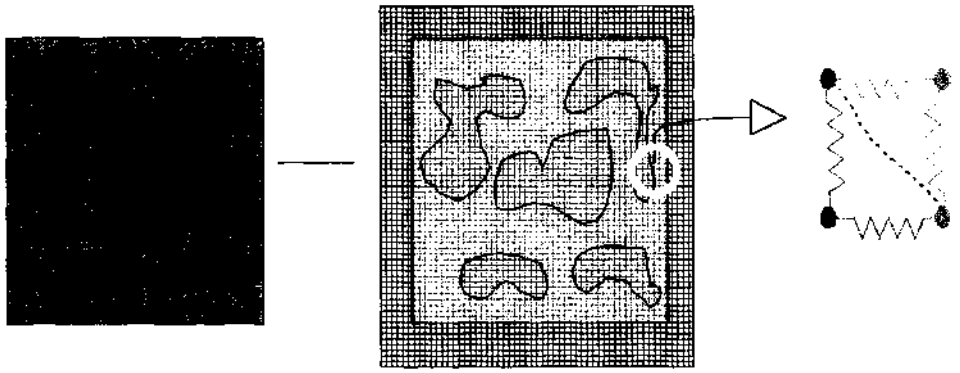


Figure 2. Mapping data from morphological studies onto an LSM. The local properties of a material are assigned, depending on the local polymer concentration and the presence of nanoparticles, obtained from the output of the polymer-processing simulation.

filler particle is represented by several nodes of the resistor network. Thus, the nanoparticles can be characterized by a macroscopic conductivity. In the next section, we describe the simulations of both mechanical and electrical properties in greater detail. We also describe the model we use to characterize the optical properties of the composites.

3. MODELS FOR DETERMINING MACROSCOPIC PROPERTIES OF NANOCOMPOSITES

3.1. Mechanical Properties: Lattice Spring Model

One variety of the LSM is the Born LSM [27]; in this model, the extension of a spring is energetically penalized through a central force constant, k, while the restriction in rotational motion is imposed through the introduction of a noncentral force constant, c. We begin by describing the Born LSM for simulating elastic deformation and then describe a method for extending this model to capture viscoelastic behavior. At each iteration in the simulation, forces are applied to the boundary sites or "nodes," and the equilibrated nodal displacements throughout the system are calculated. The method for obtaining stresses and strains from these displacements and forces are further detailed below.

As noted, in the LSM a material is represented by a network of springs, which occupy the nearest and next-nearest neighbor bonds of a simple cubic lattice (in a 3D simulation). The elastic energy associated with a site or node m in the lattice is taken to be:

$$E_m = (1/2)\sum_n (\mathbf{u}_m - \mathbf{u}_n) \bullet \mathbf{M}_{mn} \bullet (\mathbf{u}_m - \mathbf{u}_n) \tag{13}$$

where the summation is over all the neighboring nodes, n, connected to m by a spring. The term $\mathbf{u}_m$ is the displacement of node m from its original position, and $\mathbf{M}_{mn}$ is a symmetric matrix that sets the elastic properties of the springs, through k and c, the central and noncentral force constants, respectively.

It has been shown that this system of springs obeys, to first order in displacements, the equations of linear elasticity theory for an isotropic elastic medium [27]. The elastic constants for this medium can be determined from the elements of the matrices $\mathbf{M}_{mn}$ [27]. These elastic constants are the Young's modulus, E, and Poisson's ratio, ν, which are of the form [27]

$$\mathrm{E} = [5k(2k + 3c)]/(4k + c) \qquad \nu = -(k - c)/(c + 4k) \tag{14}$$

A "force constant" is initially assigned to each node. Nodes that lie within the boundaries of the particles or defects are assigned different force constants than nodes that lie within the bulk matrix. The force constants for the bonds (that is, the spring constants) are averaged from the "force constants" assigned to the associated nodes.

The harmonic form of the energy results in forces linearly dependent upon the displacement of the nodes. If forces are applied to the nodes, and the spring constants specified,

then the nodal displacements can be obtained through a set of sparse linear equations. These equations are solved by using a conjugate gradient method to find the equilibrium configuration that corresponds to no net force at each node [27].

Typically, in purely elastic LSMs, a stress or strain is applied to the system boundaries, and the system is equilibrated, giving the next nodal displacements. The applied stress may then be incremented iteratively, and the system subsequently equilibrated, thereby allowing the simulation to proceed through a series of equilibrium states. Because the system is relaxed to the minimum energy configuration at each iteration, each iteration is considered "quasi-static."

As noted before, the equilibrium state of the system is determined by finding the displacements that result in no net forces. The stress and strain tensors are calculated from these forces and displacements. The strain tensor can be obtained through a finite difference approximation of the displacement field. In particular, a central difference approximation can be used to obtain this strain tensor [27]:

$$\delta_x \mathbf{u}_{(i,j,k)} = \left(-\mathbf{u}_{(i+2,j,k)} + 8\mathbf{u}_{(i+1,j,k)} + 8\mathbf{u}_{(i-1,j,k)} + \mathbf{u}_{(i-2,j,k)}\right)/12h \tag{15}$$

where $u_{(i,j,k)}$ is the displacement field at coordinates i, j, k, and h is the initial distance between adjacent nodes; alternatively forward or backward approximations are considered at system boundaries. The stress tensor is directly obtained from the forces acting on a node (the center of a cubic unit cell) [46],

$$\sigma_{ij} = \sum_m F_m n_{ij}^m / \mathrm{A}. \tag{16}$$

Here, $\sum_m$ represents a sum over the surfaces of an elementary cubic cell, F_m is the force on any surface m of the cubic cell, n_{ij}^m is a unit vector either normal or parallel to the surface m, and A is the surface area. The scalar stress and strain values correspond to the normal stress (σ_{xx} where x is the tensile direction) and strain components in the tensile direction.

To assess the effective reinforcement provided by the particles within the composite, we determine the relative quantities $(F - F_0)/F_0$, where F is the field in question and F_0 is the homogeneous response of the unreinforced polymeric matrix. The average strain in the system can be determined through the average nodal displacements at the system boundaries, in the tensile direction. The average strain and the applied stress can then be used to calculate the Young's modulus (stress of a material divided by its strain). This allows the global stiffness of this locally heterogeneous material to be determined.

Finally, we note that viscoelastic effects can be included in the LSM; although we do not discuss such simulations in this review, we have carried out viscoelastic LSM calculations on filled polymer systems and refer the reader to Ref. [47] for further details. To include viscoelasticity, a Kelvin unit is placed in series with each spring. A Kelvin unit is an elastic component that is connected in parallel with a viscous component, such as a dashpot (see Fig. 3) [47]. This configuration allows an instantaneous elastic deformation to be obtained while enabling full recovery of strain upon removal of the applied stress. The total recovery of strain is particularly important in simulations of fracture, where in the wake of the crack, the sustained stresses reduce significantly.

In the viscoelastic model, a relaxation time is introduced through the incorporation of viscous deformation. Investigations of the dynamics of elastic oscillations or elastic waves have been undertaken using LSMs [48, 49] and similarly using a Lattice Boltzmann model [50], but it should be noted that the time scale of interest here is considered to be significantly longer than the time scale for the damping of elastic waves. The simulation of a viscoelastic system can therefore be considered as proceeding through a sequence of equilibrium states, where the equilibrium bond lengths are no longer constant but incorporate viscous deformations. The viscous deformations do not vary significantly between iterations in these simulations [47], and therefore the information from the previous iteration is considered adequate for describing the equilibrium bond lengths in the current iteration.

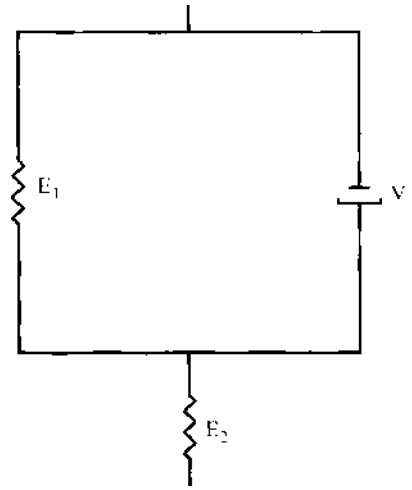


Figure 3. A Kelvin unit consists of an elastic component that is connected in parallel with a viscous component. Viscoelasticity may be incorporated into the LSM by considering a Kelvin unit in series with each spring.

3.2. Electrical Properties: Finite Difference Model

In addition to investigating the mechanical properties, we examined the electrical behavior of the nanocomposites [16]. The conservation of current, **J**, in a closed electrical circuit is written as $\nabla \cdot \mathbf{J} = 0$. Ohm's law relates the current to the electric field, $\mathbf{J} = G\mathbf{E}$, where G is the conductivity. Thus, we can write in the bulk

$$\nabla \bullet (G\mathbf{E}) = 0 \tag{17}$$

Because the electric field is the gradient of the electric potential, Φ, we obtain the following equation for an inhomogeneous system

$$\nabla \bullet (G\nabla\Phi) = 0 \tag{18}$$

Taking a finite difference approximation results in the following set of discrete equations

$$\sum_j G_{ij}|\Phi_j - \Phi_i| = 0 \tag{19}$$

where the summation is over the neighboring nodes on a square or cubic lattice, G_{ij} is the conductance between nodes i and j, and Φ_i is the potential at node i. Equation (19) is simply a combination of Ohm's law and Kirchoff's law of current conservation for a square network of resistors [29]. This system of equations is solved with constant voltages applied to the boundaries of the system in a specified direction.

Just as in the micromechanical studies, the morphological information serves as input to the electrical model. Conductivities G_i are assigned to the lattice nodes j at the beginning of the simulation; the specific value at a node depends upon its location, that is, if it is situated in a polymer domain or a nanoparticle. To assign a conductivity to the bond connecting two neighboring sites, we take the sites to be in series. In other words, the conductivity between nodes i and j is given by (1) $G_{ij} = [1/(2G_i) + 1/(2G_j)]^{-1}$. The particles are assigned a higher conductance than the polymers.

Much as in the mechanical studies, in these conductivity studies we enlarge the size of the simulation box relative to the size of the lattice in morphological studies by a specific factor (for example, a factor of 16); in this way, we more accurately capture the properties of each element and, consequently, the macroscopic properties of the system. With the voltage being applied at the system boundaries, we obtain the current flow through this heterogeneous system. In this manner, we obtain the global conductance as a function of the composition of the nanocomposite.

3.3. Optical Properties: Finite Difference Time Domain Method

The FDTD methodology [31] involves the discretization of Maxwell's curl equations of electromagnetism in three dimensions [51]. The differential form of Maxwell's equations is:

$$\nabla x H = \varepsilon(\partial E/\partial t) + \sigma E$$
$$\nabla x E = -\mu(\partial H/\partial t) - \sigma^* H \tag{20}$$

where E is the electric field, H is the magnetic field, ε and μ are the permittivity and permeability, respectively, and σ and σ^* are the respective electric and magnetic conductivities. Yee [52] introduced an efficient time stepping methodology for discretizing these equations, where the electric fields are given at the edges of the "Yee" cell, the magnetic fields are given at the faces of the Yee cell (see Fig. 4), and the magnetic fields are calculated half a time step later than the electric fields. In other words, the equations are solved in a leap-frog manner; that is, the electric field is solved at a given instant in time, then the magnetic field is solved at the next instant in time, and the process is repeated over and over again. In the study described in Section 4.4, we used the results from the SCF/DFT calculation in the FDTD simulation and thus chose the dimensions of the Yee cell to correspond to the SCF/DFT grid of points [30]. In this manner, the propagation of electromagnetic waves can be simulated in systems that possess nanoscale domains.

These equations must be solved with the appropriate boundary conditions. To simulate free space boundary conditions, the perfectly matched layer (PML) was introduced by Berenger [53]. The PML was designed to absorb electromagnetic waves without reflection and thereby negate the system size effects that arise from the use of a finite computational domain (see Fig. 5). The basic principle arises from the reflectance of a wave passing from a medium A into a medium B being given by

$$\Gamma = (\eta_A - \eta_B)/(\eta_A + \eta_B) \tag{21}$$

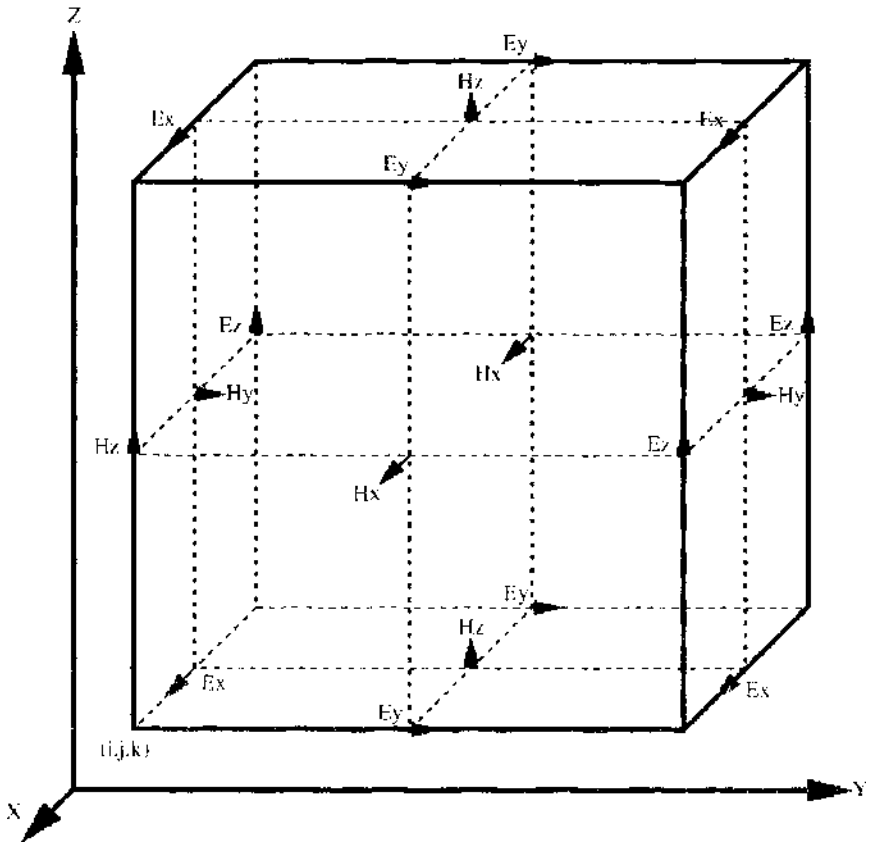


Figure 4. Positions of the field components within a Yee cell. The electric fields are given at the edges of the Yee cell, and the magnetic fields are given at the faces of the Yee cell.

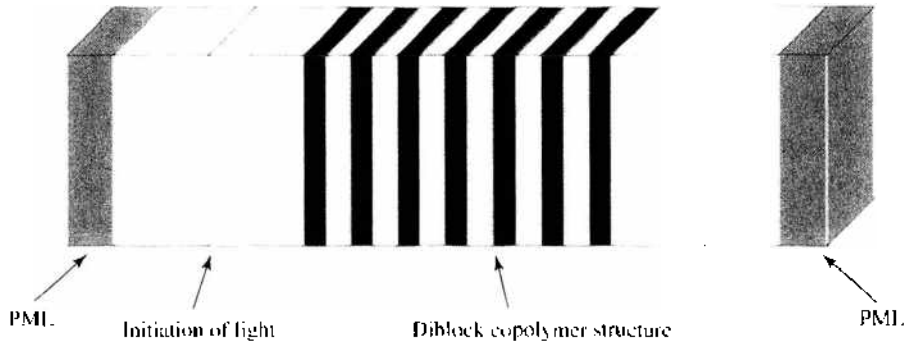


Figure 5. The structure of the FDTD simulation domain. Light is emitted toward both the filled diblock copolymer structure and the PML layer to the left of the figure. Upon reaching the PML layer to the left, all waves are absorbed without reflection. Upon reaching the filled diblock copolymer structure, some frequencies are reflected toward the PML layer to the left, and some frequencies are transmitted toward the PML layer to the right.

where $\eta = \sqrt{\mu/\varepsilon}$ is the impedance. If ε changes with μ between medium A and medium B such that η remains constant, then the wave will pass through the interface between the two media without reflection. If a series of layers is matched in this manner, then a wave will pass through the layered media without impediment. However, once the wave reaches the end of this layered medium, the wave will still be reflected from the simulation boundary and propagate back into the computational domain. To simulate an absorbing boundary, the wave must be absorbed as it propagates through the PML structure. This is accomplished through the introduction of anisotropic electric and magnetic conductivities. This is equivalent to introducing complex μ and ε, the real part dictating the propagating part of the wave and the imaginary part dictating the attenuating component [54].

Through the appropriate choice of parameters, the system will theoretically absorb incident waves without reflection. The conductivities must increase from zero at the inner interface of the computational domain to a maximum value $\sigma_{\max}$ at the outer edge of the layer at the system boundaries. All numerical computations reported here have been performed with conductivities of the form

$$\sigma(\rho) = \sigma_{\max}(\rho/\delta)^{\eta} \tag{22}$$

where ρ is the distance into the PML media, δ is the thickness of the PML media (corresponding to 200 cells), and η is taken as 3.5.

To carry out our studies, we first obtain the local volume fraction of the A-phase, B-phase, and particles from the SCF/DFT calculation [30]. These local volume fractions are then used to estimate the local dielectric constants of the media in the FDTD simulation. We employ a simple rule of mixture, linearly weighting the contributions of the different components from the respective volume fractions (that is, $\varepsilon(r) = \phi_A\varepsilon_A + \phi_B\varepsilon_B + \phi_P\varepsilon_P$ where ε_i characterizes the permittivity of the material i). In the current simulations, the dielectric constants for the respective blocks are taken to be 2.53 and 2.28; these numbers fall in the range that is typical for polymeric materials [55, 56]. The dielectric constant of the particles is taken to be 10.2, which corresponds to the value for cadmium selenide [56]. The spatially distributed dielectric constants are assigned to the center of mesh cells. The material properties at the nodes of a Yee cell are then averaged from the values in adjacent mesh cells.

We focus on cases where the system exhibits a lamellar structure, as determined through the SCF/DFT simulation [30]. Periodic boundary conditions are applied to the system boundaries in the y and z directions (reflecting the translational invariance along y and z in the lamellar structure). The periodicity of the SCF/DFT results along x (perpendicular to lamellar structure) is used to construct 48 layers of the lamellar structure in the x-direction, which are then sandwiched between two homogeneous regions. The size of the system, including the PML boundaries, is 3401 × 51 × 51, although there are no variations in the y and z directions in the lamellar systems studied here. To minimize the reflectance of incident light from the interface between the homogeneous regions and the layered region, the homogeneous regions are assigned a dielectric constant equal to that of the pure B phase material, and the

layered region begins and ends in the middle of the B domain. This ensures that the reflection of light when radiated on the multilayered structure depends upon the morphology of the layered structure. To establish a correlation between our simulation parameters and experimental values, and thus set up a physical length scale in our system, we equate the width of a lamellar domain obtained from the SCF/DFT for the pure symmetric diblocks to 50 nm [55].

A differentiated Gaussian pulse, encompassing a range of frequencies, is propagated toward the periodic structure of the diblock copolymer (see Fig. 5). Light of normal incidence (in the x-direction) when reflected or transmitted from the structure will always propagate in the x-direction due to the 1D nature of the structure. The fast Fourier transform (FFT) of the resultant transmitted and reflected signals then reveals the frequencies that were forbidden to propagate within the structure. The transmitted and reflected spectra can be presented relative to the FFT of the original differentiated Gaussian pulse to give the normalized transmittance and reflectance. In the current simulations, the diblock copolymer structure is assumed to be nonabsorbing, and therefore all light is either reflected by the structure or transmitted through it (that is, reflectance + transmittance = 1).

4. RESULTS AND DISCUSSION

The studies described next are presented in the following order. We first focus on a mixture of AB diblock copolymers and spherical nanoparticles. We couple the CH/BD and LSM models to examine how the self-assembly of the diblocks affects the dispersion of the spheres within the system and how these dispersed particles affect the mechanical behavior of the polymeric matrix [18].

In the following study, we use the CH/BD model to investigate the spatial organization of rodlike particles in a phase-separating polymer blend [16]. This study illustrates how the CH/BD technique can be adapted to model particles with high aspect ratios. In addition, the example reveals how immiscible blends can be exploited to direct the assembly of the rods within the polymer matrix. Using the results of this morphological study, we not only use the LSM to determine the extent to which the rods reinforce the matrix material but also use the FDM to calculate the electrical conductivity in these rod-filled materials. The latter calculations are particularly relevant to understanding the electrical performance of polymeric materials that contain nanotubes.

The beneficial effects of nanorod additives are also examined in the third study presented here, where we use the SCF/DFT model to capture the structure of rod-filled diblock copolymers [3]. Unlike the CH/BD model, which yields the dynamics of the structural evolution, the SCF/DFT is a purely thermodynamic model and yields the equilibrium morphology of the systems. This third study demonstrates another example of how morphological information can be coupled to the LSM simulations to obtain the mechanical response of the material.

The SCF/DFT technique is also used in the fourth study, where spherical nanoparticles are blended with diblock copolymers. In this case, we couple the structural information to the FDTD model to obtain the optical properties of the composite [30]. These studies are useful in designing photonic bad-gap materials.

These examples were chosen to demonstrate the general approach outlined in Fig. 1. Furthermore, the calculations highlight the fact that we do not make *a priori* assumptions about the structure of the material; rather, this structure is allowed to evolve or self-assemble based on the polymer–particle interactions, the architecture of the polymers, and the geometry of the particles. The examples also indicate the range of properties that can be determined once we obtain the structure of the material. Finally, the studies pinpoint the clear relationship between the structure and properties of nanocomposites.

4.1. Properties of Mixtures of Diblock Copolymers and Spherical Nanoparticles

As noted previously, we first focus on a mixture of AB diblock copolymers and spherical nanoparticles. One reason for focusing on mixtures of diblocks and nanoparticles is that the self-assembly of the diblocks can be exploited to direct the distribution of the nanoparticles

within the mixture [9, 10, 19] and thus achieve a degree of control over the morphology of the system. For example, if the particles are preferentially wetted by the A blocks of a system of AB diblocks, the particles will localize within the A domains of the microphase separated melt. In recent computational studies that involve our hybrid CH/BD model, we showed that in the presence of the AB diblocks, these A-like particles in fact form a percolated network at a significantly lower volume fraction (essentially half) than would be required in a homogeneous material [19] (that is, homopolymer melt). In this study [18], one of our aims is to determine how these percolating networks act to reinforce the copolymer matrix. More generally, we seek to tackle the issue of structure–property relationships for polymeric nanocomposites by relating the copolymer architecture, the wetting interactions between the copolymer and particles, the structure of the mixture, and the mechanical behavior of the resulting material.

We use the CH/BD hybrid method in order to determine the structural evolution of the particle-filled copolymer melt. The approach allows us to specify the architecture of the chains and the nature of the polymer–particle interactions. The output of the CH/BD simulation then serves as the input to the LSM, the micromechanical model. By combining the CH/BD and LSM models, we can determine how the structural evolution, or the history of the material, affects the mechanical response [57]. Furthermore, we do not have to make *ad hoc* assumptions about the distribution of particles in the system; this distribution evolves naturally from the self-assembling interactions between the different components. Through the LSM, we can carry out three-dimensional (3D) simulations that include as many as 1564 particles. In particular, these investigations represent the first 3D studies on the mechanical properties of such extensive filled copolymer systems [18]. The results allow us to determine how changes in the nature of the components influence the macroscopic properties of the composite.

We initially investigate the effects of varying the diblock copolymer architecture and particle volume fraction on the morphological and mechanical behavior. The effects of size and polydispersity in the particle system are also investigated. In the following section, we first describe our findings for diblock copolymers that are filled with particles of uniform size and then discuss the results for diblocks that contain binary particle mixtures.

4.1.1. Uniform Particle Size

In the following simulations, the parameters M and M_p (in Eqs. (5) and (10), respectively) are set equal to 1. The simulation box is 64^3 lattice sites in size, and periodic boundary conditions are applied along all three directions. Three different diblock copolymers are considered, corresponding to the following parameters: (1) $\Gamma = 0.004$ and $F = 0.0$, (2) $\Gamma = 0.016$ and $F = 0.0$, and (3) $\Gamma = 0.004$ and $F = -0.2$ [see Eq. (6)]. The parameter Γ is inversely proportional to N^2, where N is the degree of polymerization of the copolymer. Thus, an increase in Γ corresponds to a decrease in the domain size. Varying F from 0.0 to -0.2 changes the composition from a 50:50 to 40:60 AB diblock copolymer. Through these variations in Γ and F, we can examine the respective effects of altering the copolymer domain size and composition on the structure of the nanocomposite. The particles are preferentially wetted by the A blocks. A range of particle volume fractions, varying from 5% to 25%, are also considered, where the particle radius is 3 unit lengths (6 unit lengths in the LSM). This range in volume fraction of particles corresponds to a variation in the number of particles from 116 to 580 in the simulations.

The morphology of a filled diblock copolymer system at late times ($t = 50{,}000$) is presented in Fig. 6. The parameters of the diblock copolymer are $\Gamma = 0.004$ and $F = 0.0$, and the volume fraction of particles is 20%. The isosurface of the diblock copolymer, at an order parameter of zero (midway between phase A and phase B), is colored dark gray, while the regions where a negative order parameter intersects the system boundaries (termed *isocaps*) are colored light gray. In other words, the light gray regions mark the B phase, and the transparent regions indicate the A phase. The particles are colored black and are clearly confined within the transparent A phase of the diblock. The system shows elements of lamellar ordering on a short scale; however, the lamellae are interconnected, and the overall morphology

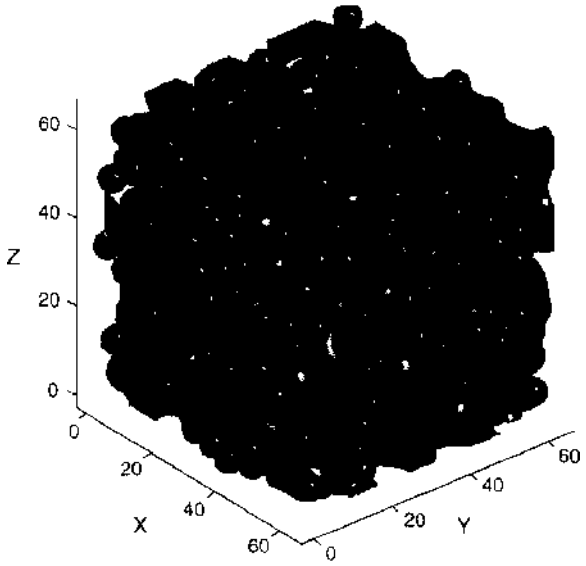


Figure 6. Three-dimensional morphology of a filled diblock polymer system. An isosurface between the A and B components is dark gray, isocaps are light gray, and the particles are black. The particles are confined within the transparent A phase of the diblock copolymer.

is closer to a bicontinuous structure. Although the system will tend toward the thermodynamic limit of a perfect lamellar phase, the time scales for reaching this state through a dynamic model are prohibitively large. In experimental systems, similar morphologies are found because again it takes long times to reach perfectly ordered phases, and the system can get kinetically trapped.

To quantify the confinement of nanoparticles within the diblock copolymer domains, the particle correlation function is presented in Fig. 7. The particle correlation function adopted in this study is defined as $g(r) = V(\sum_i \sum_{j \neq i} \delta(r - r_{ij}))/(4\pi r^2 N_p^2)$, where V is the volume of the system and N_p is the number of particles. The results are averaged over three independent runs. For clarity, only two diblock copolymer systems are shown ($\Gamma = 0.004$ and $\Gamma = 0.016$ at $F = 0.0$), as the third exhibited similar results. There is only one discernible peak at a distance of 6 unit lengths, which corresponds to the diameter of the particles.

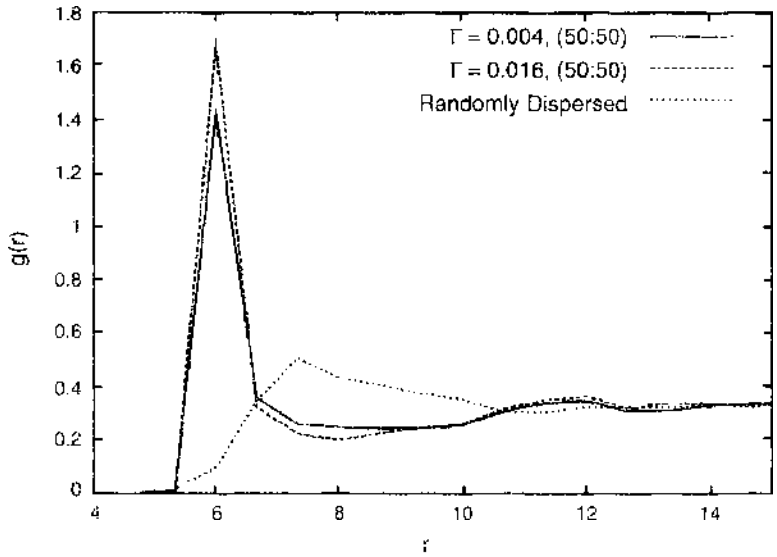


Figure 7. The pair correlation function for particles confined in two diblock copolymer systems and for randomly dispersed particles. The particles within the diblock copolymer systems display short-range order, which is not present in the homogeneous system.

This reveals that the system of particles exhibits strong short-range order but does not display long-range order. The particles are forced to lie within close proximity of each other due to the confinement within the diblock copolymer, but long-range order is suppressed because of the tortuous structure of the diblock domains. For comparison, the particle correlation function for an equivalent number of randomly dispersed particles, which exhibit no such confinement, is also presented. As can be seen, there is no local ordering of the particle positions, and the peaks observed in the diblock copolymer systems are no longer present.

To assess the consequences of such morphological variations on the resultant mechanical properties of the macroscopic material, we now use the output from our hybrid CH/BD simulation as the input for the LSM. The elastic deformation of the structures is undertaken, with both the force constants of the A and B phases being set to unity, and the particles are assigned a force constant of 100. Thus, the effects of particle distribution are of primary interest in the current investigations, and the parameters are consistent with experimental values for filled polymers [58].

As noted previously, the elastic properties of the springs within an LSM simulation are assigned values, depending on whether the node is situated within a particle or the polymer matrix, as dictated by the results of the CH/BD calculation. To accurately capture the deformation fields in the vicinity of the particles within the LSM, the system size is doubled from that of the CH/BD simulation. An LSM consisting of 148^3 nodes is utilized; the central 128^3 nodes are assigned elastic properties as a function of the particle and polymer positions in the CH/BD calculation. The system is extended by 10 unit lengths in all directions, taking values from the periodicity of the CH/BD simulation, therefore ensuring that all areas of the CH/BD model are represented by bulk nodes in the LSM simulation.

The local relative strain field, as a result of the application of a constant stress at the simulation boundaries, for a system where the particles are confined within the domains of a diblock copolymer ($\Gamma = 0.004$ and $F = 0.0$), is depicted in Fig. 8a. The corresponding relative strain field for a system of randomly dispersed particles is presented in Fig. 8b. The three-dimensional strain fields are displayed as orthogonal contours through the simulation. In both systems, the volume fraction of particles is 20%. The particles are clearly apparent as the dark regions of low strain. In particular, the strain values within the particles are significantly lower than that of the matrix, due to the large disparity in elastic constants. The inability of stiff particles to deform to the same extent as the neighboring matrix results in strain concentrations at the particle–matrix interface. These strain concentrations lie along the tensile direction and emanate from the center of a particle. Perpendicular to the tensile direction, the lower deformations within the particle inhibit the deformation of the matrix and results in lower strain fields.

It is apparent from Fig. 8a that the diblock-confined particles are clustered together, whereas the particles in Fig. 8b are more randomly dispersed. It is this clustering of the confined particles that is of primary interest. To characterize the particle clusters and determine whether geometric percolation occurs, we define particles that are closer than a certain distance to be part of the same cluster. Here, we adopt a unit length in the LSM simulations as this characteristic distance. Using this definition, we find that the confined particle system in Fig. 8a forms a percolating cluster. (We note that Ginzburg, Qiu, and Balazs [19] found the percolation threshold for particles confined in a similar diblock matrix to be approximately 10%.)

The percolating structure inhibits the deformation of the entire material and results in significant reductions in the strain fields, as can be seen by the presence of the dark domains in Fig. 8a. Alternatively, the randomly dispersed system shows isolated regions of strain relaxation within the particles, but the inhibition of the neighboring matrix is less dramatic than in Fig. 8a. Consequently, the strain concentrations (shown as lighter regions) within the matrix of the randomly dispersed system are also more pronounced, as regions within the matrix attempt to deform to the same extent as domains that neighbor the scattered particles. Such areas of strain concentration are less apparent in Fig. 8a. These plots indicate that the confinement of nanoparticles within one of the domains of the bicontinuous structure leads to a continuous network of stiff material, which reduces the overall strain field within the system.

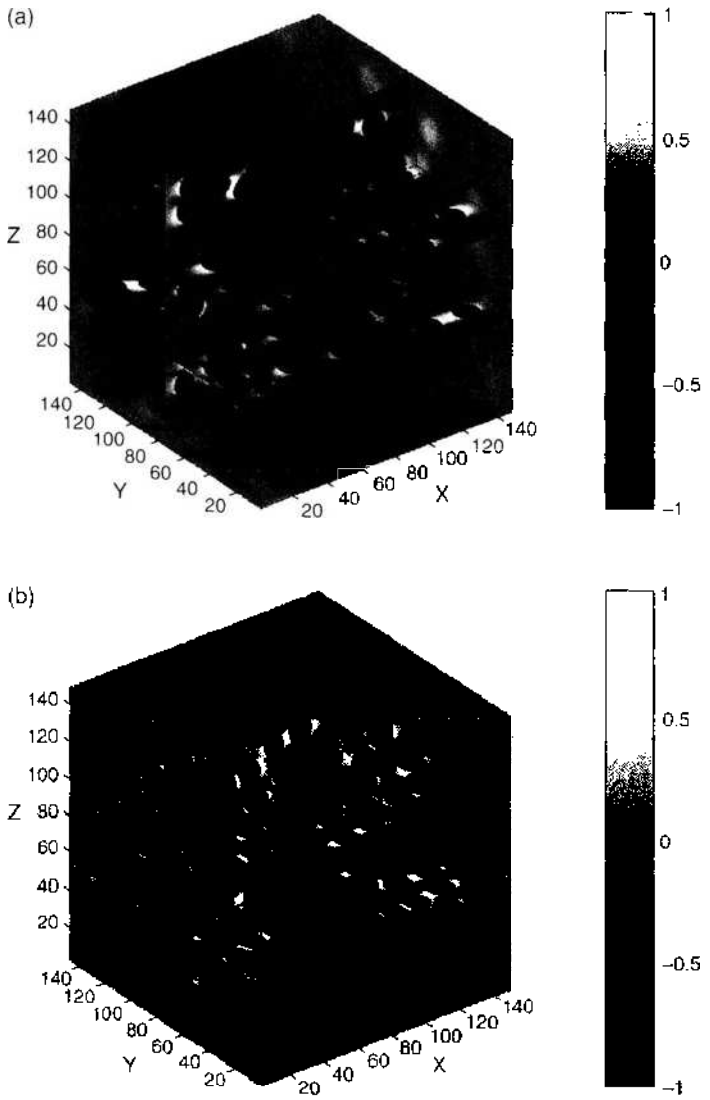


Figure 8. The relative normal strain fields (see text for details) for (a) a system where the particles are confined within the domains of a diblock copolymer and (b) a system consisting of randomly dispersed particles. Light regions depict regions of strain concentrations, and dark regions indicate regions of low strain.

To quantify the deformation of these confined and randomly dispersed particle systems, the cumulative distribution functions of the local strain fields are plotted in Fig. 9. The cumulative distribution function is defined as the probability that the field in the system takes a value less than or equal to a specific amount. A comparison between confined and randomly dispersed particle systems is made for particle volume fractions varying from 5% to 25%. The lower strains are invariably associated with the stiffer particles, whereas the regions of higher strains correspond to the matrix. At 5%, there would appear to be little difference between the two systems, because the confined particles do not percolate at such a low value. At higher particle volume fractions, the disparity between the two systems becomes more apparent, with the confined particle systems possessing significantly lower strain fields. As noted previously, the geometric percolation inhibits the local strain fields and therefore stiffens the composite material.

In Fig. 10, we plot the percentage increase in the Young's modulus relative to the unreinforced polymer for the various systems described. This parameter is a measure of the macroscopic mechanical properties of these composites. The results are averaged over three independent runs, with the error bars indicating the standard deviation. The three cases involving particles confined within the domains of diblock copolymers are significantly stiffer than the system containing randomly dispersed particles. There is no clear difference

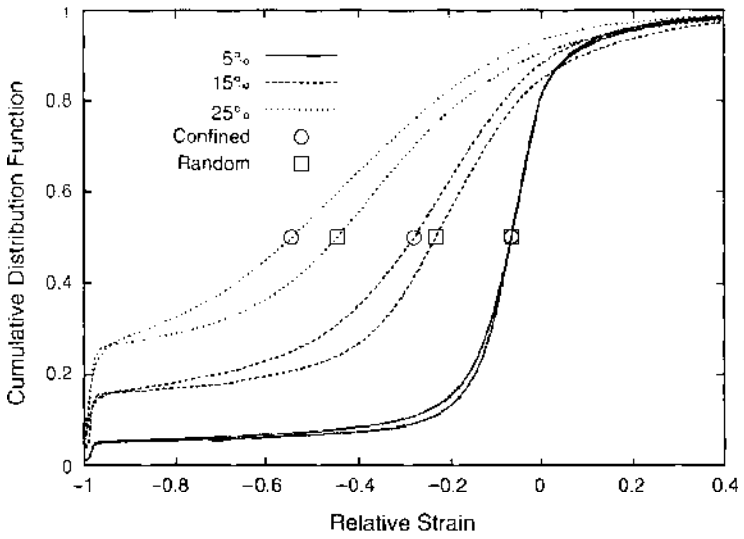


Figure 9. The cumulative distribution function of the relative local strain field for systems consisting of randomly dispersed particles and systems where the particles are confined within the domains of a diblock copolymer. Curves are shown for 5%, 15%, and 25% particle volume fractions, and as the volume fraction of particles increases, there is an increase in the disparity between systems containing particles confined within the diblock structure and systems containing particles randomly dispersed in a homopolymer.

between the three diblock-confined particle systems, with the error bars showing a clear overlap. For the systems studied here, it is not possible to distinguish effects that diblock copolymer architecture or composition may have on the reinforcement efficiency of the nanoparticle fillers. There is, however, a significant benefit to confining the particles within the diblock copolymer domains.

4.1.2. Binary Particle Systems

We also investigate the effects of adding a binary particle mixture to the copolymer matrix. The particles in the binary mixture are chemically identical (that is, they both favor the A phase), but they differ in size. The smaller particles have a radius of 2, and the larger particles have a radius of 3 unit lengths. These studies provide insight into the role that polydispersity in particle size plays in the mechanical properties of the composite. The total

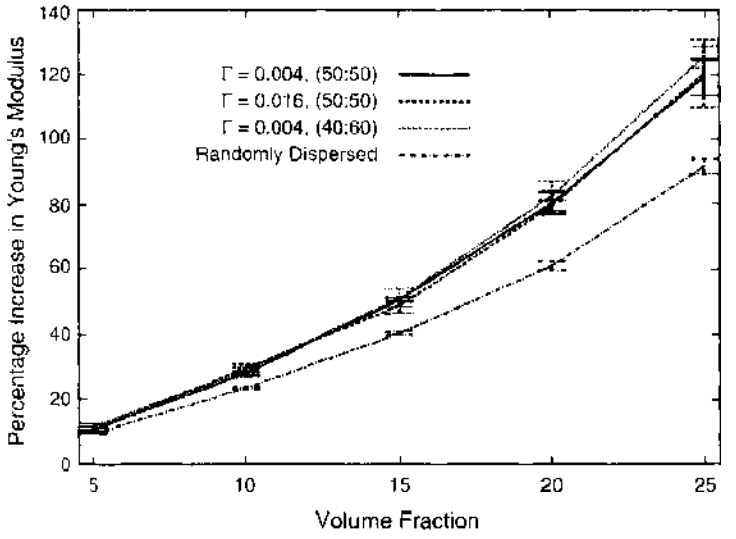


Figure 10. The percentage increase in Young's modulus as a function of particle volume fraction. Systems containing particles confined within the domains of different diblock copolymers are compared with a system consisting of particles randomly dispersed as in a homopolymer.

volume fraction of particles is held fixed at 20%, and the ratio of small to large fillers is varied between the limiting cases of all small (1564 particles) and all large (464 particles). A comparison of the results for the purely large and small fillers yields insight into the effects of particle size on the behavior of the system. In these studies, the parameters that characterize the diblock copolymers are fixed at $\Gamma = 0.004$ and $F = 0.0$.

Figure 11 shows the morphology for a system containing 10% large particles and 10% small particles. The particles are again clearly confined within the A domains of the diblock copolymer. Note that the particles selectively swell these compatible A regions, giving the diblock matrix in Fig. 11 an asymmetric appearance. However, here as in the other cases described in this section, the copolymer is a symmetric diblock.

It is also clear from Fig. 11 that the small particles can readily penetrate and localize in regions between the large particles. At a fixed particle volume fraction, decreasing the size of the particles results in an increase in the total number of particles and hence an effective increase in the particulate surface area. Consequently, there is a greater surface area available for possible polymer–particle interactions. To illustrate this point, we define V_p as the volume fraction of polymeric material (A or B) that is within a given distance (a unit length) of any particles and plot V_p as a function of the volume fraction of small particles (see Fig. 12). The results are averaged over three runs and the standard deviations were found to be negligible. As the volume fraction of small particles is increased, the particles come in contact with and affect a greater volume of the matrix.

The fraction of particles that are a part of the largest cluster, P, is plotted as a function of the volume fraction of small particles in Fig. 13. The data are averaged over three runs, with the error bars corresponding with the standard deviation. Geometric percolation occurred in all systems. Because at a fixed volume fraction there are a greater number of small particles than large ones, these fillers would be expected to cluster to a greater extent than the larger species (as the characteristic distance used to indicate clustering is not radius dependent but is fixed at one unit in our studies). This is in fact the case, with the fraction of particles in the main cluster approaching one with an increasing number of small particles. Even though small particles are more dispersed (as indicated by Fig. 12), they still cluster to a greater degree within the domain structure of the diblock copolymer (in part, because there are a greater number of them than large particles). This increase in clustering is expected to translate through to the mechanical properties.

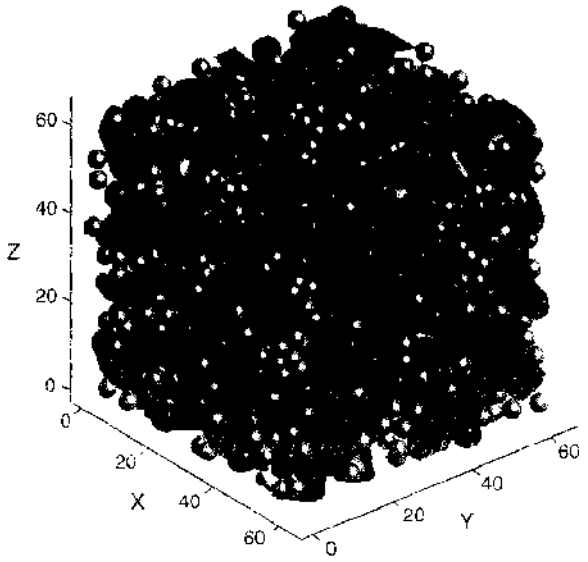


Figure 11. Three-dimensional morphology of a filled diblock copolymeric system. An isosurface between the A and B components is dark gray, isocaps are gray, and the small and large particles are colored black and light gray, respectively. Both the small and large particles are confined within the transparent A phase of the diblock copolymer.

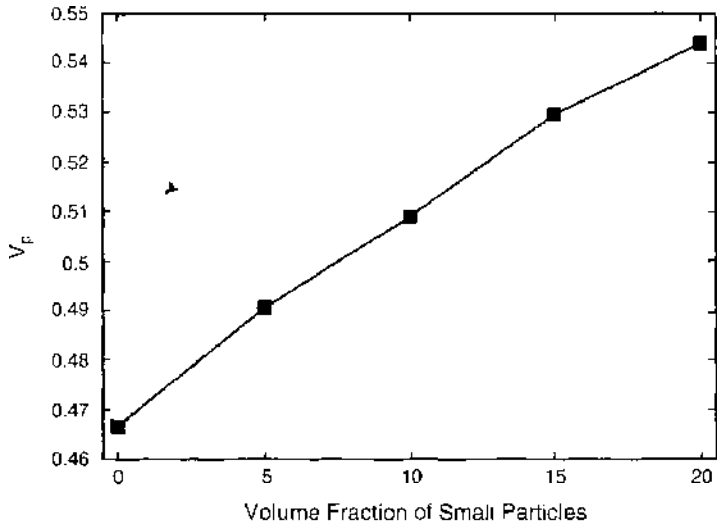


Figure 12. The volume fraction of polymeric material that is within a given distance (a unit length in the LSM) of any particles as a function of the volume fraction of small particles. This gives a measure of the amount of polymeric material that is directly affected (such as pinched) by the presence of the particles.

Figure 14 reveals the relative strain field for a system containing 10% large particles and 10% small particles. The regions of low strain, corresponding to the stiffer particles, are clearly observed, as previously. Now, however, the clustering of particles is more apparent than in Fig. 8a. A significant difference between Fig. 8a and Fig. 14 is the area over which these particles cluster. The smaller particles spread out over a greater volume of the material and inhibit the deformation of the matrix to a greater degree than in the system containing just large particles. Effectively, a larger volume of polymer matrix is trapped or surrounded by the particles and therefore less capable of deforming.

Quantitatively, the effects of particle size can be seen in Fig. 15, which depicts the cumulative distribution function of the local relative strain, as the volume fraction of particles varies from being 20% large to 20% small. The data is averaged over three independent runs. The plots show that 20% of the system possesses lower strains because of the 20% of stiff particles present within the composite; however, in the upper 80% of the system, a gradual trend is observed. The deformations in the matrix are increasingly inhibited as the volume fraction of smaller particles is increased.

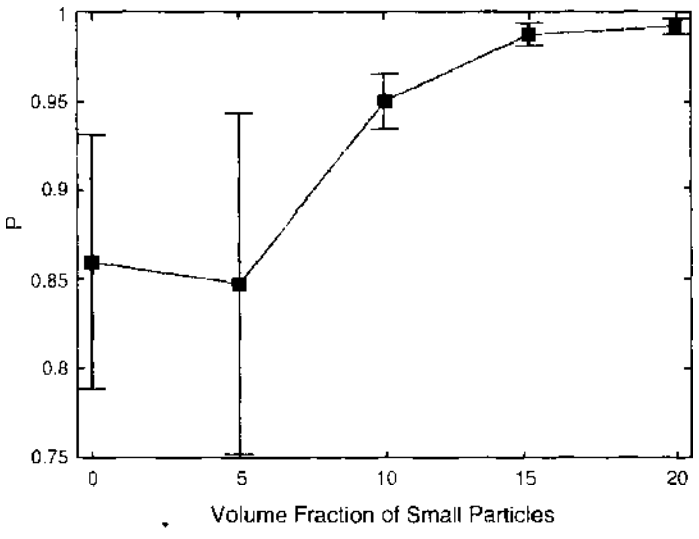


Figure 13. The fraction of particles that makes up the largest cluster as a function of the volume fraction of small particles. The overall volume fraction of small and large particles is maintained at 20%. In the system containing all small particles, nearly all of the particles are connected.

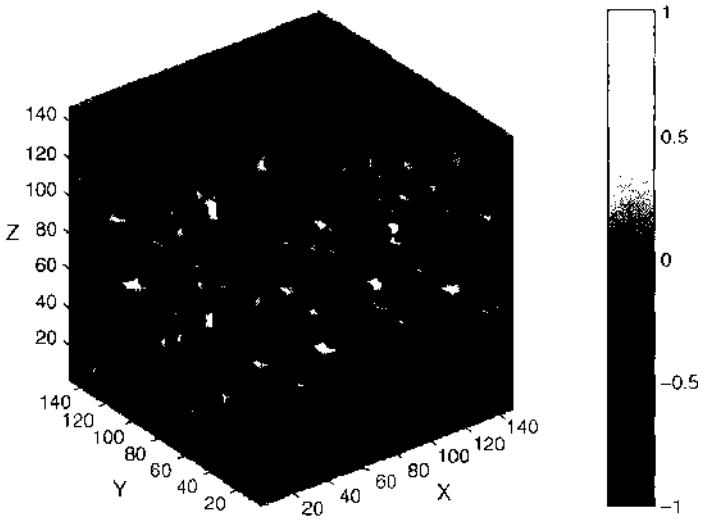


Figure 14. The relative normal strain fields (see text for details) for a system containing a 10% volume fraction of small particles and a 10% volume fraction of large particles, both confined within the domains of a diblock copolymer. Light regions depict regions of strain concentrations, and dark regions indicate regions of low strain.

The lower strain fields, due to the decrease in particle size, have a direct impact on the Young's modulus of the macroscopic material. The percentage increase in Young's modulus is plotted in Fig. 16, as a function of the volume fraction of small particles. The data is averaged over three runs, and the error bars represent the standard deviations. An increase in Young's modulus of more than 30% is observed as the particle size is altered from all large to all small. This is attributable to an increase in the total particle surface area, a greater degree of clustering, and an increase in the volume of polymeric material that is effectively trapped by the particles. These effects result in lower strains throughout the system and, an increase in the global Young's modulus.

4.1.3. Summary

Through a combination of numerical techniques, we were able to interrelate the structure and micromechanical behavior of the copolymer/nanoparticle composites. Through the CH/BD calculations, we could determine the effects of the microphase separation of the

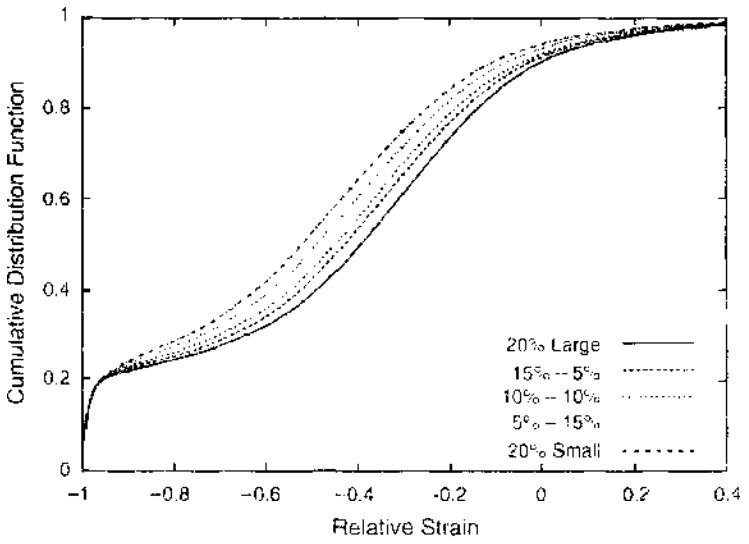


Figure 15. The cumulative distribution function of the relative local strain field for systems containing various volume fractions of small and large particles. The overall volume fraction of particles is maintained at 20%. The volume fraction of small and large particles varies from 20% small particles, through 10% small particles and 10% large particles, to 20% large particles.

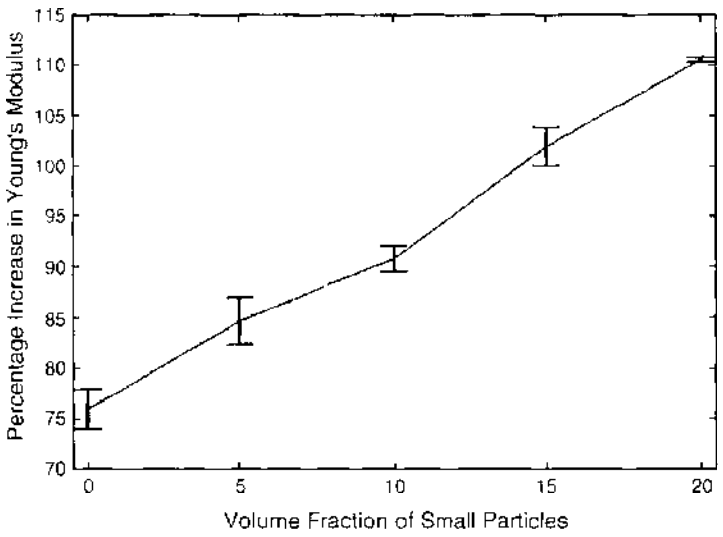


Figure 16. The percentage increase in Young's modulus as a function of small particle volume fraction. The combined volume fraction of small and large particles is maintained at 20%. As the system varies from all large particles to all small particles, there is an increase in the Young's modulus.

diblocks on the spatial distribution of the mobile particles. Through the LSM, we could capture the elastic deformation of the resultant hybrid material. Furthermore, we could investigate the behavior of systems that contain up to 1564 particles. For randomly dispersed fillers, simulations involving a relatively low number of particles can be sufficient to describe the overall stiffness of the material and thus be large enough to encompass a representative volume element (RVE) of the composite. However, the tortuous spatial arrangement of particles confined in diblock copolymers introduces an additional length scale, that of the domain size. To determine the mechanical behavior of such complex materials, it is important to consider the morphology of a sufficiently large system that captures both the unique structural characteristics of the copolymer domains and the particles' spatial arrangement, which is, in part, templated by these diblock domains.

The LSM has proven to be ideally suited for simulating the micromechanics of such large systems. Although the utility of the LSM in analyzing two-dimensional RVEs has recently been reported [59], it would appear that this technique might also prove useful in similar three-dimensional analyses.

Through the selective incorporation of nanoparticles into the domains of a diblock copolymer, three-dimensional bicontinuous nanoparticle structures were formed. As the volume fraction of particles was increased, geometric percolation of the particles occurred, and the particles effectively formed a rigid network throughout the system. The deformations within the polymer matrix are significantly suppressed by the presence of this rigid nanostructural network, and the global stiffness of the material is notably increased. For materials containing randomly dispersed spheres, rods, and platelets, the rods and platelets offer superior reinforcement over the spheres [47]. Therefore, the mechanical properties of diblock copolymers filled with such high aspect ratio particles may prove to be of particular interest, as we discuss in more detail in subsequent sections.

Varying the size of the monodispersed particles and introducing bidispersity in the particle size were shown to exert an appreciable influence over both the morphology of the diblock copolymer and the resultant mechanical properties of the solid material. Systems containing small fillers exhibited a greater degree of clustering between the particles. This behavior could explain the increased stiffness that was observed in the corresponding macroscopic material. Also of considerable consequence is the volume of polymeric material that is effectively trapped between neighboring nanoparticles. For a fixed volume fraction of particles, as the particle size is decreased, the number of particles increases. Consequently, the volume of material in which the particles are dispersed increases, and the deformation of a greater volume of interparticle polymeric material is inhibited.

We conclude that the inclusion of nanoparticles into a bicontinuous diblock copolymer structure results in a significant increase in the reinforcement efficiency of the fillers. As polymeric nanocomposites become increasingly important, such confinement effects will play a dominant role in optimizing their mechanical behavior and can be exploited to expand the utility of these materials in a range of applications.

4.2. Behavior of Nanorods in Binary Polymer Blends

In the next investigation, we make two modifications to the CH/BD model to consider the behavior of another class of filled polymeric systems. First, we set $\Gamma = 0$ and thereby model a blend of immiscible homopolymers. Through these studies, we examine how the phase separation of the blends affects the spatial organization and dispersion of the nanoparticles within the polymeric matrix. Second, we modify the aspect ratio of the nanoparticles in order to consider rodlike particles. These rodlike fillers can represent carbon fibers, ceramic whiskers, or nanotubes.

Nanoscale rods are becoming increasingly important additives in the fabrication of polymer composites [60, 61]. The high aspect ratio (and high surface area) of nanorods results in polymer nanocomposites that can possess superior mechanical properties relative to polymers reinforced with an equivalent volume fraction of spherical inclusions [47]. Recently, Huynh, Dittmar, and Alivisatos [60] blended inorganic nanorods and polymers to fabricate solar cells that possess greater efficiencies than conventional organic photovoltaic cells. In this application, nanorods were preferable, in part because they naturally provide a direct pathway for charge transport. Peng et al. [22] showed that immersing nanorods into a binary phase-separating blend can drive the rods to form percolating networks at relatively low volume fractions. These extensive networks can potentially improve both the mechanical and electrical properties of the polymer/nanorod mixtures.

To facilitate the design of such high-performance materials, it is useful to correlate the microstructure of the filler network and the macroscopic properties of the polymeric composite. To establish these correlations, in this section we simulate the morphological evolution of a binary blend that contains nanorods and the mechanical and electrical properties of the resultant nanocomposite.

We first present the morphology results from the CH/BD simulations. We vary the volume fraction of the nanorods, which are incorporated into a 30:70 AB polymer blend. The rods are preferentially wet by the minority A phase. These morphological studies are very similar to those carried out by Peng et al. [22]. Here, however, our aim is to extend these studies by relating the microstructure of the mixture to the macroscopic properties of the nanocomposite. With this aim in mind, we contrast the local deformation fields and the global Young's modulus for two separate nanorod/polymer mixtures. In the first case, the rods are introduced into the binary, phase-separating blend. In the second case, the rods are dispersed in a homogeneous matrix (such as a homopolymer); here, we neglect the Cahn-Hilliard equation and simply numerically solve the equation of motion [Eq. (12)] for the rods. We find that the localization of the nanorods into the minority phase of the phase-separating blend leads to substantial improvements in the reinforcement efficiency of the rods. Finally, we compare the local current densities for these two cases and find a relative increase in the electrical conductivity in the system where the nanorods are selectively incorporated into the minority phase of the blend.

4.2.1. Morphology: Formation of Supramolecular Networks

The distribution of nanoscale inclusions within a polymer matrix can yield an appreciable influence over the macroscopic properties of the nanocomposite. We first consider the effects of incorporating the nanoscale rods into the minority A phase of the 30:70 AB blend. In particular, we examine the effects of increasing the volume fraction of these rods on the microstructure of the complex mixture. The rods have a width W of one lattice site and a length L of 13 lattice sites. The size of the two-dimensional simulation box is 256×256 lattice sites. In Figs. 17a–c, we present the respective morphologies of nanocomposites that contain rod volume fractions of 2%, 4%, and 6%. The morphologies represent the system

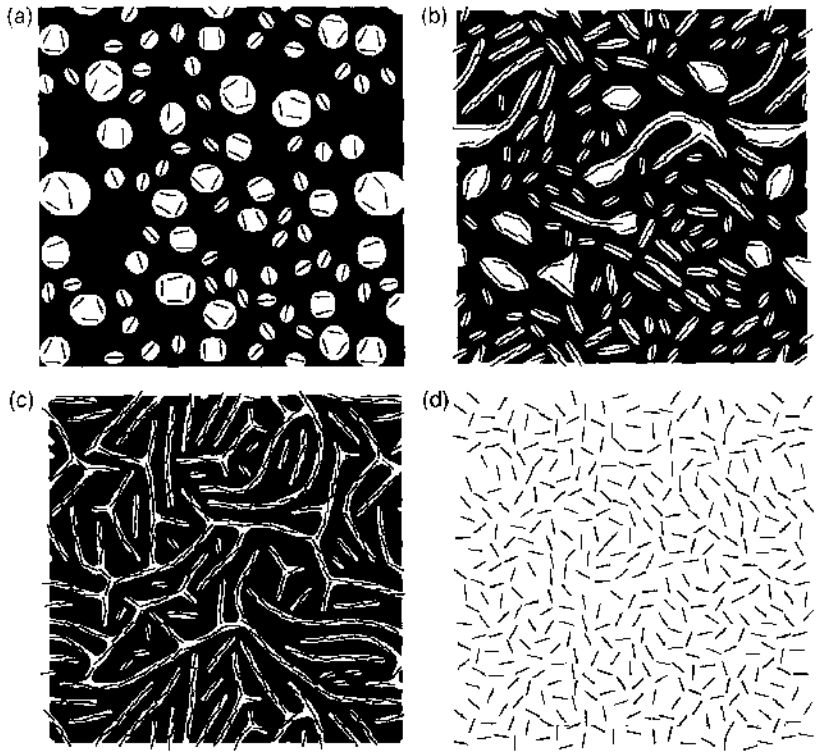


Figure 17. The morphology of nanorod/polymer blend system for (a) 2%, (b) 4%, and (c) 6% volume fraction of nanorods. White regions are the minority phase A, black regions are the majority phase B, and the rods are black rectangles. For comparison, (d) depicts the morphology for a system of 6% volume fraction of nanorods within a homogeneous matrix.

at the late stage of domain growth ($t = 600{,}000$). The minority A phase is shown as white domains, and the majority B phase is shown as black domains. The rods are shown as black rectangles. As a basis of comparison, Fig. 17d shows the structure of a 6% volume fraction of rods in a homogeneous matrix (in the absence of the phase-separating blend).

Figure 17a depicts the structure of the system that is loaded with a 2% volume fraction of rods. The rods are clearly located within the minority A phase. The A domains are dispersed throughout the system, with smaller domains containing a single rod and larger domains containing multiple rods. The shapes of the smaller domains are perturbed from circular to ellipsoidal by the presence of the rods. The larger domains appear less affected by the presence of these fillers.

The consequences of increasing the volume fraction of the rods from 2% to 4% can be seen in Fig. 17b. Again the A domains are dispersed throughout the system, but now the domain shapes are appreciably different from those shown in Fig. 17a. Smaller domains containing a single rod are still ellipsoidal, but now domains that contain more than one rod have a tendency to be elongated. The coupling between the wetting interactions and the rod–rod repulsion has led to the rods being aligned end-to-end within the A phase. Furthermore, the A phase has been stretched by the presence of the rods. This cooperative behavior leads to the formation of "stringlike," reinforced A domains. Further increases in the volume fraction of rods enhance these effects.

Figure 17c shows the morphology for a system containing 6% volume fraction of rods. Increasing the volume fraction of rods has increased the elongation of the minority domains to the extent that they now interconnect and form continuous domains across the system. The rods are "corralled" into these elongated domains to such an extent that they percolate throughout these domains [22]. The percolation of these nanoscale inclusions and the continuous nature of the minority A phase result in a system that exhibits double percolation.

Within a homogeneous matrix, the 6% volume fraction of rods is uniformly dispersed in the system, as shown in Fig. 17d. This volume fraction is below the critical value at which the rods display a nematic ordering [22], and the system is in the isotropic phase.

This uniform distribution is in distinct contrast to the highly organized structure seen in Fig. 17c; the electrical and mechanical benefits of having the nanorods distributed in this controlled manner are discussed next.

4.2.2. Mechanical Properties

Before presenting our results on the mechanical properties of the rod-filled polymers, we note that there have been previous computational studies on the micromechanical behavior of materials containing rodlike particles; these studies were primarily focused on the properties of fiber-reinforced composites. The most prevalent computational technique for simulating fiber-reinforced composites is the finite element method (FEM) [62–67]. Complex nonlinear behavior, such as plastic deformations [64–66] and fracture [63], can be examined through the use of such FEM simulations. However, the FEM simulations are sufficiently computationally intensive that the time requirements for the calculations limit the range of materials that can be investigated. In FEM studies, fiber reinforcements are often considered to be structurally periodic, simulated through the use of unit cells. The fibers are therefore often arranged unidirectionally in an equally spaced array; however, clustering effects have been investigated through the use of staggered cells [64, 65]. In such models, the clustering effects are also assumed to be unrealistically regular, though recently multi-inclusion unit cell elastic FEM simulations containing up to 25 fibers have been carried out [67]. To extend the size of the simulations that can be investigated, alternative, computationally less intensive techniques may need to be considered.

Termonia [68–70] utilized a simple finite difference approach that enabled the three-dimensional simulation of systems with rodlike inclusions. The elastic deformation of cubic and parallelepiped multi-inclusion systems were simulated [69]. Through the incorporation of a stochastic fracture criterion, the model was extended to the failure of such multifiber systems [70].

The LSM has also been utilized in the simulation of fiber-reinforced composites. Monette et al. [71, 72] simulated the two-dimensional deformation and fracture of a single fiber, whose aspect ratio and interfacial properties were varied [71]. A transition between fiber fracture at larger aspect ratios and interfacial decohesion at lower aspect ratios was observed. Ostoja-Starzewski [73] investigated the size of the representative volume element of randomly oriented fiber composites by using a two-dimensional LSM. The one-dimensional fibers consisted of the orthogonal linkage of one-dimensional stiff bonds. Recently, we used the LSM technique to examine the behavior of three-dimensional, multiparticle systems and compared the properties of composites containing rods with systems containing spherical and platelet inclusions [47].

In the LSM simulations presented here, we assume that the rods are 10 times stiffer than the polymeric matrix (both the A and B components are assumed to have the same stiffness) and investigate the effects of rod distribution on the local and global elasticity of the nanocomposite. In particular, we fix the volume fraction of rods at 6% and compare the local deformations in a system where the rods are corralled by a polymer blend into percolating pathways with a system where the rods are allowed to evolve in a homogeneous matrix. We also compare the global Young's modulus as a function of the rod volume fraction for these two distinct systems. To carry out these calculations, we apply a stress field at the LSM boundaries in the x-direction, thus defining the tensile direction.

The local relative normal stress field is defined as $(\sigma_{xx} - \sigma_0)/\sigma_0$ where σ_{ij} is the stress tensor and σ_0 is the normal stress field of an unreinforced polymer, that is, a homogeneous material that does not contain rods. Figure 18 shows these stress fields for rods dispersed in a homogeneous matrix (Fig. 18a) and in a phase-separating blend (Fig. 18b). Stress concentrations within the rods can be seen as lighter regions, corresponding to higher stresses. The darker regions correspond to lower stresses and can be seen within the polymer matrix surrounding the rods. Rods oriented in the tensile direction (x-direction) have a greater effect on the stress field than rods oriented normal to x; in particular, rods oriented along x exhibit greater stress concentrations and cause lower stresses within the neighboring polymeric material. This can be seen in Fig. 18b, which depicts the stress field for the phase-separating system. In regions where the local domains are aligned in the tensile direction, the

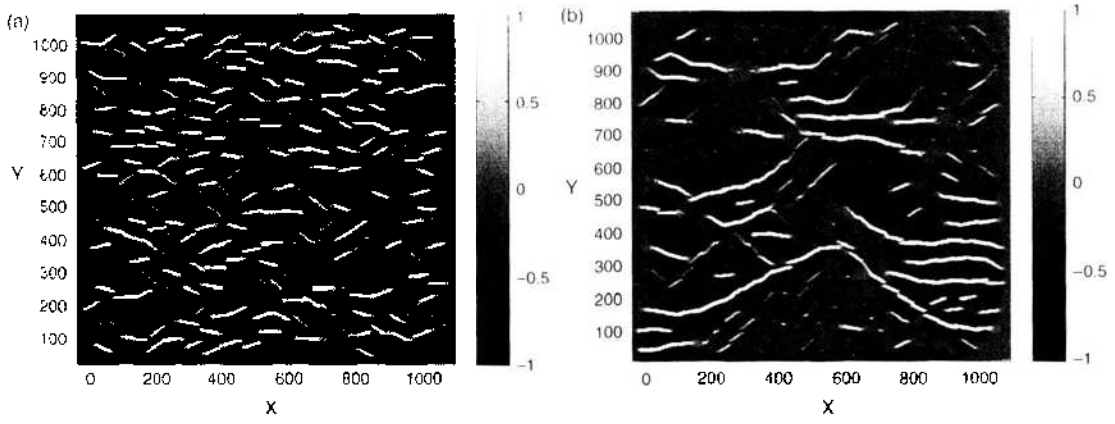


Figure 18. The relative normal stress field contours for (a) 6% nanorods in a homopolymer and (b) 6% nanorods in a phase-separating system. Light regions indicate stress concentrations, and dark regions depict stress relaxation.

rods are clearly visible as regions of high stress; however, domains that are aligned normal to the tensile direction exhibit lower stress perturbations (at the ends of the rods). Because the stiff rods prohibit the matrix from deforming, the polymeric material that is confined between neighboring rods exhibits particularly low stress. In other words, there is efficient stress transfer from the softer polymer matrix to the stiffer rods.

The local relative normal strain field is defined as $(u_{xx} - u_0)/u_0$, where u_{ij} is the strain tensor and u_0 is the normal stain field of an unreinforced polymer. These strain fields are presented in Fig. 19. The positions of the stiffer rods are apparent from the dark regions of lower strains. What is of most interest is the manner in which the matrix is prohibited from deforming. Figure 19a shows the local strains for a homogeneous system, where the rods are dispersed uniformly throughout the system. Rods aligned with the tensile direction induce strain concentrations within the polymer matrix at the edges of the rods and induce strain reductions at the sides of the rods. In contrast, rods that are aligned perpendicular to the tensile direction show little perturbation of the strain field around these inclusions. Figure 19b depicts the local strain field for a polymer blend with the rods confined within the minority phase. Regions where the domains are aligned with the tensile direction induce severe strain reductions as the local ordering of the polymer domains, and hence the corralled rods, prohibits the deformation of substantial regions of the polymer matrix. These regions of low strain have a macroscopic influence.

Figure 20 shows the percentage increase in the global Young's modulus relative to an unfilled (no rods), homogeneous material; the data are averaged over three independent realizations from the CH/BD calculations (at fixed volume fraction of rods), and the error bars represent the standard deviation. The solid line depicts data taken from the rod-filled

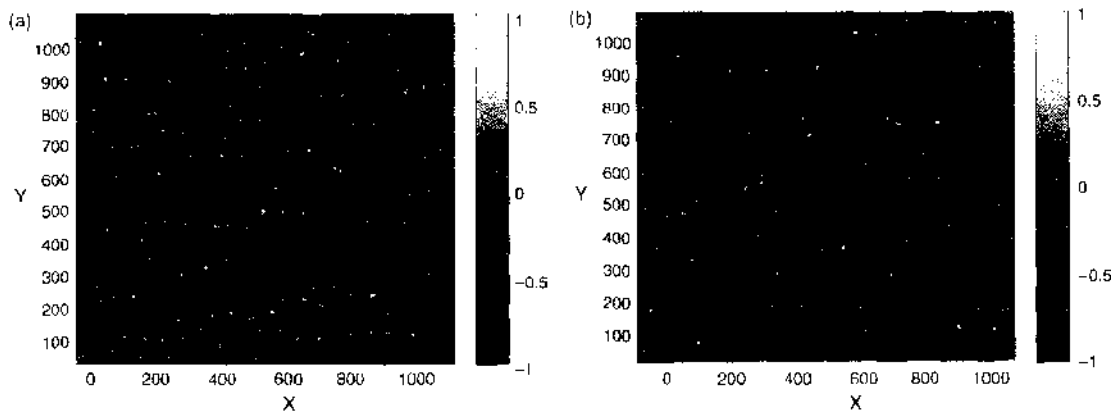


Figure 19. The relative normal strain field contours for (a) 6% nanorods in a homopolymer and (b) 6% nanorods in a phase-separating system. Light regions indicate strain concentrations, and dark regions depict low deformation.

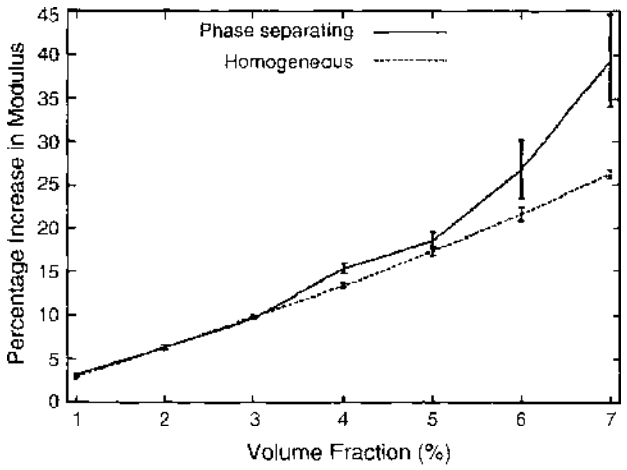


Figure 20. The percentage increase in Young's modulus as a function of nanorod volume fraction. Systems where the rods are corralled into percolating pathways by the phase-separating polymer blend are compared with systems where the rods evolve isotropically within a homopolymer.

phase-separating systems, and the dashed line represents data from the equivalent rod-filled homogeneous material. At lower volume fractions of rods, there is little difference between the two cases, as the rods do not appreciably affect the polymer blend morphology, and the polymer blend does not dramatically perturb the distribution of the rods. As the volume fraction of rods is increased, the polymer blend begins to form elongated domains, and when these domains are elongated in the tensile direction, significant improvements in the reinforcement efficiency of the rods ensue. This is especially the case at higher rod concentrations, where the system becomes doubly percolating and the rods form a continuous backbone of stiff material throughout the nanocomposite.

Next, we consider the electrical benefits of loading nanorods into a phase-separating system as opposed to a homogeneous system.

4.2.3. Electrical Properties

The conductivity of the rods is taken to be 10 times greater than that for the polymer matrix. We assess how the rod distribution affects both the local electrical behavior and the global conductivity. Figure 21 shows the current density for systems containing 6% of the rods; in Fig. 21a, the rods are dispersed in a homogeneous matrix, and in Fig. 21b the rods are corralled into percolating pathways by the surrounding polymer blend. The current density at the ith node is defined as $I_i = \sum_j |G_{ij}(\Phi_j - \Phi_i)|$ and reflects the amount of current flowing through a given node. In Fig. 21a, the current density is greater in the higher-conductance

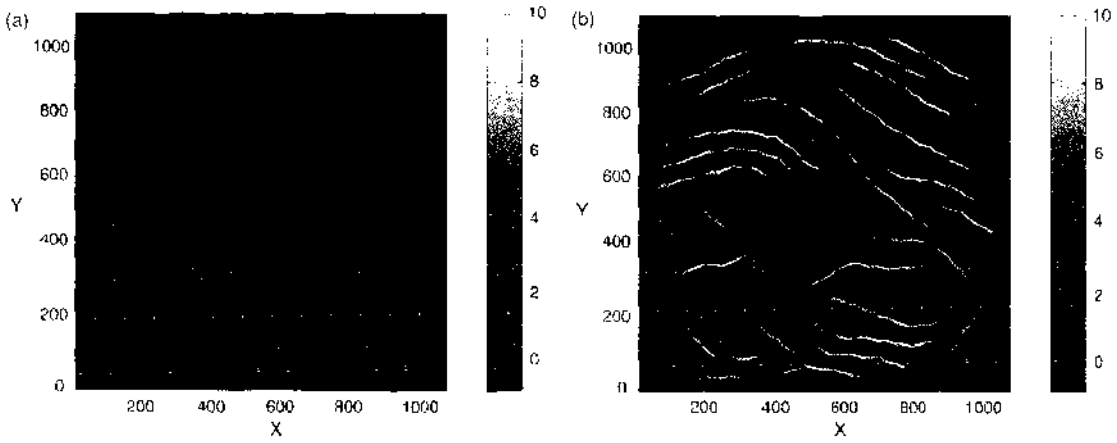


Figure 21. Current density contours for (a) 6% nanorods in a homopolymer and (b) 6% nanorods in a phase-separating system. Light regions indicate high current flow, and dark regions depict low current flow.

rods and is especially high for rods that are oriented in the direction of potential gradient (x-direction). The current density within the rods is appreciably higher in Fig. 21b than in Fig. 21a. The current flowing along the rods is greater when the rods are oriented in the same direction as the potential gradient. The percolating nanorods provide a direct pathway for electrical conductivity across the system. The effect that the percolating structure has on the global conductivity is considered next.

Figure 22 depicts the percentage increase in conductivity relative to an unfilled homogeneous material, and the data are averaged over three independent runs, with the error bars reflecting the standard deviation. At lower volume fractions of rods, there is no benefit to incorporating them into a phase-separating polymer blend relative to adding them to a homogeneous polymer matrix. As the volume fraction of rods is increased, the elongated domains in the phase-separating system corral the rods into pathways, along which electrical transport is facilitated. As these domains span the system and the morphology becomes bicontinuous rather than dispersed, the increase in conductivity of the rod-filled polymer blend is roughly 8% greater than that of the rod-filled homogeneous system. This demonstrates a clear advantage to incorporating the rods into the phase-separating system. It is anticipated that a greater disparity between the two systems would be observed if there were a more significant difference in the conductivities of the rods and polymers.

4.2.4. Summary

The selective inclusion of nanorods into the minority phase of a phase-separating polymer blend results in the emergence of complex networks of both nanorods and polymers [22]. The nanorods stretch and perturb the domains of the polymer blend, and the polymer blend confines and corrals the nanorods, producing elongated domains that are reinforced by these fillers. At a critical volume fraction of rods, the polymeric minority domains elongate to the extent that they coalesce with neighboring domains and form a continuous structure. The continuity of the minority domains and the percolation of the nanorods within these domains results in a system that is doubly percolating.

We compare the mechanical behavior of such doubly percolating composites with materials where the rods are uniformly dispersed in a homogeneous matrix. In the former case, the nanorods form a continuous backbone of stiff, reinforcing material, which spans the system and results in a significantly stiffer nanocomposite. We also consider the electrical benefits of percolating nanorod networks for systems where the nanorods are more conductive than the polymer matrix. The nanorods provide a continuous path of high conductance across the composite, thus facilitating electrical transport.

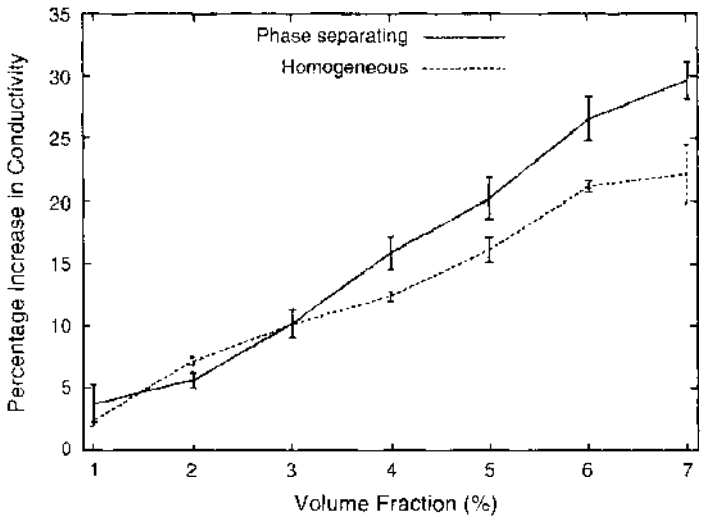


Figure 22. The percentage increase in conductivity as a function of nanorod volume fraction. Systems where the rods are corralled into percolating pathways by the phase-separating polymer blend are compared with systems where the rods evolve isotropically within a homopolymer.

The electrical and mechanical benefits of including nanorods are found to be direction dependent, with rods oriented in the tensile direction imparting greater stiffness and rods oriented with the potential difference gradient improving the conductivity. Controlling the direction in which these rods are oriented (for example, through the application of an electric field during processing) could lead to even greater improvements in macroscopic properties.

4.3. Properties of Mixtures of Nanorods and Diblock Copolymers

Having considered the behavior of the nanorods in a polymer blend, we now examine the properties of nanorods in diblock copolymers [3]. This system is of interest because the blending of rodlike particles and block copolymers can potentially yield a system where the rods effectively form "columns" that provide significant reinforcement of the soft matrix material. For example, if the rods were preferentially wetted by one block of an AB diblock, the microphase separation of the copolymer could promote the localization of rods into cylindrical cores or lamellar layers that extend throughout the material. The extensive particle-filled domains can enhance the mechanical behavior of the entire system. To design such self-reinforcing materials, it is important to be able to predict *both* the morphology of the copolymer/particle mixture *and* the macroscopic behavior of that specific composite. In this section, we use the SCF/DFT methodology for calculating the structure of rod-filled diblock copolymers and couple this method with the LSM for simulating the micromechanical behavior of composites.

We note there is considerable interest in the morphological and mechanical behavior of mixtures of block copolymers and clay platelets [74–76]. The DFT adopted in the approach presented here allows us to consider such platelet architectures and consequently to use the modified SCF/DFT to examine the self-assembly of thin platelets and diblocks. The combined SCF/DFT-LSM approach would then allow us to investigate the mechanical properties of these nanocomposites.

With respect to studies of diblock and rodlike particles, Sevink et al. [77] used a dynamic density functional theory (DDFT) to examine the morphology of a symmetric diblock copolymer melt that contains thin parallel rods, which are infinite in length. In the later model, the behavior of the diblocks is described by the same SCF terms as in our method; however, the particles are immobile and thus do not self-assemble because of the presence of the copolymers. In effect, in the Sevink et al. model, the particles act as an additional stationary surface in the system. Nonetheless, the immobile rods are found to have a significant effect on the structure of the neighboring diblocks [77]. As we describe later, this is also observed in the case of our finite-length, mobile rods.

Next, we describe the effects that varying the rod aspect ratio and the rod–block interaction energies have on the self-assembly of the hard and soft components and on the mechanical properties of the resultant hybrid system.

4.3.1. Morphological Studies

As we noted previously, we simplify the calculation (to make it computationally tractable) by constraining all the parallel particles to be oriented along one of three possible directions, namely, the x, y, or z directions. For each set of $(\sigma_x, \sigma_y, \sigma_z)$, we carry out two distinct simulations, allowing the particles to be oriented in or out of the plane of the mesophase. We subsequently compare the resultant free energies of the simulations and determine the equilibrium morphology as the one with the lowest free energy.

4.3.1.1. Selective Interactions We first consider the case where the particles have a selective interaction between the A and B blocks; namely, we focus on the case where the particles are preferentially wetted by the A chains. In this case, the χ parameter between the particles and the A block is zero, $\chi_{Ap}N = 0$, and $\chi_{Bp}N = \chi_{AB}N = 20$. The particle volume fraction is fixed at $\phi_p = 0.15$. In all the examples considered here, $\sigma = 0.1R_0$ and σ_L is systematically varied; in this way, we consider particles of fixed cross section but different length (or aspect ratio σ_L/σ). The invariant polymerization index $\bar{N} = a^6\rho_0^2N$ is set to 1000, and the fraction of A sites per chain is given by $f = 0.30$.

As a basis of comparison, we initially compared the behavior of small cubes ($\sigma = \sigma_l = 0.1R_0$) in the diblock matrix with the behavior of comparably sized spheres in the same diblocks (fixed f). The density profiles for the cube-containing systems were comparable to profiles for the diblock/sphere mixtures [3]. We then modified σ_l to simulate rodlike particles. Figure 23 shows the two-dimensional density profiles for this system when $\sigma_l = 0.6R_0$, and it clearly reveals that the system forms a lamellar morphology, where the particles are localized within the A domains.

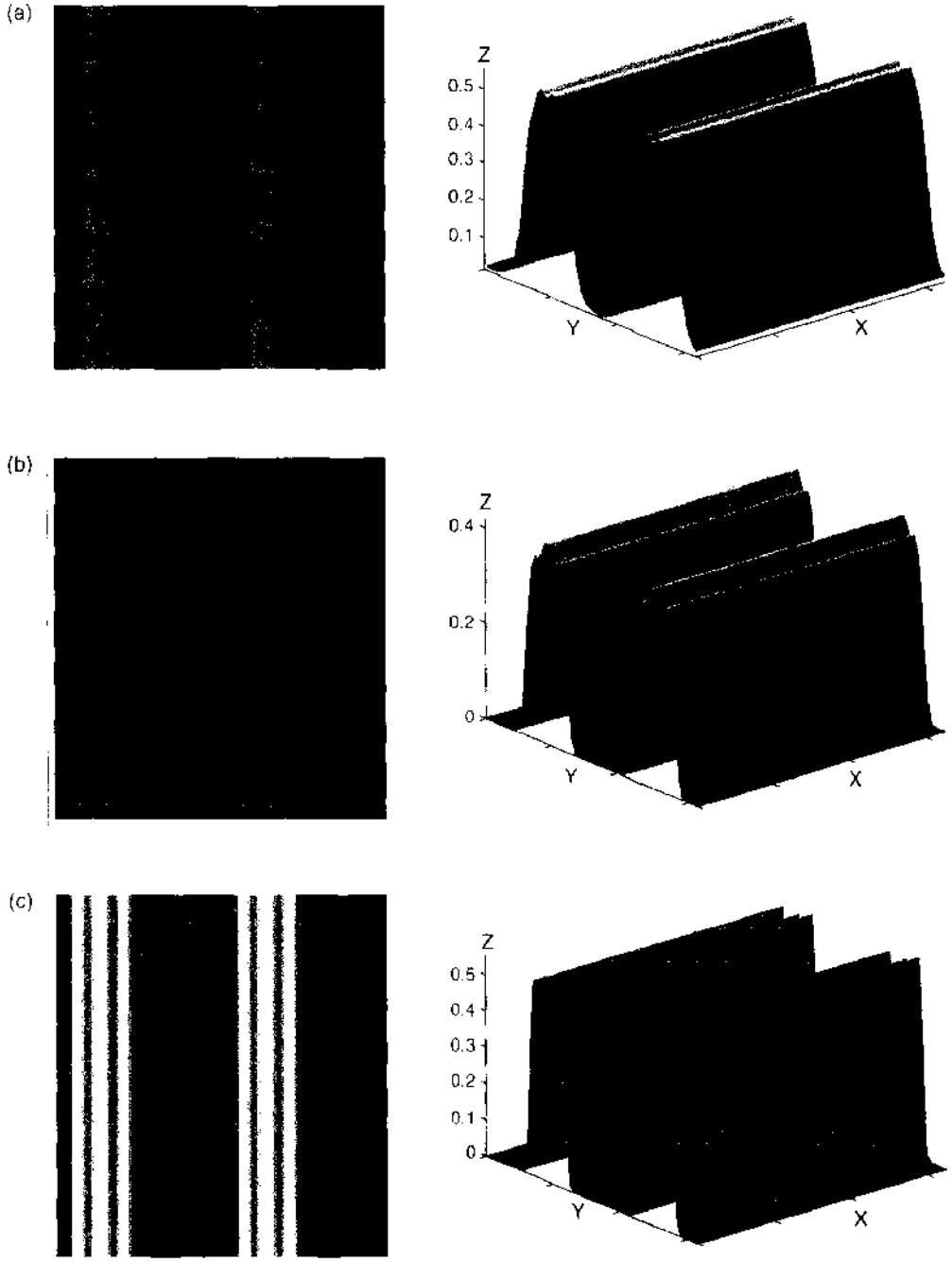


Figure 23. Lamellar phase of the composite. The particle size is given by $(\sigma, \sigma, \sigma_l) = (0.1, 0.1, 0.6)R_0$, and the particles are preferentially wetted by the A-block, where $\chi_{AB}N = 20$, $\chi_{BP}N = 20$, $\chi_{AP}N = 0$. The volume fraction of particles is 15%, and the diblock composition is characterized by $f = 0.3$. The images on the left are the two-dimensional density plots: the light regions mark the presence of a given species, and the dark regions indicate the absence of that species. The images on the right are surface plots. The distributions depicted are (a) the A-block volume fraction, (b) the particle density, and (c) the particle centers.

The details of the particle distribution are more readily visualized in the one-dimensional density profiles for the system shown in Fig. 24. The long axes of the particles are oriented perpendicular to, or coming out of, the page. Within the A domains, three peaks in the particle center distribution are clearly visible, indicating that essentially three particles lie parallel to each other to fill the A phase. Figure 25 shows similar profiles for the shortest rods, $\sigma_L = 0.2R_0$, and the longest rods, $\sigma_L = 1.2R_0$. A comparison of these density profiles shows that as the rods increase in aspect ratio (and hence size) from the $\sigma_L = 0.2R_0$ case to the $\sigma_L = 1.2R_0$, the particles become more localized in the center of the A phase. Similar effects are seen in mixtures of diblocks and spherical A-like particles when the size of the spheres is increased [8]. This phenomenon can be understood by considering both the translational entropy of the particles and the conformational entropy of the chains. The smaller particles possess a higher degree of translational entropy than the larger ones, and the chains do not have to undergo significant stretching to circumvent the smaller solids; thus, the smaller particles are distributed in a relatively uniform manner within the A domain. In mixtures containing the larger rods, the conformational entropy of the chains plays a significant role. If the larger particles were uniformly distributed in the A region, these rods would exclude a significant volume that could not be occupied by the A subchains. In effect, the conformations that could be sampled by the A blocks would be diminished, and the entropy of the system would be lowered. On the other hand, if the larger rods are localized in the region *between* the A subchains (in the center of the domain), the conformational entropy of the chains is not significantly affected. Localized in the center, the particles do lose translational entropy; however, the loss in translational entropy of the large particles is small compared with the corresponding loss in the conformational entropy of the A chains. Consequently, the large rods become localized in the center of the A phase. In practical terms, the results indicate that it is possible to control the distribution of A-like rods within the A region by tailoring the length (aspect ratio) of the particles.

4.3.1.2. Nonselective Interactions More dramatic changes in the structure of the composite can be seen if we examine the same parameters (f, $\bar{N}$, ϕ_p, σ, σ_L) as previously given but vary the polymer–particle interactions. We now focus on the case where the particles have a nonselective interaction with both the A and B blocks; here, we set $\chi_{AP}N = \chi_{BP}N = 0.02$ and maintain $\chi_{AB}N = 20$. Figure 26 shows the morphology of the system for $\sigma_L = 0.6R_0$. In contrast to the image in Fig. 23, the system now displays a cylindrical morphology. Figure 27 shows that a cylindrical morphology is also observed for $\sigma_L = 0.2R_0$, whereas a lamellar structure is observed for $\sigma_L = 1.2R_0$. Thus, increasing the aspect ratio of the particle drives a transition from a cylindrical to lamellar mesophase.

In the case of the smaller solids, the particles have significant translational entropy and are dispersed relatively uniformly in the A and B domains. (This principle is illustrated in the top plot in Fig. 28b for a slightly different f.) A small concentration is localized at the

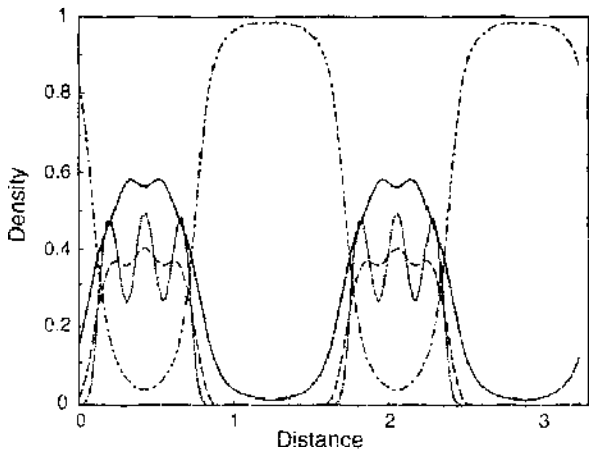


Figure 24. One-dimensional density profiles for the system in Fig. 23. The solid lines represent φ_A, the dot-dashed lines mark φ_B, the dashed lines indicate φ_P, and the dotted lines indicate μ_P, the particle center distribution.

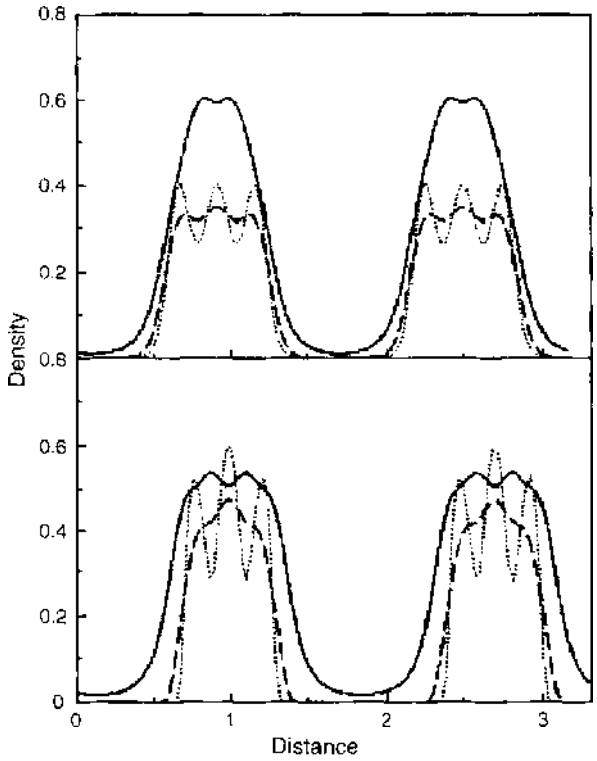


Figure 25. One-dimensional density profiles that reveal the location of particles for $\sigma_L = 0.2R_0$ in the top plot and $\sigma_L = 1.2R_0$ in the bottom plot. The parameters are the same as in Fig. 23. The solid lines represent φ_A, the dashed lines indicate φ_P and the dotted lines indicate ρ_P, the particle center distribution.

A/B interface and effectively reduces the interfacial tension between the two blocks. On the other hand, translational entropy plays a smaller role in the behavior of the larger particles; now the particles are highly localized at the A/B interfacial regions (see the bottom plot in Fig. 28b). Confined at the interface, the particles swell the polymer domains; however, because the width of the A domains is smaller than the B regions, the A phase is more strongly affected and swollen to a relatively greater degree. This selective swelling shifts the effective composition of the mixture, essentially enhancing the value of f and thereby stabilizing the lamellar phase.

Again, the results provide practical guidelines for tailoring the structure of the rod/diblock mixture. In the case of nonselective particles, an increase in the aspect ratio of the rods can be harnessed to promote transitions between the mesophases (for example, from cylindrical to lamellar at small f).

4.3.1.3. Comparison of Particle Distributions in Selective and Nonselective Cases

We shift our focus to the case of $f = 0.41$ because in this case both the selective and nonselective particles for the range of σ_L considered here form a composite with a lamellar morphology. By comparing the one-dimensional density profiles for the selective and nonselective cases for both $\sigma_L = 0.2R_0$ and $\sigma_L = 1.2R_0$, as shown in Fig. 28, we can clearly see that the particle distribution within the lamellar phase is dependent on both the aspect ratio of the solids and interaction parameters between these particles and the matrix. By contrasting the selective and nonselective cases at fixed σ_L, one can see that the distribution of particles can be manipulated by tailoring the particle–polymer interactions. Experimentally, these interactions can be varied by coating the surface of the particles with different chains [78].

As we noted in Section 1, Sevink et al. [77] used a dynamic density functional method to examine the morphology of A_8B_8 chains in the presence of fixed (immobile) rods of infinite length. Thus, in contrast to the studies presented here, the location of rods is not affected by the self-assembly of the copolymer matrix. Furthermore, only a small number of rods were considered; the number of rods was varied from one to six. Because of significant differences between the cases studied, it is not meaningful to make a quantitative comparison

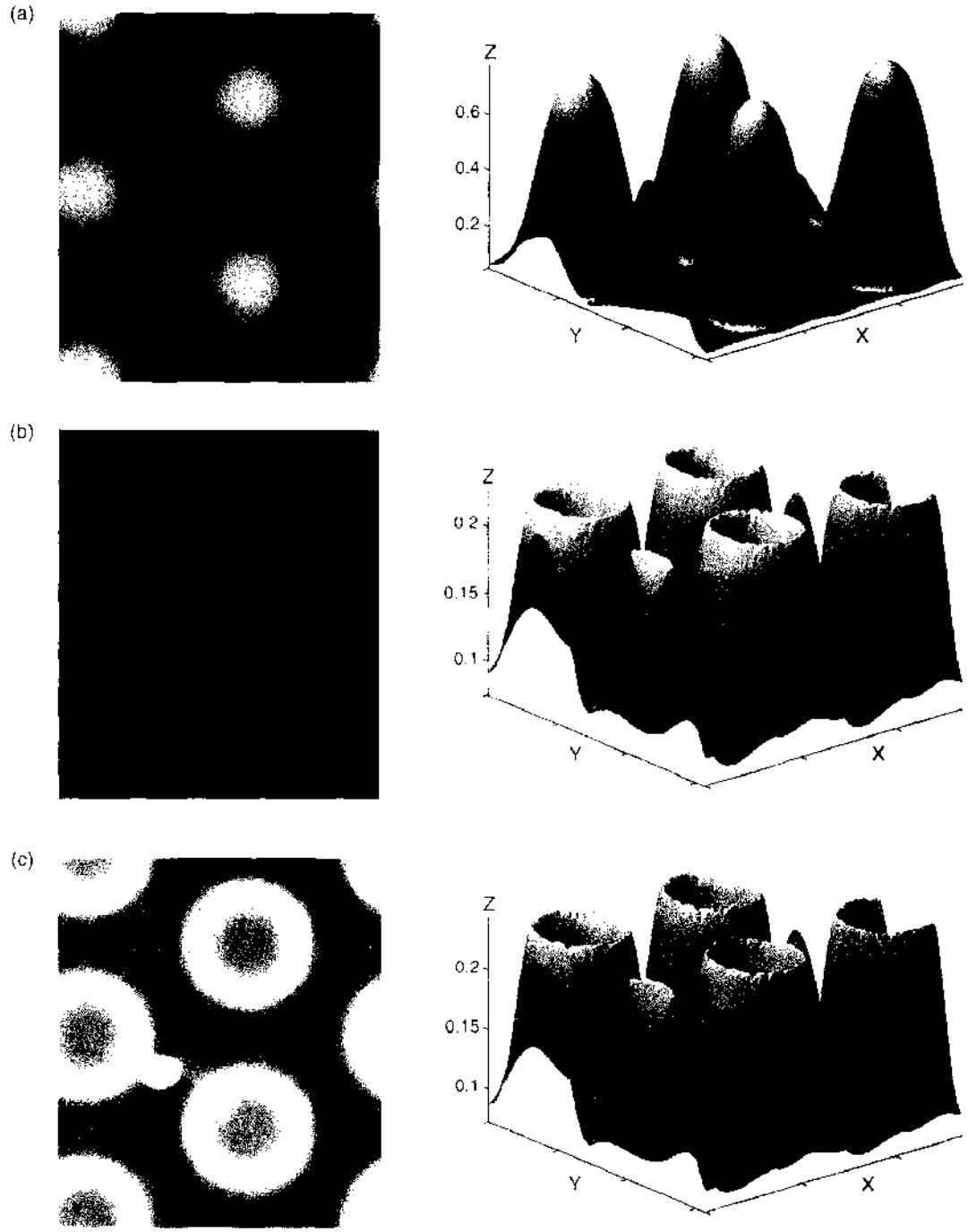


Figure 26. Cylindrical phase of the composite. The particle size is given by $(\sigma, \sigma, \sigma_f) = (0.1, 0.1, 0.6)R_0$, and the particles are nonselective, with $\chi_{AB}N = 20$, $\chi_{Af}N = \chi_{Bf}N = 0.02$. The volume fraction of particles is 15%, and the diblock composition is characterized by $f = 0.3$. The images on the left are the two-dimensional density plots; the light regions mark the presence of a given species, while the dark regions indicate the absence of that species. The images on the right are surface plots. The distributions depicted are (a) the A-block volume fraction, (b) the particle density, and (c) the particle centers.

between the respective findings. Nonetheless, it is worth noting that Sevink et al. [77] find that even the presence of a few immobile fillers can perturb the geometry of the system. In particular, the six rods cause a local bending of the lamellae in the vicinity of the rods.

Of particular interest is understanding the relative role of the interaction energies and the aspect ratio of the particles on the mechanical behavior of the composite. To obtain insight into the influence of these parameters on the macroscopic behavior, we now turn to the results of the LSM simulations on the structures described previously.

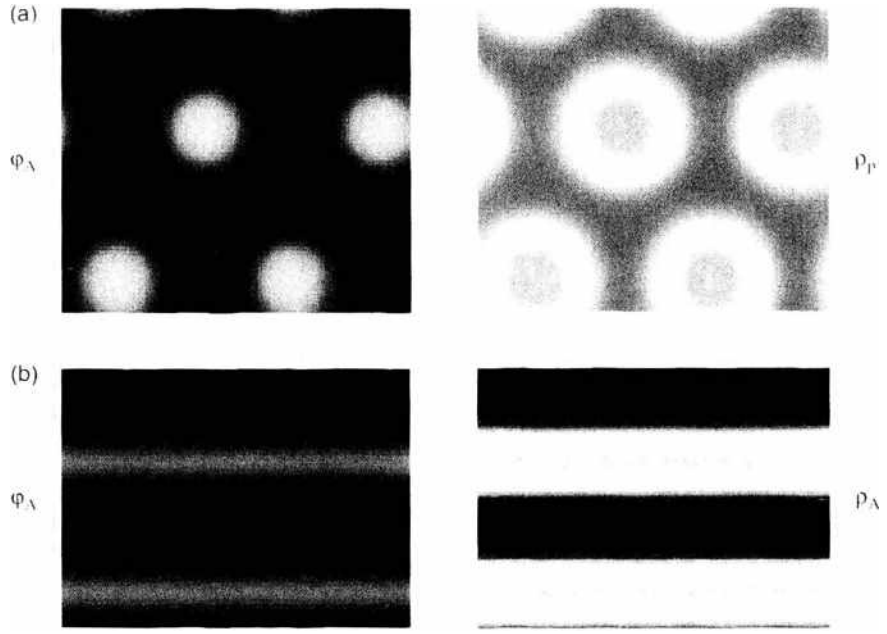


Figure 27. Two-dimensional density plots for $\sigma_f = 0.2R_0$ in (a) and $\sigma_f = 1.2R_0$ (b). The remaining parameters are the same as in Fig. 26. The system is shown to form (a) a cylindrical structure for the shorter rod and (b) a lamellar morphology when the rod length is increased.

4.3.2. Micromechanical Studies

As noted previously, the elastic properties of the springs within the LSM simulation are assigned values depending upon the phase in which they appear, as dictated by the output of the SCF/DFT method. The LSM simulation box in these simulations is 100 × 50 × 50 lattice sites in size, with the longer dimension corresponding to the direction of both the applied tensile stress and the orientation of the rods. Just as in the SCF/DFT model, all the rods in

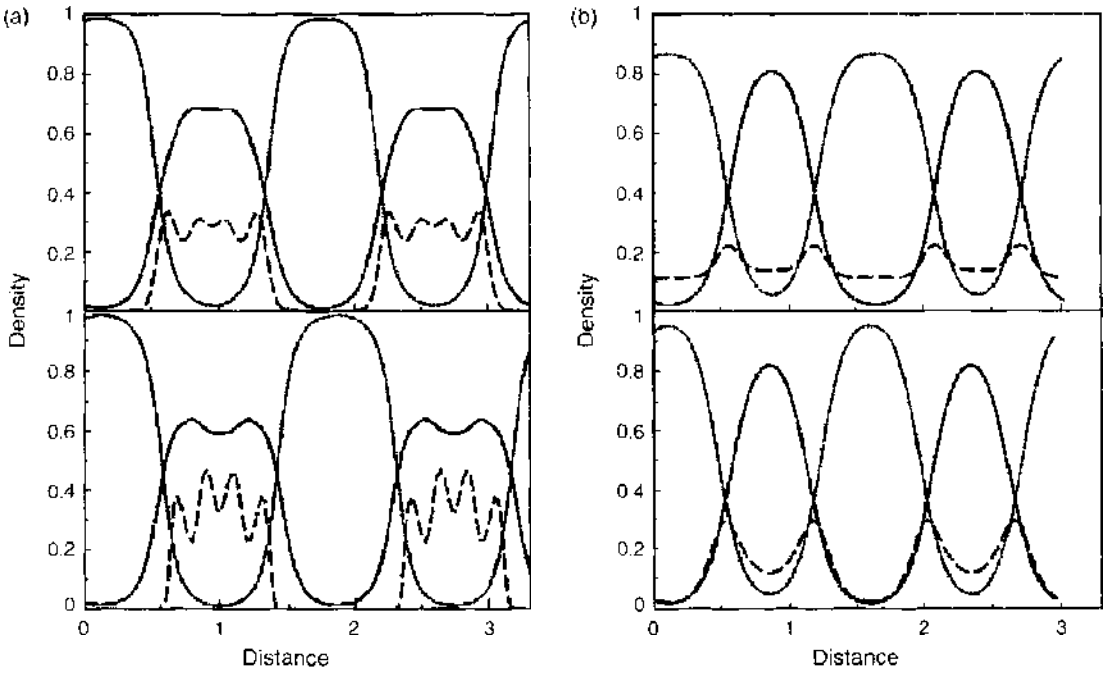


Figure 28. Comparison of the particle density profiles for selective cases in (a) and nonselective cases in (b). The top images are for $\sigma_f = 0.2R_0$ in the top plots and $\sigma_f = 1.2R_0$ in the bottom plots. Here, the diblock composition is characterized by $f = 0.41$.

the LSM are aligned in one direction and are aligned parallel to each other. To limit possible surface effects within the simulation, only the central region of $90 \times 40 \times 40$ lattice sites contains rods. The spatial dimensions are normalized such that the unit length in the LSM is equivalent to the width of the rod in the SCF/DFT model. The distribution of particle centers that is calculated in the SCF/DFT simulation is utilized in the random positioning of the rods in a LSM simulation. Recall that the SCF/DFT calculations do not involve an integer lattice, while the undeformed lattice in the LSM is initially a regular, cubic lattice; thus, we must incorporate a linear interpolation scheme to precisely determine the particle distribution in the LSM. Specifically, the probability distribution, p_i, for each node i in the LSM is linearly interpolated from the SCF/DFT values of the particle center distribution ρ_p, and subsequently normalized. In particular, the probabilities are given by

$$p_i = \frac{\rho_p(i)}{\sum_i \rho_p(j)} \tag{23}$$

where $\rho_p(i)$ is the linearly interpolated particle distribution function at node i and the summation is over all the nodes within the system. The cumulative distribution function is defined as being

$$C_i = \sum_{j=1}^{i} p_j \tag{24}$$

A random number, R, is chosen in the interval [0, 1]. A node i, which satisfies the condition $C_{i-1} < R < C_i$, is chosen to represent the center of a rod. This guarantees the selection of a node at random, while enabling the distribution of rods throughout the system to be consistent with that of the SCF/DFT simulation. These rods are not allowed to touch or overlap (a new value of R is chosen if this condition is not satisfied). Once a central node is selected, bonds in the [100] direction about this node are assigned a stiffness 100 times that of the neighboring matrix. In this manner, we capture the structure of the three-dimensional distribution of a large number of rods (up to 10,800 in the lower aspect ratio runs). We note that in this description, the rods are associated with the [100] bond, and thus have no relative width. Unlike our previous simulations [16], where the rods had a finite width, we revert to this coarse-grained description of the rodlike particles in order to incorporate a larger number of these fillers.

We now assume that the material is rapidly quenched to form an elastic solid and hence the spatial distribution of the particles and of the polymeric domains is the same as in the thermodynamically stable melt described previously. Through the LSM, we can readily obtain the stress and strain distribution throughout this composite. Recall that we plot the relative quantities $(F - F_0)/F_0$, where F is the field in question and F_0 is the response of the pure matrix material. Figure 29a shows the relative strain in the material that contains the selective particles with $\sigma_L = 0.2R_0$, and Fig. 29b shows the relative stress for the same system. Corresponding images are shown for the $\sigma_L = 1.2R_0$ case in Figs. 30a and 30b.

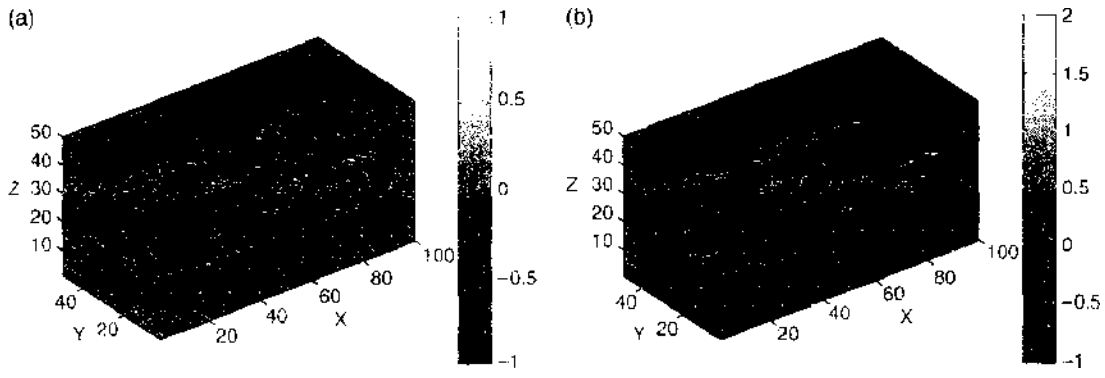


Figure 29. The relative elastic fields for a system containing selective inclusions whose aspect ratio (defined as σ_L/σ) is equal to two. Both the (a) relative normal strain fields and (b) relative normal stress fields are presented.

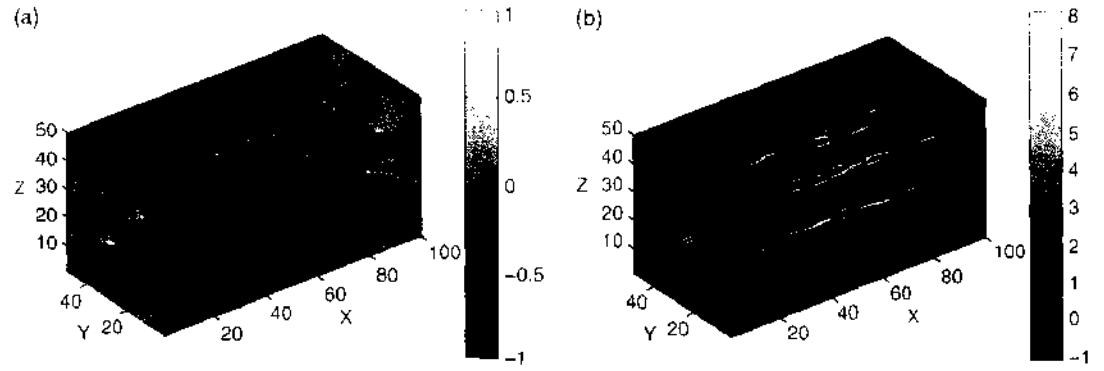


Figure 30. The relative elastic fields for a system containing selective inclusions whose aspect ratio (defined as σ_l/σ) is equal to 12. Both the (a) relative normal strain fields and (b) relative normal stress fields are presented.

(The total volume fraction of particles is held fixed in these systems at $\phi_p = 0.15$; thus, the system had a greater *number* of small particles than large particles.) Note that the LSM simulations reflect the spatial localization of the selective particles (that is, confinement in the A domains). In the case of the nonselective particles, Figs. 31 and 32 display the local elastic fields for both the short and long particles, respectively. A glance at these figures shows that the nonselective particles are more uniformly distributed within the matrix than the selective solids, capturing the spatial arrangements found in the SCF/DFT calculation.

In the following discussion, we first focus on the mechanical properties of the system that involves the selective particles, shown in Figs. 29 and 30. The most remarkable feature of these plots is the significant changes in the elastic properties as the length of the particles is increased. The long, stiff particles cannot deform as much as the softer matrix; thus, the rods reduce the overall strain in the material, especially in the vicinity of the particles. This is clearly visible in Fig. 30a, in which the matrix exhibits significant strain relaxation. Figure 29a, which contains the shorter rods, also shows strain relaxation in the bulk, but to a lesser degree than that seen for the longer rods. Furthermore, the figure shows strain concentrations, which occur at the particle edges lying perpendicular to the tensile direction. These edge effects are significantly more prevalent for the small particles (at fixed ϕ_p, there are a greater number of shorter rods than longer ones) and can ultimately lead to an increased probability of debonding between the small rods and the polymers. In both Figs. 29a and 30a, the reduced relative strain indicates that the presence of the rods enhances the stiffness of the material.

The mechanical properties of a composite also depend on the efficient transfer of stress from the matrix to the solid particles [71]. Figures 29b and 30b show the relative stress for the short and long rods, respectively. As anticipated, the stress concentrations are greater in the longer rods, and consequently the stress relaxation is greater in the matrix containing these $\sigma_l = 1.2R_0$ additives. In other words, the efficiency of stress transfer is greater in the

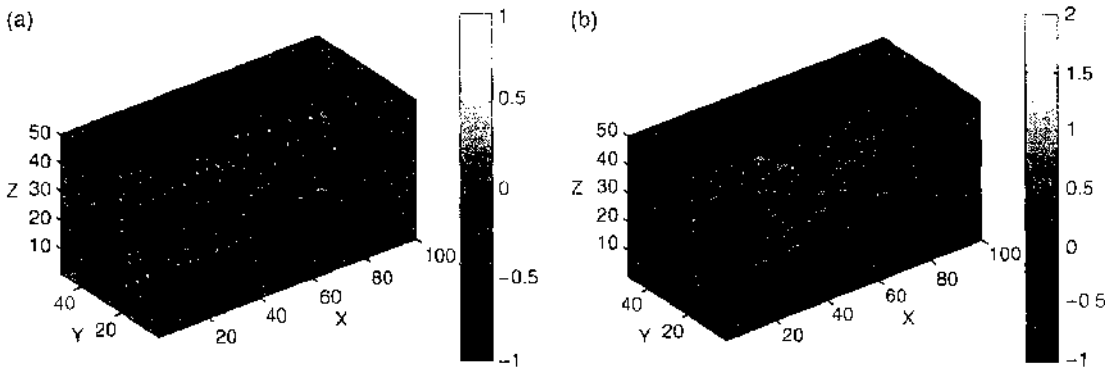


Figure 31. The relative elastic fields for a system containing nonselective inclusions whose aspect ratio (defined as σ_l/σ) is equal to two. Both the (a) relative normal strain fields and (b) relative normal stress fields are presented.

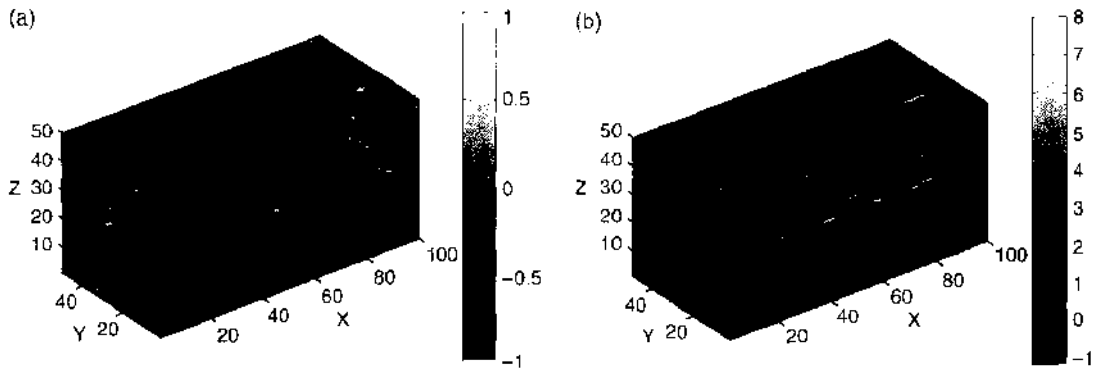


Figure 32. The relative elastic fields for a system containing nonselective inclusions whose aspect ratio (defined as σ_L/σ) is equal to 12. Both the (a) relative normal strain fields and (b) relative normal stress fields are presented.

$\sigma_L = 1.2R_0$ case than the $\sigma_L = 0.2R_0$ example. It is important to note, however, that the large stress concentrations in the longer rods imply that these inclusions will more readily fracture than the shorter species.

The same general conclusions apply to the behavior of the composites containing the short and long nonselective rods (see Figs. 31 and 32). In Fig. 33, we plot the cumulative distribution function (CDF) for the relative stress for various rod lengths in the case of the nonselective particles. The CDF indicates the percentage of the material that exhibits a relative stress less than the specified amount; the plot provides a quantitative description of the qualitative observations that can be deduced from Figs. 31b and 32b. As the length of the rods is increased, the curves in Fig. 33 shift to the left, indicating that the stress in the overall composite is decreased. The upper right hand of the curve indicates that a small fraction of the material—namely, the rods themselves—exhibits greater stress concentrations as the rod length is increased. As can be deduced from Figs. 29b and 30b, and the previous discussion, quantitatively similar plots can be drawn for the case of selective particles.

What emerges from a comparison of Figs. 29 and 30 with Figs. 31 and 32 is the fact that for the rod lengths and volume fraction considered here, there is no substantial difference between the mechanical properties of the composites that contain the selective or nonselective particles. In effect, the nanoscale differences in the distribution of particles in the two cases (at this relatively high ϕ_p of 15%) do not contribute to dramatic differences in macroscopic behavior. The overriding effect on the macroscopic properties is the length

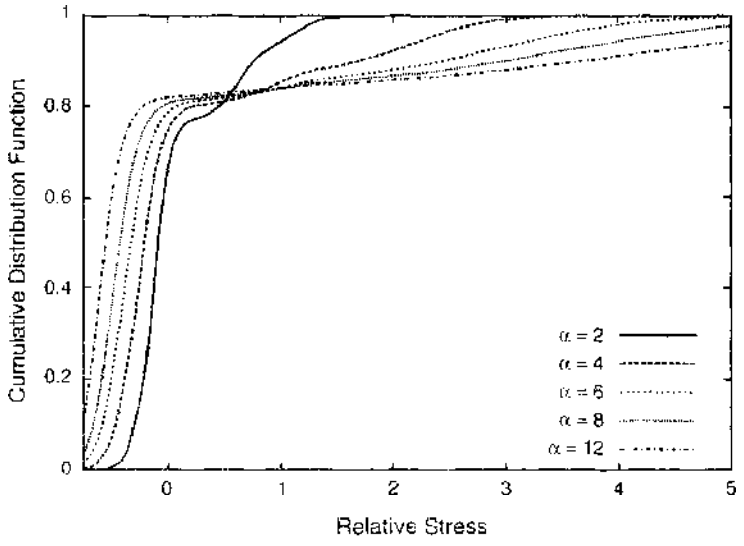


Figure 33. The cumulative distribution function of the local relative stress field in systems containing nonselective inclusions whose aspect ratio varies between 2 and 12. The higher stresses correspond to stress concentrations within the rods.

(aspect ratio) of the rods. This conclusion can be readily seen in Fig. 34, where we plot the percentage increase in the Young's modulus for the selective and nonselective cases as a function of rod length. Each point represents an average over three different LSM calculations (using the same SCF/DFT as input but different random numbers to determine the location of the particle centers). What is striking is the dramatic increase in the modulus due to the presence of the particles and how this quantity increases with the increasing aspect ratio of the rods.

What is intriguing, however, is the small (but statistically significant) difference in the values for the selective and nonselective cases at large aspect ratio, with the selective particles yielding slightly less reinforcement than the nonselective species. The plot indicates that such differences in behavior become more pronounced at higher aspect ratios. We can hypothesize that the situation is in some ways analogous to the behavior of intercalated and exfoliated polymer/clay composites. That is, as the selective rods get longer, they become more and more localized and crowded in the center of the A domain (see Fig. 28). Consequently, only small regions of the material can benefit from the reinforcing properties of these rods. In contrast, the long nonselective rods remain more uniformly distributed, even as the rod length is increased. Here, more extensive regions of the matrix are affected by the presence of the solids, and the overall stiffness of the material is enhanced. In a similar vein, at fixed volume fractions, exfoliated platelet particles provide better reinforcement than the intercalated species.

4.3.3. Summary

We determined how the self-assembly of the copolymer matrix affects the spatial distributions of the rodlike fillers and how the fillers affect the mesophase of the mixture. We then predicted the elastic properties of the resultant solid composite. In these studies, we did not consider the effects of varying the elastic properties for the different blocks, though this could be readily done by assigning different spring constants to the A and B components. In addition, we do not allow for the locally anisotropic behavior of the polymer at the domain interface as a consequence of chain stretching. However, local variations in polymeric elastic moduli are assumed to be relatively small in comparison with the differences between the polymer matrix and the rodlike particles. It is this elastic mismatch that is primarily responsible for the increase in the global Young's modulus and the reinforcement efficiency of the rodlike nanoparticles. We have accounted for the effects of varying the aspect ratio and spatial distribution of the fillers. The large increases in Young's modulus reported here as a consequence of incorporating rodlike fillers are far in excess of what might be expected because of the local variations in polymer composition and chain-stretching effects.

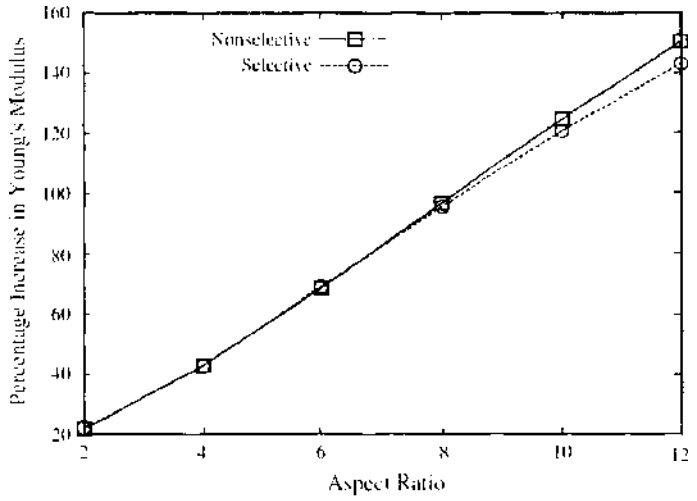


Figure 34. The percentage increase in Young's modulus as a function of particle aspect ratio for equivalent particle volume fractions. Systems containing both selective and nonselective inclusions are compared. The effects of aspect ratio are significant and the percentage increase in Young's modulus ranges from 20% to 140%.

To carry out these studies, we modified the density functional component of our previous SCF/DFT [4–10] model to incorporate solid, parallelepiped particles. In this way, we can examine the influence of rodlike fillers or platelets on the self-assembly of block copolymers. In the present study involving rodlike particles, we find that for fixed values of χ, the aspect ratio of the rods influences the spatial distribution of these fillers within the copolymer matrix. Furthermore, in the case of the nonselective particles, changes in the length of the rods can be exploited to induce phase transitions between different mesophases. Here, we observed that increasing the length of these nonselective rods drives a transition from a cylindrical to lamellar morphology.

The results of these SCF/DFT studies then served as the input to the LSM, a micromechanical simulation. We focused on a case where both the selective and nonselective particles yielded a composite with a lamellar morphology. Within this mesophase, the high aspect ratio, selective particles were largely localized between the A brushes of the A domain. On the other hand, the nonselective rods of comparable length were localized at the interface between the A and B domains. Nonetheless, the Young's moduli of the two systems were comparable. Thus, for the values of σ_f and ϕ_p considered here, these differences in the spatial distribution of the particles did not have a substantial influence on the overall mechanical properties of the composites. The dominant effect was the aspect ratio of the rods; the high aspect ratio fillers yielded a dramatic increase in the Young modulus relative to the pure diblock.

4.4. Structure and Optical Properties of Nanoparticle-Filled Diblock Copolymers

In the next study, we again utilize the SCF/DFT methodology to determine the self-assembled structure of particle-filled diblock copolymers. However, in this investigation, we focus on spherical nanoparticle fillers. Our aim is to correlate the morphology of these nanocomposites to the optical properties of the system [30]. Such correlations are vital for designing novel polymer/nanoparticle photonic crystals. In photonic crystals, the dielectric constant of the material is a spatially periodic function. In such systems, forbidden frequency bands can exist for incident electromagnetic waves. These photonic band gaps are caused by the interference of waves as they are reflected and scattered by regularly spaced regions of varying dielectric properties.

Block copolymers have the potential to form promising photonic crystals because they self-assemble into spatially periodic materials [36]. Because of limitations on the molecular weight of the polymer species (and therefore the domain sizes) and the dielectric contrast between different blocks, the range of forbidden frequencies and the size of the photonic band gaps are severely restricted. Thomas et al. [55, 79] have successfully swollen domain sizes through the addition of homopolymers in order to shift photonic band gaps from the ultraviolet to the visible region of the spectrum. It was also shown that the width of the band gap could be increased through the selective sequestering of inorganic nanocrystals within one of the phases of the microphase-separated diblock copolymers [79–81].

Our recent computational investigations into the self-assembly of nanoparticles and diblock copolymers have provided significant insight into the possible morphologies of such structures [1–10]. Here, we extend this work by combining the SCF/DFT morphological studies with a computational model for the propagation of light and thereby determine the optical properties of the self-assembled composite [30]. In the present study, we assume that the diblock copolymer structures are perfectly aligned, periodic lamellae. Such regular structures can more readily be attained in thin films [80]. We find that even for this simple 1D photonic band gap material, a rich variety of phenomena can be observed.

4.4.1. Relation of Morphology to Optical Properties

The initial case that we consider involves a symmetric diblock copolymer without the addition of the fillers. The volume fraction profiles corresponding to the A and B phases are presented with respect to distance in the x-direction in Fig. 35a. As is typical in diblock copolymers, the dielectric contrast between the two phases is relatively small. If we assume that the transition

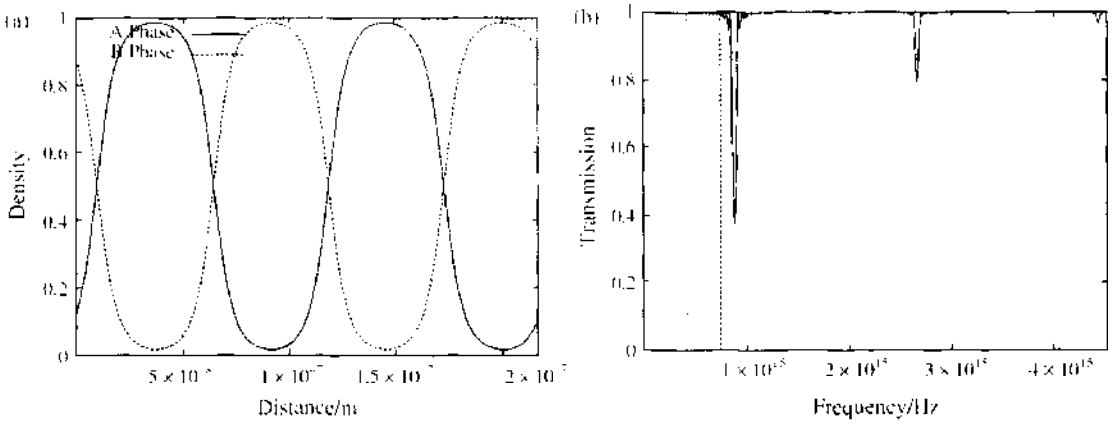


Figure 35. Symmetric unfilled diblock copolymer; $\chi_{AB}N = 20$; (a) volume fraction profiles of constituents (A blocks and B blocks of the diblock copolymer) and (b) transmittance spectra (a transmission of 1 corresponds to the total transmission of light, and a transmission of 0 corresponds to the total reflection of light).

between the A and B phases is instantaneous (that is, a sharp change in the dielectric contrast between A and B), we can use a simple theory that takes into consideration the optical paths of the domains to estimate the frequencies at which reflection will occur [82] and, consequently, transmission will be reduced. The mth order frequency is given by

$$f_m = cm/(2(d_A n_A + d_B n_B) \tag{25}$$

where c is the speed of light, d is the width of the domain, and n is the refractive index of the domain; the subscripts A and B correspond to the A and B phases, respectively. For this case, the principal frequency ($m = 1$) is estimated from theory to be 0.89×10^{15} Hz, which corresponds to a wavelength of 337 nm (ultraviolet).

In our calculations, we use the volume fraction profiles in Fig. 35a to assign the appropriate values of the dielectric constants in the FDTD simulation; consequently, the variation in the dielectric constant between the A and B phases is relatively smooth. Nonetheless, in our studies $\chi_{AB}N = 20$, and thus there is only a relatively small overlap between these domains.

Figure 35b depicts a FDTD calculation for the transmission of light as a function of the frequency of the incident light. In this plot, a value of 1 in the vertical axis corresponds to 100% transmission. We do, in fact, observe a decrease in transmission in the region corresponding to the principal frequency of reflection predicted from Eq. (25).

Also of interest is the absence of even orders of reflection. The relative intensities of the various orders depend upon the ratio between the optical thicknesses of the two domains (where $n_i d_i$ is the optical thickness of species $i = A, B$). The variable

$$g = n_A d_A/(n_A d_A + n_B d_B) \tag{26}$$

can be used to estimate these relative intensities [82]. A value of (1/2) corresponds to equivalent optical distances, as is the case for a symmetric diblock with little dielectric contrast, and this results in the suppression of all even orders. This can be seen in Fig. 35b as the high degree of transmittance in regions that correspond to even multiples of 0.89×10^{15}.

Because of the small dielectric contrast in this system, the reflectance is small, and the frequency band over which this reflectance occurs is also small. To increase the magnitude of the reflectance and the size of the band gap, the dielectric contrast between the domains must be increased. We next consider a melt of symmetric diblock copolymers that contains A-like particles. In particular, we set $\chi_{AP} = 0$ and $\chi_{BP}N = 20$. Here, and in all subsequent cases, we fix $\chi_{AB}N = 20$. In addition, the radius of the particle, R_p, is set to $0.15R_0$, where R_0 is the root mean-squared end-to-end distance of the diblock. The total volume fraction of particles is fixed at 10% ($\phi_p = 0.1$).

As can be seen from the volume fraction profiles in Fig. 36a, these fillers become localized within the A domains, which are swollen by the particles. The transmission for this system

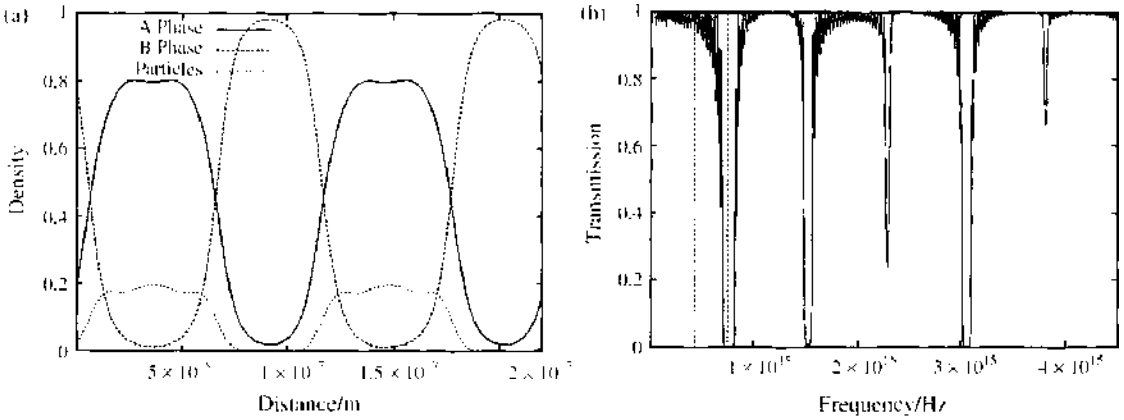


Figure 36. Symmetric filled diblock copolymer with particles localized in A phase: $\chi_{AB}N = 20$, $\chi_{BP}N = 20$, $\chi_{AP}N = 0$: (a) volume fraction profiles of constituents (A blocks and B blocks of the diblock copolymer and nanoparticles) and (b) transmittance spectra (a transmission of 1 corresponds to the total transmission of light and a transmission of 0 corresponds to the total reflection of light).

is presented in Fig. 36b. This case may be distinguished from the previous unfilled case in a number of ways. First, the reflectance of light at certain frequencies is 100% (that is, transmission is 0%), as opposed to roughly 60% in the previous system. Second, the width of the band gaps is larger, especially those corresponding to the higher orders of reflection. Third, the principal frequency of reflectance now overlaps the optical range (marked by dashed vertical lines in Fig. 36b), which corresponds to light in the visible spectrum. The dielectric constant associated with the A phase is now increased because of the presence of the particles. This behavior, combined with the swelling of the A phase, gives rise to a larger optical path, $d_A n_A$. From Eq. (26), it can be seen that this has the effect of decreasing the frequency at which principal reflectance occurs, as is the case for the simulation results in Fig. 36b.

The following and more interesting case involves a symmetric diblock copolymer melt that contains nonselective particles. Specifically, we set $\chi_{AP}N = \chi_{BP}N = 20$. Since $\chi_{AB}N$ is also equal to 20, all of the species are equally incompatible. The volume fraction profile for this χ_{AB} system is given in Fig. 37a. As can be seen, the sizes of the A and B phases are comparable, and the particles are confined at the A/B boundary. In such a mutually incompatible system ($\chi_{AB} = \chi_{AP} = \chi_{BP}$), the minority component is expelled to the interface [83] primarily because it is energetically more favorable for the A component to be surrounded by A and, similarly, for the B component to be immersed in B. At the A/B interface, the

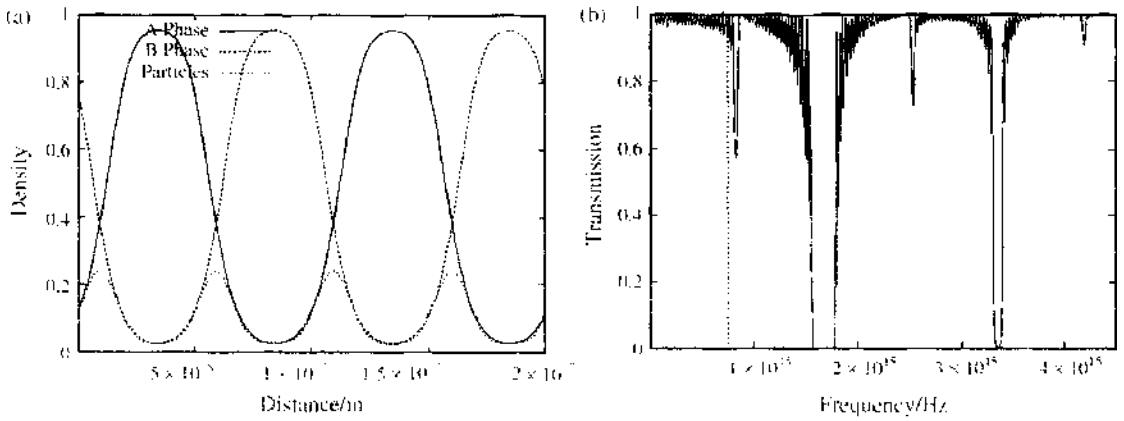


Figure 37. Symmetric filled diblock copolymer with particles localized at interface: $\chi_{AB}N = 20$, $\chi_{AP}N = \chi_{BP}N = 0.02$: (a) volume fraction profiles of constituents (A blocks and B blocks of the diblock copolymer and nanoparticles) and (b) transmittance spectra (a transmission of 1 corresponds to the total transmission of light, and a transmission of 0 corresponds to the total reflection of light).

particles increase the dielectric heterogeneity of the structure. The optical consequences of this morphology can be seen in the transmittance in Fig. 37b. The gap associated with the principal frequency corresponding to the periodic structure of the diblock copolymer is narrow, and a reflectance of only 40% is observed. The more defined and much wider band gap occurs where the second order of reflection is expected to occur. This is because the relative differences between the A and B phases are small, and the principal frequency is now associated with the polymer-particle layers. The principal dielectric contrast occurs between the particles at the interface and the unfilled polymer domains. Hence, the periodicity of high dielectric contrast is now roughly half of that associated with the diblock copolymer structure. The A and B domains are slightly different, however, and small decreases in the transmittance are still observed at frequencies associated with the diblock copolymer periodicity.

We now consider the case where the diblock copolymer is no longer symmetric but the system still displays a lamellar morphology. Figure 38a depicts the volume fraction profiles of a 37:63 diblock copolymer with the A-like particles embedded in the A phase. In this case, $\chi_{AP}N = 0$ and $\chi_{BP}N = 20$. Despite the swelling of the minority phase by the confinement of the particles, the B phase is still thicker. The transmission for this system is presented in Fig. 38b. The presence of the particles increases the dielectric contrast within the system, and, as in the filled systems above, wide band gaps are present with reflectances of 100%. The principal frequency of reflectance is clearly associated with the dielectric contrast between the filled A phase and the unfilled B phase. As in Fig. 36b, the principal frequency band gap crosses from the violet to the ultraviolet regions of the spectrum (either side of the dashed line on the right). An interesting consequence of increasing the dielectric constant within the minority phase is that the optical path of the A phase is also increased. This results in optical paths in both phases being comparable, and, like the unfilled symmetric system, the even orders of reflectance are suppressed.

In the final case, we investigate the inclusion of nonselective fillers at the interface of a nonsymmetric (37:63) AB diblock copolymer. Here, $\chi_{AP}N = \chi_{BP}N = 20$. The volume fraction profile associated with this system is presented in Fig. 39a. Although the particles are clearly localized in the interfacial regions, there is also a small volume fraction of these fillers in the minority A domain. Because the B block is longer, the overall contact energy would be greater if the particles were to leak into the B rather than the A phase. The effect that this has on the optical properties is revealed in Fig. 39b, which depicts the transmittance for this system. Because of particles in A, the dielectric contrast between the A and B phases is substantial, and therefore the principal band gap is associated with the periodicity of the diblock copolymer and not the periodicity of the polymer-interface regions, as was the case in the equivalent symmetric system (see Fig. 37). It is also worth noting that the principal band gap occurs outside the visible range of the spectrum. An interesting consequence of the

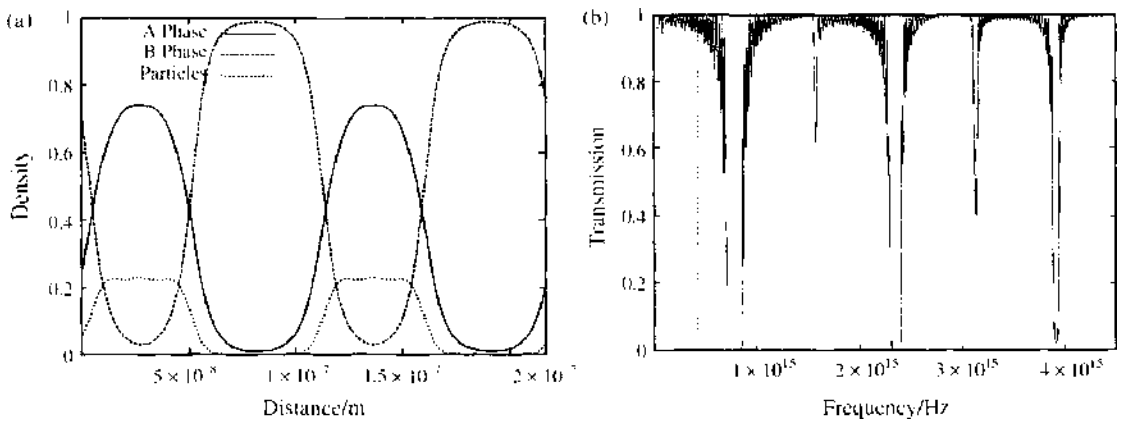


Figure 38. Asymmetric filled diblock copolymer (37:63) with particles localized in A phase: $\chi_{AB}N = 20$, $\chi_{BP}N = 20$, $\chi_{AP}N = 0$: (a) volume fraction profiles of constituents (A blocks and B blocks of the diblock copolymer and nanoparticles) and (b) transmittance spectra (a transmission of 1 corresponds to the total transmission of light and a transmission of 0 corresponds to the total reflection of light).

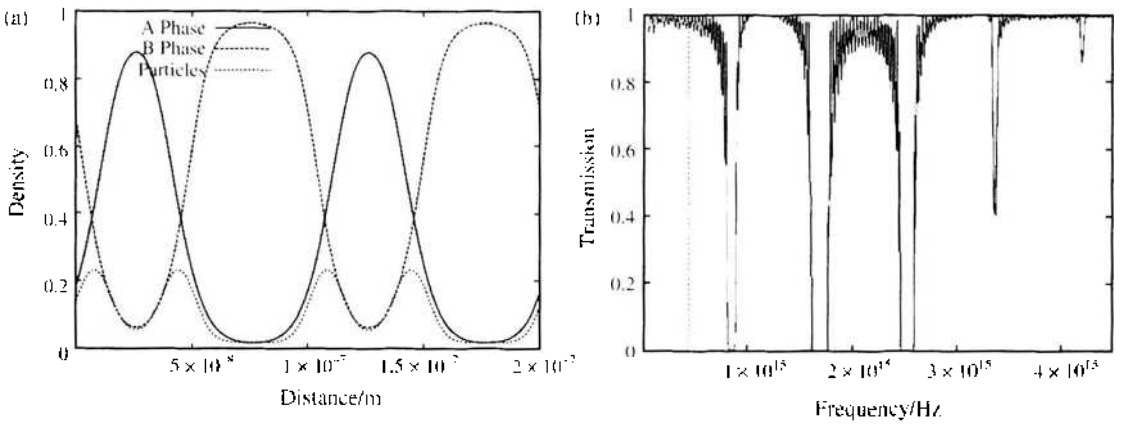


Figure 39. Asymmetric filled diblock copolymer (37:63) with particles localized at interface: $\chi_{AB}N = 20$, $\chi_{AP}N = \chi_{BP}N = 0.02$: (a) volume fraction profiles of constituents (A blocks and B blocks of the diblock copolymer and nanoparticles) and (b) transmittance spectra (a transmission of 1 corresponds to the total transmission of light, and a transmission of 0 corresponds to the total reflection of light).

clustering of particles at the interface is a decrease in the domain size. (A greater particle volume fraction at the A/B boundary results in a reduction of diblock junction points across this interface. This, in turn, results in a lower number of chains within a given domain and hence a reduction in domain size.) The decrease in domain size has the effect of increasing the principal frequency of reflection into the ultraviolet range of the spectrum.

4.4.2. Summary

The combination of techniques described here allowed us to determine both the morphology and the optical properties of filled diblock copolymer systems. We found that the addition of particles result in more defined band gaps, with 100% reflectance and a wider frequency gap. In addition, the increase in optical distances results in a decrease in the principal frequencies of reflectance. Furthermore, by combining the SCF/DFT and FDTD methods, we can see that changes in the chemical nature of the particles (selective versus nonselective) result in different spatial distributions of the particles and, hence, significantly different optical properties. In effect, this hybrid approach allows researchers to determine how choices made in the components (polymers, particles) affect the macroscopic performance of the material.

The frequencies at which light was reflected were within the violet range of the visible spectrum for certain systems containing particles. In the current simulations, the size of one lamellar domain within the pure symmetric diblock system was assumed to correspond to 50 nm. Note, however, that domain sizes up to 100 nm are possible [80], and therefore the values of the frequencies could essentially be halved. This would result in band gaps in the infrared and near-infrared ranges of the spectrum, which would be useful in the telecommunications industry.

5. CONCLUSIONS

A specific focus of our research has been to develop fundamental relationships among the following: (1) characteristics of the components in filled blends and copolymers, (2) microstructure of the mixture, and (3) macroscopic properties of the material. To establish correlations between items (1) and (2), we took advantage of our CH/BD and SCF/DFT models. Through these models, we can readily assess how variations in the features of the particles and polymers affect the morphology of the complex mixtures. Effects of temperature and pressure can be examined with both our CH/BD and SCF/DFT models, while the effects of imposed flows can readily be investigated through the CH/BD method. Thus, we can determine how processing conditions influence the final microstructure of the system.

To develop relationships among items (1) through (3), we first combined our morphological models with the LSM, which allows us to simulate the deformation of the mixtures and

determine the macroscopic material response. In other words, the output from the CH/BD or SCF/DFT models served as the input to the LSM. In subsequent studies, we combined the output of morphological studies with FDM to obtain the electrical conductivity of the material and with the FDTD to obtain the optical properties of the nanocomposite.

By integrating the morphological and mechanical, electrical, or optical models, we can isolate how specific modifications in the geometry or properties of the components affect the macroscopic behavior. Thus, we can establish how choices made in the components affect the ultimate performance of the system. Finally, we can explore vast regions of the design space to rapidly assess promising avenues from unprofitable excursions. Such investigations would be far more challenging and costly through experimental explorations. Thus, these studies can result in the efficient development of materials with optimal properties for specific applications.

In future studies, we will extend our theories to explicitly include the presence of chains or "hairs" that are grafted to the surface of the nanoparticles. Particles used in experimental studies are commonly coated by such hairs. In formulating a theory for coated particles, we recently modified the SCF/DFT formalism to model particles that contain just one anchored hair [84], and already with this extension, we observed unique behavior in the self-assembly of the species. In particular, we investigated the self-assembly of amphiphilic molecules that are composed of a hard, spherical, nanoparticle "head" and an attached polymeric "tail." We considered the case where the tail is A-like and the incompatible head is B-like, as well as the situation where an AB diblock is tethered to a B-like head. Because of the steric interactions between the solid head groups, the equilibrium morphology of a melt of the AB tadpoles was found to be significantly different from that of an AB diblock melt having the same effective composition. Similar observations were found through molecular dynamics studies on nanoparticles that contained short oligomeric tails [85]. In addition, we found that the ABB species (an AB diblock tethered to a B head) organize in a way that is distinct from the AB tadpoles (with a pure A chain tethered to a B head).

We also observed that the single-tailed tadpoles display distinct interfacial activity when blended with diblock copolymers. If the heads were covered with multiple, uniformly distributed hairs, the species would preferentially localize in the hair-compatible phase. However, with a single hair, the tadpole behaves like a surfactant, with a head localized in one phase and the tail in the other.

With the morphological output from our studies on the self-assembly of the tadpoles and tadpole/diblock mixture, an intriguing question is how the mechanical behavior of these systems would differ from that of a mixture of uncoated nanoparticles and diblocks. The physical attachment of the chains to the nanoparticles could contribute to enhancing the mechanical properties of the nanocomposites. Such studies will be carried out in the future by coupling the results of these SCF/DFT calculations to the LSM simulations. The findings can provide guidelines for creating novel nanostructured composites with improved structural integrity.

By coupling these SCF/DFT studies with the FDTD calculations, we can also determine if a linear copolymer that is anchored to a solid nanoparticle with a specific, desired index of refraction would display novel optical properties. These intriguing questions will form the basis of future work.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Army Research Office, the Department of Energy, and the National Science Foundation.

REFERENCES

1. J. Y. Lee, Z. Shou, and A. C. Balazs, *Macromolecules* 36, 7730 (2003).
2. J. Y. Lee, Z. Shou, and A. C. Balazs, *Phys. Rev. Letts.* 91, 136103 (2003).
3. Z. Shou, G. A. Buxton, and A. C. Balazs, *Composite Interfaces* 10, 343 (2003).
4. A. C. Balazs, *Current Opinion in Solid State and Materials Science* 7, 27 (2003).
5. J. Y. Lee, R. B. Thompson, D. Jasnow, and A. C. Balazs, *J. Chem. Soc., Faraday Discussions* 123, 121 (2003).

6. J. Y. Lee, R. B. Thompson, D. Jasnow, and A. C. Balazs, *Phys. Rev. Letts.* 89, 155503 (2002).
7. R. B. Thompson, J. Y. Lee, D. Jasnow, and A. C. Balazs, *Phys. Rev.* E66, 031801 (2002).
8. J. Y. Lee, R. B. Thompson, D. Jasnow, and A. C. Balazs, *Macromolecules* 35, 4855 (2002).
9. R. B. Thompson, V. V. Ginzburg, M. Matsen, and A. C. Balazs, *Macromolecules* 35, 1060 (2002).
10. R. B. Thompson, V. V. Ginzburg, M. Matsen, and A. C. Balazs, *Science* 292, 2469 (2001).
11. M. W. Matsen and F. S. Bates, *Macromolecules* 29, 1091 (1996) and references therein.
12. P. Tarazona, *Mol. Phys.* 52, 81 (1984).
13. G. J. Vroege and H. N. W. Lekkerkerker, *Rep. Prog. Phys.* 55, 1241 (1992) and references therein.
14. M. R. Bockstaller, Y. Lapetnikov, S. Margel, and E. L. Thomas, *JACS* 125, 5276 (2003).
15. S. W. Yeh, K. H. Wei, Y. S. Sun, U. S. Jeng, and K. S. Liang, *Macromolecules* 36, 7903 (2003).
16. G. A. Buxton and A. C. Balazs, *Molecular Simulation* 80, 249 (2004).
17. M. H. Cheng, C. Yeung, V. V. Ginzburg, and A. C. Balazs, *J. Chem. Phys.* 118, 9044 (2003).
18. G. A. Buxton and A. C. Balazs, *Phys. Rev.* E67, 031802 (2003).
19. V. V. Ginzburg, F. Qiu, and A. C. Balazs, *Polymer* 43, 461 (2002).
20. V. V. Ginzburg, C. Gibbons, F. Qiu, G. Peng, and A. C. Balazs, *Macromolecules* 33, 6140 (2000).
21. A. C. Balazs, *Current Opinion in Colloid and Interface Science* 4, 442 (2000).
22. G. Peng, F. Qiu, V. V. Ginzburg, D. Jasnow, and A. C. Balazs, *Science* 288, 1802 (2000).
23. A. C. Balazs, V. V. Ginzburg, F. Qiu, G. Peng, and D. Jasnow, *J. Phys. Chem.* B104, 3411 (2000).
24. V. V. Ginzburg, G. Peng, F. Qiu, D. Jasnow, and A. C. Balazs, *Phys. Rev.* E60, 4352 (1999).
25. V. V. Ginzburg, F. Qiu, M. Paniconi, G. Peng, D. Jasnow, and A. C. Balazs, *Phys. Rev. Letts.* 82, 4026 (1999).
26. K. Yurekli, A. Karim, E. J. Amis, and R. Krishnamoorti, *Macromolecules* 36, 7256 (2003).
27. G. A. Buxton, C. M. Care, and D. J. Cleaver, *Model Simul. Mater. Sc.* 9, 485 (2001) and references therein.
28. P. D. Beale and P. M. Duxbury, *Phys. Rev.* B37, 2785 (1988).
29. S. Kirkpatrick, *Phys. Rev. Letts.* 27, 1722 (1971).
30. G. A. Buxton, J. Y. Lee, and A. C. Balazs, *Macromolecules* 36, 9631 (2003).
31. S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, *Phys. Rev.* B54, 11245 (1996).
32. S. A. Palkar, N. P. Ryde, M. R. Schure, N. Gupta, and J. B. Cole, *Langmuir* 14, 3484 (1998).
33. J. T. Krug II, E. J. Sanchez, and X. S. Xie, *J. Chem. Phys.* 116, 10895 (2002).
34. G. A. Buxton and A. C. Balazs, *Phys. Rev.* B 69, 054101 (2004).
35. T. Ohta, H. Nozaki, and M. Doi, *J. Chem. Phys.* 93, 2664 (1990).
36. F. S. Bates and G. H. Fredrickson, *Phys. Today* 52, 32 (1999).
37. J. A. Cuesta, *Phys. Rev. Letts.* 76, 3742 (1996).
38. J. A. Cuesta and Y. Martinez-Raton, *J. Chem. Phys.* 107, 6379 (1997).
39. J. A. Cuesta and Y. Martinez-Raton, *Phys. Rev. Letts.* 78, 3681 (1997).
40. F. Drolet and G. H. Fredrickson, *Phys. Rev. Letts.* 83, 4317 (1999).
41. Y. Bohbot-Raviv and Z.-G. Wang, *Phys. Rev. Letts.* 85, 3428 (2000).
42. J. W. Cahn and J. E. Hilliard, *J. Chem. Phys.* 28, 258 (1958).
43. J. W. Cahn, *J. Chem. Phys.* 42, 93 (1965).
44. Y. Oono and M. Bahiana, *Phys. Rev. Letts.* 61, 1109 (1988).
45. F. Qiu, V. V. Ginzburg, M. Paniconi, G. Peng, D. Jasnow, and A. C. Balazs, *Langmuir* 15, 4952 (1999).
46. L. Monette and M. P. Anderson, *Model. Simul. Mater. Sci. Eng.* 2, 53 (1994).
47. G. A. Buxton and A. C. Balazs, *J. Chem. Phys.* 117, 7649 (2002).
48. P. D. Beale and D. J. Srolovitz, *Phys. Rev.* B37, 5500 (1988).
49. A. Politi and M. Zei, *Phys. Rev.* E63, 6107 (2001).
50. B. Chopard and P. O. Luthi, *Theoretical Comp. Sci.* 217, 115 (1999).
51. J. C. Maxwell, *Phil. Trans. Roy. Soc. Lond.* 155, 459 (1865).
52. K. S. Yee, *IEEE Trans. Antennas Prop.* AP-14, 302 (1966).
53. J. P. Berenger, *J. Comput. Physics* 114, 185 (1994).
54. J. P. Berenger, *J. Compat. Physics* 127, 363 (1996).
55. A. M. Urbas, R. Sharp, Y. Fink, E. L. Thomas, M. Xenidou, and L. J. Fetters, *Adv. Mater.* 12, 812 (2000).
56. D. R. Lide (ed.) "CRC Handbook of Chemisty and Physics," 83rd ed. CRC, New York, 2003.
57. G. A. Buxton and A. C. Balazs, *Interface Science* 11, 175 (2003).
58. N. G. McCrum, C. P. Buckley, and C. B. Bucknall, "Principles of Polymer Engineering," Oxford University Press, Oxford, 1997.
59. M. Jiang, K. Alzebdeh, I. Jasiuk, and M. Ostoja-Starzewski, *Acta Mech.* 148, 63 (2001).
60. W. U. Huynh, J. J. Dittmar, and A. P. Alivisatos, *Science* 295, 2425 (2002).
61. C. A. Mitchell, J. L. Bahr, S. Arepalli, J. M. Tour, and R. Krishnamoorti, *Macromolecules* 35, 8825 (2002).
62. A. S. Carrara and F. J. McGarry, *J. Compos. Mater.* 2, 222 (1968).
63. S. R. Nutt and A. Needleman, *Scripta Metall.* 21, 705 (1987).
64. T. Christman, A. Needleman, and S. Suresh, *Acta Metall.* 37, 3029 (1989).
65. A. Levy and J. M. Papazian, *Metall. Trans.* A21, 411 (1990).
66. V. Tvergaard, *Acta Metall. Mater.* 38, 185 (1990).
67. A. A. Gusev, *Macromolecules* 34, 3081 (2001).
68. Y. Termonia, *J. Mater. Sci.* 22, 504 (1987).
69. Y. Termonia, *J. Mater. Sci.* 22, 1733 (1987).
70. Y. Termonia, *J. Mater. Sci.* 25, 4644 (1990).
71. L. Monette, M. P. Anderson, S. Ling, and G. S. Grest, *J. Mater. Sci.* 27, 4393 (1992).

72. L. Monette, M. P. Anderson, and G. S. Grest, *J. Mater. Sci.* 28, 79 (1993).
73. M. Ostoja-Starzewski, *Mech. Mater.* 31, 883 (1999).
74. J. Groenewold and G. Fredrickson, *Eur. Phys. J.* E5, 171 (2001).
75. J. X. Ren, A. S. Silva, and R. Krishnamoorti, *Macromolecules* 33, 3739 (2000).
76. Y.-H. Ha and E. L. Thomas, *Macromolecules* 35, 4419 (2002) and references therein.
77. G. J. A. Sevink, A. V. Zvelindovsky, B. A. C. van Vlimmeren, N. M. Maurits, and J. G. E. M. Fraaije, *J. Chem. Phys.* 110, 2250 (1999).
78. K. Tsutsumi, Y. Funaki, Y. Hirokawa, and T. Hashimoto, *Langmuir* 15, 5200 (1999).
79. A. C. Edrington, A. M. Urbas, P. DeRege, C. X. Chen, T. W. Swager, N. Hadjichristidis, M. Xenidou, L. J. Fetters, J. D. Joannopoulos, Y. Fink, and E. L. Thomas, *Adv. Mater.* 13, 421 (2001).
80. Y. Fink, A. M. Urbas, M. G. Bawendi, J. D. Joannopoulos, and E. L. Thomas, *J. Lightwave Tech.* 17, 1963 (1999).
81. M. Boekstaller, R. Kolb, and E. L. Thomas, *Adv. Mater.* 13, 1783 (2001).
82. T. Alfrey, Jr., E. F. Gurnee, and W. J. Schrenk, *Poly. Eng. Sci.* 9, 400 (1969).
83. C. Huang and M. Olvera de la Cruz, *Phys. Rev.* E53, 812 (1996).
84. J. Y. Lee, R. B. Thompson, A. C. Balazs, and R. Hill, *Macromolecules* 37, 3536 (2004).
85. Z. Zheng, A. Mark, M. A. Horsch, M. H. Lamm, and S. C. Glotzer, *Nano Letters* 3, 1341 (2003).

CHAPTER 4

Monte Carlo Simulations and Self-Consistent Field Theory for Thin Polymer Films

Marcus Müller

University of Wisconsin–Madison, Madison, Wisconsin, USA and Institut für Physik, WA 331, Johannes Gutenberg Universität, Mainz, Germany

CONTENTS

1. Introduction 152
2. Coarse-Grained Models 153
3. Monte Carlo Simulation and Self-Consistent Field Technique 156
   3.1. Simulation Techniques for One-Component Polymer Liquids 156
   3.2. Self-Consistent Field Technique for One-Component Polymer Liquids 160
4. Wetting of Polymer Liquids 170
   4.1. Locating the Wetting Transition in Monte Carlo Simulations 170
   4.2. Nanodroplets on Layered Substrates 175
   4.3. Wetting on a Polymer Brush 180
   4.4. Adsorption on Laterally Structured Substrates 183
5. Binary Polymer Brushes: Self-Assembled Structures with Switchable Wetting Properties 184
6. Incompressible Binary Polymer Films 189
   6.1. Generalization of the Computational Techniques to Binary Polymer Blends 189
   6.2. Wetting Transition in Incompressible Binary Polymer Blends 194
   6.3. Interplay Between Wetting and Phase Behavior in Confined Geometry 195

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (151–202)

7. Outlook 198
References 199

1. INTRODUCTION

Thin polymer films find ample applications as dielectric materials in microelectronics, protective coating, adhesives and lubricants [1–5]. As the length scale of devices becomes smaller and smaller, effects of boundaries and surfaces become increasingly important. In many cases, the stability of the film against rupture or the morphology of the film in a multicomponent system is of key importance. These properties are determined by a variety of factors, which mirror structure on vastly different timescales and length scales. For illustration, let us consider a one-component polymer film on a solid substrate.

On the atomistic scale, specific interactions between the polymer and the substrate dominate the behavior—there might be chemisorption or physisorption of small portions of the polymer onto the substrate. In principle, atomistic modeling and quantum chemical calculations can provide much insight into these local structural properties (e.g., the specific adsorption of chain ends to surfaces [6]). It is important to realize that the energy scale for these interactions is on the order of eV—this is the largest energy scale to be considered. It is large compared to thermal fluctuations ($k_B T \approx 0.025$ eV) or typical liquid–vapor interface tensions ($\gamma = 10^{-22}$ J/Å$^2 \approx 0.006$ eV/Å^2). Therefore, these interactions have a substantial influence on the stability of thin films, but by the same token, they are very sensitive to specific chemical substances and preparation methods (e.g., the cleaning of surfaces [7]).

On the mesoscopic scale, which is larger than a few atomistic monomeric units, other interactions play an important role in polymeric materials: the conformational entropy of the extended molecules and the van der Waals attractions between constituents; the corresponding energy scale $k_B T$, where k_B is Boltzmann's constant and T denotes the temperature, is characteristic for soft-condensed matter and it is about two orders of magnitude smaller than the energy of a chemical bond.

On even larger length scales, one can describe the system by the location of the liquid–vapor interface of the polymer film. The interaction of this interface with the substrate per unit area is described by the interface potential g [8]. The latter quantity is of paramount importance for the wetting properties of the polymer film. The typical free-energy scale only is a small fraction of the interface tension γ.

Thus, as one proceeds to longer length scales, the effective interactions become softer and the concomitant timescales longer. Obviously, a seamless description that spans all the different length scales is computationally not feasible, but much progress has been achieved in specific areas. For instance, combining Car–Parinello density-functional calculations with molecular dynamics simulation of a coarse-grained polymer model, the conformations of polycarbonate in the vicinity of a nickel surface have been studied [6]. The quantum mechanical density-functional calculations provided inspiration for the interaction potentials that were used as input for the simulation of the coarse-grained model to study the behavior on larger length scales. This is one example of bridging the length scales by combining different techniques tailored to address problems on specific length scales. In the following text, we focus mainly on coarse-grained, particle-based, and field theoretic models because they can address phenomena on the length scale of a few nanometers. We investigate the relation between these two approaches, and on the other hand, we will discuss them in context to experimental systems.

We will describe the application of coarse-grained models to study the structure and thermodynamics of one- and two-component polymer films. We do not focus on deriving these models for a specific system but, rather, investigate the generic properties on the length scale from nanometers to a micrometer. Contact to experimental realizations is established via a few coarse-grained parameters (e.g., Hamaker constant, molecular extension, Flory–Huggins parameter or interface/surface tensions). More details of molecular architecture can be incorporated (e.g., stiffness via bending and torsional potentials, chain branching), and coarse-grained models can investigate the qualitative dependence of the behavior on

larger length scales (e.g., contact angles and morphology of phase separation in films) on the molecular architecture. They also are an ideal testing bed for phenomenological concepts. Moreover, coarse-grained models can investigate systems containing interfaces and self-assembled structures that cannot be studied by atomistic simulations with present-day computers [9].

Simulation of polymer films is a rapidly developing research area. We have made no attempt to be comprehensive, and we have drawn examples from our own work. We restrict ourselves to properties of fluid films in thermal equilibrium and do not discuss other interesting properties like the glass transition [10] and the mechanical properties [11] of thin films or the kinetics of phase separation [12] or dewetting [13, 14].

2. COARSE-GRAINED MODELS

Rather than "deriving" a coarse-grained model for a specific system, one can ask the question of which interactions are necessary to bring about a specific behavior on large length scales. Including only those relevant interactions, one obtains a computationally effective, minimal model and learns a great deal about the ingredients that cause the qualitative behavior on large length scales. If need be, these models can be successively refined to incorporate more structural details.

Coarse-grained models have a longstanding history in polymer physics. They were the first to investigate the structure of polymers in good solvent, where polymer conformations adopt self-avoiding walk statistics [15]. In this case, the use of a coarse-grained model can be formally justified by the self-similar structure on a large range of length scales from the statistical segment length b to the polymer's end-to-end distance R_e. The structure of polymer solutions corresponds to the critical behavior of a field theory with n-component order parameter in the limit $n \to 0$ [16]. Similar to critical phenomena in the theory of phase transitions, the universal behavior on large length scales does not depend on the detailed implementation of the relevant interactions—connectivity along the polymer and repulsion between segments (excluded volume)—on short length scales.

In the following text we shall describe two popular coarse-grained models for dense polymer melts and solutions. In these coarse-grained models, one lumps a small number of monomeric units together to form an effective segment. In addition to the connectivity of monomeric units along the backbone and the excluded volume of the monomeric units, a finite energy scale describes the attraction between segments of the same type or a repulsion between segments of different types in a mixture. Comparing this finite energy to the thermal energy $k_B T$, we introduce a temperature T in our model.

Two representations will be employed in the following: the bond-fluctuation model [17] and a bead-spring off-lattice model [18, 19]. In the bond-fluctuation model (cf. Fig. 1a) monomeric units are represented by unit cubes on a three-dimensional simple cubic lattice.

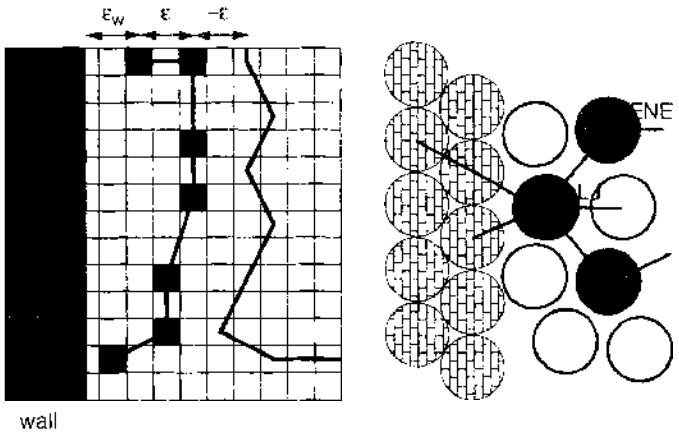


Figure 1. Schematic representation of two coarse-grained polymer models. Left, bond fluctuation model, right, bead-spring model. See text for further explanations.

Each monomeric unit blocks the eight corners of the unit's cell from double occupancy. This represents the excluded volume of the monomeric units. Monomeric units are connected by one of 108 bond vectors with lengths 2, $\sqrt{5}$, $\sqrt{6}$, 3, and $\sqrt{10}$ in units of the lattice spacing a [namely, (2, 0, 0), (2, 1, 0), (2, 1, 1), (2, 2, 1), (3, 0, 0) and (3, 1, 0) and all symmetry-related ones]. This represents the connectivity along the backbone of the polymer. We shall use this model to investigate the phase behavior of a binary polymer blend in Section 6. The mixture comprises two species, denoted A and B. Monomeric units of the same species attract each other via a square-well potential that is extended over the nearest 54 lattice sites. This roughly corresponds to the first neighbor shell in the density pair correlation function. Each contact between monomers of the same type decreases the energy by an amount ϵ, and each contact between monomers of different types increases the energy by the same amount. The ratio between this energy scale and the typical energy of thermal fluctuations k_BT determines the phase behavior.

In the bead-spring model (cf. Fig. 1b) monomeric units are represented by Lennard–Jones interaction centers that live in three-dimensional continuum space. Let r denote the distance between two monomeric units, and then their interaction is given by

$$V_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] - V_c \qquad \text{for } r < r_c \tag{1}$$

and $V_{LJ}(r) = 0$ for distances larger than the cut-off r_c. V_c is a constant chosen such that $V_{LJ}(r_c) = 0$. In our simulations we use $r_c = 2\sqrt[6]{2}\sigma$. The Lennard–Jones potential is characterized by two parameters, the energy scale ϵ and the monomeric size σ. The harsh repulsive contribution mimics the segmental excluded volume, whereas the $1/r^3$ term describes the van der Waals–like attraction between monomers. Monomeric units that are nearest neighbors along the chain are bonded via a finitely extensible, nonlinear elastic (FENE) potential

$$V_{FENE}(r) = -15\epsilon(R_0/\sigma)^2 \ln\left(1 - \frac{r^2}{R_0^2}\right) \tag{2}$$

This interaction represents the connectivity along the backbone of the polymer. We represent the wall or substrate as an impenetrable structureless wall, but the interaction between the monomers and the wall is long-ranged.

$$V_{wall}(z) = A_w\left[\left(\frac{\sigma}{z}\right)^{9} - \left(\frac{\sigma}{z}\right)^{3}\right] \tag{3}$$

where z denotes the distance from the planar substrate. These interactions stem from the Lennard–Jones type of interactions between the constituents of the polymer fluid and the substrate. The strength A_w plays the role of the Hamaker-constant.

Mapping atomistic representations onto these coarse-grained models, one aims at roughly retaining the molecular geometry, which is characterized by the radius of gyration of the polymers and the distance between neighboring chains. The latter quantities can be conceived as the "thickness" of the polymer. In an atomistic model, the van der Waals radii of the chemical repeat units strongly overlap, whereas the bond length and the size σ of a monomeric unit are comparable in the coarse-grained description. In an atomistic model, a torsional potential imparts some persistence onto the direction along the backbone of the polymer, and in the coarse-grained model, typically torsional potentials are omitted and the monomeric units are flexibly bonded together. Therefore, one can roughly estimate that a monomeric unit of the coarse-grained model corresponds to 3–5 chemical repeat units in an atomistic model. These considerations can be made more quantitatively by comparing inter- and intramolecular pair correlation functions between atomistic models and coarse-grained models and adapting the potentials in the coarse-grained model to faithfully represent their atomistic counterparts [20, 21].

In the same spirit, one can roughly identify a lattice unit a of the bond fluctuation model with a length scale of 2–3 Å. Given that the typical linear dimensions of the lattice used in Monte Carlo simulations are on the order of 200 a, this model can explore the behavior on

scales of up to 50 nm. The length scale σ of the Lennard–Jones model corresponds to about 5 Å. Because a typical simulation cell has a linear extension of about $20\,\sigma$, the model captures the behavior on the length scale of about 10 nm. Of course, we would like to emphasize that these values are just very rough estimates: The system size that can be investigated within reasonable computer resources depends on the specific question and algorithm. The coarse-graining procedure sketched above has been carried out for specific substance, but we shall not pursue this interesting issue further [21]. Rather than modeling a specific polymer, we shall concentrate on the generic properties of polymeric liquids.

In both models, the interactions between nonbonded monomeric units comprise two qualitatively different contributions: the harsh, short-ranged repulsion determines the packing of the monomers and gives rise to a local fluid structure. The fluid structure manifests itself in pronounced oscillations in the density pair correlation function at short distances and high densities. The packing effects significantly contribute to the compressibility of the liquid. They are somewhat stronger in the off-lattice model than in the bond fluctuation model, but they are nevertheless present in the lattice model because of the size difference between monomers and vacancies. Given the fact that a monomer in the coarse-grained model corresponds to a small number of repeat units in a chemically realistic model, one expects that the effective interaction between a small group of repeat units in a chemically realistic model is softer than the harsh repulsion in the coarse-grained models, and that the coarse-grained models systematically overestimate the structuring of the fluid.

The longer-ranged attractions set the temperature scale. In a one-component system, they mimic the effect of the solvent. Rather than considering the solvent explicitly, one can conceive these monomer–monomer interactions as a result of integrating out the degrees of freedom corresponding to the solvent. This yields a faithful description of thermodynamic equilibrium properties of incompressible polymer+solvent mixtures, but it might lead to quite different dynamical properties [22] and fails even for the statics if the system is not incompressible [23]. These effective attractions between monomeric units cause an isolated chain to collapse from a self-avoiding structure at high temperature to a dense globule at low temperature. For an infinitely long chain, this collapse transition occurs at the Θ temperature. Below this temperature, a multichain system (a polymer solution) phase separates into liquid and vapor [16].

The functional form of the potential in the off-lattice model mimics van der Waals interactions. Typically, the interaction is cut off at a finite distance for sake of computational efficiency. The value of the cutoff does affect the equation of state of the polymer liquid and the liquid–vapor coexistence quantitatively, but these differences can be accounted for by tail corrections to the chemical potential or pressure. The cutoff can also modify qualitative features of the wetting behavior, and we shall discuss this issue further in Section 4.1.

Both models are well-suited to investigate the universal properties of polymeric materials on the mesoscopic length scale: Each one offers distinct advantages and drawbacks. The major advantage of lattice models is their computational efficiency. Taking advantage of the lattice structure, energies can be calculated faster, and more complicated Monte Carlo moves are more easily formulated and more efficiently implemented on the lattice than in continuum space. As computer simulations of polymeric materials are computationally much more exacting than simulations of simple liquids, computational speed is a major issue. One disadvantage of lattice models is the possibilities of lattice artefacts. These are particularly important for phases with nematic order or at very high densities. Lattice artefacts are strongly reduced in the bond fluctuation model compared with simple lattice models that represent monomeric units by a single lattice site. The large number of bond vectors (108) allows for 87 distinct bond angles and provides a much finer approximation of continuum space than simple lattice models. In this sense, the bond fluctuation model is a compromise between simple lattice models and a continuum description (cf. also Ref. [24] for a comparison between "fine-grained" lattice models and continuum models).

The major advantage of models in continuum space is the presence of forces. This presence allows for molecular dynamics simulations in which one propagates the particle positions according to Newton's law of motion and obtains a realistic description of the dynamics. These simulations are also well suited for parallel computations, as one can partition the

calculation of forces between particles onto different processors. Forces can also be used to calculate the pressure via the virial expression, and the velocity correlations can be employed to determine viscosities.

In general, one should use the model and computational technique that is computationally the most efficient if there are not artefacts in the model that alter the qualitative physics. Therefore, we will discuss the stability of one-component polymer films in the framework of the bead-spring model. Earlier studies of the bond fluctuation model [25–27] have shown that under very bad solvent conditions, the polymers form a very dense liquid, the properties of which are dominated by the structure of the underlaying lattice and that is extremely difficult to equilibrate. Also in the off-lattice model, the liquid condenses into a dense structure at low temperatures or high pressures, which, in turn, exhibits interesting glassy behavior. However, at moderate melt like densities, the bond fluctuation model is very efficient, and we shall use it to discuss the properties of binary polymer films in Section 6.

3. MONTE CARLO SIMULATION AND SELF-CONSISTENT FIELD TECHNIQUE

3.1. Simulation Techniques for One-Component Polymer Liquids

In the following we describe Monte Carlo simulations to investigate the structure and thermodynamics of thin polymer films [28]. To illustrate the simulation technique, we will focus on one-component polymer films within the bead-spring model. We consider polymers that consist of N monomers. We work in the grandcanonical ensemble (i.e., we fix the volume V of our simulation cell, the temperature T and the chemical potential μ of the polymers). The number of polymers n fluctuates. This ensemble is particularly advantageous: First, we directly determine the relation between the chemical potential and the density, and we can efficiently determine the liquid–vapor coexistence and the properties of the liquid–vapor interface (see following). Second, the longest relaxation time of the system corresponds to the decay of density fluctuations. For instance, the equilibration of the thickness of wetting layers at the surface in the canonical ensemble would require an exchange of polymers between the vapor phase and the liquid layer at the wall via polymer diffusion, whereas in the grandcanonical ensemble, polymers are removed at one location and inserted at another. Moreover, to act as a particles reservoir, the vapor phase in a canonical simulation would have to be enormously large to accommodate sufficient polymers to observe, for instance, the dependence of the wetting layer on the monomer wall interaction.

To realize the grandcanonical ensemble, two types of elementary Monte Carlo steps or moves can be distinguished: canonical moves that relax the spatial conformations of the polymers, and moves that alter the number of polymers. In our simulations, canonical moves consist of local random displacements of monomers and slithering snakelike movements [29]. In the latter moves, one chooses an end monomer at random and tries to attach it to the opposite end of the molecule. These moves relax the global chain conformation a factor of the order N faster than local random displacements. Of course, they become inefficient when the liquid is so dense that one cannot insert even a single monomer, but we shall use only moderately dense liquids in our simulations. In very dense liquids, molecular dynamics might offer advantages to equilibrate the polymer conformations and molecular dynamics simulations can be integrated into a Monte Carlo simulation via a hybrid algorithm [30]. In addition, other moves that alter the connectivity of monomers have been efficiently implemented and are well suited to equilibrate dense polymer liquids [31].

To insert and remove polymers, one can use the configurational bias algorithm [32–34]. This scheme uses a biased insertion method to "grow" a polymer successively into the system. At each step, a small number (typically 25) of segment positions is examined, and a position for inserting of the next monomer along the chain is chosen according to its Boltzmann weight. This choice biases the insertion towards nonoverlapping/low-energy chain conformations. Once a chain has been grown, the bias in the construction is compensated in the Monte Carlo lottery.

The key quantity in our simulations is the probability distribution $P_{GC}(n|\mu, V, T)$ of the number of chains in the simulation cell at a given chemical potential, volume, and temperature. A typical example for a bulk system in which we employ periodic boundary conditions in all three directions is presented in Fig. 2a. We plot the negative logarithm of the distribution, which corresponds to the canonical free energy, as a function of the monomer number density $\phi = nN/V$. The free energy exhibits two minima, which correspond to the vapor and the liquid. As they have equal weight, the two phases coexist at this value of the chemical potential [35, 36].

One way to obtain this equal weight criterium for phase coexistence is to relate the statistical weight in each peak of the distribution $P_{GC}(n)$ to the pressure $p = -\Omega(\mu, V, T)/V$, where $\Omega = -k_B T \ln \mathcal{Z}_{GC}$ is the grand potential

$$\begin{aligned} \frac{p^{(V)}V}{k_B T} &= \ln \sum_{n\in(V)} \frac{e^{\mu n/k_B T}}{n!} \int \mathcal{D}[\{\mathbf{R}\}] \exp\left(-\frac{E[\{\mathbf{R}\}]}{k_B T}\right) \\ &= \ln \mathcal{Z}_{GC} + \ln \sum_{n\in(V)} P_{GC}(n|\mu, V, T) \end{aligned} \tag{4}$$

$$\text{with} \qquad \mathcal{Z}_{GC} = \sum_{n=0}^{\infty} \frac{e^{\mu n/k_B T}}{n!} \int \mathcal{D}[\{\mathbf{R}\}] \exp\left(-\frac{E[\{\mathbf{R}\}]}{k_B T}\right)$$

where the first summation is only extended over all number of polymers n, which correspond to the vapor phase (V). $\mathbf{R}_{i,j}$ denotes the coordinate of the jth monomer of chain i, $\{\mathbf{R}\}$ is the set of all polymer coordinates and specifies a configuration of polymers, $\mathcal{D}[\{\mathbf{R}\}]$ denotes a sum over all polymer conformations, and E is the sum of the Lennard–Jones and FENE interactions in the multichain system. A similar expression holds for the pressure in the liquid phase (L). If the system is large and the transition of first order, the peaks in P_{GC} are describable by Gaussian distributions. The wings of the distribution yield only an exponentially small contribution to the pressure, such that the detail of how to divide the n-range into the two phases—vapor (V) and liquid (L)—imparts only a negligible error. In practice, we divide the distribution around the average value [i.e., we define

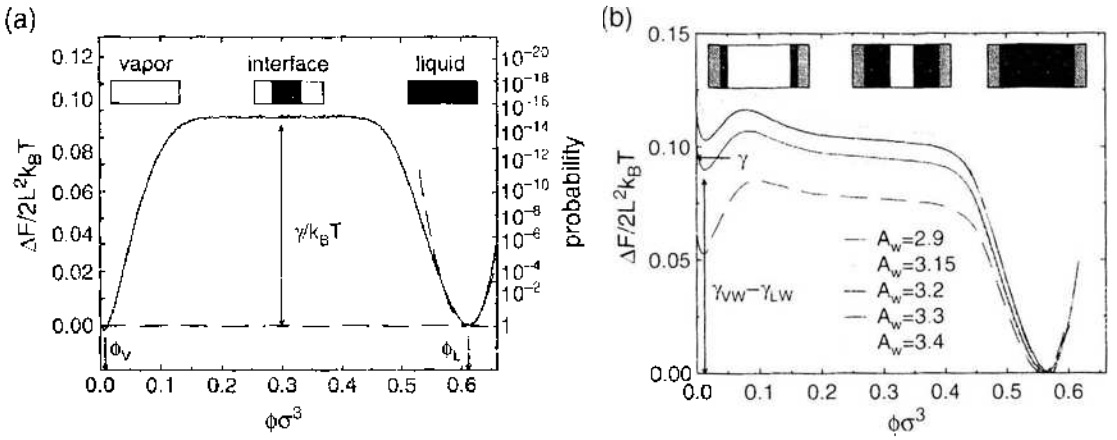


Figure 2. (a) Illustration of the simulation technique for a polymer liquid of $N = 10$ monomeric units at temperature $k_BT/\epsilon = 1.68$ and $\mu_{coex}/\epsilon = 106.897$. A cuboidal system geometry $13.8\sigma \times 13.8\sigma \times 27.6\sigma$ with periodic boundary conditions in all three directions is used. The solid line corresponds to the negative logarithm of the probability distribution $P(\phi)$ in the grand canonical ensemble. The two minima correspond to the coexisting phases, and the arrows on the ϕ-axis mark their densities. The height of the plateau yields an accurate estimate for the interfacial tension γ. The dashed line is a parabolic fit in the vicinity of the liquid phase employed to determine the compressibility. The typical system configurations are sketched schematically. (b) Free energy as a function of the density of a system, which is confined between walls of attractive strength A_w. The grand canonical simulations at $k_BT/\epsilon = 1.68$ and coexistence chemical potential in the bulk are performed in a geometry $13.8\sigma \times 13.8\sigma \times 27.6\sigma$. The curves are shifted such that the free energy of the liquid phase vanishes. The horizontal arrow on the left marks the value of the interfacial tension γ, whereas the vertical arrow marks the difference in the surface tension between the vapor/wall and the liquid/wall for $A_w = 3.15$. Typical system configurations are sketched schematically. Adapted from Ref. [28], M. Müller and L. G. MacDowell, *Macromolecules* 33, 3902 (2000).

$n^* = \sum_n nP_{GC}(n|\mu_{coex}, V, T)$]; all systems with lower than n^* polymers belong to the vapor phase, and all configurations that contain more polymers are identified with the liquid phase. This definition can also be applied in the vicinity of the critical point.

In the grandcanonical ensemble, liquid and vapor coexist at a fixed temperature T and chemical potential μ_{coex} if the pressure in both phases is equal: $p^{(V)}(\mu_{coex}) = p^{(L)}(\mu_{coex})$. Using Eq. (4), one obtains the equal-weight-rule of phase coexistence

$$\sum_{n<n^*} P_{GC}(n|\mu_{coex}, V, T) = \sum_{n>n^*} P_{GC}(n|\mu_{coex}, V, T). \tag{5}$$

The search for phase coexistence is facilitated by histogram extrapolation techniques [37, 38]. Having collected the joint histogram $P_{GC}(n, E|\mu, T, V)$ of the number of polymers n and energy E at a given state characterized by parameters μ and T, we can extrapolate the probability distribution to neighboring values T' and μ' in parameter space via

$$P_{GC}(n, E|\mu', T', V) \sim P_{GC}(n, E|\mu, T, V) \exp\left[n\left(\frac{\mu'}{k_B T'} - \frac{\mu}{k_B T}\right) - E\left(\frac{1}{k_B T'} - \frac{1}{k_B T}\right)\right] \tag{6}$$

This is a very efficient method for accurately determining phase coexistence from near-coexistence data. The method is only reliable if the distributions that correspond to the parameters μ, T and μ', T' significantly overlap. Therefore, it works best around the liquid–vapor critical point, where fluctuations are very large and one can extrapolate the data of smaller systems over a wider parameter range than those of larger systems. Verification of this extrapolation by a control simulation is also a fine validation of the simulation program because it directly assesses the statistical weight of the configurations generated in the course of the Monte Carlo simulation.

To establish that the statistical weight of the vapor phase equals that of the liquid at coexistence, the simulation has to "tunnel," often between the two phases. The error in the weight of the phases decreases roughly as the square root of these tunneling events. Below the critical point, this tunneling between the two phases becomes increasingly difficult; the probability of finding the system in an intermediate state is strongly suppressed. The typical configurations around n^* correspond to the coexistence of the liquid and the vapor inside the simulation box. If both phases are present at roughly equal volume fraction, there will be a slab of liquid that spans the simulation cell via the periodic boundary conditions in two directions and that is separated from the vapor by two interfaces along the remaining direction. As sketched in Fig. 2a, those configurations have excess free energy because of the presence of the two liquid–vapor interfaces. The data presented are obtained from an elongated simulation box, which is advantageous for separating the two interfaces [39]. The plateau in the free energy indicates that one can change the density, and thereby vary the distance between the two interfaces, at no cost in free energy. Therefore, the interactions between the two interfaces across the liquid or vapor are negligible, and the excess free energy can be used to estimate the liquid–vapor interface free energy γ [40]

$$\frac{2\gamma L^2}{k_B T} = \ln \frac{P(n_L)}{P(n^*)} \tag{7}$$

This simulation scheme can also be used to extract the surface tension [28]. To this end, one performs a grandcanonical simulation at the bulk coexistence value of the chemical potential μ_{coex} in the presence of two walls (cf. Fig. 2b). The figure contains data for walls that attract the polymers with different strengths A_w. The free energy as a function of the density exhibits two minima that correspond to the vapor and the liquid in contact with the wall, respectively. As the liquid phase benefits more from the attraction between substrate and polymer, it has a lower free energy than the vapor. At phase coexistence in the bulk, the free-energy difference is solely the result of the difference of surface free energies between the wall and the vapor $\gamma_{W,V}$ and the wall in contact with the liquid $\gamma_{W,L}$. Therefore, we can directly read off the difference

$$\frac{2L^2\Delta\gamma}{k_B T} = \frac{2L^2(\gamma_{W,V} - \gamma_{W,L})}{k_B T} = \ln \frac{P_L}{P_V} \tag{8}$$

Thus, the probability distribution in the grandcanonical ensemble provides a wealth of information: From the equal-weight rule, one obtains the density of the coexisting liquid and vapor and the chemical potential of coexistence. From the fluctuations of the density in each phase, one estimates the compressibility of the liquid and the vapor. The "valley" between the peaks yields an estimate of the interface tension, and in the presence of walls, one obtains the surface free-energy difference.

As those configurations that contain interfaces are strongly suppress. the probability of finding the system in such is state is vanishingly small in the grandcanonical ensemble; in this specific example, we find $P(n^*) \sim 10^{-14}$. Hence, a straightforward Monte Carlo sampling is not feasible. To sample states with a useful rate, a reweighting technique has been devised to overcome this problem [41]. To this end, one adds to the original Hamiltonian $E[\{\mathbf{R}\}]$ of the system a weight function $k_B T w(n)$ that only depends on the number of the particles in the system, but otherwise not on the detailed configuration. Therefore, different configurations containing the same number of particles are still sampled with the Boltzmann weight that corresponds to the canonical ensemble. The probability distribution $P(n)$, however, is modified to

$$E_{\mathrm{RW}}[\{\mathbf{R}\}] = E[\{\mathbf{R}\}] + k_B T w(n) \qquad \text{and} \qquad P_{\mathrm{RW}}(n) = P_{\mathrm{GC}}(n) \exp[-w(n)] \tag{9}$$

Hence, if one chooses the weight function $w(n) \approx \ln P_{\mathrm{GC}}(n)$, the simulation will sample all states with roughly equal probability. The crux is that $P_{\mathrm{GC}}(n)$ is just the result of the simulation and is not known *a priori*. Several schemes to overcome this dilemma have been proposed.

First, histogram-reweighting techniques [37] alleviate this problem by performing a sequence of weighted simulations and extrapolations starting at a point at which barriers are small and the system explores a broad range of n (i.e., close to the liquid–vapor critical point). Those results can then be extrapolated to lower temperatures via Eq. (6), where barriers are larger. More sophisticated methods combine results of multiple histograms [38]. Second, weight factors can be obtained from the transition probabilities between macrostates [42, 43]. Third, multicanonical recursion [44] conducts a series of short trial runs. After each run, $w(n)$ is adjusted until the simulation can access all relevant states. Fourth, the weight function can also be self-adjusted during the simulation [45–48]. One starts with an initial guess of the weight function $w(n)$ and incrementally increases its value by Δw; each time, a state with n particles is visited in the course of the simulation. This procedure tends to "push" the system out of states that it has frequently visited and allows it to explore all pertinent states. If the histogram of visited states is approximately flat (i.e., all states have been visited with nearly equal probability), the increment Δw is decreased and the histogram of visited states reset. This method works in many practical cases, however, it requires some fine-tuning of the initial value of the increment Δw. Moreover, for any $\Delta w > 0$, detailed balance [49] is violated and separation of statistical and systematic errors becomes difficult. Fifth, alternatively, one could use successive umbrella sampling [50] to generate the probability distribution $P_{\mathrm{GC}}(n)$. To this end, one divides the interval of particle numbers into smaller windows that overlap at their boundaries. Each window is sampled consecutively, and the results are linked together at the boundaries of the window. Ideally, the statistical error for a given total amount of computation time is independent from the window size. Hence, choosing a small windows size is beneficial because the difference in the $P_{\mathrm{GC}}(n)$ within a single window becomes small and one can simulate without weight function or use a rather crude estimate for $w(n)$ (e.g., a polynomial extrapolation of the results from previous windows). Choosing a window size that is too small, however, might inflict sampling difficulties.

In principle, one can also measure the interface free energy by monitoring the anisotropy of the pressure tensor across the interface [51, 52]. Close to liquid–vapor coexistence the pressure is very small, but in our model this small value stems from a cancellation of a large positive contribution resulting from the Lennard–Jones interactions and a large negative contribution caused by the bonding interactions. Accurate measurement of the interface or surface tension via the pressure tensor would require extremely good statistics [28]. For our parameters, the error from the analysis of the anisotropy of the pressure tensor is at

least one order of magnitude larger than from the reweighting method discussed above (using a comparable computational effort). This is an agreement with the results of independent studies [53]. As the error of the pressure tensor results from a cancellation of rather noisy data, it might not be proportional to the value of the tension (but, rather, to the value of the virial of the Lennard–Jones or FENE forces), and the method might become more useful when the tension is larger [28, 53].

In addition to these thermodynamic properties, computer simulations also provide detailed information about the structure of the polymeric fluid both in bulk as well as at interfaces and surfaces. These simulations are typically performed in the canonical ensemble (i.e., at a fixed number of particles).

3.2. Self-Consistent Field Technique for One-Component Polymer Liquids

3.2.1. General Formalism

In principle, computer simulations provide exact results on the statistical mechanics of a specific model apart from statistical errors and finite sizes effects, which can be controlled via sophisticated and efficient techniques. Nevertheless, they are computationally very demanding and, in particular, for applications that involve an interplay of phenomena on different length scales the computational resources, they are often not sufficient to investigate a wide range of parameters. Therefore, it is useful to explore approximate, but computationally less demanding approaches.

One promising and widely used approach is the self-consistent theory. In the following text, we use the notation of self-consistent theories [54–58], although a completely equivalent description can be obtained using the density-functional theory formalism [59–63]. The starting point of the self-consistent field calculations is the canonical partition function $\mathcal{Z}$ of n polymers in a volume V [28, 64, 65]:

$$\mathcal{Z} \sim \frac{1}{n!} \int \mathcal{D}[\{\mathbf{R}\}]\, \mathcal{P}[\{\mathbf{R}\}] \exp\left(-\frac{F_{ex}[\hat{\phi}(\mathbf{r})]}{k_B T} \right) \tag{10}$$

$\mathcal{P}[\{\mathbf{R}\}]$ denotes the probability distribution of chain conformations, F_{ex} is the excess free energy functional, and $\hat{\phi}$ denotes the microscopic monomer density generated by the specific polymer configurations; that is

$$\hat{\phi}(\mathbf{r}|\{\mathbf{R}\}) = \sum_{i=1}^{n} \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{R}_{i,j}) \tag{11}$$

where the first sum runs over all n polymers and the second sum runs over all N monomers of a chain. $\mathcal{P}[\{\mathbf{R}\}]$ is given by

$$\mathcal{P}[\{\mathbf{R}\}] = \mathcal{P}_{\text{melt}}[\{\mathbf{R}\}] \exp\left(-\int \mathrm{d}^3\mathbf{r}\, \frac{V_{\text{wall}}(\mathbf{r})\, \hat{\phi}(\mathbf{r})}{k_B T} \right) \tag{12}$$

where $\mathcal{P}_{\text{melt}}[\{\mathbf{R}\}]$ is the probability distribution of chain conformations in the melt (cf. following) and $V_{\text{wall}}(\mathbf{r})$ is the external interaction potential between a monomeric unit and the surface or surfaces. To quantitatively compare simulations and self-consistent field calculations, one has to find a suitable approximation to the interaction free-energy functional F_{ex}, and to solve the partition function [see Eq. (10)]. In this paragraph, we describe the latter problem and defer the former to the next subsection.

The derivation of a mean field approximation to the partition function Eq. (10) follows the development of self-consistent theories of polymers [54–58]. Introducing auxiliary fields U and Φ, we rewrite the multichain partition function in terms of noninteracting chains

$$\begin{aligned} \mathcal{Z} &\sim \int \mathcal{D}U \mathcal{D}\Phi \exp\left(-n \ln n + n - \frac{F_{ex}[\Phi]}{k_B T} + \int \mathrm{d}^3\mathbf{r}\, U(\mathbf{r})\Phi(\mathbf{r}) + n \ln \zeta V \right) \\ &= \int \mathcal{D}U \mathcal{D}\Phi \exp\left(-\frac{\mathcal{H}[U, \Phi]}{k_B T} \right) \end{aligned} \tag{13}$$

$\mathcal{Q}[U]$ is the single-chain partition function in the external field U; that is

$$\mathcal{Q}[U] = \frac{1}{V}\int \mathcal{D}[\mathbf{R}_1]\mathcal{P}[\mathbf{R}_1]\exp\left(-\int \mathrm{d}^3\mathbf{r}\, U(\mathbf{r})\hat{\phi}(\mathbf{r}|\mathbf{R}_1)\right)$$
$$= \frac{1}{V}\int \mathcal{D}[\mathbf{R}_1]\mathcal{P}[\mathbf{R}_1]\exp\left(-\sum_{j=1}^{N} U(\mathbf{R}_{1,j})\right) \tag{14}$$

and

$$\frac{\mathcal{F}[U,\Phi]}{Vk_BT} = \frac{\phi_{\mathrm{av}}}{N}\ln\frac{\phi_{\mathrm{av}}}{e\mathcal{N}} + \frac{F_{\mathrm{ex}}[\Phi]}{Vk_BT} - \frac{1}{V}\int \mathrm{d}^3\mathbf{r}\, U(\mathbf{r})\Phi(\mathbf{r}) - \frac{\phi_{\mathrm{av}}}{N}\ln\mathcal{Q}[U] \tag{15}$$

where $\phi_{\mathrm{av}} = \frac{nN}{V}$ is the average monomer density.

The functional integral Eq. (13) cannot be evaluated explicitly, and we resort to a saddle point approximation. At this stage, the problem of many interacting molecules is replaced by that of a single molecule in an external field. The latter mimics the effect of the other molecules and depends, in turn, on the local density. The self-consistent relation between the fields and the densities are obtained by minimizing the functional $\mathcal{F}[U,\Phi]$. Those values are denoted by lowercase letters.

$$u(\mathbf{r}) \equiv U_{\min}(\mathbf{r}) = \frac{\delta}{\delta\Phi}\frac{F_{\mathrm{ex}}[\Phi(\mathbf{r})]}{k_BT} \tag{16}$$

and

$$\phi(\mathbf{r}) \equiv \Phi_{\min}(\mathbf{r}) = -\frac{\phi_{\mathrm{av}}V}{N}\frac{\delta}{\delta U}\ln\mathcal{Q}[U]$$
$$= \frac{\phi_{\mathrm{av}}V}{N}\frac{\int \mathcal{D}[\mathbf{R}_1]\mathcal{P}[\mathbf{R}_1]\hat{\phi}(\mathbf{r}|\mathbf{R}_1)\exp(-\int \mathrm{d}^3\mathbf{r}'\, u(\mathbf{r}')\hat{\phi}(\mathbf{r}'|\mathbf{R}_1))}{\int \mathcal{D}[\mathbf{R}_1]\mathcal{P}[\mathbf{R}_1]\exp(-\int \mathrm{d}^3\mathbf{r}'\, u(\mathbf{r}')\hat{\phi}_1(\mathbf{r}'|\mathbf{R}_1))}. \tag{17}$$

The last expression identifies $\phi(\mathbf{r})$ as the Boltzmann average of the density that a single chain in the external field $u(\mathbf{r})$ produces; it is also the estimate for the monomer density profile of the polymer fluid.

Given an approximation for F_{ex} as a function of the density $\phi(\mathbf{r})$, Eqs. (16) and (17) form a closed set of self-consistent equations for the densities and fields $\phi(\mathbf{r})$ and $u(\mathbf{r})$ Once the density profile $\phi(\mathbf{r})$ and the effective field $u(\mathbf{r})$, which solve the self-consistent equations, have been obtained, all single-chain properties (e.g., segment profiles, orientation) can be calculated from the properties of a single chain in the external field u. Substituting the saddle point values u and ϕ into the free-energy functional [Eq. (15)], we obtain the free energy. The chemical potential μ is given by

$$\frac{\mu}{k_BT} \equiv \frac{1}{k_BT}\frac{\delta F[\phi, u[\phi]]}{\delta\phi(\mathbf{r})} = \frac{1}{N}\ln\frac{\phi_{\mathrm{av}}}{N} - \frac{1}{N}\ln\mathcal{Q}[u[\phi]] \tag{18}$$

and the grand potential, Ω, is given by

$$\frac{\Omega}{Vk_BT} \equiv \frac{F}{Vk_BT} - \frac{\mu\phi_{\mathrm{av}}}{k_BT} = -\frac{\phi_{\mathrm{av}}}{N} + \frac{F_{\mathrm{ex}}[\phi]}{Vk_BT} - \frac{1}{V}\int \mathrm{d}^3\mathbf{r}\, u(\mathbf{r})\phi(\mathbf{r}) \tag{19}$$

One could also choose the grandcanonical ensemble from the outset. Then one would fix μ instead of ϕ_{av}, and Eq. (17) would be replaced by

$$\phi(\mathbf{r}) = -V\exp(\mu/k_BT)\frac{\delta}{\delta U}\mathcal{Q}[U] \tag{20}$$

The numerical procedure of the self-consistent field calculations is straightforward but computationally intensive. Given an initial guess for the field $u(\mathbf{r})$, the monomer density profile is calculated from Eq. (17). This procedure is iterated until Eq. (16) is fulfilled. Computing the density profile given the field is the most computationally intense part of

the calculations. It is often repeated in the self-consistent scheme. Two major schemes have been devised: First, the sum $(\int \mathcal{D}_1[\mathbf{R}]\mathcal{P}_1[\mathbf{R}])$ over the single-chain conformations is approximated by a partial enumeration over a large number of explicit chain conformations [28, 66–69]. Ideally, these conformations have been extracted from a computer simulation of the liquid phase, or they are obtained from a molecular modeling approach. Typically, 10^6–10^7 single-chain conformations are employed in our calculations. The longer the chains and the stronger the changes of the chain conformations in the spatially inhomogeneous field, the more conformations have to be considered. The sample size should be large enough that sufficiently many chains significantly contribute to the Boltzmann weight. The enumeration over the chain conformations conveniently is performed in parallel. To this end, a small fraction of single-chain conformations is assigned to each processor. Then, each processor calculates the Boltzmann weight of its conformations, the corresponding density profile, and the weight of its fraction of chain conformations. Subsequently, the total density profile is constructed by summing the weighted results of all processors. Typically, 64 or 128 processors have been employed in parallel, and a self-consistent field calculation of a profile takes a few minutes on a CRAY T3E.

The advantage of this scheme is that is can deal with arbitrary chain architectures and incorporates the details of the molecular structure on all length scales. (i.e., not only the Gaussian behavior on large length scales but also the conformational statistics on the length scale of bond vectors). These properties become important if the length scale of the spatial inhomogeneity (e.g., interfacial width) is comparable to the statistical segment length $b \equiv R_e/\sqrt{N}$, or the size of a monomeric unit σ. This is the typical situation for liquid-vapor interfaces, and it also might occur in blends of very incompatible polymers.

Second, if interfaces are not extremely narrow, the conformational statistics of the polymer can be described by the Gaussian chain model on the length scale of the variation of the density profile. A Gaussian chain is completely characterized by its end-to-end distance R_e. In this case, one can calculate the properties of a single chain in an external field analytically. To this end, one introduces the end-segment distribution function $q(\mathbf{r}, s)$ that describes the probability of finding the end of a chain of length sN at position $\mathbf{r}$, given that the starting point is uniformly distributed in space.

$$q(\mathbf{r}, t) = \int_0^t \mathcal{D}[\mathbf{R}_1]\mathcal{P}[\mathbf{R}_1]\,\delta(\mathbf{r} - \mathbf{r}(t)) \exp\left[-N \int_0^t \mathrm{d}\tau\, u(\mathbf{r}(\tau))\right] \tag{21}$$

The Wiener measure $\mathcal{P}[\mathbf{R}] \sim \exp[-\frac{3}{2R_e^2}\int_0^1 \mathrm{d}s\,(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}s})^2]$ denotes the probability distribution of the chain conformations. The end-segment distribution function of a single Gaussian chain in an external field obeys a modified diffusion equation

$$\frac{\partial q}{\partial t} = \frac{R_e^2}{6}\Delta q - Nuq \qquad \text{with} \qquad q(\mathbf{r}, t = 0) = 1 \tag{22}$$

The single-chain partition function $\mathcal{Q}$ and the monomer density can be expressed in terms of q

$$\mathcal{Q} = \frac{1}{V}\int \mathrm{d}\mathbf{r}\, q(\mathbf{r}, 1) = \frac{1}{V}\int \mathrm{d}\mathbf{r}\, q(\mathbf{r}, s)q(\mathbf{r}, 1 - s) \qquad \forall\, 0 \le s \le 1 \tag{23}$$

$$\phi(\mathbf{r}) = \frac{\bar{\phi}}{\mathcal{Q}}\int_0^1 \mathrm{d}t\, q(\mathbf{r}, t)q(\mathbf{r}, 1 - t) \tag{24}$$

The modified diffusion equation can be solved many times faster than the partial enumeration over explicit chain conformations. Hence, this method should be preferred if the local chain architecture does not matter. Equation (22) can be solved efficiently in real space, using a Crank–Nicholson scheme or pseudospectral algorithms [70]. If one is interested in self-assembled phases with long-ranged periodicity (e.g., self-assembled structures), it is advantageous to expand the spatial dependencies of the fields and densities in eigenfunctions of the Laplace operator, which possess the symmetry of the phase under consideration [71].

Then the modified diffusion equation becomes a linear matrix equation for the Fourier coefficients.

Alternatively, other chain models have been used. Some applications of the self-consistent theory have used the wormlike chain model [72]. It captures the crossover between the stiff, rodlike behavior of a polymer on short length scales and the Gaussian statistics on large length scales in the melt. In this case, the end-segment distribution q depends on the spatial coordinate $\mathbf{r}$ as well as on the orientation $\mathbf{u}$, defined by the tangent vector of the space curve at position $\mathbf{r}$ [73, 74]. The wormlike model is a compromise between the computationally intense, partial enumeration scheme and the simple Gaussian chain model. It captures some non-Gaussian structure on short length scales and still allows for a analytical treatment.

3.2.2. An Approximation for the Excess Free-Energy Functional $F_{ex}[\phi]$

An important ingredient of the self-consistent field calculations is the excess free-energy functional $F_{ex}[\phi]$. Naively, one could depart from a particle-based model (e.g., the bead-spring model of Section 2) and identify the excess free-energy functional via

$$F_{ex}[\hat{\phi}] = \frac{1}{2}\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\,\hat{\phi}(\mathbf{r})V_{LJ}(|\mathbf{r}-\mathbf{r}'|)\phi(\mathbf{r}') \tag{25}$$

and use

$$\mathcal{P}[\{\mathbf{R}\}] \sim \exp\left[-\frac{1}{k_B T}\sum_{i=1}^{n}\sum_{j=1}^{N-1} V_{\mathrm{FENE}}(|\mathbf{R}_{ij}-\mathbf{R}_{ij+1}|)\right] \tag{26}$$

Then, the partition function [Eq. (10)] of the self-consistent field theory before the saddle-point approximation and the partition function of the bead-spring model would coincide. Unfortunately, the self-consistent theory invokes a saddle-point approximation. The effect of this approximation is twofold. On the one hand, fluctuations that determine the local structure of the polymeric fluid (i.e., packing effects and chain structure on the scale of the monomer size) are disregarded. On the other hand, fluctuations on the length scale of the whole molecule (i.e., density fluctuations in the vicinity of the critical point and capillary waves at interfaces) are omitted.

In many circumstances, the saddle-point approximation becomes better for the latter, long wavelength fluctuation if one molecule interacts with many neighbors. This can be characterized by the invariant degree of polymerization $\bar{N}$

$$\bar{N} \equiv \left(\frac{\phi R_e^3}{N}\right)^2 \tag{27}$$

where R_e denotes the end-to-end distance of the molecule, and $\sqrt{\bar{N}}$ is the number of molecules in the volume that corresponds to the extension of a chain. In a dense melt, excluded volume interactions are screened and the conformational statistics on large length scales is Gaussian $R_e^2 \sim N$, and, hence, the square of the number of neighbors grows like $\bar{N} \sim N$.

Unfortunately, there is no such parameter that controls the strength of the fluctuation on the monomeric scale (fluid structure). For instance, if we used Eq. (26), the self-consistent field theory would predict the chain conformations in the bulk (i.e., a spatially homogeneous system) not to depend on density or temperature. They would not agree with the simulation data for any temperature or density. The bond length would be determined by the FENE potential, but the repulsion of bonded monomeric units via the Lennard–Jones potential would not be accounted for. Clearly, such a literal identification of the ingredients of the partition function [Eq. (10)] is not useful for a comparison between Monte Carlo simulations and self-consistent theory. Therefore, we identify $\mathcal{P}[\{\mathbf{R}\}]$ so as to reproduce the chain conformations characteristic of a spatially homogeneous, concentrated solution or melt that coexists with the vapor phase. Any dependence of chain conformations on temperature or density cannot be described; the change of chain conformations in the vicinity of a surface or an interface, however, is captured by our calculations (cf. Fig. 3).

By the same token, we shall not identify F_{ex} with the interaction energy but with an (unknown) density functional chosen as to reproduce (at least approximatively) the local

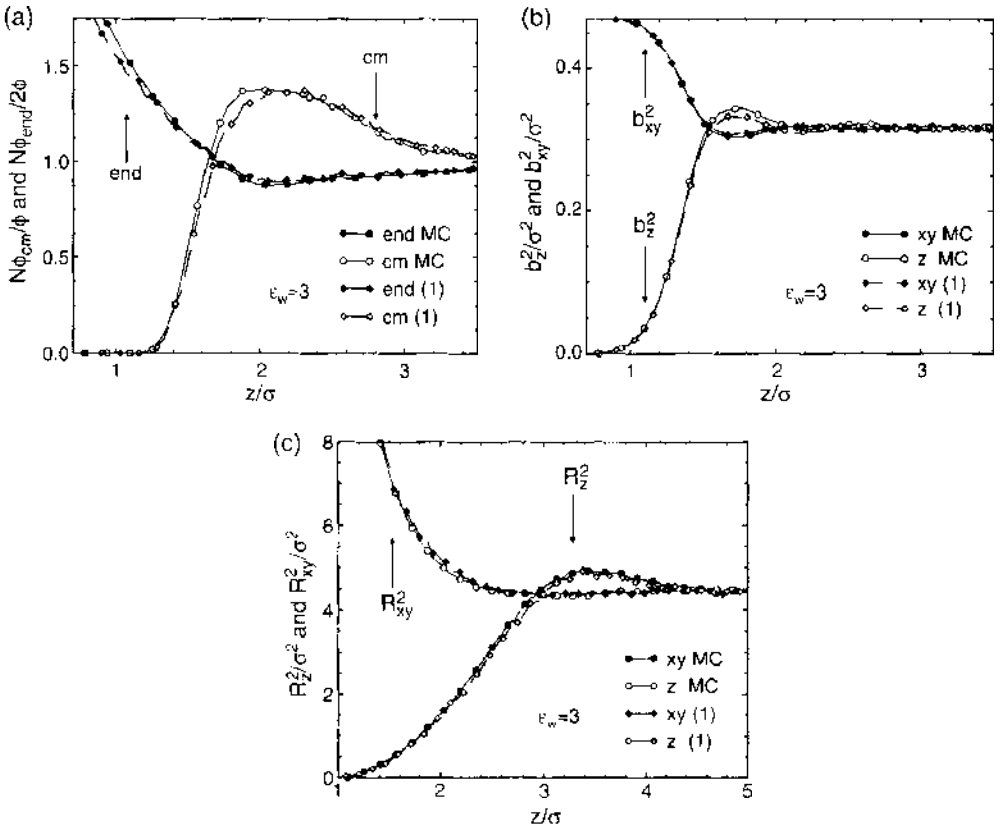


Figure 3. (a) Relative density of chain ends and chains' center of mass as a function of the distance from the wall at $k_BT/\epsilon = 1.68$ and $A_w = 3$. (b) Parallel (xy) and perpendicular (z) components of the bond vector for the same parameters. (c) Parallel and perpendicular components of the end-to-end vector for the same parameters. Adapted from Ref. [65]. M. Müller, et al., *J. Chem. Phys.* 118, 2929 (2003).

fluid structure on minimization. Unfortunately, the results of the calculation depend sensitively on the specific choice for F_{ex}. In the following [28, 64, 65], we use a weighted density approximation. Previous density functional calculations [75] and integral equations [76, 77] for Lennard–Jones monomeric fluids close to attractive walls demonstrated that the interface properties may be described in terms of two different interactions. On the one hand, the harsh repulsive interactions, whose range is set by the effective hard core diameter of the monomers, determine the packing of the monomeric units. On the other hand, the attraction, whose range is set by the range of the attractive Lennard–Jones potential (i.e., $r_{max} = 2 \cdot \sqrt[6]{2}\sigma$) do not modify the packing to a large degree but may considerably affect the value of the free energy. Because the effects of attraction and repulsion are qualitatively different, an explicit decomposition of the functional into contributions from a harsh repulsion and a weak attraction turns out to be important for obtaining quantitative agreement with the Monte Carlo results.

We use the same decomposition of the excess free-energy functional for our polymer model [65]

$$\frac{F_{ex}[\phi(\mathbf{r})]}{k_BT} = \int d^3\mathbf{r}\, \phi(\mathbf{r}) \left\{ g_{hc}\left[\bar{\phi}_{hc}(\mathbf{r})\right] + g_{att}\left[\bar{\phi}_{att}(\mathbf{r})\right] \right\} \tag{28}$$

where g_{hc} and g_{att} represent the free energies per monomer resulting from the repulsive interaction and the attractive interaction, respectively, where and $\bar{\phi}_{hc}(\mathbf{r})$ and $\bar{\phi}_{att}(\mathbf{r})$ are weighted densities, defined by

$$\begin{aligned} \bar{\phi}_{hc}(\mathbf{r}) &= \int d^3\mathbf{r}'\, w_{hc}(\mathbf{r} - \mathbf{r}')\phi(\mathbf{r}') \\ \bar{\phi}_{att}(\mathbf{r}) &= \int d^3\mathbf{r}'\, w_{att}(\mathbf{r} - \mathbf{r}')\phi(\mathbf{r}') \end{aligned} \tag{29}$$

$w_{hc}(r)$ and $w_{att}(r)$ are (as yet undetermined) weighting functions that satisfy the normalization condition

$$\int d^3\mathbf{r}\, w_{hc}(\mathbf{r}) = \int d^3\mathbf{r}\, w_{att}(\mathbf{r}) = 1. \tag{30}$$

Phenomenologically, the weighting functions w_{hc} and w_{att} parameterize the spatial extent of the monomer–monomer interactions. The theory is specified by the two thermodynamic functions of state g_{hc} and g_{att} and the two weighting functions $w_{hc}(r)$ and $w_{att}(r)$. Approximate expressions for these four functions are discussed in the next paragraph.

Note that within the weighted density approximation, we can obtain a closed form expression for $u(\mathbf{r})$

$$\begin{aligned} u(\mathbf{r}) = {} & g_{hc}\big[\bar{\phi}_{hc}(\mathbf{r})\big] + \int d^3\mathbf{r}'\, w_{hc}(\mathbf{r}-\mathbf{r}')\phi(\mathbf{r}')\frac{dg_{hc}}{d\bar{\phi}_{hc}}(\mathbf{r}') \\ & + g_{att}\big[\bar{\phi}_{att}(\mathbf{r})\big] + \int d^3\mathbf{r}'\, w_{att}(\mathbf{r}-\mathbf{r}')\phi(\mathbf{r}')\frac{dg_{att}}{d\bar{\phi}_{att}}(\mathbf{r}') \end{aligned} \tag{31}$$

Substituting the saddle point values of the fields [cf. Eq. (31)] into Eq. (15), we obtain

$$\begin{aligned} \frac{F}{Vk_BT} = {} & \frac{\phi_{av}}{N}\ln\frac{\phi_{av}}{N} - \frac{\phi_{av}}{N}\ln e_z[u] - \frac{1}{V}\int d^3\mathbf{r}\, d^3\mathbf{r}'\, w_{hc}(\mathbf{r}-\mathbf{r}')\frac{dg_{hc}}{d\bar{\phi}_{hc}}\phi(\mathbf{r})\phi(\mathbf{r}') \\ & - \frac{1}{V}\int d^3\mathbf{r}\, d^3\mathbf{r}'\, w_{att}(\mathbf{r}-\mathbf{r}')\frac{dg_{att}}{d\bar{\phi}_{att}}\phi(\mathbf{r})\phi(\mathbf{r}') \end{aligned} \tag{32}$$

Because we consider only one-dimensional profiles across interfaces or in the vicinity of surfaces, we replace integrals of the form $\int d^3\mathbf{r}\, w(\mathbf{r})\phi(\mathbf{r})$ by $\int dz\, w_z(z)\phi(z)$ with

$$w_z(z) = 2\pi\int_0^\infty d\rho\, \rho w\big(\sqrt{z^2+\rho^2}\big) \tag{33}$$

3.2.2.1. Thermodynamic Properties of the Homogeneous System As input into the theory, we require the thermodynamic properties of the spatially homogeneous fluid. Because we divide the free-energy functional into contributions from the repulsive and attractive parts, we require estimates for the two free-energy densities $g_{hc}(\phi)$ and $g_{att}(\phi)$. The thermodynamics of the homogeneous system are completely determined by the free-energy densities and are independent from the weighting functions. First, we use thermodynamic perturbation theory (TPT1) [78–83] to calculate the total free-energy density $g(\phi) \equiv g_{hc}(\phi) + g_{att}(\phi)$, and we subsequently consider the free-energy density caused by repulsive interactions. Then, the attractive contribution to the free-energy density is obtained from $g_{att} = g_{TPT1} - g_{hc}$.

We briefly summarize the salient features of Wertheim's perturbation theory of first order [78–80]; details regarding the implementation of the TPT1 theory to this model are given in Refs. [84, 85]. Using the monomer fluid of Lennard-Jones particles as a reference system, we determine the total free-energy of the polymer solution by calculating the effect of the bonding potential V_{FENE} as a thermodynamic perturbation. The free-energy difference between the polymer solution and the monomeric reference system (LJ) is given by

$$\begin{aligned} \frac{F-F_{LJ}}{k_BT} &= -\ln\frac{\frac{1}{(nN)!}\int\mathcal{N}[\{\mathbf{R}\}]\exp\left(-\frac{E_{LJ}[\{\mathbf{R}\}]+E_{FENE}[\{\mathbf{R}\}]}{k_BT}\right)}{\frac{1}{(nN)!}\int\mathcal{N}[\{\mathbf{R}\}]\exp\left(-\frac{E_{LJ}[\{\mathbf{R}\}]}{k_BT}\right)} \\ &= -\ln\frac{(nN)!}{n!} - \ln\left\langle\exp\left(-\frac{E_{FENE}[\{\mathbf{R}\}]}{k_BT}\right)\right\rangle_{LJ} \end{aligned} \tag{34}$$

where $E_{LJ}[\{\mathbf{R}\}]$ denotes the sum over all Lennard–Jones interactions and $E_{FENE}[\{\mathbf{R}\}]$ represents the sum over the $n(N-1)$ bonding potentials. The last average $\langle\cdots\rangle_{LJ}$ refers to the

reference system of Lennard–Jones monomers. To first order the last term can be approximated in the form (TPT1)

$$\left\langle \exp\left(-\frac{E_{\mathrm{FENE}}[\mathbf{r}]}{k_B T}\right)\right\rangle_{\mathrm{LJ}} \approx \left[\frac{1}{V}\int \mathrm{d}^3\mathbf{r}_{12}\; g_{\mathrm{LJ}}(\mathbf{r}_{12}) \exp\left(-\frac{V_{\mathrm{FENE}}(\mathbf{r}_{12})}{k_B T}\right)\right]^{n(N-1)} \equiv \left[\frac{\delta}{V}\right]^{n(N-1)} \tag{35}$$

where $g_{\mathrm{LJ}}(\mathbf{r}_{12})$ denotes the pair correlation function of the Lennard–Jones monomeric reference fluid. This approximation yields the following relation between the free-energy density of the Lennard–Jones monomer fluid and the polymer solution

$$\frac{F}{V k_B T} = \frac{\phi}{N}\ln\frac{\phi}{N} + \frac{F_{\mathrm{LJ}}}{V k_B T} - \phi\ln\phi - \phi\left(1-\frac{1}{N}\right)\ln\delta \equiv \frac{\phi}{N}\ln\frac{\phi}{N} + \phi g(\phi) \tag{36}$$

In principle, we could determine the pair correlation function in a separate Monte Carlo (MC) simulation of the monomer fluid. However, the self-consistent field (SCF) calculations require the knowledge of the g_{LJ} for many different densities and temperatures. Moreover, the bonded distance in our model is slightly smaller than the typical distance between Lennard–Jones monomers, such that we would have to generate very accurate data for g_{LJ} at small interparticle distances for the evaluation of the above integral. We determined the structure and the thermodynamics of the reference system from an integral equation approach. Two closures to the Ornstein–Zernike equations of the reference system have been considered: the mean spherical approximation (MSA) and the reference hypernetted chain (RHNC). Both closures were found to give similar results [84].

The applicability of TPT1 is mainly limited by two conditions: first, being a mean field theory, TPT1 does not properly account for critical density fluctuations and overestimates the critical temperature of the polymer fluid; second, the perturbative treatment of the effect of the bonding potential is poor if the structure of the Lennard–Jones reference fluid differs strongly from the structure of the polymer solution. In the concentrated regime, the fluid structure is determined by the density and is only weakly perturbed by the bonding potential. Hence, TPT1 gives a good description of the melt. Close to the critical point of the reference fluid [$T_c(N=1) \approx 1$] and at low densities, the structure of the monomer and polymer fluids differs more strongly, so that the agreement is less satisfactory. In addition, TPT1 neither predicts the density dependence of the polymer structure nor properly distinguishes between inter- and intramolecular interactions.

In spite of these caveats, TPT1 yields for our model a good description of the binodals (not too close to the critical point) and the equation of state without any adjustable parameter. Most important, it also provides the free-energy for the unstable homogeneous state inside the miscibility gap without any *ad hoc* extrapolation procedure. In addition, the scaling of the critical point parameters with chain length has been investigated [84]. TPT1 agrees with the scaling predictions of the Flory–Huggins model in the long chain-length limit [86], which are correct because the critical point exhibits tricritical behavior in the limit $N \to \infty$, and TPT1 captures qualitatively some of the deviations observed in simulations of short chains [84]. Moreover, this approach can be generalized to mixtures [85].

The same TPT1 scheme can also be used to obtain the free-energy of the repulsive interaction $g_{\mathrm{hc}}(\phi)$, which we approximate by using those of a system of tangent hard-sphere chains of diameter d_{hc}. We obtain the hard core diameter d_{hc} using the Barker–Henderson [87] recipe; that is

$$d_{\mathrm{hc}} = \int_0^{r_{\mathrm{min}}} \mathrm{d}r\left[1 - \exp(-V_{\mathrm{WCA}}/k_B T)\right] \tag{37}$$

where the integration is extended to the minimum $r_{\mathrm{min}} = \sqrt[6]{2}\sigma$ of the Lennard–Jones potential and $V_{\mathrm{WCA}}(r) = V_{\mathrm{LJ}}(r) + V_{\mathrm{LJ}}(r_{\mathrm{min}})$ for $r < r_{\mathrm{min}}$. For the hard core contribution, we also

use the thermodynamic perturbation theory, which is known to be quantitatively accurate for the equation of state of hard-chain fluids [78, 80].

For tangent hard chains, Eq. (36) simplifies to

$$g_{\rm hc}(\phi) = \frac{F_{\rm ex}^{\rm hs}}{k_B T n N} - \left(1 - \frac{1}{N}\right) \ln \delta_{\rm hs} \tag{38}$$

where $F_{\rm ex}^{\rm hs}/nN$ is the excess free energy per particle of a reference fluid of non-bonded hard spheres, and $\delta_{\rm hs}$ is the contact value of the pair correlation function of the reference fluid. The contact value $\delta_{\rm hs}$ can be obtained from the hard-sphere pressure using the virial theorem; that is, $\delta_{\rm hs} = 3/(2\pi\phi d_{\rm hc}^3)[p/(k_B T\phi) - 1]$. Using the Carnaham–Starling equation of state [88] for hard spheres, we obtain [78]

$$g_{\rm hc}(\phi) = \frac{4\eta - 3\eta^2}{(1-\eta)^2} - \left(1 - \frac{1}{N}\right) \ln \frac{1 - \eta/2}{(1-\eta)^3} \tag{39}$$

where $\eta = \frac{\pi d_{\rm hc}^3}{6}\phi$ denotes the packing fraction.

The difference $g - g_{\rm hc}$ yields then the attractive contribution $g_{\rm att}$. Over the pertinent range of densities, the attractive contribution resembles a linear function of density. This behavior is expected from a van der Waals theory and shows that the decomposition into repulsive and attractive contributions is physically reasonable. In our numerical calculations, we expand the free energy per particle in a fifth-order polynomial in the density.

3.2.2.2. Choice of Weighting Functions The other ingredient in the theory is the choice of weighting functions. In previous work on hard chains [89] and Lennard–Jones chains [28], it has been shown that the theoretical predictions are rather insensitive to the choice of weighting functions. This suggests that some very simple choices for the weighting functions might be adequate. Some guidance in the choice of weighting functions can be obtained from the relation between the second functional derivative of $F_{\rm ex}$ with respect to $\phi(\mathbf{r})$ and the direct correlation function

$$\left.\frac{\delta^2 F_{\rm ex}}{\delta\phi(\mathbf{r})\delta\phi(\mathbf{r}')}\right|_{\phi(\mathbf{r})=\phi(\mathbf{r}')=\phi} = -k_B T c(|\mathbf{r}-\mathbf{r}'|) \tag{40}$$

where $c(|\mathbf{r}-\mathbf{r}'|)$ is the direct correlation function of the homogeneous fluid. Using Eqs. (16) and (31), we obtain

$$\begin{aligned}
&\frac{1}{k_B T}\left.\frac{\delta^2 F_{\rm ex}}{\delta\phi(\mathbf{r})\delta\phi(\mathbf{r}')}\right|_{\phi(\mathbf{r})=\phi(\mathbf{r}')=\phi} \\
&\quad = +2w_{\rm hc}(\mathbf{r}-\mathbf{r}')\frac{{\rm d}g_{\rm hc}}{{\rm d}\bar{\phi}_{\rm hc}} + \int {\rm d}\mathbf{r}''\, w_{\rm hc}(\mathbf{r}-\mathbf{r}'')w_{\rm hc}(\mathbf{r}''-\mathbf{r}')\phi\frac{{\rm d}^2 g_{\rm hc}}{{\rm d}\bar{\phi}_{\rm hc}^2} \\
&\qquad + 2w_{\rm att}(\mathbf{r}-\mathbf{r}')\frac{{\rm d}g_{\rm att}}{{\rm d}\bar{\phi}_{\rm att}} + \int {\rm d}\mathbf{r}''\, w_{\rm att}(\mathbf{r}-\mathbf{r}'')w_{\rm att}(\mathbf{r}''-\mathbf{r}')\phi\frac{{\rm d}^2 g_{\rm att}}{{\rm d}\bar{\phi}_{\rm att}^2}
\end{aligned} \tag{41}$$

where we have assumed our weighting functions $w_{\rm att}$ do not depend on density.

Our first choice, which we will refer to as approximation (1), is to separate the direct correlation function into two contributions: now $c(r) = c_0(r) + c_{\rm att}(r)$, where the first term is the result of repulsive interactions and the second is the result of attractions. In the low-density limit $w(r) \sim c(r)$, we can identify $w_{\rm hc}(r) \sim c_0(r)$ and $w_{\rm att}(r) \sim c_{\rm att}(r)$ [65]. The polymer reference interaction site model (P-RISM) theory [90] is used to calculate the direct correlation functions [65]. The theory requires, as input, the single-chain structure factor for the chain fluid, which one obtains using the semiflexible chain model with the semiflexibility adjusted so that the chain size is the same as in the simulations. The P-RISM equations are solved with the Percus–Yevick closure [65]. The full interaction is employed to calculate $c(r)$, and only the repulsive interaction potential is employed to calculate $c_0(r)$.

A much simpler *ansatz* is to ignore the fluid–fluid correlations altogether and assume that the weighting function merely has the same range as the interaction potential. Our second choice, which we refer to as approximation (2), considers the following

$$w_{\text{att}}(\mathbf{r}) \sim \begin{cases} V_{\text{LJ}}(r_{\min}) & \text{for } r < r_{\min} \equiv \sqrt[6]{2}\sigma \\ V_{\text{LJ}}(r) & \text{for } r_{\min} < r \end{cases} \tag{42}$$

$$w_{\text{hc}}(\mathbf{r}) \sim (1 - |\mathbf{r}|/d_{\text{hc}})\Theta(d_{\text{hc}} - |\mathbf{r}|) \tag{43}$$

where Θ is the Heavyside step function, and the weighting functions are normalized as before. The simple form reproduces the qualitative features of the P-RISM results for the direct correlation function.

In both approximations, $w_{\text{hc}}(r)$ is a roughly linear function of r for distances less than the hard core diameter and almost zero for larger distances. On the other hand, $w_{\text{att}}(r)$ is roughly constant for distances less than the hard core diameter and proportional to the interaction potential for larger distances. These results are quite similar to what is seen in density functional theories of simple liquids.

3.2.3. Comparison Between Monte Carlo Simulations and Self-Consistent Field Calculations

In the following text, we present a quantitative comparison between the self-consistent field scheme outlined above and the Monte Carlo simulations discussed in Section 3.1. This comparison serves to explore the potential and limitations of the self-consistent field calculations.

We begin our comparison with the density profile in the vicinity of a wall that attracts monomers at temperature $k_B T/\epsilon = 1.68$. This temperature is well below the Θ point $k_B T_\Theta/\epsilon = 3.3(2)$ for our model, and the polymer solution phase separates into a liquid of density $\phi\sigma^3 = 0.611$ and a vapor (with $\phi\sigma^3 = 0.0083$). The temperature was chosen because the vapor density is negligible but the density of the liquid phase is not too large to allow for an efficient insertion and deletion of chains via the configuration bias algorithm.

The density profile in the vicinity of the wall is shown in Fig. 4. It is very similar to what one would expect for a fluid of nonbonded monomers. The density vanishes at the impenetrable surface, and on the distance of a few monomeric length scales σ, it reaches the density of the liquid in coexistence with the vapor. Note that this width is comparable to the extension of a monomeric unit and that on this short length scale, the chain conformations cannot faithfully be described by the Gaussian chain model. As the attractive strength of the wall A_w

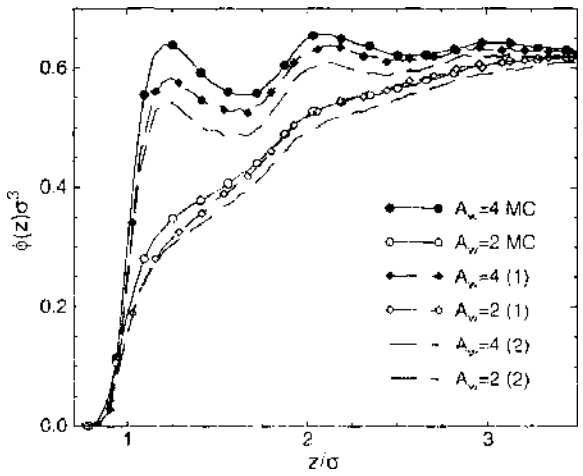


Figure 4. Monomer number density profile at an attractive wall, $A_w = 2$ and 4, for temperature $k_B T/\epsilon = 1.68$. Thick lines with circles correspond to Monte Carlo results, dashed lines with diamonds show the results of weighting functions (1), and dashed lines present the results obtained with weighting functions (2). Adapted from Ref. [65], M. Müller, et al., *J. Chem. Phys.* 119, 2929 (2003).

is increased, the liquid approaches the surface, and oscillations in the density profile gradually build up. These qualitative features are well described by the self-consistent field theory. The details of the packing effects at the wall depend sensitively on the excess free-energy functional employed in the self-consistent field calculations. As can be observed in Fig. 3, the scheme outlined above provides a reasonably good description of the density profile [65].

In Fig. 3 we discuss polymeric aspects of the fluid in contact with the surface. As the molecules are extended, they can use their internal degrees of freedom to minimize the free-energy costs the surface imparts to them. In the left panel, we show the relative density of the chain ends and the center of masses. The chain ends are enriched at the surface, and there is a concomitant depletion region further away from the wall. Likewise, the density of the center of masses is depleted in the ultimate vicinity of the surface [28].

The packing of the chains close to the surface goes along with their orientation. As expected, the chains align parallel to the surface, and the alignment effect is larger the larger the length scale (i.e., the orientation of bond vectors is fairly small and decays faster than the orientation of the end-to-end vector of the polymer). The self-consistent field calculations even capture fine details of these orientational profiles, such as the overshooting of the chain extension perpendicular to the surface at intermediate distances. The excellent account of the configurational statistics of the polymer in the vicinity of the surface is largely independent from the choice of the excess free-energy functional, but it is the result of the partial enumeration scheme. Note that there are no orientations of bond vectors and no dependence of the parallel chain extension on the distance from the wall within the framework of the Gaussian chain model.

In addition to structural quantities, both Monte Carlo simulations as well as self-consistent field calculations provide information about the thermodynamics of spatially inhomogeneous systems. In Fig. 5 we present the liquid–vapor interface tension γ, obtained from grandcanonical Monte Carlo simulations and self-consistent field calculations. The interface tension γ vanishes at the critical point and increases as we reduce the temperature and the miscibility gap opens. As expected from any mean field theory, the self-consistent field calculations overestimate the critical temperature, as they cannot cope with the strong critical fluctuations of the density in the ultimate vicinity of the critical point. The tension vanishes like $\gamma \sim (T_c - T)^{\mu}$, where the critical exponent μ adopts its mean field value 3/2. Accordingly, the self-consistent field calculations also overestimate the interface tension, but the relative deviations become smaller at lower temperatures. This is a very fine test of the free-energy functional, as it involves estimates of the free energy of hypothetical, homogeneous states inside the miscibility gap. In fact, the estimates for the interface tension depend

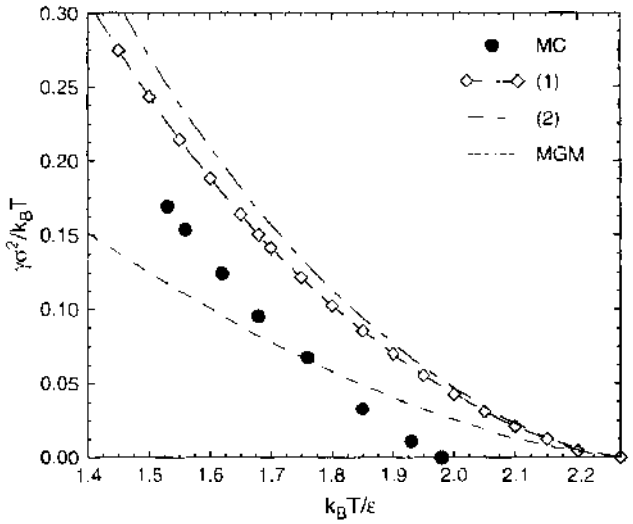


Figure 5. Interface tension as a function of temperature. Circles denote the result of grandcanonical Monte Carlo simulations [28]; the dashed line with diamonds depicts the calculations with weighting functions (1), the dashed line corresponds to weighting functions (2), and the thin dashed line shows the result of a single weighting function (MGM) [28]. Adapted from Ref. [65], M. Müller, et al., *J. Chem. Phys.* 118, 2929 (2003).

very sensitively on the choice of $F_{ex}[\phi]$. Whereas the versions (1) and (2) yield good agreement with the Monte Carlo simulation data, the version (MGM), in which we use the same free energy density but only a single weighting function, yields significantly worse results. In particular, it underestimates the tension at lower temperatures.

4. WETTING OF POLYMER LIQUIDS

4.1. Locating the Wetting Transition in Monte Carlo Simulations

4.1.1. Contact Angle of Microscopic Droplets

As a first application, let us discuss the wetting behavior of polymeric liquids. The most intuitive way to locate the wetting transition is to observe the contact angle of a drop. This is a routine experimental procedure, in which one determines the contact angle of a macroscopic drop with great precision. Below the wetting transition temperature T_{wet} there is only a microscopically thin layer of liquid at the surface, and the excess material forms a drop. Far away from the three-phase contact between the substrate, the liquid and the vapor the liquid–vapor interface makes a contact angle θ with the plane of the substrate. Young's equation [91] describes the balance of lateral forces at the contact line and relates the difference between the surface tensions $\Delta\gamma$ and the interface tension γ to the contact angle θ.

$$\cos\theta = \frac{\Delta\gamma}{\gamma} \tag{44}$$

Observing the dependence of the contact angle on temperature, solvent pressure, or other external parameters tells a great deal about the tensions. On increasing the temperature, the contact angle typically decreases, and it vanishes at the wetting transition temperature. Above T_{wet}, a macroscopically thick liquid film will be stable on the surface.

Although in principle the simulation of droplets is straightforward [92], some care must be taken. In a grandcanonical ensemble, for example, a droplet is unstable, and one will rather find the system forming a microscopically thin, homogeneous liquid film below T_{wet}, but no drop. To observe the droplet as a stable state, one must constrain the number of particles to a fixed value (i.e., work in the canonical ensemble).

The chosen geometry of the simulation box and the total number of molecules turn out to be crucial for the droplets to be stabilized [92], because drop and vapor will exchange particles, and the chemical potential will be shifted away from the liquid–vapor coexistence value in the bulk. On the one hand, if the volume of the simulation cell is much larger than the droplet, the simulation cell will mimic the effect of a grandcanonical reservoir, and rather than forming a droplet the particles will distribute homogeneously in the volume. On the other hand, the surface gives rise to an adsorption of liquid and shifts the chemical potential by an amount $\sim 1/L$ (L, being the linear extension of the simulation cell) away from the coexistence value of the chemical potential in the bulk.

By carefully choosing the number of particles, a drop can be the thermodynamically stable configuration in a simulation cell of finite size. The properties of the nanoscopic droplet that can be observed in the simulations, however, differ considerably from the macroscopic drops studied experimentally [92].

Let us consider $T < T_{wet}$. The shape of the droplet (cf. Fig. 6a) is determined by the interaction of the liquid–vapor interface with the substrate. This interaction can be phenomenologically characterized by the interface potential $g(l)$ that describes the free-energy it costs to place a unit area of the liquid–vapor interface a distance l away from the substrate [8]. Following common notation, $g(l) \to 0$ for $l \to \infty$ (i.e., there is no interaction between the substrate and the liquid–vapor interface if they are infinitely apart). The generic form of the interface potential in the vicinity of a first-order wetting transition is presented in Fig. 6b. $g(l)$ exhibits a minimum at finite, microscopic film thickness, that is separated by a free-energy barrier from the minimum of $g(l)$ for $l \to \infty$. Let l_{min} denote the thickness of the stable microscopic film on the substrate; then the difference in the surface free energies is given by $\Delta\gamma = g(l_{min}) + \gamma$. Below the wetting transition temperature T_{wet}, the value of the interface potential at its minimum is negative, $g(l_{min}) < 0$. On increasing the temperature,

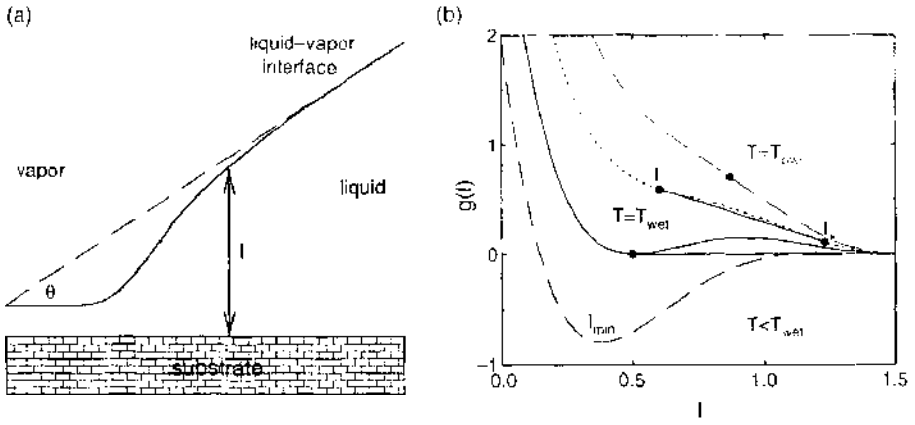


Figure 6. (a) Illustration of the three-phase contact. (b) Sketch of the interface potential. For explanations, see text.

$g(l_{\min})$ increases, and right at T_{wet}, a film of thickness $l_{\min} < \infty$ and an infinitely thick, macroscopic film have the same free energy $g(l_{\min}) = 0$. On increasing the temperature, we make the thickness of the polymer layer jump from a microscopic value to a macroscopic one, and a first-order wetting transition occurs.

Slightly above the wetting transition temperature, there is a coexistence between two microscopic films of different thickness—the so-called prewetting coexistence. Those two coexisting film thicknesses, l_- and l_+, are related by a double tangent construction in Fig. 6b. Recently, it has been suggested that this prewetting transition is related to anomalously low friction [93]. This surface phase coexistence occurs at slight undersaturation compared to the liquid vapor coexistence in the bulk, which corresponds to the slope of the double tangent. On approaching T_{wet} from above, the thickness of the thinner layer decreases and the thickness of the thicker layer diverges. On increasing the temperature, the difference between the thicknesses of the two layers decreases and vanishes at the prewetting critical point at T_{pwc}.

The macroscopic contact angle θ depends only on a single value of the interface potential, and it only determines the asymptotic behavior of the liquid–vapor interface far away from the substrate (as indicated by the dashed line in Fig. 6a). The profile of the droplet probes, however, all distance l between the interface and the substrate, and therefore it contains much more information in principle. In the vicinity of the three-phase line, the shape of the drop will differ significantly from the behavior of a macroscopic drop. The length scale on which those deviations occurs is set by the range of the interface potential.

Intriguingly, one can even observe "flat droplets" above the wetting transition temperature in the range $T_{\text{wet}} < T < T_{\text{pwc}}$. In this interval, there are nonstable finite film thicknesses. If the number of particles in a finite-sized simulation box is such that a homogeneous film would have a film thickness between the microscopically thin and thick layers, l_- and l_+, coexisting on the prewetting line, it will phase separate laterally into a domain with thicknesses l_- and l_+. Of course, the formation of such structures is not in contradiction to Young's equation. On the contrary, "flat droplets" have a finite height and not a spherical cap shape; hence, if one measured the contact angle θ of a (laterally) large "flat droplet" far away from the three-phase contact line, one would obtain $\theta = 0$. If the lateral extension of the droplet is comparable to l_+ (note that $l_+ \to \infty$ for $T \to T_{\text{wet}}^-$), however, one cannot observe the flat portion of the domain. There is a gradual crossover from "true spherical" droplets for $T < T_{\text{wet}}$ to those "flat droplets" slightly above the wetting transition temperature. Hence, those "flat droplets" are not easily recognized as lateral phase separation due to prewetting coexistence.

In Fig. 7 we present radially averaged drop profiles, which have been extracted from simulations at $k_B T/\epsilon = 1.68$ and various attractive strengths of the surface. The box considered had a size of $68\sigma \times 68\sigma \times 68\sigma$ and contained 1000 polymer molecules of length $N = 10$.

For the weakest attraction, $A_{ss} = 2.6$, shown in left panel, a well-defined droplet forms. From the profile we read off an apparent contact angle $\theta_{\text{drop}} = 81(4)°$. This value is comparable with the contact angle of a macroscopic drop as estimated by Young's equation (cf. following), which was found to be $\theta = 70°$. As we increase the value of the attractive

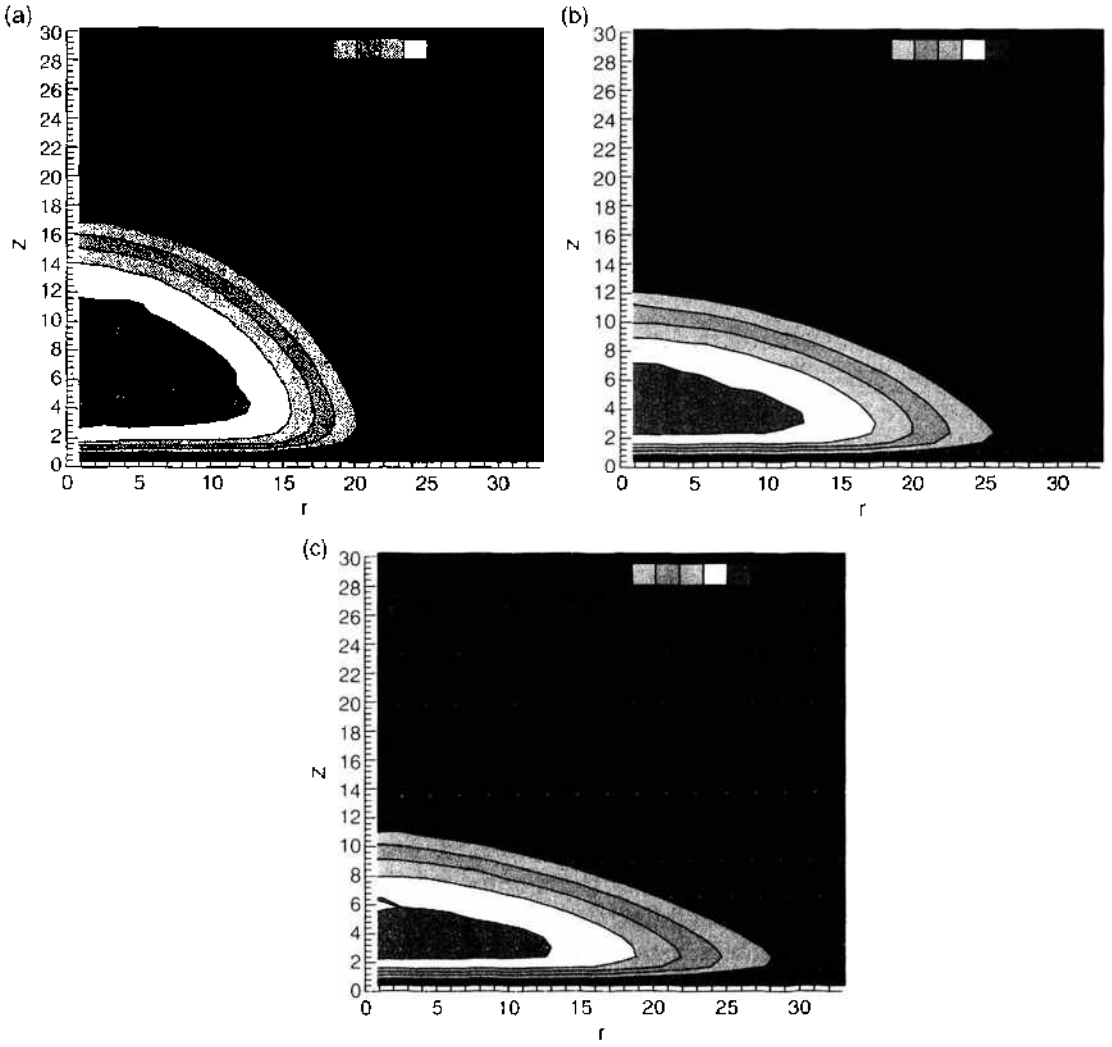


Figure 7. Radially averaged density profile of droplets for the various strengths of the monomer–wall attraction: $A_w = 2.6$ (a) 3.2 (b), and 3.4 (c). Adapted from Ref. [92]. L. G. MacDowell, et al., *Colloids and Surfaces A* 206, 277 (2002).

strength of the substrate A_w, the apparent contact angle decreases gradually. For $A_w = 3.2$ (middle panel), we obtain $\theta_{drop} = 52(4)°$, and for $A_w = 3.4$ (right panel), we estimate $\theta_{drop} = 44(4)°$. Note that, qualitatively, the shape of the small droplet is similar for all three values of A_w, and from an analysis of the apparent contact angle, one would estimate the wetting transition to occur significantly above $A_w = 3.4$. However, the accurate location of the wetting transition is $A_w = 3.22$, and the contact angle of a macroscopic drop at $A_w = 3.2$ is only $\theta = 12°$. The droplet observed for $A_w = 3.4$ is therefore a "flat droplet" whose height is dictated by the equilibrium film thickness of the prewetting coexistence. Note that these considerations of microscopic droplets in the canonical ensemble also might have consequences for polymeric droplets in nanofluidic devices.

Unlike the experimental situation, which relies on the observation of macroscopic drops, the apparent contact angle of microscopic drops yields only a very inaccurate estimate of the macroscopic behavior. In particular, it tends to severely overestimate the contact angle in the vicinity of the wetting transition [92].

4.1.2. Measurement of the Liquid Film Thickness in the Grandcanonical Ensemble

The crux of estimating the wetting transition from the contact angle is twofold: First, the measurement relies on the observation of a spatial inhomogeneity (drop), and only for very large drops does the apparent contact angle agree with the macroscopic one. Second, the size

and geometry of the simulation box matters, because the droplet and the surrounding vapor exchange particles (e.g., the vapor can condense on the surface), such that the system is not exactly at the liquid–vapor coexistence chemical potential. Both issues can be avoided in the grandcanonical ensemble.

The simplest approach consists of performing grandcanonical simulations at the liquid–vapor coexistence chemical potential and to monitor the thickness $l_{\min}$ of the liquid film on the surface. In the nonwet situation, we have either $A < A_{\text{wet}}$ or $T < T_{\text{wet}}$, depending on whether the transition is approached by changing A_w or T, respectively. In this situation, a thin, liquid film is stable at the wall. On approaching a first-order wetting transition, the layer thickness diverges discontinuously from a microscopic value to a macroscopic one. Above the wetting transition, a macroscopically stable film is stable. Thus, starting below the wetting transition with a finite liquid layer, one can increase A_w or T and estimate the location of the transition by the divergence of the film thickness. At a first-order wetting transition, there are, however, two caveats: first, even above the wetting transition, a thin, liquid layer might be metastable (because of the free-energy barrier that separates $l_{\min}$ from the macroscopically thick film), and thus, one estimates the location of the wetting spinodal instead of the wetting transition. At a strong first-order wetting transition, this might result in a significant overestimation of the location of the wetting transition. Second, if one uses a simulation cell with two symmetric walls, as we do in our simulations, and further sets the chemical potential to its coexistence value, the most stable configuration consists of a simulation cell filled with liquid. The monitored state made of two liquid films adsorbed on the wall is metastable, and it becomes thermodynamically stable only at a lower value of the chemical potential (capillary condensation). This metastability of the starting configuration severely limits the length of the simulation runs. Indeed, using such a scheme, we estimate the divergence of the wetting layer for $k_BT/\epsilon = 1.68$ to be at $A_w = 3.4$ [28]. As expected, this value is larger than the accurate location of the wetting transition $A_{\text{wet}} = 3.22$.

The first caveat might be overcome by monitoring the probability distribution $P(l)$ of the layer thickness l. This gives direct information about the interface potential $g(l)L^2 = -k_BT \ln P(l)$, where L^2 denotes the surface area. At a first-order wetting transition, the free-energy barrier between the thin and the (macroscopically) thick layer might be overcome by reweighting techniques.

The second problem could be alleviated by choosing an attractive and a repulsive wall, thus stabilizing a single liquid–vapor interface in the simulation cell. Ideally, one could choose antisymmetric surface interactions such that the contact angle of a liquid drop on the wall equals the contact angle of a bubble at the opposite wall, and no capillary condensation would occur. Although this is a promising route for symmetric mixtures (cf. Section 6), it might not be straightforward if the coexisting phases—such as the liquid and the vapor phases—are not related by symmetry.

4.1.3. Young's Equation

By exploiting the ability of the grandcanonical ensemble to measure free-energy differences, the problems with capillary condensation and metastability may be avoided altogether [28, 69]. In one simulation with standard periodic boundary conditions, one calculates the liquid–vapor interface tension, as described previously. In another set of simulations, one considers a cell made of two parallel walls and periodic boundary conditions in the remaining directions. By monitoring the density fluctuations, one directly obtains the interface tension γ and the difference of the surface tension $\Delta\gamma$ (cf. Fig. 1).

In Fig. 8a, we plot $\cos\theta = \Delta\gamma/\gamma$ at $k_BT/\epsilon = 1.68$ as a function of the attractive strength A_w of the surface. Unlike the measurement of the effective interface potential $g(l)$, one does not regard the free energy of a liquid film bound to the wall, and hence, one need not choose the system size large enough that both the interactions of the liquid–vapor interfaces with the walls and among themselves can be neglected. The ratio $\Delta\gamma/\gamma$ crosses 1 with a finite slope; a fact that indicates that the wetting transition is of first order. The intersection point yields the accurate estimate $A_{\text{wet}} = 3.22$ for the wetting transition at temperature $k_BT/\epsilon = 1.68$.

At small values of A_w, the ratio $\Delta\gamma/\gamma$ approaches the value -1 with a very small (possibly vanishing) slope. This marks the drying transition at a hard, impenetrable wall. In accord

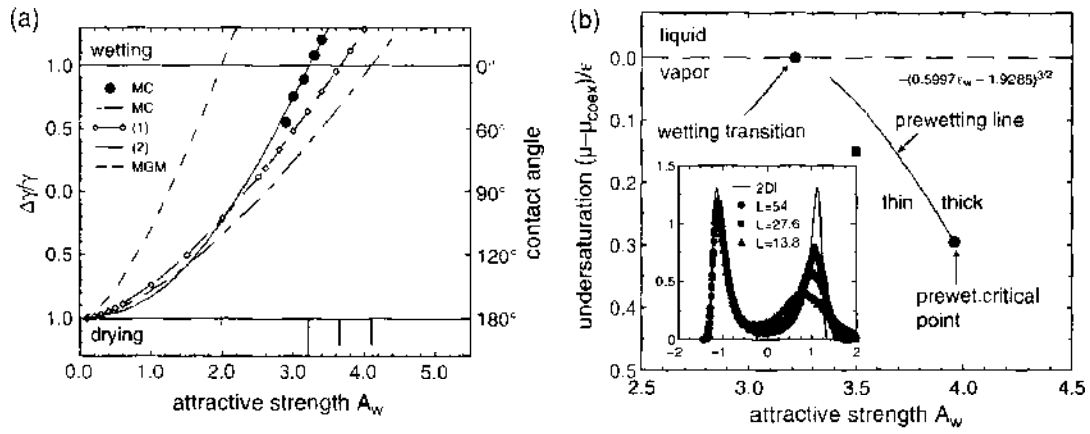


Figure 8. (a) Ratio between the difference in the surface tension $\Delta\gamma = \gamma_{WV} - \gamma_{WL}$ and the interface tension γ as a function of the attraction between monomers and wall at $k_BT/\epsilon = 1.68$. The location of the wetting transitions are shown at the bottom. Filled circles and the thick solid line represent the MC results. The dashed line with open diamonds corresponds to the calculations with weighting functions (1), the thick, dashed-dotted line depicts the results with weighting function (2), and the thin, dashed line is the result of using a single weighting function (MGM). From Ref. [65]. (b) Wetting and prewetting at $k_BT/\epsilon = 1.68$. The horizontal, dashed line marks the liquid vapor coexistence, and the solid curve presents the results for the prewetting line. A fit of the prewetting line according to the Clapeyron equation [Eq. (46)] is shown. The wetting transition and the prewetting critical point are marked. The filled square marks the simulation point used for patterned substrates in Fig. 11. The inset displays the probability distribution of the layer thickness normalized to zero mean and unit variance for the system size $L/\sigma = 13.8$, 27.6 and 54 at $T = 1.68$ and for our estimate of the prewetting critical point $A_w = 3.96$. At this critical point, the normalized distribution depends only on the universality class, and the corresponding function for the two-dimensional Ising model is also shown. Adapted from Ref. [28], M. Müller and L. G. MacDowell, *Macromolecules* 33, 3902 (2000).

with observations of Lennard–Jones monomer fluids [94, 95], this transition is very weakly first order, and we cannot rule out a second-order transition. A careful study of finite size effects is necessary [95] to accurately pinpoint the location of the drying transition and its order.

To investigate the wetting behavior of the polymer liquid in the SCF theory, we calculate the surface free energy as a difference between the grandcanonical potential of the system in the presence of a wall and that of a homogeneous system

$$\frac{\gamma}{k_BT} \equiv \frac{\Omega - \Omega_{\text{hom}}}{k_BTA} = \frac{(F - \mu\phi_{\text{av}}V + pV)D}{Vk_BT} \tag{45}$$

where Ω_{hom} is the grand canonical free energy of a homogeneous system and A denotes the area of the wall. The results of the self-consistent field calculations are compared with the Monte Carlo simulations in Fig. 8a. If one uses a simple weighted density functional (MGM), one obtains only qualitative agreement. The versions that decompose the interaction free-energy density into a short-ranged, repulsive and a longer-ranged, attractive contribution yield a better, quantitative agreement with the simulation data.

4.1.4. Extrapolating the Prewetting Line toward Coexistence

If the wetting transition is of first order, there will be a concomitant jump in the layer thickness above the wetting temperature away from the coexistence curve in the A_w–μ plane. At this prewetting line, a thin, absorbed layer coexists with a thick layer. In the vicinity of the prewetting line, the probability distribution of the layer thickness is bimodal, and we can accurately locate the prewetting coexistence by the equal-weight rule (cf. Section 3.1). The prewetting line is presented in Fig. 8b and it terminates in the prewetting critical point. Here, the difference in the layer thicknesses vanishes. This transition is believed to belong to the two-dimensional Ising universality class [96]. Because the correlation length of density fluctuations parallel to the surface diverges at this point, effects of the finite lateral extension of the simulations cell have to be carefully investigated. The normalized probability distribution of the layer thickness at our estimate for the prewetting critical point is presented in

the inset of Fig. 8b. As a result of field mixing effects [97], the distribution is asymmetric. As expected from a correction to the leading finite size scaling behavior, on increasing the system size, we decrease the asymmetry of the distribution. Adjusting the attractive strength of the substrate to map the normalized distribution onto the universal scaling function of the two-dimensional Ising universality class, we obtain the estimate $A_{\mathrm{pwc}} = 3.96(6)$ for the prewetting critical point. This value is also marked in the wetting phase diagram Fig. 8b.

The approach of the prewetting to the coexistence curve has been studied by Hauge and Schick [98]. Using a Clapeyron equation along the prewetting coexistence line, one obtains for the approach of the prewetting line toward the bulk coexistence curve [98].

$$\left.\frac{\mathrm{d}\Delta\mu}{\mathrm{d}A}\right|_{\mathrm{prewet}} \approx -\frac{N(9/l_{\mathrm{thin}}^{8} - 3/l_{\mathrm{thin}}^{2})}{l_{\mathrm{thick}}} \sim -\frac{1}{l_{\mathrm{thick}}} \sim \Delta\mu^{1/3} \tag{46}$$

Above T_{wet}, the layer thickness increases like $l_{\mathrm{thick}} \sim |\Delta\mu|^{-1/3}$ (complete wetting). (If we used only short-ranged forces, there would only be a logarithmic divergence.) Integration of Eq. (46) yields $\Delta\mu|_{\mathrm{prewet}} \sim (A_w - A_{\mathrm{wet}})^{3/2}$ [98]. This dependence can be used to extrapolate the prewetting line toward the coexistence curve and to estimate the location of the wetting transition. The approach of the prewetting line in the MC simulations and the SCF calculations is presented in Fig. 8b. The extrapolation agrees very well with the estimate obtained from the Young equation.

To summarize, the self-consistent field calculations describe the conformational properties of polymers in the vicinity of surfaces and interfaces quantitatively, but thermodynamic properties and the details of density oscillations in the vicinity of the surfaces depend on the approximation for the interaction free-energy functional F_{ex}. Two versions of the the self-consistent field calculations have been discussed that use a weighted density-functional decomposing the interactions into short-ranged, repulsive and longer-ranged, attractive contributions. Having assessed the accuracy of the self-consistent approach by quantitatively comparing it to Monte Carlo simulations, we demonstrate its usefulness by discussing in the following sections some applications of the self-consistent field calculations to slightly more complicated but experimentally relevant situations.

4.2. Nanodroplets on Layered Substrates

Dewetting experiments can probe the subtle features of the interface potential. The minimum of the interface potential is related to the contact angle θ via $g_{\mathrm{min}} = \gamma(\cos\theta - 1)$, and the kinetics of the dewetting process tells a great deal about the curvature of the interface potential: If the curvature is negative, the spinodal dewetting occurs that is characterized by a dominant lateral length scale. If the curvature is positive, the film is locally stable, and dewetting (if it occurs) proceeds via nucleation of holes [13].

A standard experimental system is polystyrene (PS) on a silicon (Si) wafer that is coated by a thin oxide layer of thickness d_{ox}. Such a situation is sketched in Fig. 9a. Experiments by Müller-Buschbaum et al. [101] reveal a rather intriguing dewetting behavior, as illustrated in Fig. 9. At high temperature, $T = 165°\mathrm{C}$, the polymer dewets the substrate and forms drops. These macroscopic drops sit on a mesoscopic film of thickness $l^* = 2.2(6)$ nm. On cooling below the glass transition temperature $T_g \approx 100°\mathrm{C}$, this mesoscopic film between the macroscopic drops dewets and forms nanodroplets of diameter ~70 nm. As the polymer film becomes glassy, these nanodroplets do not coalesce further into macroscopic drops, and one observes a dewetting morphology with two distinct drops sizes: macroscopic drops and nanodroplets.

Assuming that the dewetting morphology observed in the experiments is a thermodynamic equilibrium phenomena, we can rationalize this two-stage dewetting process by an interface potential that exhibits two minima—one that corresponds to the mesoscopic film that is stable at high temperatures, and one that corresponds to a microscopic, vanishingly small thickness $l \approx 0$, which is stable at lower temperatures. How can we explain the occurrence of two minima in the interface potential?

In general, the interface potential comprises a long-ranged and short-ranged contribution. The long-ranged forces stem from the van der Waals interactions inside the fluid and between

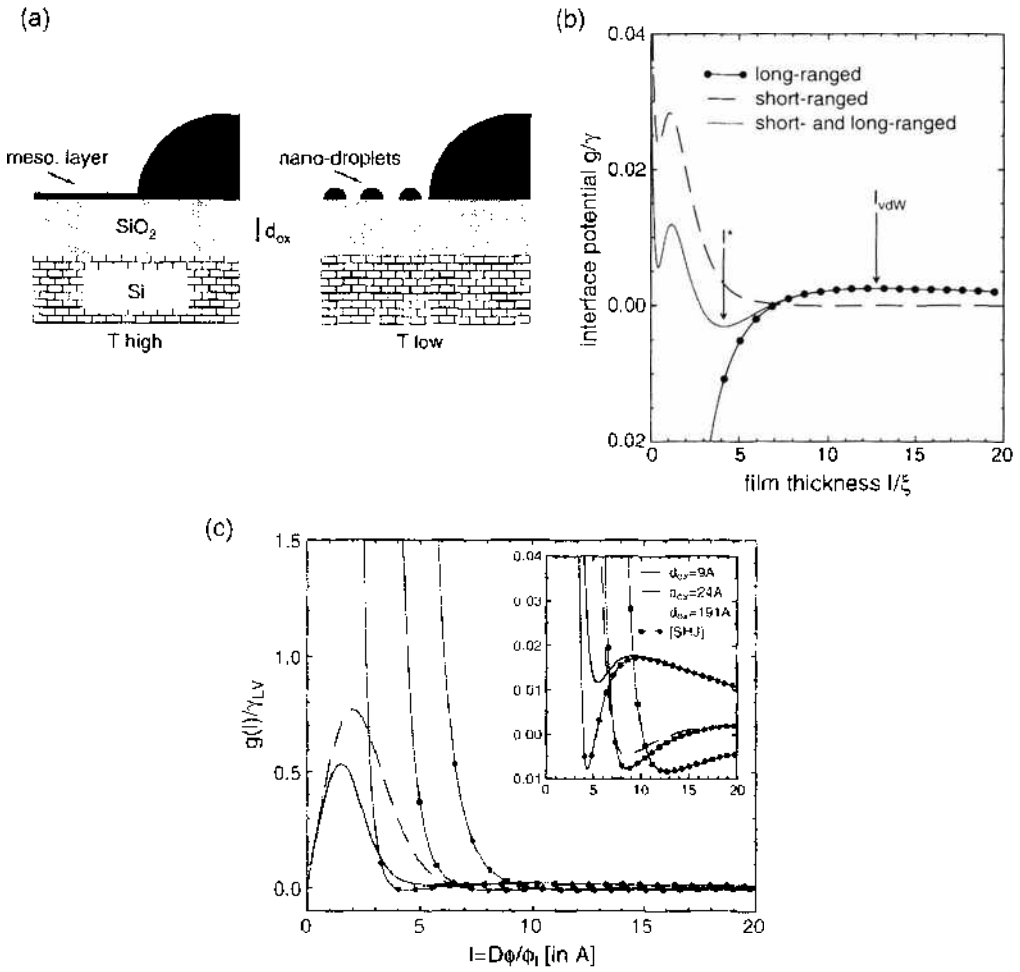


Figure 9. (a) Sketch of the polymer film on top of a bilayer substrate, consisting of Si and a thin oxide layer. The variable l denotes the thickness of the polymer film, and d_{ox} the thickness of the oxide layer. At high temperature, macroscopic drops sit on top of a mesoscopic film. At low temperature, the mesoscopic film becomes unstable and breaks up into smaller drops. (The relative size of the nanodroplets and the macroscopic drops is not to scale.) (b) Schematic illustration of nanodewetting: Weak but long-ranged van der Waals interactions determine the interface potential at long distances. They repel the interface at large distances $l > l_{vdW}$ and attract the interface at intermediate distances. Strong but short-ranged contributions to the interface potential arise from the distortion of the density profile in the vicinity of the substrate. These forces tend to stabilize a microscopically thin film, and they repel the interface at intermediate and large distances. The interplay between van der Waals and short-ranged interaction might result in two competing minima. (The relative magnitude of the short- and the long-ranged forces is not to scale). (c) Effective interface potential $g(l)$ for various oxide layer thicknesses obtained from self-consistent field calculations. The contact interaction V_c has been chosen such that the minima at $l \approx 0$ and $l \to \infty$ have almost equal free energy: $V_c = 3.26\epsilon$ for $d_{ox} = 9\text{Å}$, $V_c = 1.4\epsilon$ for $d_{ox} = 24\text{Å}$, and $V_c = 0.89\epsilon$ for $d_{ox} = 191\text{Å}$. The dashed-dotted line, labeled [SHJ], corresponds to the results of Ref. [99]. The inset presents an enlarged view of the minimum at the mesoscopic film thickness. Adapted from Ref. [100]. M. Müller, et al., *J. Chem. Phys.* 115, 9960 (2001).

the fluid and the wall. The Hamaker constant $A_{ij} > 0$ parameterizes the strength of attractive, long-ranged interactions between atoms of species i and j: $V(\mathbf{r}_{ij}) = -A_{ij}/(\pi^2\phi_i\phi_j r^6)$, where ϕ_i and ϕ_j denote the number densities of species i and j, respectively. The contribution of these long-ranged attractions to the energy E in a situation that is depicted in Fig. 9a is given by

$$E(l) \equiv -\frac{A_{\text{PSSi}}}{\pi^2}\int_{-\infty<z<-d_{ox}} d^3\mathbf{r}\int_{0<z'<l} d^3\mathbf{r}'\,\frac{1}{(\mathbf{r}-\mathbf{r}')^6} - \frac{A_{\text{PSO}_x}}{\pi^2}\int_{-d_{ox}<z<0} d^3\mathbf{r}\int_{0<z'<l} d^3\mathbf{r}'\,\frac{1}{(\mathbf{r}-\mathbf{r}')^6}$$
$$-\frac{A_{\text{PSPS}}}{2\pi^2}\int_{0<z<l} d^3\mathbf{r}\int_{0<z'<l} d^3\mathbf{r}'\,\frac{1}{(\mathbf{r}-\mathbf{r}')^6} \qquad (47)$$

where A_{PSPS}, A_{PSSi}, and A_{PSOx} denote the Hamaker constants inside the polymer fluid, between the polymer and the substrate, and between the polymer and the O_x coating layer. The integrations are extended over the volumes of the materials. The factor 1/2 in the last contribution takes account of the double counting of interaction pairs in the polymer layer. We have assumed a rectangular shape for the density profile (sharp-kink approximation). This is a reasonable assumption for large film thicknesses, where the major contribution to the interface potential comes from the bulk-like liquid layer on the substrate.

Both in the Monte Carlo simulations and in the self-consistent field calculations, the interactions inside the fluid are strictly short-ranged. Therefore, when we want to describe a specific experimental system, the monomer-wall interaction A_w also has to account for the cutoff in the Lennard–Jones interactions between the fluid particles (tail correction), and the monomer–wall interaction comprises both the long-ranged portion of interactions between the monomeric units and the wall and those among the monomeric units inside the fluid. (In principle, there are also contributions from the Si–Si interaction and the Si–O_x interactions, but these do not depend on the thickness l of the polymer film and can therefore be omitted.) The integrals are elementary. Keeping only l dependent terms, we obtain, for the long-ranged contribution to the interface potential, $g(l) \equiv E(l)/L^2$ [102]

$$g_{lr}(l) = \frac{1}{12\pi}\left(-\frac{A_{PSPS} - A_{PSOx}}{l^2} + \frac{A_{PSSi} - A_{PSOx}}{(l + d_{ox})^2}\right) \tag{48}$$

Specifically for PS on an oxide-coated Si wafer, the Hamaker constant of the oxide coating is smaller than the Hamaker constants of the polymer and the substrate, and both terms in the equation above tend to cancel. In this case, the long-ranged part of the interface potential alone exhibits a maximum at a film thickness $l_{vdW} \sim d_{ox}$ of a few nanometer, as sketched in Fig. 9b.

The expression (48) is based on the sharp-kink approximation and holds only for large film thicknesses l. It cannot, however, be applied for film thicknesses on the order of the width of the liquid–vapor interface: the density profile of a thin film is nontrivial (cf. Fig. 4) and differs from the rectangular profile assumed in the sharp-kink approximation (most notably, the density gradually vanishes in the vicinity of the surface); and the interaction between the monomeric units and the surfaces at short distances is not describable by $1/r^3$-type attractions. These two facts lead to an unphysical divergence of this approximation for the long-ranged contribution at small distances. The long-ranged contribution has to remain constant in the limit $l \to 0$ because there is no fluid to interact. Therefore, the long-ranged contribution has be cut off at a microscopic distance.

The other contribution to the interface potential stems from the distortion of the liquid–vapor interface profile resulting from the presence of the substrate. Its characteristic length scale is set by the decay of the interface profile in the wings of the profile (i.e., the correlation length of density fluctuations in the bulk). For a polymer melt, this is a microscopic length scale on the order of a few Ångstroms. The strength of the short-ranged contribution is set by the interface tension γ. The short-ranged contribution to the interface potential is also shown in Fig. 9b. In the vicinity of the wetting transition, it exhibits a minimum for small film thicknesses that is separated by a barrier from the plateau for $l \to \infty$.

Balancing short- and long-ranged contributions, one finds that the short-ranged contribution dominates for

$$l \ll l_{cross} = \xi \ln\left(\frac{\gamma}{g_{lr}(l_{cross})}\right) \tag{49}$$

Hence, short-ranged forces are important for extremely thin polymer films or when the Hamaker constant, which controls the strength of the long-ranged interaction, is small. The latter fact holds true for the specific experiment because the long-ranged interactions of the Si and O_x have opposite signs, and this cancellation leads to a very weak effective (i.e., thickness-dependent Hamaker constant).

Adding both contributions to the interface potential, we obtain two minima: the minimum at microscopic film thickness is only caused by short-ranged forces. The corresponding

barrier signals the free-energy costs of forming a (hypothetical) homogeneous film that is thinner than ξ. The second minimum, at a mesoscopic film thickness $l^* < l_{vdW}$, arises from the interplay of both contributions: The short-ranged part repels the liquid–vapor interface from the substrate, but it rapidly decays as the film thickness grows. The long-ranged contribution attracts the liquid–vapor interface for thickness $l < l_{vdW}$.

This qualitative explanation relies on two assumptions: first, the phenomenological form of the short-ranged contribution neither specifies the length scale ξ of the short-ranged interaction, nor its strength. Additional contributions that might arise from the entropy loss of the polymers because of the confinement into a thin film are neglected. Second, the way the long-ranged contribution to the interface potential has to be cut off at small distance remains unclear—it depends on the details of the density profile in the vicinity of the substrate. In the following text, we use self-consistent field calculations, which do not invoke these assumptions, to corroborate the explanation of the experimental finding. Starting from the interactions between the liquid and the substrate, we calculate the detailed form of the density profile and the free energy of the liquid film.

To compare the self-consistent field calculations to the experiments, we chose the parameters of our computational model to be comparable to the experiments on PS. The Lennard–Jones chains of N beads do not capture details of 8.7 k PS on an atomistic scale, nor do we adjust the equation of the state of our model to experimental data for the pVT behavior of PS. Therefore, we cannot expect agreement between our computational model and the experiments on all quantities simultaneously, and we identify the model parameters to match quantities pertinent to the wetting behavior.

We use chain length $N = 84$, which corresponds to the number of repeat units in the experiment [101]. As the number of single-chain conformations increases exponentially with chain length, the partial enumeration scheme would require a very large number of single-chain conformations to capture the orientation and deformation of the chains at the substrate and the liquid–vapor interface. Therefore, we use the Gaussian chain model, for which the single-chain partition function can be calculated exactly numerically. The statistical segment length $b = 1.3\sigma$ is chosen so as to reproduce the chain extension of the bead-spring model. The length scale in the calculations is set by the the radius of gyration $R_g = b\sqrt{N/6}$. In the calculations, $R_g \approx 4.8\sigma$ while the experimental value is $R_g = 2.54$ nm. Hence, we identify $\sigma = \sqrt{6}R_g/[b/\sigma]\sqrt{N} = 0.522$ nm. Not including any architectural details on the segmental length scale, we use a local density functional [i.e., $w_{hc}(r) = w_{att}(r) = \delta(r)$] in junction with the TPT1 equation of state (employing the MSA closure for the monomer reference fluid). The density of PS at $T = 413$ K is $\phi = 1\text{g/cm}^3$. This corresponds to $6.9 \cdot 10^{-5}$ molecules per A^3, or a segment density of $\phi_l = 6.9 \cdot 10^{-5}(\sqrt{6}R_g/A)^3/b^3\sqrt{N} = 0.83\sigma^{-3}$ in Lennard–Jones units. We adjust the temperature k_BT/ϵ such that the density of the liquid (in coexistence with the vapor) matches the experimental density. This yields $k_BT/\epsilon \approx 0.96$ in Lennard–Jones units (or $\epsilon = 0.57 \cdot 10^{-20}J$). The critical temperature $T_c \approx 2.96\epsilon/k_B T$ for chain length $N = 84$ corresponds to 1300° K, which is well beyond the thermal stability limit. For these parameter values, the self-consistent field calculations yield, for the interface tension, the value $\gamma = 0.36\epsilon/\sigma^2$. Using the estimates for σ and ϵ, we find this value to be about a factor of four smaller than the experimental data $\gamma = 0.031 \cdot 10^{-20}J/A^2$ at $T = 140$°C. If we used a nonlocal density functional with different weighting functions for attractive and repulsive interactions, the self-consistent field calculation would yield a larger interface tension (cf. Fig. 5). We emphasize, however, that no experimental input about the pVT behavior has been used to parameterize the free-energy density, and no specific information about the atomistic structure of PS has entered the calculation. As a result, we consider the self-consistent field result for the interface tension γ as an estimate of a typical interface tension of a polymer liquid in coexistence with its vapor, rather than a prediction for the specific system at hand.

The long-ranged part of the effective interface potential has been extracted from a careful analysis of dewetting experiments [99]. Although the experiments of Seemann et al. employ shorter chains (2k PS) [99], we do not expect the van der Waals interactions (or the phase

diagram far below the Θ temperature) to exhibit a strong dependence on chain length. The experimental data can be parameterized by

$$\frac{g_{\text{exp}}(l)}{\gamma} = A\left(-\frac{1}{l^2} + \frac{r}{(l+d_{\text{ox}})^2} + \frac{c(d_{\text{ox}})}{4l^8}\right) \tag{50}$$

with parameters $A = (A_{\text{PSPS}} - A_{\text{PSOX}})/12\pi\gamma = 1.88A^2$, $r = (A_{\text{PSSI}} - A_{\text{PSOX}})/(A_{\text{PSPS}} - A_{\text{PSOX}}) = 6.9$, and $c(d_{\text{ox}}) = 3.5 \cdot 10^5 A^6$ for an oxide layer of $d_{\text{ox}} = 2.4$ nm [99]. The first two terms represent the long-ranged interaction according to Eq. (48). The value of the coefficients are compatible with estimates obtained from the Hamaker constants of the pure materials [101].

To describe the experimental finding, the description of the interface potential has been augmented by an additional (long-ranged) term, which decays like $1/l^8$. This term dominates g_{exp} at small distances, and its strength has been adjusted to yield the measured contact angle of the polymer drops for each thickness of the coating layer. The form of the potential has a single minimum at short distances.

Using the sharp-kink approximation, we can calculate the monomer–substrate interactions that would correspond to the interface potential above:

$$\frac{V_{\text{wall}}(z)}{\gamma} = A_{2\,\text{layers}}\left(\frac{1}{l^3} - \frac{r}{(l+d_{\text{ox}})^3} - \frac{c}{l^9}\right) \tag{51}$$

with $A_{2\,\text{layers}} = 2A/\phi_l$.

So that the inaccuracy of our model and the approximation of our self-consistent field calculations in predicting the interface tension do not upset the comparison with the experiment, we choose the strength of monomer–wall potential V_{wall} such that the dimensionless ratio of the van der Waals contribution to the effective interface potential and the interface tension takes the experimental value. Using $A_{2\,\text{layers}} = 2A/\phi_l$ with $A = 1.882A^2 = 0.069\sigma^2$, we find $A_{2\,\text{layers}} = 0.167\sigma^5$. The additional term l^{-9} in the interface potential could be (formally) conceived as a strong $1/r^{12}$ attraction between the polymer segments and the constituents of the substrate, but it should instead be considered as an effective description of the attraction between the polymer and the substrate at short distances.

The monomer–wall potential [Eq. (51)] is appropriate for intermediate and large distances of l, which is the actual range of film thickness the dewetting experiments are able to probe. If we used Eq. (51) for all film thicknesses, we would seriously underestimate the contact angle of the polymer drops on the substrate. Unfortunately, details of the monomer–wall potential at short distances are not known. We expect V_{wall} to be repulsive in the ultimate vicinity of the wall, because of the short-ranged repulsion (hard core interaction) between the polymer segments and the constituents of the oxide layer. At short distances, there is an attraction between the polymer segments and the wall. This contribution to the free energy is partially the result of van der Waals interactions between polymer and oxide and partially caused by specific interactions between PS and the SiO_x layer on the atomic scale. This short-ranged behavior of V_{wall} cannot be faithfully modeled in the framework of our coarse-grained model; an accurate modeling of the liquid and the surface on the atomistic scale would be required.

As a minimal model, we use the monomer–wall potential [Eq. (51)] up to distances at which $V_{\text{wall}} < -V_c$. For smaller distances, we set $V_{\text{wall}} = -V_c$ and $V_{\text{wall}} = \infty$ for $z < 0$ (i.e., inside the oxide layer). The value of the contact potential V_c determines the contact angle of droplets on the bare substrate; the contact angle of drops on the mesoscopic film is independent from V_c. The values V_c used in our calculations (cf. Fig. 9) are comparable to the energy scale of thermal fluctuations k_BT but about two orders of magnitude lower than the energy of a chemical bond—it would be very difficult for an *ab initio* calculation to predict those small free energies accurately. One should instead conceive of this quantity as an effective, mesoscopic parameter that has been adjusted to reproduce the experimentally observed behavior. A rather small change in the contact interaction V_c can change the wetting behavior—this points to a strong dependence of the wetting behavior on the preparation of the surface (e.g., the cleaning process [7]).

The dependence of the interface potential $g(l)$ on the thickness of the oxide layer is shown in Fig. 9c. Indeed, the interface potential exhibits two minima in agreement with the phenomenological considerations above. The local minimum at mesoscopic thicknesses is stabilized by an increase in the thickness of the oxide layer, and its thickness l^* grows. The thicker the oxide layer, the further out is the mesoscopic minimum and the more the short-ranged interactions g_{sr} have decayed. In fact, for a thick oxide layer ($d_{ox} = 19.1$ nm), our calculations are by construction in good agreement with the mesoscopic film thickness $l^* = 1.3(2)$ nm and the contact angle of 7.5° measured by Seemann et al. [99]. If the thickness of the oxide layer d_{ox} becomes too small, however, the mesoscopic film becomes unstable with respect to a macroscopically thick film or a microscopically thin film. This is an agreement with experiments [101]: A thin oxide layer ($d = 0.9$ nm) cannot stabilize a mesoscopic film, and no nanodroplets can be observed.

Hence, the self-consistent field calculations corroborate the qualitative explanation that the field formation of nanodroplets mirrors the interplay between short-ranged and long-ranged contributions to the interface potential. The calculations also illustrate the relative magnitude of both contributions.

4.3. Wetting on a Polymer Brush

Grafting or adsorbing chains to the substrate is another route to tune the wettability [103–106]. The grafted polymers may "tie down" [104] the liquid film and prevent rupture. Their stabilizing effect is twofold: Grafted chains might inhibit dewetting kinetically, or they modify the interface potential as to form a thermodynamically stable wetting layer. We shall discuss only the latter effect and constrain ourselves to the most symmetric case of an oligomeric liquid on top of a chemically identical brush.

According to Young's equation [91] the polymer liquid will wet the brush if

$$\gamma_{Bl} - \gamma_{Bl} + \gamma \leq 0 \tag{52}$$

Typically, the tension at the liquid–vapor γ and brush-vapor γ_{Bl} interface are large but of comparable magnitude, whereas the free-energy cost γ_{Bl} of the interface between the brush and the liquid of free chains is small. Hence, wettability is controlled by a subtle balance of all three contributions.

Much effort has been focused on the calculation of the surface tension γ_{Bl} between the brush and the liquid, which stems from the interplay between the translational entropy the liquid gains by penetrating into the brush and the configurational entropy loss that follows when the chains of the brush are swollen by the free chains. Scaling considerations [107, 108] and self-consistent field calculations [109–113] of incompressible melts with purely repulsive interactions have been applied with success to the manner in which a concentrated solution of free chains penetrate into the brush. At low grafting densities, the penetration is large and γ_{Bl} is negative, but there is little penetration at large grafting densities, and γ_{Bl} is positive. This dependence on the grafting density partially rationalizes the experimental observation that a polymer liquid does not wet a brush of identical monomers at high grafting densities [105, 114, 115]—a phenomenon termed "autophobicity."

The properties of the liquid–vapor or brush–vapor interface, however, cannot be described by models that invoke an incompressibility constraint, and the effect of long-ranged van der Waals interactions between the liquid and the substrate is often neglected. In our self-consistent field calculations, we capture compressibility effects and can calculate all three interface tensions within a unified framework. Moreover, the self-consistent field scheme [64, 116] captures the chain conformations on all length scales, whereas the parameters of a hypothetical, noninteracting Gaussian chain might be difficult to identify for experimental or simulational realizations of brushes [117]. In the following text, we illustrate the qualitative behavior at temperature $k_BT/\epsilon = 1.68$ and chain length $N = 10$ using the simplified interaction-free functional (MGM, cf. Section 3.2). The average area per grafted chain is denoted as Σ. The first monomer of the grafted chains is placed a distance $\Delta z = 1.2\sigma$ away from the wall, whereas free chains sample all spatial positions. We investigate the wetting

properties under conditions in which the liquid coexists with its vapor in the bulk (i.e., we do not discuss the prewetting transitions).

Three regimes of the reduced grafting density R_e^2/Σ can be distinguished.

4.3.1. Low Grafting Density

In Fig. 10a, we present the interface potential as a function of the average area Σ per grafted chain in the absence of long-ranged interactions $A_w = 0$ (i.e., a hard repulsive wall). For low grafting densities, the minimum of $g(l)$ occurs close to the wall. The film thickness that corresponds to this minimum is independent of the amount of grafted chains, but its depth decreases as we increase the grafting density. Free chains penetrate into the collapsed brush (cf. Fig. 10c), but the attraction between the brush and the polymer liquid is not strong enough to make the liquid wet the substrate. An attractive long-ranged interaction $A > 0$ is necessary to bring about a first-order wetting transition at A_{wet}. As we increase the grafting density R_e^2/Σ, the wetting transition occurs at smaller values of A_{wet}. The dependence of the wetting transition on the grafting density is presented in Fig. 10b.

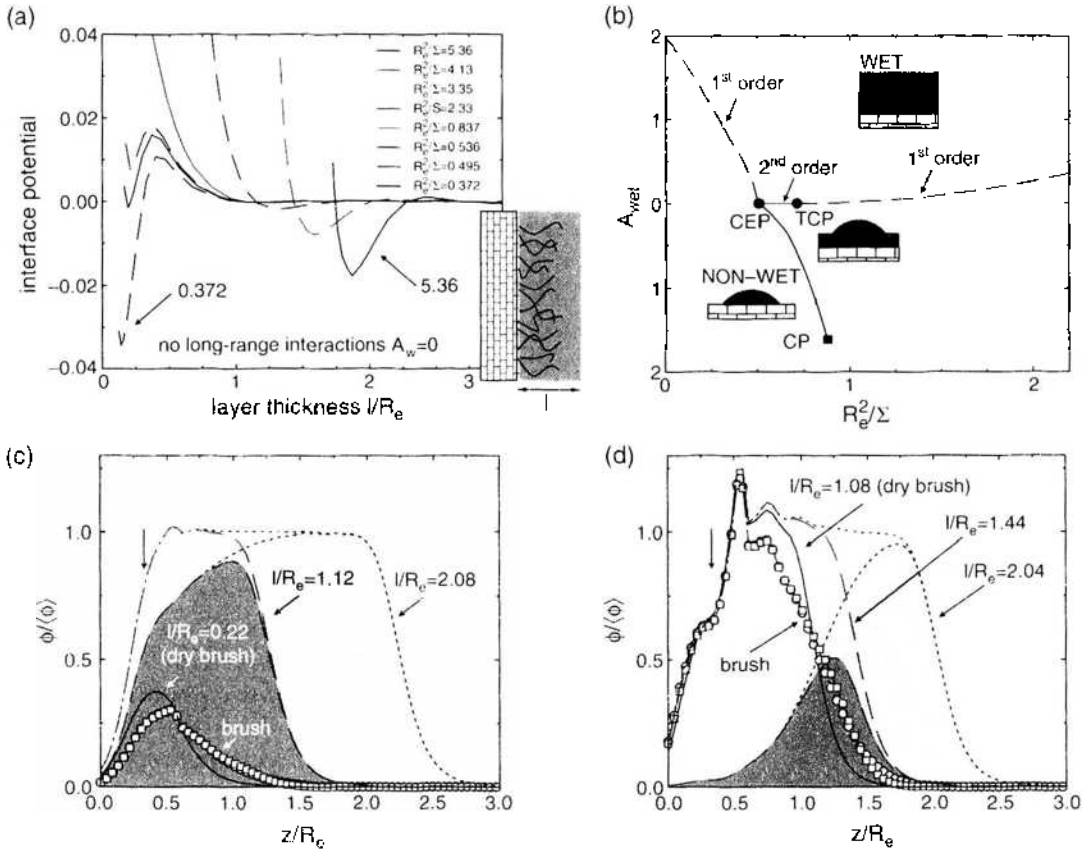


Figure 10. (a) Interface potential $g(l)\sigma^2/k_BT$ for various grafting densities R_e^2/Σ as indicated in the key. There are no long-ranged interactions ($A_w = 0$) between the substrate and the monomers. (b) Wetting phase diagram in terms of the effective Hamaker constant A_{wet} and the grafting density R_e^2/Σ. Dashed lines denote first-order wetting transitions, the solid curve denotes transitions between a microscopic thin and a mesoscopic thick layer, and the horizontal line at $A_{wet} = 0$ marks second-order wetting transitions. The line of transitions between the thin and thick layers terminates in a critical endpoint (CEP) at low grafting densities and in a critical point (CP) at high grafting densities. Second-order and first-order wetting transitions at high grafting densities are separated by a tricritical point (TCP). (c, d) Density profiles at intermediate $R_e^2/\Sigma = 0.69$ (c) and high $R_e^2/\Sigma = 3.35$ (d) grafting densities for $A_w = 0$. Film thicknesses are indicated in the key; the lowest value corresponds to a dry brush. The first monomer of each grafted chain is excluded from the profiles; its position is indicated by the vertical arrow. Thick lines present the total density profile $\phi = \phi_g + \phi_f$. For the two larger thicknesses, lines with symbols represent the profiles of grafted chains, and shaded areas correspond to free chains. Adapted from Ref. [64]. M. Müller and L. G. MacDowell, *Europhys. Lett.* 55, 221 (2001).

4.3.2. Intermediate Grafting Density

On increasing the grafting density further, the brush creates enough attraction to make the fluid wet the substrate at $A_{wet} = 0$. The line of first-order wetting transitions terminates in a critical endpoint (cf. panel 10b). For larger grafting densities, wetting transitions are of the second order (i.e., the thickness of the wetting layer increases continuously as $A \to A_{wet} = 0^-$ is approached). For negative values $A < 0$, the liquid cannot wet the substrate, and the contribution of the brush to the effective interface potential is short-ranged, and cannot outweigh the long-ranged interactions for large (macroscopic) film thicknesses. The behavior resembles the experimental observations ("frustrated wetting") of alkanes at the water–air interface [118–121]. The short-ranged interactions favor a thick liquid layer, but the long-ranged interactions $A < 0$ inhibit the growth of a macroscopic liquid layer. The line of first-order wetting transitions at low grafting densities $A_{wet} > 0$ continues into the regime of second-order wetting in the form of a transition between a microscopically thin and a mesoscopically thick layer at $A_{tt} < 0$. This behavior is different from the prewetting behavior because it occurs at liquid–vapor coexistence $\Delta\mu = 0$ and not at undersaturation $\Delta\mu < 0$. For $A < A_{tt}$, a drop sits on top of a microscopically thin layer, while for $A_{tt} < A < 0$, the drop sits on top of a mesoscopically thick layer. Such a jump in the film thickness at coexistence, which precedes the second-order wetting transition, is also in agreement with the experiments of alkanes at the water–air interface [118–121]. The line of thin–thick transitions A_{tt} terminates in a critical point at which the difference between the microscopically thin and the mesoscopically thick layer vanishes.

4.3.3. High Grafting Density

On increasing the grafting density even further, a stable minimum of the interface potential in the absence of long-ranged interactions $A_{ll} = 0$ develops. The larger the grafting density, the deeper the minimum, and the further away from the substrate it occurs. The density profiles at a high grafting density are shown in Fig. 10d. Without any free chains, the brush–vapor interface resembles a narrow liquid–vapor interface. Adding a small amount of free chains, we allow the collapsed chains of the brush to relax, and the free chains allow the brush profile to adopt a broader profile, which resembles the profile of a brush in contact with a melt. Those free chains are localized at the brush–vapor interface. The free chains cannot penetrate into the brush, because its density is comparable to the density of the liquid that coexists with the vapor. Increasing the amount of free chains does not significantly alter the liquid–vapor interface or the brush–melt interface. The additional free chains are confined into a liquid layer on top of the brush, defined by the brush–melt interface and the liquid–vapor interface. The addition of more free chains than necessary to relax the brush–vapor profile increases the free energy, because these chains, confined into the layer on top of the brush, have a reduced configurational entropy. There is a minimum in the interface potential, which shifts to larger thickness as the grafting density is increased. Thus, in the absence of long-ranged attraction, the liquid does not wet the dense brush of identical chains. This autophobicity has been observed on polymer brushes and crosslinked polymer networks. The minimum also becomes deeper at higher grafting density, because the chains can penetrate less into the brush. On making the substrate attract the liquid $A_{ll} > 0$, we can overcome the autophobicity and make the liquid wet the brush. The wetting transition on the dense brush is of first order, and $A_{wet} > 0$ increases with grafting density.

Experiments [122] at intermediate grafting densities yield evidence for evidence for the stability of a thin and a mesoscopically thick film (at phase coexistence between the liquid and the vapor). More recently, experiments [123, 124] on the wetting of PS on a PS brush indeed observed two wetting transitions: at small and very large grafting density, the polymer film dewets the brush, whereas at moderate grafting density, the PS film remains stable.

At even lower temperatures, the brush in a bad solvent does not collapse into a laterally homogeneous dense layer, as we have assumed implicitly in our one-dimensional self-consistent field calculations, but the chains gather into clusters that have a lateral distance comparable to the molecular extension [125]. Such lateral structure formations shall be discussed in Section 5.

4.4. Adsorption on Laterally Structured Substrates

Another way to modify the wetting and adsorption properties [126] is to exploit the influence of geometrical surface patterns. Bauer and Dietrich have investigated the wetting behavior of a planar substrate containing a stripe [127–129], that consists of a chemically different material than the remaining substrate. A particularly intriguing adsorption behavior is found in the vicinity of the prewetting coexistence of the stripe material, where at $\Delta\mu < \Delta\mu_{pw} < 0$ the infinitely extended stripe material would be covered by a thin liquid layer, whereas for $\Delta\mu_{pw} < \Delta\mu < 0$, a thick (but finite) layer would build up. Similar to capillary condensation, the transition between the thin and thick adsorption layers on the stripe is rounded off, because the stripe is quasi-one-dimensional, and shifted away from the prewetting line of the infinite stripe. The magnitude of the shift depends both on the widths of the stripe and on the adsorption properties of the substrate bordering the stripe [129]. Making the substrate more attractive to the fluid, one induces a change from a thin liquid layer adsorbed on the stripe to a thick film, and vice versa. This phenomenon is termed "morphological transition" or "interface bending/unbending transition" (cf. sketch on the left-hand side of Fig. 11), because it implies a transition between two different density profiles (a snapshot showing the two different states is shown on the right-hand side of Fig. 11).

To observe such a morphological transition in our Monte Carlo simulations, we have to adjust the wetting properties of the stripe and the neighboring substrate carefully. In our simulations, we use periodic boundary conditions in the two lateral directions such that

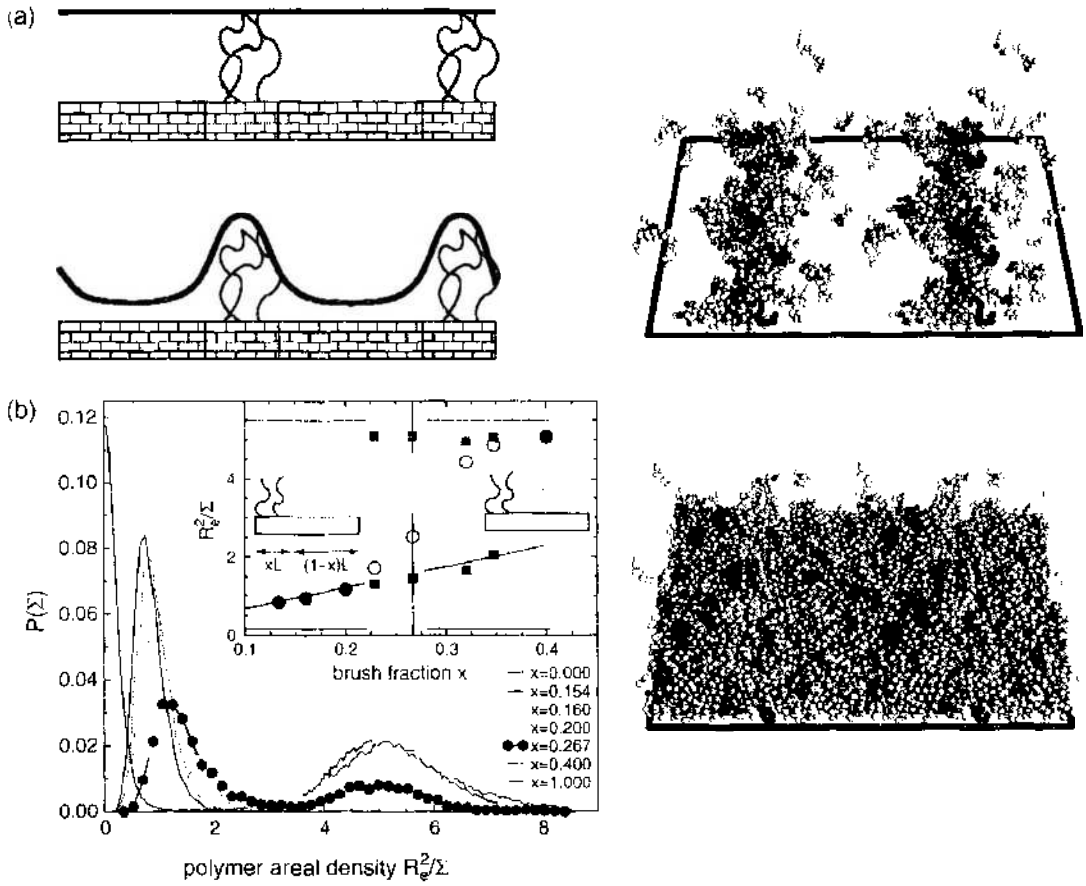


Figure 11. (a) Sketch of the two coexisting phases at the interface bending/unbending transition. (b) Probability distribution of the total thickness as a function of the fraction of the polymer brush $x = L_B/(L_A + L_B)$ (cf. sketch in the inset) at temperature $k_BT\epsilon = 1.68$; $A_w = 3.5$, and undersaturation $\mu - \mu_{coex} = -0.15\epsilon$. (cf. Fig. 8). Around $x \approx 0.27$ there is a morphological transition. The dashed lines mark the adsorption on the stripe and on the brush, as well as the linear superposition. The panels on the right present snapshots of configurations at $x = 0.266$. The simulation box and three periodic images are shown. The upper and lower panels correspond to $R_e^2/\Sigma = 1.38$ and 5.67, respectively. Adapted from Ref. [126], M. Müller and L. G. MacDowell, *J. Phys. Condens. Matter* 15, R609 (2003).

we look at alternating stripes A and B. We again work at temperature $k_BT/\epsilon = 1.68$ and undersaturation $\Delta\mu = -0.15\epsilon$ and use long-ranged attractions between the fluid and the wall according to Eq. (3) with $A_w = 3.5$. Under these conditions, a thin liquid layer adsorbs on the stripe A (cf. Fig. 11), and this stripe is close to its prewetting transition. For the second stripe B (or substrate), we use the same long-ranged attraction between the fluid and the wall, but we increase the fluid adsorption by grafting chains to the surface with $R_g^2/\Sigma = 0.837$. The self-consistent field calculations in Fig. 10 indicate that the adsorbed amount on the brush B can be tuned easily, because brush B exhibits a second-order or very weakly first-order wetting transition.

In the following we vary the width of the (first) stripe L_A and fix the width of the brush-covered second stripe to $L_B = 8\sigma$. In the inset of Fig. 11 we present the adsorbed amount per area as a function of the fraction $x = L_B/(L_A + L_B)$ of the brush. For $x < 0.2$ the adsorption increases linearly with the brush fraction. Around $x \approx 0.27$ the average adsorption develops an S-shaped dependence on the brush fraction x, and for $x > 0.35$ the adsorbed amount on the patterned substrate almost reaches the value of the brush substrate. The average adsorption on the substrate is more than twice as much as the linear superposition would indicate. In the side panels, we show two snapshots at the same brush fraction $x = 0.266$. In the upper snapshot, there is hardly any adsorption on stripe A, and the liquid condenses onto the brush-covered stripes B. The total adsorption is what one would expect to achieve by adding the adsorption on stripe A and brush B, weighted by their areal fraction. In the lower snapshot, the liquid forms a uniform thick layer on the substrate, and the total adsorption is comparable to the adsorption on an infinitely extended brush B. The snapshots illustrate the two states separated by the morphological transition. Because the system is two-dimensional (i.e., neighboring stripes of the same type are coupled), the transition can be truly first order. During the simulation, the system of finite size switches from one state to another; therefore, we do not observe a sudden increase of the adsorption at the morphological transition. It is instructive to look at the probability distribution of the adsorption. For small and large values of x, the distribution $P(\Sigma)$ has a single peak centered around the average value of the adsorption. Close to the transition, however, the distribution is bimodal; Each peak corresponds to the two states at the morphological transition. Around $x \approx 0.27$, both peaks have equal weight, and this criterium yields an estimate for the location of the transition. In the inset, we have also indicated the average adsorption that corresponds to the individual peaks. The lower value follows the weighted average of stripe A and brush B, whereas the higher value is close to the adsorption of brush B. This gives an estimate for the adsorption one would observe in a larger system. When studying the ability of the stripe to modify the wetting behavior of the substrate, it is interesting to compare the above mentioned results with a similar adsorption simulation in which the chains are grafted uniformly on the substrate, rather than forming stripes. For example, the stripe with an areal fraction $x = 0.267$ may be uniformly distributed to form a brush with grafting density $1/\Sigma = x/\Sigma_{\text{stripe}}$. In this case, we find that the stable state of the system consists of a thick layer of polymers, with total adsorption of $R_g^2/\Sigma = 4.9$ polymers. The adsorption onto a patterned surface is slightly larger $R_g^2/\Sigma = 5.1$.

5. BINARY POLYMER BRUSHES: SELF-ASSEMBLED STRUCTURES WITH SWITCHABLE WETTING PROPERTIES

Instead of patterning the surface on the molecular scale "by hand," one can employ self-assembly to create structures on the nanometer scale. In many applications, one uses diblock copolymers (i.e., linear polymers that consist of two distinct blocks or more complicated molecular architectures; e.g., multiblock copolymers or graft copolymers [130]. The two monomeric species want to separate to reduce the energetically unfavorable interactions between unlike species. The connectivity along the backbone of the molecule, however, prevents them from segregating into macroscopical domains, and they phase separate into periodic microstructures with a characteristic length scale set by the molecular extension. Those microstructures have many fascinating thermodynamic and rheological properties. Most notable, those systems are promising candidates for templating nanostructures. In this

section, we briefly describe the self-assembly that occurs in mixed polymer brushes [131, 132]. If one grafts two incompatible polymers with their ends irreversible onto a surface and thereby immobilizes the grafting points, one also can avoid macroscopic segregation of the two species, and by the same token, the chains phase separate on the length scale of the molecular extension [133].

Surfaces covered with those mixed brushes have proven to be practically useful. Combining polymers with different properties (e.g., hydrophilic and hydrophobic species), one can fabricate structured surfaces with tunable wetting and adhesion properties [134]. Exposing the brush to a hydrophilic solvent, the hydrophilic chains segregate to the top of the brush and make the surface hydrophilic after the solvent has quickly evaporated. Likewise, the exposure of the mixed brush to a hydrophobic solvent caused the hydrophobic component to enrich at the brush's outmost layer. After the solvent has evaporated, the brush-coated substrate is hydrophobic. This change in the wettability of the substrate is reversible. Although a flat substrate coated with a mixed polymer brush leads only to a change of the contact angle by a few tens of degrees, this change in the wettability can strongly be amplified by using a rough substrate and results in quite a significant control over the surface properties [134, 135].

Measurements of the contact angle of a liquid on top of the brush provide only information over a large macroscopic patch of the surface. To investigate the structure and the molecular conformations on the nanometer scale, special experimental techniques [132] have to be employed (e.g., x-ray photon emission spectroscopy, XPEEM). It is on this length scale that coarse-grained models are most valuable, and self-consistent field calculations can provide a detailed picture of the structure and the thermodynamics.

Two limiting types of morphologies into which chains arrange to avoid unfavorable interactions between different species can be envisioned [132]: (1) The two species segregate perpendicular to the substrate (layered phase, sandwich-like structure), with one species being enriched at the substrate while the other segregates to the top of the brush. The mixed brush remains laterally homogeneous. This homogeneity is advantageous for designing surfaces with reversibly tunable wettability. (2) Alternatively, the two species can laterally self-assemble into two-dimensional structures with a well-defined lateral length scale, which is on the order of the molecules' extension.

To investigate the universal behavior on the length scale larger than a few monomeric units, we model both types of chains—A and B—as ideal Gaussian walks with end-to-end distance R_e. Restricting ourselves to the large length scale properties of the molecular conformations, we also disregard packing effects and employ a very simple, local density functional for the interactions of the two monomer species, which takes the form of a virial expansion in the two monomer number densities ϕ_A and ϕ_B

$$\frac{F_{ex}[\phi_A, \phi_B]}{k_B T} = \int d\mathbf{r} \left\{ \frac{v_{AA}}{2}\phi_A^2 + \frac{v_{BB}}{2}\phi_B^2 + v_{AB}\phi_A\phi_B \right\} \tag{53}$$

$v = (v_{AA} + v_{BB} + 2v_{AB})/4$ characterizes the average strength of the excluded volume interaction between monomers. The equation $\tilde{\chi} = (2v_{AB} - v_{AA} - v_{BB})/2v$ denotes the mutual attraction(repulsion) between (un)like monomers. It is related to the commonly used Flory–Huggins parameter via $\chi = v\tilde{\chi}$. Increasing $\tilde{\chi}$, we increase the incompatibility between the two species and reduce the solvent quality. The quantity $\zeta = (v_{AA} - v_{BB})/2v$ specifies the solvent selectivity toward the A and B species. It is obvious that this crude approximation neither includes any local structure of the monomeric units nor faithfully describes the thermodynamic properties of a compressible binary polymer mixture. It should just be regarded as a minimal model that captures some of the universal, qualitative properties of the structure and the thermodynamics of binary brushes, but not as a faithful model of a specific experimental realization.

Even though our computational model is extremely simplified, it still poses a challenge to describe the immobility of the grafting points within the framework of the self-consistent field theory. To this end, we formally identify each chain by its grafting point $\mathbf{r}_0$ and regard the systems as a mixture consisting of infinitely many components. Within the framework of

the self-consistent theory, the free energy takes the form $F = E - T(S_A + S_B)$, where S_A and S_B denote the conformational entropies of A polymers and B polymers, respectively. The configurational entropy of all A-polymers, S_A, is given by

$$S_A = k_B \sum_{A\mathbf{r}_0} \left\{ \ln \mathcal{Q}_{A\mathbf{r}_0} + \int \mathrm{d}\mathbf{r}\, w_A \phi_{A\mathbf{r}_0} \right\} \tag{54}$$

where $\mathcal{Q}_{A\mathbf{r}_0}[w_A]$ is the single-chain partition function of the A polymer (cf. Section 3.2), which is grafted at position $\mathbf{r}_0$, in an external field w_A. The sum runs over the grafting points $\mathbf{r}_0$ for all A chains. The variable $\phi_{A\mathbf{r}_0}$ denotes the monomer density of the A polymer that is grafted at $\mathbf{r}_0$. The fields and densities satisfy the self-consistent set of equations

$$w_A = \frac{\partial F_{\mathrm{ex}}}{k_B T \partial \phi_{A\mathbf{r}_0}} = v\left\{ \frac{2-\tilde{\chi}}{2}\phi_A + \frac{2+\tilde{\chi}}{2}\phi_B \right\} \qquad \text{and} \qquad \phi_{A\mathbf{r}_0} = -\frac{1}{\mathcal{Q}_{A\mathbf{r}_0}} \frac{\partial Q_{A\mathbf{r}_0}}{\partial w_A} \tag{55}$$

In accord with intuition, the field w_A is independent from the grafting point $\mathbf{r}_0$ (i.e., the interaction an A segment experiences does not depend on the grafting point). Similar expressions hold for S_B, w_B, and $\phi_{B\mathbf{r}_0}$.

To calculate the single-chain partition function and monomer density, it is again useful to define end-segment distributions $q_{A\mathbf{r}_0}(\mathbf{r}, t)$ and $\tilde{q}_A(\mathbf{r}, t)$. There are two distinctions compared to the case of a polymer film at a surface. First, because of the inequivalence of the two chain ends, there are two types of propagators: $q_{A\mathbf{r}_0}(\mathbf{r}, t)$ describes the probability that a chain of length tN, grafted at $\mathbf{r}_0$, terminates at $\mathbf{r}$, whereas $\tilde{q}_A(\mathbf{r}, t)$ is the probability that a chain of length tN, starting anywhere in the bulk, terminates at $\mathbf{r}$. Second, formally, there is one propagator $q_{A\mathbf{r}_0}(\mathbf{r}, t)$ for each grafted chain.

Both end-segment distributions satisfy the diffusion equation $\frac{\partial q_A}{\partial t} = \frac{R_0^2}{6}\Delta q_A - Nw_A q_A$ [cf. also Eq. (22)]. One chain end is grafted at $\mathbf{r} = \mathbf{r}_0$ and, hence, $q_{A\mathbf{r}_0}(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)$, and the other chain end is free, $\tilde{q}_A(\mathbf{r}, 0) = 1$. The monomer density $\phi_{A\mathbf{r}_0}$ and single-chain partition function $\mathcal{Q}_{A\mathbf{r}_0}$ can be expressed as

$$\phi_{A\mathbf{r}_0}(\mathbf{r}) = \frac{N}{\mathcal{Q}_{A\mathbf{r}_0}} \int_0^1 \mathrm{d}t\, q_{A\mathbf{r}_0}(\mathbf{r}, t)\tilde{q}_A(\mathbf{r}, 1-t) \quad \text{with}$$

$$\mathcal{Q}_{A\mathbf{r}_0} = \int \mathrm{d}\mathbf{r}\, q_{A\mathbf{r}_0}(\mathbf{r}, s)\tilde{q}_A(\mathbf{r}, 1-s) \qquad \forall\, 0 \le s \le 1 \tag{56}$$

We obtain the total A monomer density ϕ_A by summing over all grafting points of A chains

$$\phi_A(\mathbf{r}) = N \int_0^1 \mathrm{d}t\, q_A(\mathbf{r}, t)\tilde{q}_A(\mathbf{r}, 1-t) \tag{57}$$

with $q_A = \sum_{A\mathbf{r}_0} q_{A\mathbf{r}_0}/\mathcal{Q}_{A\mathbf{r}_0}$. Because q_A is a linear combination of $q_{A\mathbf{r}_0}$, it also obeys the diffusion equation with initial condition

$$q_A(\mathbf{r}, 0) = \sum_{A\mathbf{r}_0} \frac{\delta(\mathbf{r} - \mathbf{r}_0)}{\tilde{q}_A(\mathbf{r}_0, 1)} \tag{58}$$

where we have employed Eq. (56) with $s = 0$. Hence, to calculate the density we do not have to calculate the propagator $q_{A\mathbf{r}_0}$ for each grafting point individually, but we can simply use the effective propagator for the total A monomer density. Of course, the density distribution of the 0th segment reproduces the grafting points $\phi_A(\mathbf{r}, t = 0) = q_A(\mathbf{r}, 0)\tilde{q}_A(\mathbf{r}, N) = \sum_{A\mathbf{r}_0} \delta(\mathbf{r} - \mathbf{r}_0)$.

The equations above describe a binary brush with an arbitrary distribution of grafting points within self-consistent theory. At this stage, correlations and inhomogeneities of the grafting densities could be explicitly incorporated into the description. In the following text, however, we assume that the grafting is laterally homogeneous and denote the relative concentration of A chains by Φ, also replacing the summation over the grafting points by an

integral $\sum_{A\mathbf{r}_0} \to \Phi\sigma \int d\mathbf{r}_0\, \delta(x_0)$. We measure the degree of stretching [136] by the dimensionless parameter

$$\frac{1}{\delta^3} = \frac{3}{2}\left[\frac{vN^2}{R_e^3}\right]^2 \left[\sigma R_e^2\right]^2 \tag{59}$$

Intriguingly, the two dimensionless parameters σR_e^2, which measures the overlap of the chains inside the brush, and the Fixmann parameter vN^2/R_e^3, which describes the strength of the excluded volume interaction, do not enter the self-consistent theory independently, but only as a product. We define a rescaled A monomer density $\tilde{\phi}_A \equiv vN\delta\Phi\phi_A$ and measure all length scales in units of R_e.

Qualitative insight into the physical origin of phase separation can be gained in the limit of strong stretching $\delta \to 0$, where the situation is similar to the binary brush in the melt [133]. Then, the dominant contribution to the free energy stems from the interplay between excluded volume interactions and chain stretching (similar to the one-component brush). Using the fact that the height of the brush scales like $h \sim R_e/\sqrt{\delta}$, we find that this contribution is of the order k_BT/δ per chain. It controls the density profile. On increasing the incompatibility $\tilde{\chi}$, there is a transition from the disordered phase to the "ripple" phase to reduce the energy of mixing. For $\delta \to 0$, only the composition $\rho_A \equiv \phi_A/(\phi_A + \phi_B)$ becomes laterally inhomogeneous, but the total density, $\phi = \phi_A + \phi_B$, is not affected by the phase transition. The energy of mixing is on the order of $\tilde{\chi}k_BT/\delta$ per chain. Lateral demixing reduces the entropy of mixing of the free chain ends, which is on the order of k_BT per chain. At the transition, this entropy loss is comparable to the energy of mixing, and we obtain $\tilde{\chi}_{\text{dis}-\text{ripple}} \sim \delta$ (in agreement with our self-consistent field calculations). As the incompatibility increases or the stretching decreases, the decoupling of density and composition breaks down. Figure 12 illustrates the variation of the total density at intermediate stretching and a rather large incompatibility. (Small values of stretching δ correspond either to a small overlap of grafted chains $\sigma R_e^2 \ll 1$ or to a solvent of marginal quality $vN^2/R_e^3 \ll 1$; in the former case, correlations of the grafting points are very important in experiments, and in the latter case they are not.) In this parameter region, we investigate the interplay

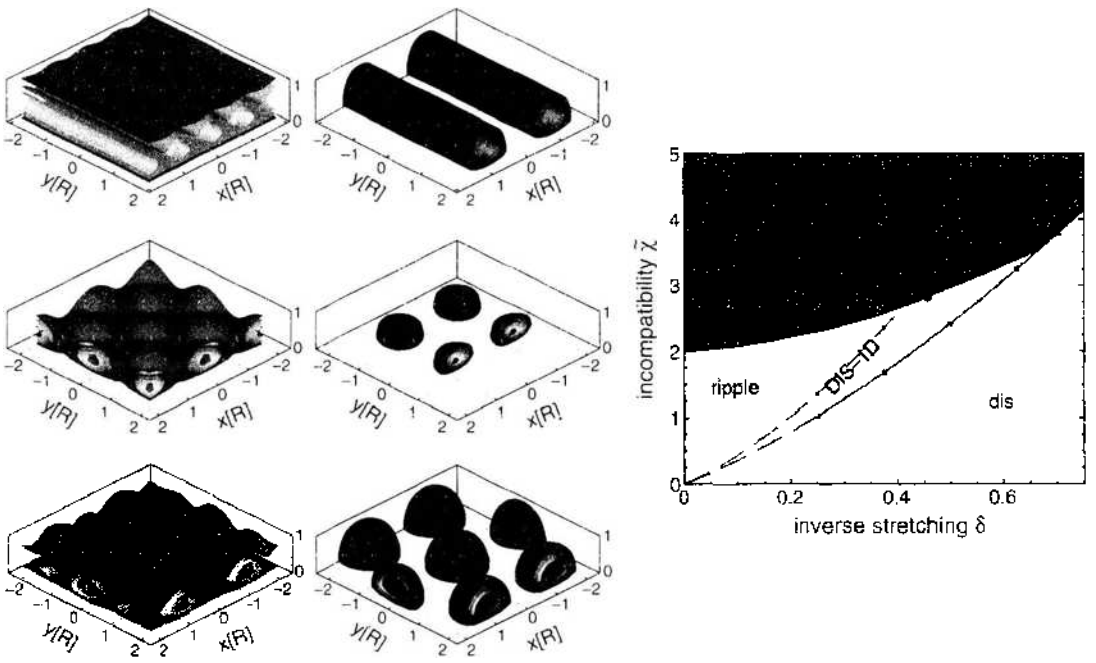


Figure 12. Left-hand side: Contour plots of the total density and the density of the A-component for the "ripple" morphology (upper panel), the checkerboard-like "dimple S" morphology (middle panel), and the hexagonal "dimple" structure (bottom panel). Right-hand side: Phase diagram of a symmetric brush ($\Phi = 1/2$) as a function of inverse stretching δ and incompatibility $\tilde{\chi}$. Adapted from Ref. [131], M. Müller, *Phys. Rev. E* 65, 030802(R) (2002).

between the energy of mixing and conformational entropy via numerical self-consistent field calculations. Additional morphologies are found to be stable, and lateral and perpendicular segregation occur simultaneously.

First, in Fig. 12, we investigate the phase behavior for symmetric composition ($\Phi = 1/2$) as a function of stretching δ and incompatibility $\tilde{\chi}$. The relative stability of five phases is explored: the disordered phase (dis) that is neither laterally nor perpendicularly segregated; the layered phase (one-dimensional, 1D), in which the symmetry between A and B is spontaneously broken and the two components segregated perpendicular to the surface; the "ripple" phase in which the two components laterally segregated into symmetrical cylinders; and two "dimple" phases. In one phase, denoted as "dimple S," both species segregate symmetrically into clusters, which arrange as on a checkerboard. In the other phase ("dimple A" or "dimple B"), one of the components—A or B—segregates into clusters, which form an hexagonal lattice in the lateral direction, whereas the other component is less dense and fills the space between the clusters. The symmetry between A and B is broken. The (total) density profiles of the laterally structured phases are shown in Fig. 12. The size of the lateral repeat units is about $1.9R_e$ for the ripple and symmetric dimple phases and $2.2R_e$ for the hexagonal dimple phase in the range of $\tilde{\chi}$ close to the order–disorder transition. Note that the length scale of the spatial ordering is somewhat larger than in diblock copolymers; $1.32R_e$ at the onset of ordering in the lamellar phase [137].

The phase diagram is presented in Fig. 12. On increasing the incompatibility, we find a second-order transition from the disordered phase to the "ripple" phase. As discussed above, the incompatibility $\tilde{\chi}$ at which the "ripple" structure forms increases linearly with δ in the limit of strong stretching $\delta \to 0$. For small values of δ and stronger incompatibility ($\tilde{\chi} > 2$, bad solvent), we encounter a transition from the "ripple" phase to an hexagonal "dimple" phase. For larger values of δ, we first encounter a transition from the "ripple" phase to the symmetrical "dimple S" phase, and at even larger incompatibilities we find a transition to a hexagonal "dimple" phase. There is a triple point at $\delta = 0.46$ and $\tilde{\chi} = 2.83$, where the "ripple," the symmetrical, and the hexagonal "dimple" phases coexist. At the triple point, [illegible] present the contour plots of the density and composition in Fig. 12.

Our calculations omit a repulsive, third-order virial coefficient. For $\tilde{\chi} > 2$ the solvent is bad for the pure components, but the average segment interaction would be repulsive for $\tilde{\chi}|\Phi - 1/2| < 1$ if the brush was homogeneously mixed. The density inside the "dimple" structure at large $\tilde{\chi}$ is limited by chain stretching rather than by packing effects. Including third-order virial coefficients in our calculations, we have explicitly verified that the qualitative features of the phase behavior remain unaltered. A third-order virial coefficient would stabilize phases with smaller density variations (i.e., the disordered phase would gain stability with respect to the "ripple" phase, and the transitions from the "ripple" to the dimple phases would shift to slightly larger values of $\tilde{\chi}$).

The phase diagram for symmetrical composition as a function of the incompatibility $\tilde{\chi}$ and selectivity ζ of the solvent is presented in Fig. 13 for moderate stretching, $\delta = 0.46$. The solvent properties have a profound influence on the self-assembled structures [138]. On increasing the incompatibility in a nonselective solvent $\zeta = 0$, we find a transition from a laterally homogeneous phase (dis) to the "ripple" phase. As we decrease the solvent quality for the A component, $\zeta < 0$, the "ripple" phase transforms to a "dimple" structure, in which the A component segregates into clusters that arrange on an hexagonal lattice. Likewise, we find a hexagonal "dimple" structure with a collapsed B brush when the solvent is poor for the B component ($\zeta > 0$). At higher incompatibility (or poorer solvent quality), only "dimple" structures are stable. If we increased δ, the region of stability of the "ripple" phase would increase and the transition from the laterally homogeneous phase to the structured phases would occur at lower incompatibility.

In the inset of Fig. 13, we present the laterally averaged composition profiles at $\tilde{\chi} = 2.4$. As we increase the selectivity, the B component segregates to the top of the brush, whereas the relative density of A is higher at the substrate. Contour plots of the composition at small and large solvent selectivity are presented in the lower panels. Note that the hexagonal "dimple" morphology exhibits in addition to the two dimensional lateral structure a pronounced perpendicular segregation. Even in the absence of a solvent preference, $\zeta = 0$,

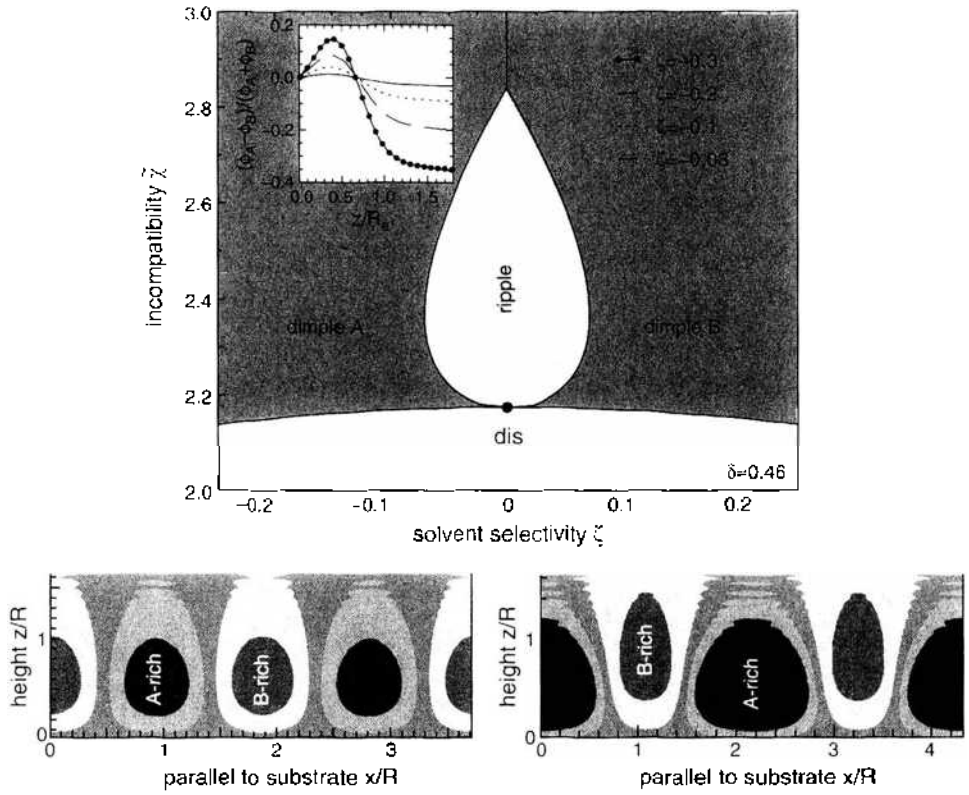


Figure 13. Phase diagram of a symmetric mixed brush as a function of solvent selectivity and incompatibility at $\delta = 0.46$. The inset shows the laterally averaged composition profile for $\chi = 2.4$ and solvent selectivities $\zeta = -0.3$, -0.2, -0.1, and -0.03. In the lower row, we show contour plots of the composition for $\zeta = -0.03$ in the "ripple" phase and $\zeta = -0.3$ in the "dimple" phase. Adapted from Ref. [132], S. Minko, et al., *Phys. Rev. Lett.* 88, 035502 (2002).

the symmetry between A and B is spontaneously broken and the hexagonal "dimple" phases exhibit a laterally averaged perpendicular segregation. Therefore, a transition between the two "dimple" phases leads to a significant variation in the surface properties, which can then experimentally be exploited to tune the surface properties [132, 135]. This perpendicular segregation can be enhanced by a solvent that selectively attracts the matrix component.

In addition to the solvent, external interactions can also trigger transition between the different morphologies of mixed binary brushes. Those interactions might comprise a preference of the grafting surface to one component or the interaction of the mixed polymer brush with an external surface (i.e., stamp, or possibly electrical fields in the case of a mixed polyelectrolyte brush). The interplay between morphology of the brush and external fields might allow for a control of practically important surface properties (i.e., wettability and adhesion) over a wide range, and we anticipate a rather rich phase behavior. We shall illustrate the effect of preferential surface interactions on the morphology of mixed polymer films for the much simpler example of a homopolymer blend in the next section.

6. INCOMPRESSIBLE BINARY POLYMER FILMS

6.1. Generalization of the Computational Techniques to Binary Polymer Blends

The fact that confining surfaces can profoundly alter the miscibility behavior is well known [139–141]. For instance, if the walls of a slitlike pore attract the fluid, it will condense in a pore at a lower pressure than in the bulk—a phenomenon termed capillary condensation. If the pore is of finite cross-section, the system is quasi-one-dimensional, and there is no true phase transition at finite temperatures [142]. The condensation transition is rounded off—the correlation length far below the critical point of the bulk is only finite, but it grows

exponentially with the cross-section of the pore. The condensation of fluids or the phase behavior of binary AB mixtures in pores, slits, and films has attracted interest from theorists and experimentalists [143] alike. In the following text, we shall illustrate these effects for an incompressible, binary homopolymer mixture (i.e., we consider a thin polymer film that contains two types of polymers—denoted A and B—on top of a substrate). The qualitative features of phase transitions with a single scalar order parameter (e.g., the composition of the mixture) in confined geometry are rather universal and are shared by polymer mixtures as well as small molecules. Symmetric binary polymer blends are, however, particularly well suited to study the interplay between wetting and miscibility, because, first, the wetting transition temperature typically is much lower than the critical temperature, where the demixing occurs in the bulk [69], and second, because fluctuations can be controlled by the degree of polymerization [68, 144] [cf. Eq. (27)]. The more extended the molecule the larger the number of neighbors it interacts with is, and the smaller the effect of fluctuations. Third, the spatial extension of the molecules also facilitates experimental investigations. Indeed, the interface localization/delocalization transition (to be discussed later) was first experimentally observed in polymer blends [145], and wetting transitions have been studied in recent experiments [146, 147]. Fourth, the vapor pressure of polymers is negligible, and evaporation effects can often be neglected. Fifth, polymers offer much control over the miscibility and wetting behavior. Changing the chain length, one can alter those properties without affecting the interactions between the monomeric units. Sixth, polymers tend to form a glass rather than a crystal at low temperatures, and hence, the wetting transition might not be preempted by a crystalline phase, as happens in many simple fluids.

In the following text, we consider a symmetric blend (i.e., a blend in which the two types of polymers contain the same number of monomers N and have the same extension R_e). The polymers involved are confined into a thin film; the bottom substrate (W) might be a silicon wafer, whereas the other surface might be the interface to the vapor (V).

In the case of a binary compressible polymer blend that consists of components A and B, the homogeneous state is characterized by two independent-order parameters, total density ϕ and composition ρ_A, or the number densities ϕ_A and ϕ_B of the two species. For instance, the dense polymer melt on the substrate is separated from its vapor of vanishingly low density by a liquid–vapor interface. The dense polymer film might be mixed, or the two components might laterally phase separate into macroscopically large domains. Those compressible mixtures exhibit a rich interplay between liquid–liquid immiscibility and liquid–vapor phase separation. In fact, six qualitatively different types of phase diagrams can be distinguished according to the classification of Konynenburg and Scott [23].

In principle, the simulation methodology and the self-consistent field technique can straightforwardly be carried over to multicomponent systems. In the simulations, one uses the grandcanonical ensemble and inserts or deletes both species via configurational bias Monte Carlo moves, with the chemical potential of each species being fixed. To establish phase coexistence, one has to adjust one chemical potential (e.g., μ_A) for a given chemical potential of the other species μ_B, temperature, and volume so as to fulfill the equal weight criterium. Similar to the one-component system, one determines the interface and surface tensions by analyzing the joint probability distribution of ϕ_A and ϕ_B [148].

In the self-consistent field calculations, one needs the excess free-energy functional as a function of two densities, ϕ_A and ϕ_B. To a first approximation, one can use a virial expansion in both densities similar to that of the previous section on mixed polymer brushes. TPT1 calculations have also been carried out for coarse-grained models of a compressible binary mixture [85]. TPT1 calculations yield more accurate estimates for the free-energy density of the spatially homogeneous state of a specific model system. In fact, both Monte Carlo techniques and self-consistent field calculations have been applied to the general case of a compressible polymer mixture or polymer + solvent systems [149]. In the following text, however, we restrict ourselves to symmetric mixtures with negligible volume change on mixing. Then, fluctuations of the composition $\phi_A - \phi_B$, are much larger and approximately decoupled from fluctuations of the density $\phi_A + \phi_B$, and we can characterize the system by a single scalar order parameter—the composition.

Symmetry between the components imparts a great simplification onto the simulation methodology. By virtue of the structural symmetry with respect to exchanging the species, $A \rightleftharpoons B$, phase coexistence occurs at $\mu_A = \mu_B$. In this case, it is convenient to choose the semigrandcanonical ensemble [150] (i.e., one fixes the volume, the total monomer density ϕ, and the exchange potential $\Delta\mu = \mu_A - \mu_B$ and observes the fluctuation of the composition ρ_A). In the Monte Carlo simulations, one can try to relabel an A-polymer into a B-polymer and vice versa without altering the chain conformations. Obviously, such a move is computationally much less demanding than configurational bias insertion/deletion tries, which, generally, would have to be performed in an asymmetric blend.

In the following text, we will focus on the interplay between wetting and phase separation in thin films. To this end, we assume that the attraction between the wall and both monomeric units is strong enough to make both the blend as well as each component wet the substrate. Then, depending on the ratio between the interface tension γ_{AB} between the segregated bulk phases and the surface tension $\gamma_{AV} - \gamma_{BV}$ of the components and the vapor, the upper film surface might be rough. The qualitative behavior is illustrated in Fig. 14. If the AB interface tension is comparable to the liquid–vapor tension, it will "drag" the film surface toward the substrate so as to reduce the length L_{AB} of the AB interface. If the liquid–vapor tension exceeds the AB interface tension by about two orders of magnitude, $\gamma_{AB} \ll \gamma_{AV}$ or γ_{BV}, however, the surface will be almost flat and the situation is equivalent to a binary mixture between two hard walls a distance D apart. This condition is closely related to the incompressibility assumption. When there is only a negligible volume change on mixing, the repulsion between the different monomeric units χ is much weaker than the attractive forces that hold the fluid together. Under these conditions, it is also reasonable to expect that γ_{AB} would be much smaller than all other tensions and comparable in magnitude to the differences $\gamma_{AV} - \gamma_{BV}$ and $\gamma_{AS} - \gamma_{BS}$.

Typically, these conditions can be realized if all binary interactions are strongly attractive, but the difference that determines the energy of mixing is small. The strongly attractive interactions inside the fluid are necessary to bring about the liquid–vapor coexistence. Unfortunately, the incompatibility between the species A and B (i.e., the energy change on mixing) results from a subtle cancellation of large negative contributions to the total energy. To avoid this complication, one can confine the blend from the onset between two hard walls and reconsider liquid A liquid B miscibility at large pressure. The large pressure reduces the volume change on mixing in the dense polymer liquid, but the binary interactions can be of

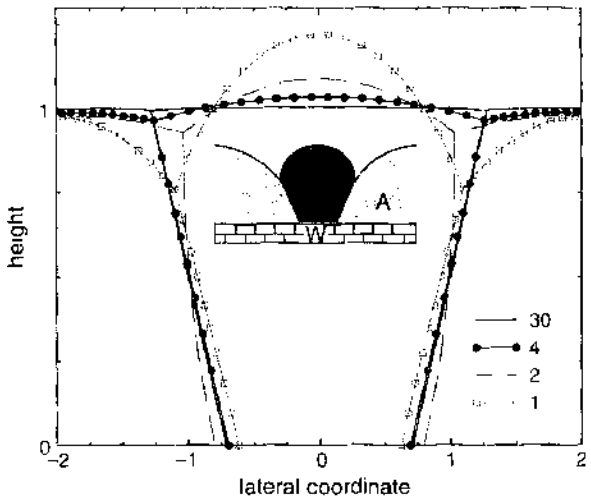


Figure 14. Interfaces in a laterally segregated two-component film. The shape of the interface is obtained by minimizing the effective Hamiltonian $H = \gamma_{AB}L_{AB} + \gamma_{AW}L_{AW} + \gamma_{BW}L_{BW} + \gamma_{AV}L_{AV} + \gamma_{BV}L_{BV}$ at fixed volume of the components. L_{ij} denotes the length of the interface between substances i and j, and γ_{ij} is the corresponding interface tension. A: A component of the mixture, B: B component of the mixture, W: wall, V: vapor. The following parameters were used: $\gamma_{AV} - \gamma_{BV} = \gamma_{AW} - \gamma_{BW} = 0.5\gamma_{AW}$, $\gamma_{AW} = \gamma_{AB}$, and γ_{AV}/γ_{AB} as indicated in the key. Adapted from Ref. [139]. M. Müller, *Comp. Phys. Comm.* 147, 292 (2002).

the same order as their difference, which determines the energy change on mixing. If there is no volume change on mixing and one neglects the differences between the liquid–vapor interface profile and the density profiles in the vicinity of the wall, both situations are equivalent, because only the local composition of the blend—the order parameter—is important. We have chosen the latter setting in our Monte Carlo simulations of the bond fluctuation model (cf. Section 2), which mimics most closely the self-consistent field calculations (cf. following). As interactions between monomeric units, we use a square-well potential that is extended over distances $r \leq \sqrt{6}$ with depths $\epsilon \equiv \epsilon_{AB} = -\epsilon_{AA} = -\epsilon_{BB}$. Simulations of interfaces between the A-rich and B-rich phases reveal only a small reduction of the total density at the center of the interface for $\phi = 1/16$, demonstrating that the model can well be approximated by an incompressible mixture [151]. In our simulations, we model the surfaces as hard structureless walls that interact with the monomeric units via a square-well potential of depth A that extends over the nearest $d_w = 2$ lattice layers. A positive value of A corresponds to an attraction for A segments and a repulsion for B segments.

Similar to Section 3.2, the free-energy functional of the self-consistent field calculations can be decomposed into two contributions $F_{ex}[\phi_A, \phi_B] = F_{hc}[\phi_A, \phi_B] + F_{att}[\phi_A, \phi_B]$, corresponding to the harsh short-ranged repulsions and the longer-ranged attractions. In a symmetric blend, the specific volumes of both types of monomeric units are identical, and we denote them by $1/\phi$. Thus, we can define the compositions $\rho_A \equiv \phi_A/\phi$ and $\rho_B \equiv \phi_B/\phi$. In the following text, we assume no volume change on mixing, and hence the liquid density ϕ does not depend on the composition.

The harsh repulsions determine the equation of state of the fluid, and we can use an expansion of the free energy around the density ϕ of the liquid

$$F_{hc}[\phi_A, \phi_B] = \frac{\kappa}{2} \int d\mathbf{r}\, (\rho_A + \rho_B - 1)^2 \tag{60}$$

In the limit of large κ, one obtains an incompressible liquid, and one can alternatively use

$$\exp(-F_{hc}[\phi_A, \phi_B]/k_B T) \sim \delta(\rho_A + \rho_B - 1) \tag{61}$$

In a binary polymer mixture, the scale of variation in the composition ρ_A is typically set by the chain extension R_e. On this length scale, the fluid is incompressible and the constraint [Eq. (61)] provides a good description. Packing effects (i.e., fluctuations of the density ϕ) occur on the scale of the monomeric units. Hence, the decoupling of composition and density fluctuations is often justified. Qualitatively, compressibility effects can be modeled by Eq. (60), but generally, the spatial extension of the monomeric units has to be accounted for via a nonlocal functional to be quantitatively accurate.

The attraction between like segments and the repulsion between unlike ones can be approximated by a bilinear expression in the local compositions, ρ_A and ρ_B

$$F_{att}[\phi_A, \phi_B] = (\phi/N)\chi N k_B T \int d\mathbf{r}\, \rho_A(\mathbf{r})\rho_B(\mathbf{r}) \tag{62}$$

where χ denotes the Flory-Huggins parameter. Note that both contributions to the excess free-energy functional are local (i.e., packing effects or spatial correlations in the composition are not explicitly captured by this approach).

Equations (61) and (62), in junction with the Gaussian chain model, constitute the standard self-consistent theory for binary polymer blends. The excess free-energy functional comprises one parameter χN. To identify this parameter, we regard the thermodynamics of the homogeneous, incompressible binary polymer blend

$$\frac{F}{(\phi/N)R_e^3(V/R_e^3)k_B T} = \rho_A \ln \rho_A + \rho_B \ln \rho_B + \chi N \rho_A \rho_B + \text{linear terms in } \rho_A \text{ and } \rho_B \tag{63}$$

This is the Flory–Huggins free energy of mixing [152]. The variable χN can be determined by a comparison of the predictions of the Flory–Huggins theory for the phase behavior in the bulk to experiment or simulation.

If the two polymer species are structurally symmetric, the changes of the molecular conformations on mixing are small (for large chain lengths) [67], and the entropy of mixing is chiefly determined by the translational contribution. Then the Flory–Huggins parameter can be identified by the energy of mixing [153, 154]. Only intermolecular interactions contribute to the energy of mixing, and the χ parameter can be related to the structure of the polymeric fluid via

$$\chi = \phi \int \mathrm{d}\mathbf{r}\, g_{\mathrm{inter}}(\mathbf{r}) \left[\epsilon_{AB}(\mathbf{r}) - \frac{\epsilon_{AA}(\mathbf{r}) + \epsilon_{BB}(\mathbf{r})}{2} \right] \tag{64}$$

where ϵ_{IJ} denote the interaction between the different monomer species. Specifically for the symmetric square-well interactions in the bond fluctuation model, the Flory–Huggins parameter takes the simple form [153]

$$\chi = \phi 2\epsilon \int_{r_{\mathrm{nn}}} \mathrm{d}\mathbf{r}\, g_{\mathrm{inter}}(\mathbf{r}) \equiv 2\phi z_c \epsilon \tag{65}$$

where z_c is the equivalent of the coordination number of the original treatment by Flory and Huggins. This identification of the Flory–Huggins parameter according to Eq. (65) has proven successful in describing the bulk phase behavior in the limit of large chain length and interface properties, and we shall also employ it for incompressible binary mixtures confined into thin films.

If the two polymer species are not structurally symmetric (e.g., they differ in stiffness or shape/packing of the monomeric units), the excess free energy of mixing also comprises entropic contributions. This is commonly observed in experimental systems, where the temperature dependence of the Flory–Huggins parameter can be described by $\chi = a + b/k_B T$, with a and b being constants. Such a temperature dependence can also stem from a temperature dependence of the chain conformations or the fluid structure, and therefore it is difficult to separate contributions to the Flory–Huggins parameter into enthalpic and entropic parts. Although the relation between the Flory–Huggins parameter and the structure of the polymeric fluid is of great practical importance [155], we shall restrict ourselves in the following to symmetric blends and use the simple estimate, Eq. (65).

To model the confinement into a film within the self-consistent field calculations, we impose the total density profile $\phi_0(z)$ across the film of width D_0. At the center of the film, the density reaches the bulk density ϕ while it smoothly decays to zero at the surfaces in a boundary region of width $\Delta_w = 0.15 R_e$ [156]

$$\frac{\phi_0(z)}{\phi} = \begin{cases} \dfrac{1 - \cos\left(\frac{\pi z}{\Delta_w}\right)}{2} & 0 \le z \le \Delta_w \\ 1 & \Delta_w \le z \le D_0 - \Delta_w \\ \dfrac{1 - \cos\left(\frac{\pi (D_0 - z)}{\Delta_w}\right)}{2} & D_0 - \Delta_w \le z \le D_0 \end{cases} \tag{66}$$

The shape of the profile and, in particular, the width of the boundary region Δ_w is chosen for computational convenience, but $\Delta_w \ll R_e$. To describe a film, we replace Eq. (61) by

$$\exp[-F_{\mathrm{hc}}(\phi_A, \phi_B)/k_B T] \sim \delta[\rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) - \phi_0(\mathbf{r})/\phi] \tag{67}$$

The density profile in the vicinity of the surfaces results from a subtle interplay between packing effects of the monomeric units, the entropy loss the extended molecules suffer at a spatial inhomogeneity, and the equation of state. Obviously, an imposed density profile ϕ_0 cannot account for these effects, and one should instead use a generalization of the methods employed in Section 3.2 to binary compressible blends for calculating the density profile in the vicinity of the surface and the concomitant surface tension. Here, we shall argue that the assumption of a density profile will impart a large, uncontrolled error onto the surface tension, but that it does not seriously affect our estimate of the free energy of mixing in confined geometry: First, by virtue of the negligible volume change on mixing, the density

profile does not depend on the local composition, and second, both types of chains suffer the same entropy loss as they pack against the wall, because they are structurally symmetric.

Sometimes an even bolder choice for the density profile is used; namely, a uniform density throughout the film and reflective boundary conditions at the surfaces. This corresponds to the limit $\Delta_w \to 0$. The advantage of this choice is mainly computational, because one does not have to resolve the rapid spatial variation in the boundary layer. The reduction of the density in the vicinity of the wall, however, also gives rise to a reduction of the incompatibility in the boundary layer. This mimics the effect of "missing neighbors" in models (e.g., the bond fluctuation model) with spatially extended interactions.

The interactions between the surfaces and the monomeric units are catered for by short-ranged interactions that extend only in the boundary layer of width Δ_w. Variables Λ_1 and Λ_2 denote the strengths at each surface in the self-consistent field calculations [156]

$$V_{\text{wall}}(z) = \begin{cases} \dfrac{4\Lambda_1 R_c \left[1 + \cos\left(\frac{\pi z}{\Delta_w}\right)\right]}{\Delta_w} & 0 \le z \le \Delta_w \\ 0 & \Delta_w \le z \le D_0 - \Delta_w \\ \dfrac{4\Lambda_2 R_c \left[1 + \cos\left(\frac{\pi (D_0 - z)}{\Delta_w}\right)\right]}{\Delta_w} & D_0 - \Delta_w \le z \le D_0 \end{cases} \tag{68}$$

The prefactor is chosen such that the strength of the monomer–wall interaction is independent from the width δ_w of the boundary region. The integrated strength of the interaction at one wall is $\int \mathrm{d}z\, V_{\text{wall}}(z)\phi_0(z) = \phi\Lambda R_c$.

6.2. Wetting Transition in Incompressible Binary Polymer Blends

Let us first consider the semi-infinite system, in which the phase that is attracted by the surface might exhibit a wetting transition. Similar to the wetting transition of a one-component polymer liquid, one can use Young's equation [91] to accurately determine the wetting transition (c.f. Section 4.1). The A-rich phase wets the surface, if $\gamma_{AW} + \gamma_{AB} = \gamma_{BW}$. Using the semigrandcanonical ensemble in junction with preweighting techniques, one can determine the probability distribution of the composition of the mixture and thus accurately measure phase equilibria and interface tension γ_{AB} as a function of temperature. In principle, one could also use the same simulation technique for a slitlike simulation cell to determine the difference in the surface tensions, $\Delta\gamma = \gamma_{BW} - \gamma_{AW}$ (c.f. Section 3.1). Because of the strict symmetry between the A-rich and B-rich phase however, we can exploit the fact that $\gamma_{AW}(A) = \gamma_{BW}(-A)$, and therefore, we can measure the difference of the surface tensions via Ref. [69]

$$\begin{aligned} \Delta\gamma(A) &= \gamma_{BW}(A) - \gamma_{AW}(A) = \gamma_{BW}(A) - \gamma_{BW}(-A) \\ &= \int_{-A}^{A} \mathrm{d}A' \frac{\mathrm{d}\gamma_{BW}}{\mathrm{d}A'} = \int_{-A}^{A} \mathrm{d}A' \frac{\langle E_{\text{wall}} \rangle_{A'}}{A'} \end{aligned} \tag{69}$$

This method is very accurate and convenient if the wetting transition is of first order. In this case, the microscopic enrichment layer of the A-component at the wall in the vicinity of $A_{\text{wet}} > 0$ is vanishingly thin (i.e., for any value of $A < A_{\text{wet}}$, the configurations are very similar). Of course, one can also formulate this thermodynamic integration in terms of an expanded ensemble, where the strength of the monomer–wall interaction A is a Monte Carlo variable.

This observation also allows us obtain an analytical estimate for the wetting transition. Because of the strict $A \rightleftharpoons B$ symmetry, both polymer species suffer identical loss in conformational entropy, such that the difference in surface free energies is dominated by the energy. If the microscopic enrichment layer of the A component did not exist at all, we could estimate the difference in surface energies per unit area to $\Delta\gamma = 2d_w A\phi$, where we have assumed that the blend is strongly segregated such that $\langle \rho_A \rangle_{\text{bulk}} \approx 0$. In this strong

segregation limit, the interface tension can be describe by $\gamma_{AB}R_c^2 = k_BT(\phi R_c^3/N)\sqrt{6\chi N}$ for large χN [157]. Hence, we obtain for the contact angle [69, 144]

$$\cos\theta = \frac{2d_w A}{k_B T b\sqrt{\chi/6}} \tag{70}$$

where $b = R_c/\sqrt{N}$ denotes the statistical segment length. The results are shown in Fig. 15. The crossing points of the γ_{AB} interface tension and the difference $\Delta\gamma$ of surface tensions yield an accurate estimate of the first-order wetting transition. Most notably, this estimate of the contact angle does not depend on the chain length N, which is in marked contrast to the critical temperature of the blend, $\chi_c N = 2$ or $T_c \sim N$. Increasing the chain length N, we can increase the critical temperature, and hence, the wetting transition temperature is well separated and much lower than the critical temperature of unmixing. This behavior contrasts with the behavior of small molecules, where wetting typically occurs close to the critical point [158].

6.3. Interplay Between Wetting and Phase Behavior in Confined Geometry

If the mixture is confined into a thin film, two limiting surface interactions can be considered: symmetric surfaces, where both surfaces of the film (e.g., the waver and the vacuum) prefer the same component with the same strength and the equivalent of capillary condensation occurs, or antisymmetric surfaces, where one surface attracts one component of the symmetric mixture with exactly the same strength as the opposite surface attracts the other component. These two opposite limits result in qualitatively different miscibility behaviors [159, 160].

In the first case—capillary condensation—the phase diagram qualitatively resembles the miscibility behavior in the bulk. The shift of the critical point of the mixture has been studied within Ginzburg–Landau theory by Nakanishi and Fischer [142, 161, 162]. Typically, the

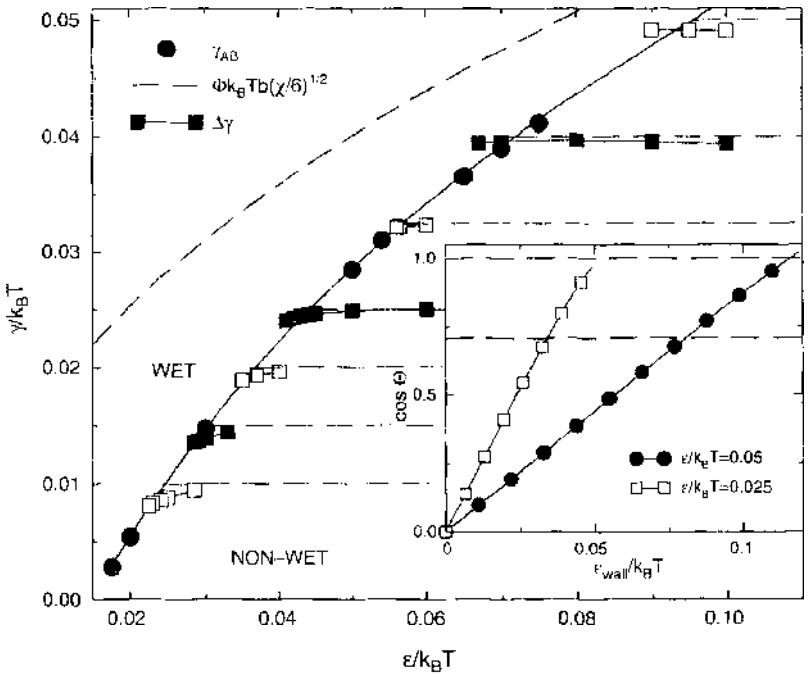


Figure 15. Interface tension γ_{AB} and difference in surface tensions $\Delta\gamma$ as a function of inverse temperature ϵ/k_BT obtained from simulations. Approximations for the interface tension $\gamma_{AB} = b\Phi\sqrt{\chi/6}$ and $\Delta\gamma = 2\Phi d_{wall}A = A/4$ in the strong segregation limit are also shown. Adapted from Ref. [69], M. Müller and K. Binder, *Macromolecules* 31, 8323 (1998). The inset shows the dependence of the contact angle of droplets of the A-component on the monomer wall interaction $A = \epsilon_w/k_BT$ for the two temperatures.

critical temperature is shifted to lower temperatures and the critical composition is enriched in the component that is attracted by the wall.

In the second case of antisymmetric boundaries [163–165], enrichment layers of A and B gradually build at the corresponding surfaces in the vicinity of the critical temperature of the bulk. These enrichment layers stabilize an interface, which is delocalized in the middle of the film. The phase behavior can be rationalized by the effective interface potential that acts on this interface. If the film thickness is sufficiently large, it results from a superposition of the interface potentials from the surfaces. This superposition is sketched in Fig. 16 in the vicinity of a first-order wetting transition. A double tangent construction relates the two coexisting phases. Far below T_{wet}, the interface is localized either at the right or the left wall (i.e., there is a two-phase coexistence between an almost pure A-rich phase and an almost pure B-rich phase). Those phases coexist laterally, and the interface between those two coexisting phases is an AB interface that runs straight across the film and makes the macroscopic contact angle with each surface (see, too, Fig. 19) [159]. In the vicinity of T_{wet}, of the semi-infinite system, the interface potential has three minima corresponding to an A-rich phase, a B-rich phase, and a phase with symmetric composition, where the interface is delocalized in the middle of the film. Slightly above this triple temperature, the prewetting transitions at each surface give rise to two two-phase coexistence regions. On the A-rich side, a surface with a thin and a surface with a thick enrichment layer of B coexist. As we reduce the temperature, the thick enrichment layer becomes thicker and the thin layer becomes thinner. When the AB interface in the phase with the thick enrichment layer reaches the middle of the film (delocalized state) a triple point forms, where this phase coexists with an A-rich and a B-rich phase (localized states). As we increase the temperature, the difference between the thin and the thick enrichment layer decreases and each two-phase coexistence terminates in a critical point. When the film thickness grows, the critical temperature converges toward the temperature of the prewetting critical point of the surfaces, and the triple point corresponds to the wetting transition of the semi-infinite system [159].

The prewetting lines of first-order wetting transitions in semi-infinite systems give rise to lateral phase coexistence in a thin film. If the film thickness becomes of the order of the thickness of the enrichment layer at the prewetting coexistence, the qualitative phase behavior is altered [144, 166]. This is illustrated in Fig. 17. As we decrease the film thickness of an antisymmetry film, the two critical points move toward the symmetry axis of the phase diagram. For one film thickness D_{tri} the two critical points merge and form a tricritical point. Reducing the film thickness further, we find only a single critical point (i.e., a second-order interface localization/delocalization transition [159, 166]).

The dependence of the phase diagram on the surface interactions within the self-consistent field calculations is presented in Fig. 18 [159]. Panel (a) presents the binodals, and (b) shows

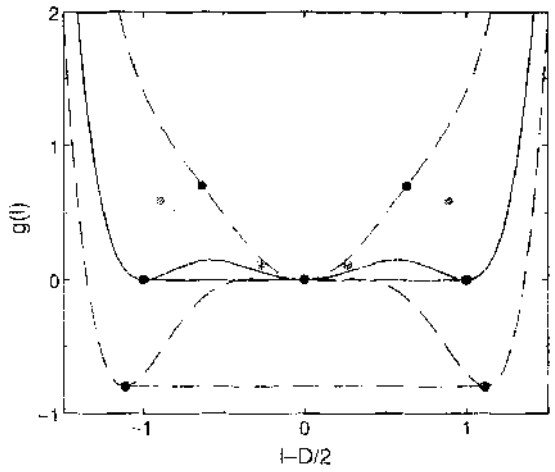


Figure 16. Sketch of the superposition of the interface potential that originates from each wall for strictly antisymmetric boundary conditions. The interface potential of a single wall in the vicinity of a first-order wetting transition is depicted in Fig. 6b.

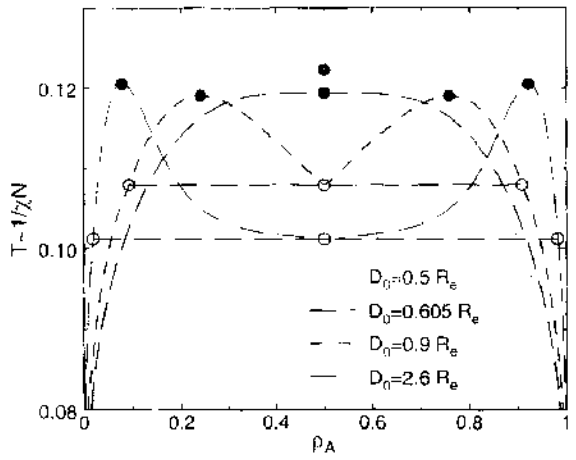


Figure 17. Phase diagram of an antisymmetric film ($A = A_1 = -A_2$) for $AN = 0.5$ and various film thicknesses D_0. For $D_0 = 2.6R_e$ and $0.9R_e$, the interface localization/delocalization transition is first order, $D_0 = 0.605R_e$ corresponds to a tricritical transition, whereas the transition is second order for $D_0 = 0.5R_e$. Adapted from Ref. [144], M. Müller et al., *Phys. Rev. E* 62, 5281 (2000).

the coexistence curves as a function of the temperature $1/\chi N$ and exchange potential $\Delta\mu$. The bottom surface attracts the A component of the symmetric mixture, while we vary the interaction at the top surface from attracting A to attracting B. If both surfaces attract the A component with the same strength, capillary condensation [141] occurs. The critical point is shifted toward higher A concentration and lower temperatures compared to the bulk ($\phi_c = 1/2$ and $\chi_c N = 2$). The coexisting phases have almost uniform composition across the film and differ in their average composition ρ_A. As we reduce the preference of the top surface for species A, the critical point and the critical composition tend toward their bulk values (i.e., the critical temperature increases and the critical composition becomes more symmetric [159]). The coexistence curve in the $1/\chi N$–$\Delta\mu$ plane approaches the symmetry axis. On making the top surface attract the other component B, we gradually change the character of the phase transition toward an interface localization/delocalization transition. In this case, there are enrichment layers of the A component at the bottom and the B component at the top, and the two coexisting phases differ in the location of the AB interface, which runs parallel to the surfaces. As the preferential interaction of the top surface increases, the critical temperature decreases and the critical composition becomes richer in A. When the coexistence curve in the $\Delta\mu - 1/\chi N$-plane intersects the prewetting line of the bottom

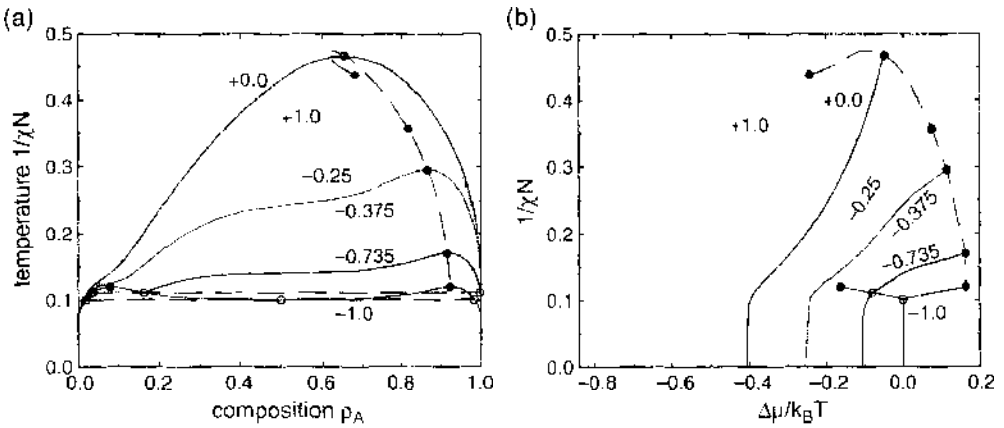


Figure 18. (a) Binodals for $D_0 = 2.6R_e$ and $A_1N = 0.5$. The ratio A_2/A_1 varies, as indicated in the key. The dashed curve shows the location of the critical points. Filled circles mark critical points; open circles/dashed horizontal lines denote three-phase coexistence for $A_2N = -0.3675$ and -0.5. The inset presents part of the phase boundary for antisymmetric boundaries. (b) Coexistence curves in the χN–$\Delta\mu$ plane. The "quasi-prewetting" lines for $\Delta\mu < 0$ and $A_2N = -0.3675$ and -0.5 are indistinguishable, because they are associated with the prewetting behavior of the surface with interaction $A_1N = +0.5$. Adapted from Ref. [159], M. Müller et al., *Europhys. Lett.* 50, 724 (2000).

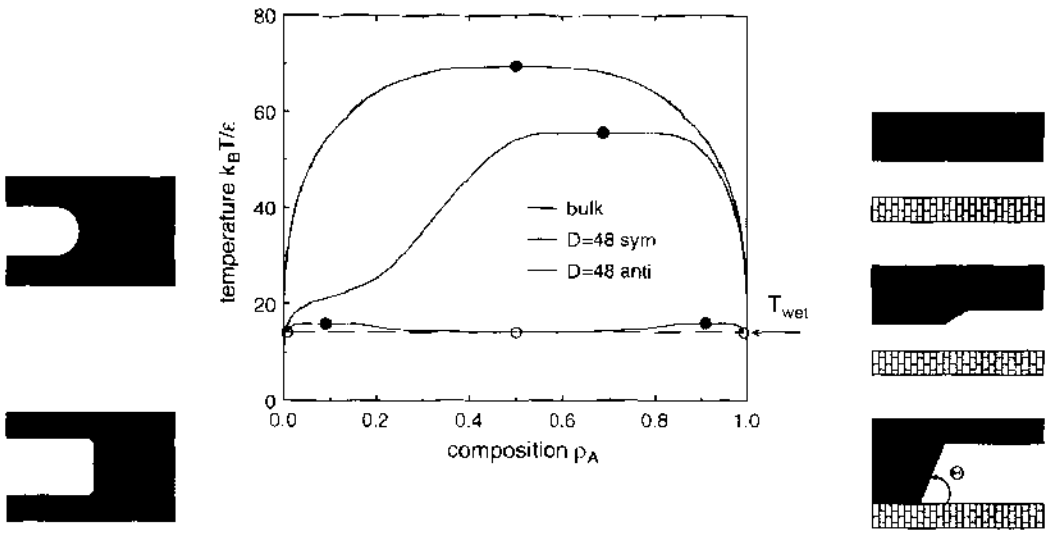


Figure 19. Phase diagram of a binary polymer blend of $N = 32$ monomers in a thin film of width $D = 48 \approx 7R_g$ (in units of the lattice spacing a). The miscibility gap in the bulk, a film with symmetric and a film with antisymmetric boundaries $A = 0.16k_BT$, is shown. The sketches depict typical configurations above and below the T_{wet} for capillary condensation (left) and interface localization.delocalization (right). Adapted from Ref. [160], M. Müller and K. Binder, *Phys. Rev. E* 63, 021602 (2001).

surface at $\Delta\mu < 0$, a triple point forms at which an A-rich phase and two B-rich phases with a thin and a thick A-enrichment layer coexist. When the bottom surface attracts A with exactly the same strength as the top surface attracts B (antisymmetric surfaces), the phase diagram becomes symmetric, as in Fig. 17. It comprises two critical points, below which there are two two-phase coexistences.

The qualitative features of the phase behavior are confirmed by Monte Carlo simulations of the bond fluctuation model (cf. Fig. 19) [160]. Fluctuations, which are ignored in the self-consistent field calculations, result in two modifications of the mean field result: in the vicinity of the critical points, we observed two-dimensional Ising critical behavior, and capillary waves renormalize the effective interface potential, which, in turn, leads to a stabilization of the delocalized state and a systematic overestimation of the triple temperature by the mean field approximation [160].

7. OUTLOOK

In this chapter, we have discussed some aspects of thin polymer films within the framework of coarse-grained models. Those models provide information on the length scale of a few nanometers—they do not explicitly account for details of the atomistic structure, and even on state-of-the-art computers they can reach only length scales of a few tens of nanometers. This length scale is characteristic for the extension of a polymer molecule R_g, and it is on this length scale where many fascinating phenomena occur in thin polymer films and where coarse-grained models can contribute to understanding of materials behavior.

Within the framework of a coarse-grained model, Monte Carlo simulations and self-consistent field calculations provide detailed information about structure and thermodynamics. One can investigate the dependence of the behavior on coarse-grained parameters (e.g., the Flory–Huggins parameter or the Hamaker constant) and thereby make rather close connection to experiments.

The self-consistent field calculations invoke approximations, both in the formulation of the free-energy functional that describes the interactions between monomeric units and in the solution of the statistical mechanics of the many-chain problem. Nevertheless, they often capture the qualitative behavior, and they are computationally fast enough to explore a wide range of parameters. Computer simulations do not invoke these approximations, and they provide essentially exact results, apart from well-controlled finite size effects and statistic

uncertainties. Comparing computer simulations and self-consistent field calculations quantitatively for the same model system, we can gauge the accuracy of the approximations involved in the self-consistent treatment. Typically, simulations are computationally much more demanding than self-consistent field calculations, but this drawback can be reduced by sophisticated techniques. Some of the advanced simulation methods have been described (e.g., reweighting techniques to determine phase equilibria and free energies of spatially inhomogeneous systems), but many other, significant techniques have been devised to make simulations more efficient. For instance, there are special techniques to deal with long-ranged electrostatic interactions [167, 168], or Monte Carlo moves that alter the chain connectivity [31] to equilibrate the macromolecular conformations faster.

Both methods—Monte Carlo simulations and self-consistent field calculations—are versatile enough to tackle a variety of interesting questions. To connect with atomistic modeling, one can use more realistic chain models both in the simulations and in the self-consistent field calculations. To predict material properties, one challenging problem consists of mapping a specific chemical substance onto the coarse-grained model. Then one also has to address the question of which interactions on the coarse-grained scale are necessary and how to transfer atomistic interactions to a coarse-grained model.

To make connection with methods on larger length scales, one has to extract parameters from coarse-grained models [e.g., the interface tension γ and the interface potential $g(l)$] that can subsequently be used in phenomenological models (e.g., effective interface Hamiltonian, Landau–Ginzburg models). In this context, it is also interesting to investigate kinetic phenomena (e.g., dewetting or kinetics of phase separation), because many materials properties depend on their history (e.g., morphology of a blend).

ACKNOWLEDGMENTS

It is a great pleasure to thank Ezequiel V. Albano, Kurt Binder, Yu-cheng Lin, Valery Luchnikov, Luis G. MacDowell, Sergej Minko, Peter Müller-Buschbaum, Manfred Stamm, and Arun Yethiraj for their collaboration. Financial support from the Deutsche Forschungsgemeinschaft under grant Mu 1674/1 as well as ample computer time at the computing center of the university of Mainz, the John von Neumann Institut für computing Jülich, and the High Performance Computing Center Stuttgart are greatfully acknowledged.

REFERENCES

1. N. Awaji, Y. Sugita, S. Ohkubo, and T. Nakanishi, *Jpn. J. Appl. Phys.* 34, L1013 (1995).
2. J. Bauer, G. Drescher, and M. Illig, *J. Vac. Sci. Technol. B* 14, 2485 (1996).
3. J. Becker, G. Grün, R. Seemann, H. Mantz, K. Jacobs, K. R. Mecke, and R. Blossey, *Nat. Mater.* 2, 59 (2003); S. Herminghaus, *ibid.* 2, 11 (2003).
4. R. Seemann, S. Herminghaus, and K. Jacobs, *J. Phys. Condens. Matter* 13, 4925 (2001).
5. P. Müller-Buschbaum, *J. Phys. Condens. Matter* 15, R1549 (2003).
6. L. Delle Site, C. F. Abrams, A. Alavi, and K. Kremer, *Phys. Rev. Lett.* 89, 156103 (2002); C. F. Abrams, L. Delle Site, and K. Kremer, *Phys. Rev. E* 67, 021807 (2003).
7. P. Müller-Buschbaum, *J. Phys.: Condens. Mat.* 15, 1549 (2003).
8. M. Schick, in "Les Houches lectures on Liquids at Interfaces" (J. Charvolin, J. F. Joanny, and J. Zinn-Justin, Eds.). Elsevier Science, Amsterdam, 1990.
9. M. Müller, K. Katsov, and M. Schick, *J. Polym. Sci. B* 41, *Polym. Phys.* 1441 (2003).
10. J. Baschnagel and K. Binder, *Macromolecules* 28, 6808 (1995).
11. S. O. Kim, H. H. Solak, M. P. Stoykovich, N. J. Ferrier, J. J. de Pablo, and P. F. Nealey, *Nature* 424, 411 (2003).
12. E. Reister, M. Müller, and U. Binder, *Phys. Rev. E* 64, 041804 (2001).
13. P. G. deGennes, *Rev. Mod. Phys.* 57, 827 (1985).
14. F. Brochard-Wyart and J. Daillant, *J. Can. J. Phys.* 68, 1084 (1990).
15. N. Madras and A. D. Sokal, *J. Stat. Phys.* 50, 109 (1988).
16. P. G. de Gennes, "Scaling Concepts in Polymer Physics." Cornell University Press, Ithaca, NY, 1979.
17. I. Carmesin and K. Kremer, *Macromolecules* 21, 2819 (1988); H.-P. Deutsch and K. Binder, *J. Chem. Phys.* 94, 2294 (1991).
18. K. Kremer and G. S. Grest, *J. Chem. Phys.* 92, 5057 (1990).
19. C. Bennemann, W. Paul, K. Binder, and B. Dünweg, *Phys. Rev. E* 57, 843 (1998).
20. F. Müller-Plathe, *Chem. Phys. Chem.* 3, 754 (2002).
21. J. Baschnagel et al., *Adv. Polym. Sci.* 152, 42 (2000).

22. R. W. Chang and A. Yethiraj, *J. Chem. Phys.* 114, 7688 (2001).
23. P. H. Van Konynenburg and R. L. Scott, *Phil. Trans. R. Soc. A* 298, 498 (1980).
24. A. Indrakanti, J. K. Maranas, A. Z. Panagiotopoulos, and S. K. Kumar, *Macromolecules* 34, 8596 (2001); A. Z. Panagiotopoulos, *J. Chem. Phys.* 112, 7132 (2000).
25. P. K. Lai and K. Binder, *J. Chem. Phys.* 95, 9288 (1991).
26. N. B. Wilding, M. Müller, and K. Binder, *J. Chem. Phys.* 105, 802 (1996).
27. V. A. Ivanov, M. R. Stukan, M. Müller, W. Paul, and K. Binder, *J. Chem. Phys.* 118, 10333 (2003).
28. M. Müller and L. G. MacDowell, *Macromolecules* 33, 3902 (2000).
29. K. Kremer and K. Binder, *Comp. Phys. Rev.* 7, 259 (1988).
30. B. M. Forrest and U. W. Suter, *J. Chem. Phys.* 101, 2616 (1994).
31. N. C. Karayiannis, A. E. Giannousaki, V. G. Mavrantzas, and D. N. Theodorou, *J. Chem. Phys.* 117, 5465 (2002); N. C. Karayiannis, V. G. Mavrantzas, and D. N. Theodorou, *Phys. Rev. Lett.* 88, 105503 (2002); V. G. Mavrantzas, T. D. Boone, E. Zervopoulou, and D. N. Theodorou, *Macromolecules* 32, 5072 (1999).
32. J. I. Siepmann, *Mol. Phys.* 70, 1145 (1990).
33. D. Frenkel, G. C. A. M. Mooij, and B. Smit, *J. Phys. Condens. Matter* 3, 3053 (1992).
34. M. Laso, J. J. de Pablo, and U. W. Suter, *J. Chem. Phys.* 97, 2817 (1992).
35. C. Borgs and R. Kotecky, *J. Stat. Phys.* 60, 79 (1990); *Phys. Rev. Lett.* 68, 1734 (1992).
36. M. Müller and N. B. Wilding, *Phys. Rev. E* 57, 2079 (1994); M. Müller and K. Binder, *Comp. Phys. Commun.* 84, 173 (1994).
37. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* 61, 2635 (1988).
38. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* 63, 1195 (1989).
39. M. Müller and M. Schick, *J. Chem. Phys.* 105, 8885 (1996).
40. K. Binder, *Phys. Rev. A* 25, 1699 (1982).
41. B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* 68, 9 (1992); B. A. Berg, U. Hansmann, and T. Neuhaus, *Z. Phys. B* 90, 229 (1993).
42. J.-S. Wang, T. K. Tay, and R. H. Swendsen, *Phys. Rev. Lett.* 82, 476 (1999).
43. G. R. Smith and A. D. Bruce, *J. Phys. A* 28, 6623 (1995).
44. B. A. Berg, *J. Stat. Phys.* 82, 323 (1996).
45. F. Wang and D. P. Landau, *Phys. Rev. Lett.* 86, 2050 (2001).
46. Q. Yan and J. J. de Pablo, *Phys. Rev. Lett.* 90, 035701 (2003).
47. U. H. E. Hansmann and L. T. Wille, *Phys. Rev. Lett.* 88, 068105 (2002).
48. B. J. Schulz, K. Binder, M. Müller, and D. P. Landau, *Phys. Rev. E* 67, 067102 (2003).
49. D. Frenkel and B. Smith, "Understanding Molecular Simulation," Academic, Boston, 1996.
50. P. Virnau and M. Müller, *J. Chem. Phys.* 120, 10925 (2004).
51. M. Rao and B. J. Berne, *Mol. Phys.* 37, 455 (1978).
52. B. Smit, *Phys. Rev. A* 37, 3481 (1988).
53. J. K. Singh, D. A. Kofke, and J. R. Errington, *J. Chem. Phys.* 119, 3405 (2003).
54. E. Helfand and Y. Tagami, *J. Chem. Phys.* 56, 3592 (1972); E. Helfand, *J. Chem. Phys.* 62, 999 (1975).
55. J. Noolandi and K. M. Hong, *Macromolecules* 14, 727 (1981); *ibid.* 15, 483 (1982).
56. K. R. Shull, *Macromolecules* 26, 2346 (1993).
57. J. M. H. M. Scheutjens and G. J. Fleer, *J. Phys. Chem.* 83, 1619 (1979); *ibid.* 84, 178 (1979); *Macromolecules* 18, 1882 (1985).
58. M. W. Matsen and M. Schick, *Phys. Rev. Lett.* 74, 4225 (1995).
59. C. E. Woodward, *J. Chem. Phys.* 94, 3183 (1991).
60. C. E. Woodward and A. Yethiraj, *J. Chem. Phys.* 100, 3181 (1994).
61. C. N. Patra and A. Yethiraj, *J. Chem. Phys.* 112, 1570 (2000).
62. S. Sen, J. M. Cohen, J. D. McCoy, and J. G. Curro, *J. Chem. Phys.* 101, 9010 (1994); *ibid.* 102, 3431 (1994).
63. C. N. Patra and A. Yethiraj, *J. Chem. Phys.* 118, 4702 (2003).
64. M. Müller and L. G. MacDowell, *Europhys. Lett.* 55, 221 (2001).
65. M. Müller, L. G. MacDowell, and A. Yethiraj, *J. Chem. Phys.* 118, 2929 (2003).
66. I. Szleifer, *Curr. Opin. Colloid Interface Sci.* 2, 416 (1997); I. Szleifer and M. A. Carignano, *Adv. Chem. Phys.* 94, 742 (1996).
67. M. Müller, *Macromolecules* 31, 9044 (1998).
68. M. Müller, *Macromol. Theory Simul.* 8, 343 (1999).
69. M. Müller and K. Binder, *Macromolecules* 31, 8323 (1998).
70. K. O. Rassmusen and G. Kalosakas, *J. Polym. Sci. B* 44, 1777 (2002); S. W. Sides and G. H. Fredrickson, *Polymer* 44, 5859 (2003).
71. M. W. Matsen and M. Schick, *Phys. Rev. Lett.* 72, 2660 (1994).
72. N. Saito, K. Takahashi, and Y. Yunoki, *J. Phys. Soc. Jpn.* 22, 219 (1967).
73. D. C. Morse and G. H. Fredrickson, *Phys. Rev. Lett.* 73, 3235 (1994).
74. F. Schmid and M. Müller, *Macromolecules* 28, 8639 (1995).
75. E. vanSwol and J. R. Henderson, *Phys. Rev. A* 43, 2932 (1991).
76. J. D. Weeks, K. Katsov, and K. Vollmayr, *Phys. Rev. Lett.* 81, 4400 (1998); J. D. Weeks, K. Vollmayr, and K. Katsov, *Physica A* 244, 461 (1997).
77. K. Katsov and J. D. Weeks, *J. Phys. Chem. B* 10, 6758 (2001).
78. M. S. Wertheim, *J. Chem. Phys.* 87, 7323 (1987).
79. J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, *Mol. Phys.* 3, 591 (1993).

80. W. G. Chapman, G. Jackson, and K. S. Gubbins, *Mol. Phys.* 65, 1057 (1988).
81. G. Stell and Y. Zhou, *J. Chem. Phys.* 91, 3618 (1989).
82. G. Stell and Y. Zhou, *J. Chem. Phys.* 96, 1504 (1992).
83. C. Vega, C. McBride, and L. G. MacDowell, *Phys. Chem. Chem. Phys.* 4, 853 (2002).
84. L. G. MacDowell, M. Müller, C. Vega, and K. Binder, *J. Chem. Phys.* 113, 419 (2000).
85. L. G. MacDowell, P. Virnau, M. Müller, and K. Binder, *J. Chem. Phys.* 117, 6360 (2002).
86. P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press, Ithaca, NY, 1954.
87. A. Barker and D. Henderson, *J. Chem. Phys.* 47, 4714 (1967).
88. N. F. Carnahan and K. E. Starling, *J. Chem. Phys.* 51, 635 (1969).
89. A. Yethiraj, *J. Chem. Phys.* 109, 3269 (1998).
90. K. S. Schweizer and J. G. Curro, *Adv. Chem. Phys.* 98, 1 (1997).
91. T. Young, *Philos. Trans. R. Soc. London* 57, 827 (1805).
92. L. G. MacDowell, M. Müller, and K. Binder, *Colloids and Surfaces A* 206, 277 (2002).
93. D. Adrienko, B. Dünweg, and O. I. Vinogradowa, *J. Chem. Phys.* 119, 13106 (2003).
94. P. Adams and J. R. Henderson, *Mol. Phys.* 73, 1383 (1991).
95. S. Bruin, M. J. P. Nijmeijer, and R. M. Crevecoeur, *J. Chem. Phys.* 102, 7622 (1995).
96. D. Nicolaides and B. Evans, *Phys. Rev. Lett.* 63, 778 (1989).
97. N. B. Wilding, *Phys. Rev. E* 52, 602 (1995).
98. E. H. Hauge and M. Schick, *Phys. Rev. B* 27, 4288 (1983).
99. R. Seemann, S. Herminghaus, and K. Jacobs, *Phys. Rev. Lett.* 86, 5534 (2001).
100. M. Müller, L. G. MacDowell, P. Müller-Buschbaum, O. Wunnike, and M. Stamm, *J. Chem. Phys.* 115, 9960 (2001).
101. P. Müller-Buschbaum, P. Vanhoorne, V. Scheumann, and M. Stamm, *Europhys. Lett.* 40, 655 (1997).
102. A. M. Cazabat, in "Les Houches lectures on Liquids at Interfaces" (J. Charvolin, J. F. Joanny, J. Zinn-Justin, Eds.), p. 374. Elsevier Science, Amsterdam, 1990.
103. R. Yerushalmi-Rozen, J. Klein, and J. L. Fetters, *Science* 263, 793 (1994).
104. G. Reiter, J. Schultz, P. Auroy, and L. Auvray, *Europhys. Lett.* 33, 29 (1996).
105. Y. Liu, M. H. Rafailovich, J. Sokolov, S. A. Schwarz, X. Zhong, A. Eisenberg, E. J. Kramer, B. B. Sauer, and S. Satija, *Phys. Rev. Lett.* 73, 440 (1994).
106. C. Yuan, M. Ouyang, and J. T. Koberstein, *Macromolecules* 32, 2329 (1999); C. Jalbert, J. T. Koberstein, A. Hariharan, and S. K. Kumar, *Macromolecules* 30, 4481 (1997).
107. L. Leibler, A. Ajdari, A. Mourran, G. Coulon, and D. Chatenay, in "Ordering in Macromolecular Systems" (A. Teramoto and T. Norisuya, Eds.), Springer, Berlin, 1994.
108. C. Gay, *Macromolecules* 30, 5939 (1997).
109. M. Carignano, R. Yerushalmi-Rozen, and N. Dan, *Macromolecules* 33, 3453 (2000).
110. P. G. Ferreira, A. Ajdari, and L. Leibler, *Macromolecules* 31, 3994 (1998).
111. J. I. Martin, Z.-G. Wang, and M. Schick, *Langmuir* 12, 4950 (1996).
112. K. R. Shull, *Macromolecules* 29, 8487 (1996).
113. M. W. Masten and J. M. Gardiner, *J. Chem. Phys.* 115, 2794 (2001).
114. T. Kerle, R. Yerushalmi-Rozen, and J. Klein, *Europhys. Lett.* 38, 207 (1997).
115. J. Jopp and R. Yerushalmi-Rozen, *Macromolecules* 32, 7269 (1999); T. Kerle, R. Yerushalmi-Rozen, and J. Klein, *Macromolecules* 31, 422 (1998).
116. J. D. McCoy, Y. Ye, and J. G. Curro, *J. Chem. Phys.* 117, 2975 (2002); Y. Ye, J. D. McCoy, and J. G. Curro, *J. Chem. Phys.* 119, 555 (2003).
117. T. Kreer, S. Metzger, M. Müller, J. Baschnagel, and K. Binder, *J. Chem. Phys.* 120, 4012 (2004).
118. K. Ragil, J. Meunier, D. Broseta, J. O. Indekeu, and D. Bonn, *Phys. Rev. Lett.* 77, 1532 (1996).
119. N. Shahidzadeh, D. Bonn, K. Ragil, D. Broseta, and J. Meunier, *Phys. Rev. Lett.* 80, 3992 (1998).
120. J. O. Indekeu, K. Ragil, D. Bonn, D. Broseta, and J. Meunier, *J. Stat. Phys.* 95, 1009 (1999).
121. E. Bertrand, H. Dobbs, D. Broseta, J. Indekeu, D. Bonn, and J. Meunier, *Phys. Rev. Lett.* 85, 1282 (2000).
122. J. H. Maas, M. A. Cohen-Stuart, F. A. M. Leermakers, and N. A. M. Besseling, *Langmuir* 16, 3478 (2000).
123. A. Vornov and O. Shafranska, *Langmuir* 18, 4471 (2002).
124. J. H. Maas, G. J. Fleer, F. A. M. Leermakers, and M. A. Cohen-Stuart, *Langmuir* 18, 8871 (2002).
125. P. K. Lai and K. Binder, *J. Chem. Phys.* 97, 586 (1992); W. Zhao, G. Krausch, M. H. Rafailovich, and J. Sokolov, *Macromolecules* 27, 2933 (1994).
126. M. Müller and L. G. MacDowell, *J. Phys. Condens. Matter* 15, R609 (2003).
127. C. Bauer and S. Dietrich, *Phys. Rev. E* 60, 6919 (1999).
128. C. Bauer, S. Dietrich, and A. O. Parry, *Euro. Phys. Lett.* 47, 474 (1999).
129. C. Bauer and S. Dietrich, *Phys. Rev. E* 61, 1664 (2000).
130. M. Park, C. Harrison, P. M. Chaikin, R. A. Register, and D. H. Adamson, *Science* 276, 1401 (1997); T. Thurn-Albrecht, J. Schotter, C. A. Kastle, N. Emley, T. Shibauchi, L. Krusin-Elbaum, K. Guarini, C. T. Black, M. T. Tuominen, and T. P. Russell, *Science* 290, 2126 (2000); M. J. Fasolka and A. M. Mayes, *Annu. Rev. Mater. Res.* 31, 323 (2001); S. Ludwigs, A. Boker, A. Voronov, N. Rehse, R. Magerle, and G. Krausch, *Nat. Mater.* 2, 744 (2003).
131. M. Müller, *Phys. Rev. E* 65, 030802(R) (2002).
132. S. Minko, M. Müller, D. Usov, A. Scholl, C. Froeck, and M. Stamm, *Phys. Rev. Lett.* 88, 035502 (2002).
133. J. F. Marko and T. A. Witten, *Phys. Rev. Lett.* 66, 1541 (1991); *Macromolecules* 25, 296 (1992).
134. A. Sidorenko, S. Minko, K. Schenk-Meuser, H. Duschner, and M. Stamm, *Langmuir* 15, 8349 (1999).

135. S. Minko, M. Müller, M. Motornov, N. Nischke, K. Grundke, D. Usov, A. Scholl, and M. Stamm, *J. Am. Chem. Soc.* 125, 3896 (2003).
136. R. R. Netz and M. Schick, *Europhys. Lett.* 37, 38 (1997); *Macromolecules* 31, 5105 (1998); J. L. Goveas, S. T. Milner, and W. B. Russel, *Macromolecules* 30, 5541 (1997).
137. L. Leibler, *Macromolecules* 13, 1602 (1980).
138. C. Singh, G. T. Pickett, E. Zhulina, and A. C. Balazs, *J. Phys. Chem. B* 101, 10614 (1997).
139. M. Müller, *Comp. Phys. Comm.* 147, 292 (2002).
140. S. Dietrich, in "Phase Transitions and Critical Phenomena" (C. Domb and J. Lebowitz, Eds.), Vol. 12, Academic, London, 1988.
141. M. E. Fisher and H. Nakanishi, *J. Chem. Phys.* 75, 5857 (1981); H. Nakanishi and M. E. Fisher, *J. Chem. Phys.* 78, 3279 (1983).
142. M. E. Fisher, *J. Phys. Soc. Japan S* 26, 87 (1969).
143. L. D. Gelb, K. E. Gubbins, R. Radhakrishnan, and M. Sliwinska-Bartkowiak, *Rep. Prog. Phys.* 62, 1573 (1999).
144. M. Müller, E. V. Albano, and K. Binder, *Phys. Rev. E* 62, 5281 (2000).
145. T. Kerle, J. Klein, and K. Binder, *Phys. Rev. Lett.* 77, 1318 (1996).
146. J. Rysz, A. Budkowski, A. Bernasik, J. Klein, K. Kowalski, J. Jedlinski, and L. J. Fetters, *Europhys. Lett.* 50, 35 (2000).
147. M. Geoghegan, H. Ermer, G. Jüngst, G. Krausch, and R. Brenn, *Phys. Rev. E* 62, 940 (2000); M. Geoghegan and G. Krausch, *Prog. Polym. Sci.* 28, 261 (2003).
148. P. Virnau, M. Müller, L. G. MacDowell, and K. Binder, *Comp. Phys. Comm.* 147, 378 (2002); P. Virnau, Dissertation, Mainz 2003.
149. K. Binder, M. Müller, L. G. MacDowell, and P. Virnau, *Adv. Polym. Sci* (in press).
150. A. Sariban and K. Binder, *Macromolecules* 21, 711 (1988); 24, 587 (1991); *J. Chem. Phys.* 86, 5859 (1987).
151. A. Werner, F. Schmid, M. Müller, and K. Binder, *Phys. Rev. E* 59, 728 (1999).
152. M. L. Huggins, *J. Chem. Phys.* 9, 440 (1941); P. J. Flory, *J. Chem. Phys.* 9, 660 (1941).
153. M. Müller and K. Binder, *Macromolecules* 28, 1825 (1995); M. Müller and K. Binder, *Comp. Phys. Comm.* 84, 173 (1994).
154. M. Müller, *Macromolecules* 28, 6556 (1995).
155. J. Dudowicz, K. F. Freed, and J. F. Douglas, *J. Chem. Phys.* 116, 9983 (2002); *Phys. Rev. Lett.* 88, 095503 (2002).
156. M. W. Matsen, *J. Chem. Phys.* 106, 7781 (1997); T. Geisinger, M. Müller, and K. Binder, *J. Chem. Phys.* 111, 5241 (1999).
157. A. N. Semenov, *J. Phys. II* 6, 1759 (1996).
158. J. W. Cahn, *J. Chem. Phys.* 66, 3667 (1977).
159. M. Müller, K. Binder, and E. V. Albano, *Europhys. Lett.* 50, 724 (2000).
160. M. Müller and K. Binder, *Phys. Rev. E* 63, 021602 (2001).
161. A. O. Parry and R. Evans, *J. Phys. A* 25, 275 (1992).
162. A. O. Parry, *J. Phys. Condens. Matt.* 8, 10761 (1996); R. Evans, *J. Phys. Condens. Matt.* 2, 8989 (1990); R. Evans, U. Marini, B. Marconi, and P. Tarazona, *J. Chem. Phys.* 84, 2376 (1986).
163. F. Brochard-Wyart and P.-G. de Gennes, *Acad. Sci. Paris* 297, 223 (1983).
164. A. O. Parry and R. Evans, *Phys. Rev. Lett.* 64, 439 (1990); *Physica A* 181, 250 (1992).
165. K. Binder, D. P. Landau, and A. M. Ferrenberg, *Phys. Rev. Lett.* 74, 298 (1995); *Phys. Rev. E* 51, 2823 (1995); E. V. Albano, K. Binder, D. W. Heermann, and W. Paul, *Surf. Sci.* 233, 151 (1989).
166. M. R. Swift, A. L. Owczarek, and J. O. Indekeu, *Europhys. Lett.* 14, 475 (1995).
167. M. Deserno and C. Holm, *J. Chem. Phys.* 109, 7678 (1998).
168. A. C. Maggs, *J. Chem. Phys.* 117, 1975 (2002); *Phys. Rev. Lett.* 88, 196402 (2002).

CHAPTER 5

Conjugated Organic Polymers: From Bulk to Molecular Wire

Ulrike Salzner
Department of Chemistry, Bilkent University, Ankara, Turkey

CONTENTS

1. Introduction 204
   1.1. Properties of Conjugated Organic Polymers 204
   1.2. Theoretical Approach 206
2. Accuracy of Theoretical Excitation Energies for Polyenes in the Gas Phase 209
   2.1. High Level *Ab Initio* Calculations 209
   2.2. Single-Electron Methods 211
   2.3. Calculation of Excitation Energies Employing Density-Functional Theory 213
   2.4. Frozen Orbital Approximation 214
3. Extrapolation to Infinitely Long Chains 217
4. Excitation Energies of Oligomers in Liquid and Solid Solutions 221
   4.1. Polyenes 222
   4.2. Thiophene Oligomers 222
   4.3. Pyrrole Oligomers 223
5. Interchain Interactions 224
   5.1. Experiment 224
   5.2. Theory 225
6. Comparison Between Theory and Experiment for Polymers 227
7. Conducting State of Conjugated Organic Polymers 229
8. Polarizability 235
9. Conductivity 237
10. Summary 239
   References 241

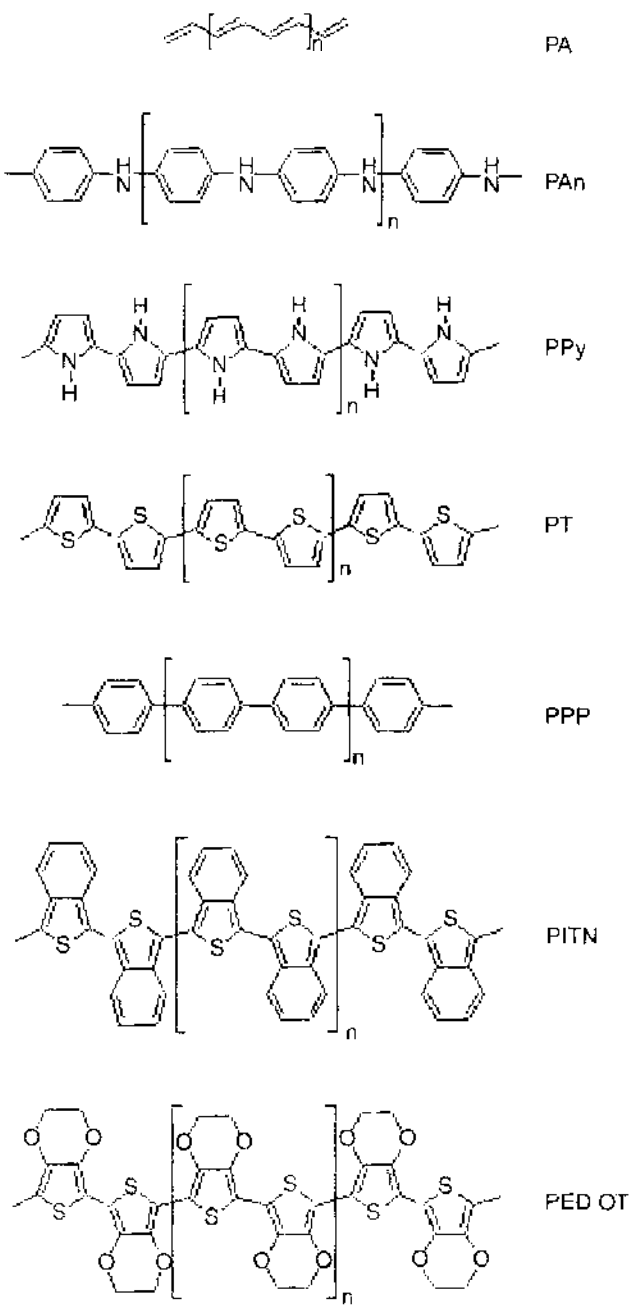
ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (203–251)

1. INTRODUCTION

1.1. Properties of Conjugated Organic Polymers

Conjugated organic polymers (COPs) combine remarkable optical and electrical properties with typical properties of plastics such as light weight, flexibility, and low cost [1, 2]. In contrast to traditional plastics, COPs are polyunsaturated organic polymers with conjugated π-systems. The π-systems give rise not only to desirable properties but also to problems with environmental stability, stiffness, low solubility, and nonprocessibility [1–3]. COPs are semiconductors with band gaps ranging from almost 0 to about 4 eV. Depending on band gaps, the full range of colors is observed, but COPs may also be transparent or black [3]. Polyacetylene (PA), the prototype of COP, consists of carbon and hydrogen only and has the formula $(CH)_x$ (Scheme 1). In 1977 it was discovered that oxidizing (or p-doping) increases the conductivity of PA by many orders of magnitude [3, 4], transforming it from an insulator into a metal. Within 10 years, the conductivity of p-doped PA was optimized and is now



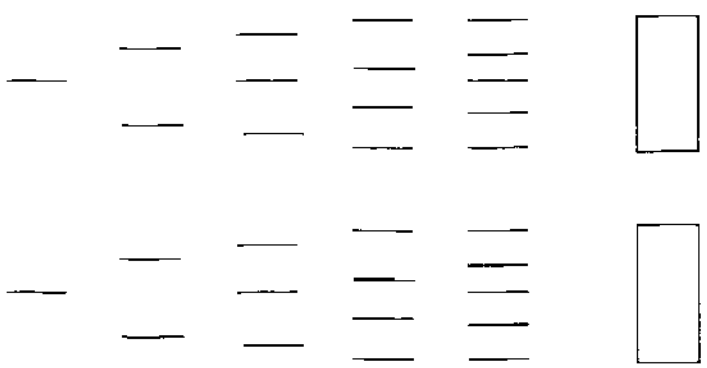
Scheme 1.

1.5×10^5 S/cm [5], close to that of copper, 5×10^5 S/cm [6]. In theoretical investigations, it was predicted that PA should be able to reach conductivities of at least 10^6 S/cm if the crystalline order of the samples is improved and the intrinsic value approached [6, 7]. The discovery of COPs opened a completely new field of research, and the number of publications dealing with conducting polymers reached several tens of thousands by 2004. In 2000, the Nobel Prize in chemistry was awarded to Shirakawa, MacDiarmid, and Heeger for the discovery and development of conducting polymers [8–10].

Although PA in its p-doped form holds the record in conductivity among all COPs, and although it has the lowest band gap of all parent (unsubstituted) systems, 1.4 eV [3], PA is not suitable for commercial applications because it is not stable enough under environmental influences and insoluble in common solvents [1]. COPs based on small aromatic ring systems like thiophene, pyrrole, benzene, and aniline are more stable and offer the possibility of modifying their properties through chemical substitution [11, 12]. Polyaniline (PAn), polypyrrole (PPy), polythiophene (PT), polyparaphenylene (PPP), polyisothianaphthene (PITN), and poly (ethylenedioxythiophene) (PEDOT) (Scheme 1) have band gaps between 1.0 and 3.5 eV and conductivities between 10^2 and 10^4 S/cm on p-doping [1, 2, 13–19]. Whether the lower conductivities compared to PA are caused by intrinsic differences or the different quality of the materials, different crystal packing or a different response to doping is not known to this date. Substituting hydrogen atoms with alkyl groups, halogens, and cyano groups can be used to increase solubility and to manipulate ionization potentials (IPs) and electron affinities (EAs) [11, 12]. Fusing ring systems to the backbone of heteroaromatic polymers as in PITN and in PEDOT was used to decrease band gaps, to prevent mislinking, and to increase planarity [11, 12].

Conductivity of COPs is caused by the delocalized nature of the π-electrons. Conductivity is therefore a property intrinsic to the individual polymer chains. This is a big difference to metals, where conductivity arises as a property of the bulk. Because properties of small metal clusters differ substantially from those of the bulk, conductivity in metals requires a minimum number of atoms. In contrast, the intrinsic conductivity [20] of fibers or single-polymer chains in COPs is expected to be proportionally higher than that measured for bulk samples, as electrons travel most easily along the chains, whereas the overall macroscopic conductivity is limited by the need for hopping between chains. Thus, with COPs, miniaturization is possible down to the single-molecule level. The width of a polymer chain is only a couple of Ångstrøms and is thus of subnanometer dimension. The length of the polymer chains is not limited in principle. Being able to manipulate individual molecules and to create nanostructured materials [20–89], self-assembled mono layers [90–92], and molecular wires [49, 92–106] is investigated with the aim of constructing a new generation of molecular electronic devices. Smaller dimensions allow for lower charge consumption and faster switching times [107], which are important properties for future computer architectures.

Optical properties also arise from the delocalized π-systems. Occupied and unoccupied conjugated π-orbitals are lying energetically between the σ-orbitals. As the size of the conjugated π-system increases, conjugated π-orbitals tend to form energy bands with decreasing energy gaps (Scheme 2). Thus, ultraviolet spectra are dominated by π–π^* transitions and show bathochromic shifts with extensions of the π-system. When a sufficient number of



Scheme 2.

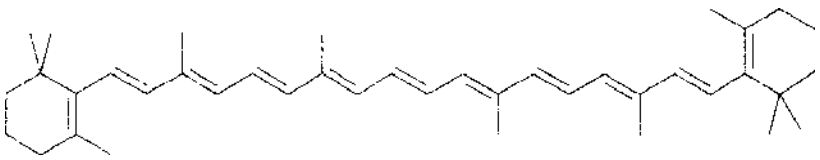
double bonds is conjugated, visible light can excite π-electrons, and the compound is colored. Excitation of electrons from the valence to the conduction band creates charge carriers, gives rise to a transient photocurrent [108–118] and opens the way to photovoltaics [41, 69, 83, 84, 113, 119–147]. The reverse process, charge injection from electrodes, leads to electron and hole pairs that travel toward one another and recombine. When the electron falls from the conduction band into the hole in the valence band, the energy difference is emitted as light. Electroluminescence is the basis of the construction of organic light emitting diodes (OLED)s [126, 148–171] and colored displays [172–175]. Other possible applications arising from electrical and optical properties of COPs are organic field effect transistors [17, 176–179], lightweight conductors, batteries [173, 180–198], supercapacitors [61, 199–203], sensors [71, 204–225], lasers [165, 226], and smart fabrics [227, 228], to name just a few.

The onset of a photocurrent in PA was observed at about 1.4 eV [109]. The coincidence of the onset of absorption in the optical spectrum and the onset of photoconductivity may be taken as proof that the band model is adequate for COPs and that free charge carriers are produced by a simple band transition. There is evidence, however, that photoexcitation in COPs differs from that in inorganic semiconductors because molecular organic solids form bound electron hole pairs rather than free charge carriers upon excitation [229, 230]. An important issue in this regard is the energy difference between a bound and a separate electron hole pair, the so-called "exciton binding energy." Exciton binding energies for COPs were claimed to be between 0 and well above 1 eV. The issue is still discussed controversially [118]. Theoretical calculations indicate that exciton binding energies decrease with increasing conjugation length and decrease when interchain interactions are accounted for [231]. The size of the exciton binding energies therefore seems to be influenced by the nature of the polymer and by the presence of impurities or defects, which influence the conjugation length. Because COPs may have rather short conjugation lengths, their properties could be closer to those of molecular crystals than to those of polymers; hence the large exciton binding energies. Maybe if the quality of the materials is improved, exciton binding energies will decrease and more band-like transitions will be observed. Therefore, the controversy in the experimental studies may be partially a result of short conjugation lengths.

To reach a permanently conducting state, COPs have to be oxidized or reduced. Because changes in electron distribution are coupled with geometry distortions in organic molecules, ionization alters the band structure. Therefore, injection of charges influences optical spectra, and therefore the color of the systems [3, 232–235]. This process is called electrochromism [41, 133, 172, 175, 236–243]. If the band gap of the neutral polymer is large, the polymer is transparent. Doping shifts the absorption into the visible range, and the polymer is colored in the conducting state. If the band gap of the neutral polymer is small, optical absorption of the doped form occurs in the infrared region, producing transparent conductors [13, 16, 157, 233, 244, 245]. To explain the effects on doping, localized defects such as solitons, polarons, and bipolarons have been invoked [3, 246–255].

1.2. Theoretical Approach

Theoretical research on conjugated π-systems [247, 256–259] started long before the discovery of the conductivity of PA, as conjugated π-systems are the crucial part of organic dyes [260]. The orange color of carotene (Scheme 3), for instance, is caused by 11 conjugated double bonds [261]. Optical properties of conjugated π-systems are important in biochemistry and are crucial for processes such as vision and light harvesting. The bathochromic shift with an increasing number of conjugated double bonds in polymethine dyes and of polyenes was investigated by Kuhn in the late 1940s [262, 263]. One important question was

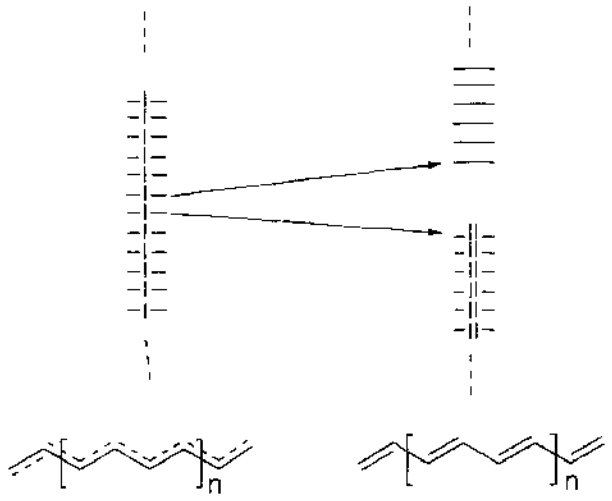


Scheme 3.

whether an infinite conjugated chain has a band gap; that is, whether an infinitely long conjugated π-system is a metal or a semiconductor. Peierls [264] showed theoretically that a one-dimensional system with equal bond distances and a half-filled energy band (singly occupied orbitals) is unstable with respect to a geometry distortion that introduces bond alternation (dimerization). In this process, half of the orbitals is lowered and the other half is increased in energy, and an energy gap opens up (Scheme 4). The total energy of the system is lowered by dimerization because the electrons doubly occupy the lower-lying orbitals. Thus, single, isolated, infinitely long PA chains are semiconductors with finite band gaps rather than metals. Bulk COPs may still be metals because interchain interactions are strong enough that the materials are not strictly one-dimensional [6]. Therefore, part of the research on COPs is devoted to band gap reduction with the aim of producing "synthetic metals" [265]. To this date, zero–band gap COPs with metallic conductivity (with decreasing conductivity on increase in temperature) have not been produced, although systems with vanishingly small electrochemical band gaps have been reported [19, 266–273].

The aim of theoretical investigation of COPs is, in principle, twofold. First, theory is employed to understand the physics of electrical and optical properties. Second, theoretical insights can be used to design new systems with specifically tailored properties [11, 12, 265]. The crucial properties of COPs are geometric and electronic structures of neutral and ionized (doped) forms, excitation energies (band gaps), order of allowed and forbidden excited states, luminescence (wavelength and life-time), IPs, EAs, polarizabilities, mobility of charge carriers, and conductivities. Ideally, theory should be able to reproduce and predict properties with sufficient accuracy to guide experiments. In general, the quality of theoretical methods should not be judged by accidental agreement with experiment. The proper approach is to first converge the basis set and then the theoretical level. This alone is usually difficult for systems the size of COPs. If experimental data are obtained in condensed phases, solvent or interchain interactions have to be included in the theoretical treatment. Finally, effects like disorder, the presence of counter ions, and so forth have to be accounted for before agreement with experiment can be claimed. Because the systems under consideration are large and some of the required properties are hard to treat, a clever choice of methods based on experience with smaller systems has to be made to achieve meaningful results. Unfortunately, good theoretical practices are not always applied, and much too often "excellent agreement" with experiment is claimed even when gas phase calculations are compared to condensed phase data. Analyzing the accuracy of available methods is one major emphasis of this chapter.

Geometries and electronic structures are easy to obtain with a number of theoretical methods. The accuracy is generally good, even with inexpensive methods. Thus, geometries can be obtained for large oligomers or for polymers. In contrast, excitation energies require high levels of theory. For small molecules in the gas phase, excitation energies can be calculated



Scheme 4.

with rather good accuracy by calculating the wave function of the excited state explicitly. For accurate determination of the order of different types of excited states and of their energy differences, highly correlated *ab initio* calculations with large basis sets are necessary. Therefore, excitation energies of large oligomers or of polymers have to be obtained at theoretical levels that are not necessarily very accurate. IPs and EAs are calculated as energy differences between neutral species and the corresponding ions. This procedure requires two calculations. IPs are comparatively easy to obtain because cations can be described with high accuracy. EAs are notoriously difficult calculate because the extra electron causes large electron repulsions and diffuse electron clouds. Therefore, large basis sets and a good treatment of the electron correlation are required for the anions. Alternatively, excitation energies, IPs, and EAs can be determined approximately with a single calculation on the ground state if the frozen orbital approximation is applied. Within this approximation, IPs are the negative energies of the highest-occupied molecular orbital (HOMO), EAs are the negative energies of the lowest-unoccupied molecular orbital (LUMO) [274], and excitation energies are HOMO-LUMO energy differences. This approach neglects all relaxation effects of the electron density on changing the number of electrons or their distribution [275]. Thus the frozen orbital approximation introduces errors but is widely applied in band structure calculations for comparing properties of chemically different conducting polymers [256, 259].

Polarizabilities are needed to understand nonlinear optical properties of COPs. Polarizabilities are so-called first-order properties. They are obtained by calculating the wavefunction in the presence of an electric field. Applying an electric field (E) to a molecule leads to a change in dipole moment (μ), which can be expanded in the following way:

$$\mu = \mu_{\parallel} + \alpha E + \beta EE + \gamma EEE + \cdots$$

The last three components are nonlinear in the field strength and are the so-called nonlinear optical properties. The first is the polarizability, followed by the first and second hyperpolarizability [276]. COPs show very large linear and nonlinear optical responses because of their extended π-systems.

Conductivities and electron mobilities in bulk COPs are probably the most complicated properties to treat theoretically because they are influenced strongly by factors beyond the intrinsic properties of the molecules. Because even the longest polymer chains are small compared to macroscopic dimensions and because bulk polymers are disordered [277], interchain hopping is crucial for transport. Conductivity of bulk polymers was investigated with classical models [6, 278, 279]. Transport through mesoscopic devices composed of individual molecules and metal contacts that act as reservoirs for electrons is a complicated problem still. The current (I) through a molecule is a highly nonlinear function of the applied voltage (V). Calculation of $I(V)$ curves requires quantum statistical methods and many-body theory [280].

Experimental data on polymers are generally obtained for films, crystals, or solutions. It is difficult to predict bulk properties of COPs not only because of theoretical problems but also because these properties depend on a variety of factors that are hard to control experimentally. For instance, optical absorption and conductivity depend on conjugation length and interchain interactions. If the band gaps of two systems are supposed to be compared, one would ideally want to do this under identical conditions: that is, equal conjugation lengths and identical crystal structures. Polymers, however, are difficult to crystallize and consist of crystalline and amorphous regions [277]. The extent of these regions crucially influences the properties. Depending on the nature of the polymer chain and the substituents, crystal structures may differ. Theoretical calculations can be carried out for three-dimensional arrays of infinite polymer chains. The results, however, correspond to ideal systems that can never be realized experimentally.

Because well-characterized polymers are difficult to make, chemical as opposed to electrochemical syntheses for oligomers of defined length were developed [34, 39, 281–284]. Monodisperse oligomers have much sharper spectral bands and allow for better comparison between theory and experiment. Oligomers can also be crystallized much more easily than polymers. Single crystals were obtained, for instance, of sexithiophene, octithiophene, and tert-butyl capped dodecahexaene [283, 285]. Monodisperse polyenes and thiophene oligomers were investigated spectroscopically in liquid solution [94, 261, 281, 282, 286–295]

and in matrices at low temperature [293, 296–307]. Interestingly, optical properties of crystals of oligomers as short as sexithiophene do not differ much from those of bulk polythiophene [308], indicating that conjugation lengths in COPs might be far from infinite, justifying the oligomer approach [309].

In this chapter the following approach is used: section 2 deals with the accuracy of quantum chemical methods for excitation energies of polyenes for which gas phase data are available. The first part shows how accurately the highest levels of theory can reproduce experimental data under the best of circumstances. Then the results with various lower level approaches and with density-functional theory (DFT) are described and compared with the high-level *ab initio* calculations. Finally the frozen orbital approximation is introduced. In Section 3, extrapolation of oligomer data is investigated. In Section 4, absorption energies in solution are analyzed. In Section 5, the influence of interchain interactions is considered. Section 6 summarizes the evolution of polymer band gaps starting from excitation energies of single chains in the gas phase. In Section 7, the conducting state of polymers is investigated. Polarizabilities are treated in Section 8. In Section 9, theoretical treatment of conductivity through mesoscopic devices is reviewed. Section 10 contains a summary of the results.

2. ACCURACY OF THEORETICAL EXCITATION ENERGIES FOR POLYENES IN THE GAS PHASE

2.1. High Level *Ab Initio* Calculations

In this section, a detailed analysis of *ab initio* calculations of polyene excitation energies in the gas phase is presented because polyene excitation energies extrapolate to the band gap of an isolated PA strand in the infinite chain limit. All of the problems discovered for these comparatively simple cases are also present in the treatment of optical properties of COPs. For COPs a number of additional problems arise that make comparison between theory and experiment even more complicated. Therefore, knowledge about the accuracy of theoretical data regarding isolated well-defined molecules is extremely useful in the assessment of the additional effects arising in condensed phases and for polydisperse materials.

The first two low-lying singlet excited states of polyenes are the most important in connection with COPs. The 1^1B_u state is optically allowed and corresponds, in the infinite limit, to the band gap in PA. Transition from the ground state to 1^1B_u involves mainly the excitation of one electron from HOMO to LUMO. The other low-lying singlet state 2^1A_g has the same symmetry as the ground state and is therefore one-photon forbidden and of very low intensity. The 2^1A_g state cannot be described by a single-electron configuration but arises from a mixture of three electronic transitions: HOMO − 1 → LUMO, HOMO → LUMO + 1, and a two-electron excitation from HOMO to LUMO [310]. Experiment shows that the allowed 1^1B_u state lies lower in energy for butadiene and hexatriene [311–313]. The crossover occurs near hexatriene and octatetraene. For all longer polyenes, the one-photon forbidden 2^1A_g state lies below 1^1B_u [288, 293, 296, 303, 304, 306, 307, 310, 314–316]. Thus, 2^1A_g is S_1 and 1^1B_u is S_2 in polyenes longer than octatetraene.

The relative order of the two excited states is important for understanding optical properties of COPs, as fluorescence occurs from the lowest excited state [310]. Absorption of energy in longer polyenes yields a $S_0 \rightarrow S_2$ transition to the allowed 1^1B_u state. The excited state quickly relaxes to the lower-lying 2^1A_g state, and fluorescence occurs from 2^1A_g. Because the 2^1A_g state decreases faster in energy than 1^1B_u on chain length increase, the $S_1 \rightarrow S_2$ gap increases and the $S_1 \rightarrow S_0$ gap decreases. Thus, nonradiative relaxation from S_2 decreases and nonradiative decay from S_1 is enhanced [310]. Fluorescence decreases gradually as the chain length increases but disappears abruptly in going from octatetraene to hexatriene [303]. The transition to 2^1A_g is visible in the electroabsorption spectrum of carotene [317]. Because polyenes are important for biological processes such as vision and light harvesting and for electronic and optical properties of COPs, excited states of polyenes have been investigated intensively, experimentally [261, 283, 288, 293, 296, 301, 303–307, 310–326] and theoretically [101, 262, 263, 327–362].

Theoretical treatment of the 1^1B_u excitation might be expected to be relatively simple, as it is dominated by a one-electron HOMO-LUMO transition. However, even the highest-level

ab initio methods such as the complete active space self-consistent field method with correlation energy treatment using second-order Møller-Plesset perturbation theory (CASPT2) [346, 347, 351], multireference configuration interaction (MRCI) [332, 345, 352], and high-level single-determinant configuration interaction [328, 329, 337, 339–341, 343] have difficulties with obtaining this excitation energy with high accuracy. For small systems such as ethene and butadiene, the 1^1B_u excited states are strongly distorted out of planarity and the electron density is diffuse [363]. The diffuse character leads to a mixing of valence and Rydberg character that is hard to describe accurately [329, 331, 332, 335, 343, 364]. Analysis of the experimental data has shown that the valence Rydberg mixing is a physical fact and not an artifact of the calculations [325]. Extensive treatment of $\sigma - \pi$ correlation is therefore required [351]. A natural choice is to start with a complete active-space self-consistent field (CASSCF) wave-function, which can account for multiconfigurational character. The CASSCF wave-function, however, was shown to be plagued for small molecules by overestimated interactions between valence and Rydberg states, and a second B_u state was incorrectly found below 1^1B_u. Low-level perturbation theory, as in the CASPT2 method, has problems recovering the correct order of the excited states. All high-level *ab initio* calculations converge toward excitation energies between 6.2 and 6.3 eV for the 1^1B_u state of butadiene. This is 0.3–0.4 eV higher than the observed absorption at 5.92 eV [311, 312].

Fortunately, valence Rydberg mixing decreases for longer polyenes, and the agreement between theory and experiment for hexatriene and longer systems improves [346]. However, even for longer polyenes, high-level treatment of nondynamic and dynamic correlation effects was proven necessary. To avoid problems with intruder states, state-averaged wave-functions have to be used [347]. Recently, a restricted active-space SCF (RASSCF) method was developed that in combination with a specially designed basis set that includes optimized 3p-orbitals, led to more satisfactory results than CASSCF [361]. Because of the multiconfigurational nature of the 2^1A_g state, theoretical treatment might be expected to be even harder. In contrast, apart from the fact that single-electron theory cannot be applied, the treatment of the 2^1A_g state proved to be less difficult. Because of the problems with determination of the 1^1B_u excitation energy, calculating the energy difference between 1^1B_u and 2^1A_g even for short polyenes remains challenging. For long polyenes, high-level treatments are still out of the question.

For checking the accuracy of theoretical calculations, reliable vertical excitation energies are required. However, assignment of the excitation peaks is difficult. At room temperature, broad absorption bands with little resolution are observed. For this reason, low-temperature techniques were applied. Absorption spectroscopy of jet-cooled butadiene, hexatriene, and octatetraene [323] indicates that vibrational progressions for longer and shorter polyenes are similar. It was concluded, therefore, that no significant out-of-plane distortion, as in ethene and in butadiene, exists for longer polyenes. The very presence of a vibrational progression, however, indicates that geometries of ground and excited states differ. The spacing between the peaks is around 0.2 eV for hexatriene and octatetraene. This coincides with the vibrational energy for stretching C–C double bonds. According to the Franck-Condon principle the 0–0 peak should correspond to the adiabatic excitation. The vertical transition should occur at a higher energy and have the highest intensity. Experimentally, the 0–0 transition can be determined more easily. Theoretical calculations, however, correspond to vertical transitions. For hexatriene, the 1^1A_g-1^1B_u transition measured with electron impact spectroscopy shows vibronic bands at 4.95, 5.13, and 5.30 eV [313]. Both the 4.95 and the 5.13 eV peaks have high intensity. The peak positions are confirmed with ultraviolet (UV) spectroscopy [323], but the relative intensities of the peaks (4.93 and 5.13 eV in the UV spectrum) are inverted. Thus, λ_{max} lies at 5.13 eV with electron impact spectroscopy and at 4.93 eV with UV spectroscopy. This shows that the higher-energy peak in the UV spectrum may correspond to a vertical transition in these systems, even if it does not have the highest intensity. Theoretical vertical transition energies for hexatriene lie between 5.0 and 5.15 eV [340, 341, 346, 351]. Because the difference obtained with different methods is as large as the experimental difference between the two lowest peaks, theory can not resolve the problem with the peak assignment. Because it is so difficult to determine the precise nature of the peaks experimentally, it is hard to make decisions about which theoretical method is the

most adequate. The difference between vertical and the adiabatic excitation was calculated to be 0.46 eV [341] and 0.39 eV [361], which is twice as large as difference between the first two vibrational peaks in the spectrum.

Octatetraene is the first system in the polyene series that shows fluorescence. If fluorescence originates from the same excited state that is reached during excitation, fluorescence and excitation peaks coincide. The presence of a Stokes shift in condensed phase octatetraene spectra was one of the indications that the 2^1A_g state lies below 1^1B_u. In the gas phase, the Stokes shift is absent and the fluorescence life time is shorter than 2 ns [315]. The lowest-energy peak of the 1^1A_g-1^1B_u vibrational progression occurs at 4.41 eV [314, 323]. This peak has the highest intensity in the gas phase [288, 303, 305, 314, 315, 324] and corresponds, therefore, to λ_{max}. For this reason, the 0–0 excitation of octatetraene and longer polyenes is usually supposed to be a vertical transition. A second peak at 0.2 eV higher energy and with substantial intensity exists in the spectra for octatetraene (and for the longer polyenes as well). For octatetraene in matrix, the intensity of this second peak is slightly higher than that of the 0–0 transition [305]. This raises doubt about the assignment of the 0–0 peak to a vertical transition in the gas phase. Theoretical results find a 0.2-eV energy lowering when the geometry of the excited state is optimized [339]. Thus, it appears that theory and experiment indicate that the 0–0 transition does not correspond to a vertical transition and that the correct vertical excitation energy is about 4.61 eV. *Ab initio* values for adiabatic and vertical excitation energies were calculated to be 4.56 and 4.76 eV, respectively, with configuration interaction (CI). With CASPT2, values of 4.42 [347, 350] and 4.66 eV [351] were obtained for the vertical transition. Thus, the difference in the theoretical results amounts to 0.34 eV. Again, the difference in the theoretical values is too large to reach a decision about the nature of the λ_{max} peak. Because the excited state of octatetraene is planar, and as valence-Rydberg mixing is small, the accuracy of the best of the available theoretical methods is somewhat disappointing.

The longest polyene for which gas phase data [288] and high-level *ab initio* (MRMP) results [351] are available is decapentaene. The 0–0 transition occurs at 4.02 eV with direct measurement and at 3.98 eV when solution data are corrected for solvent effects [288]. The second peak in the spectrum is shifted from the 0–0 transition by 0.19 eV. For the MRMP treatment, the size of the basis set and the active space had to be reduced compared to those for the shorter homologuoes. The final value for the excitation energy was obtained with an extrapolation scheme that was tested on the shorter systems. The corrected theoretical value for the vertical excitation 1^1B_u excitation energy is 4.05 eV. The 2^1A_g state is placed 0.4 eV below 1^1B_u. Geometry optimization of the excited states lowers the excitation energies and increases the splitting between the 2^1A_g and 1^1B_u states to 0.9 eV because the former is more sensitive to geometric relaxation.

The gas phase excitation energies and the most reliable theoretical values for ethene through decapentaene are summarized in Table 1 and plotted against experimental values in Fig. 1. With any given method, the errors differ depending on chain length. The general tendency is for the energy of the 1^1B_u state to be slightly overestimated for small systems. For longer polyenes, theoretical results are very close to or below the experimental values. The general trend is thus that theoretical values either approach or fall below experimental values in the long-chain-length limit. Unfortunately, calculations on larger polyenes are not available that would allow for checking this issue. Care must be taken when values are extrapolated, as the trend to underestimate excitation energies with increasing chain length will lead to a significant underestimation of the gap in the infinite limit. Because the best *ab initio* calculations produce differences of 0.2–0.35 eV among each other, usually bracketing the experimental results, and as the assignment of the spectral bands is not entirely certain, perfect agreement between theory and experiment has not been achieved even for these relatively small, isolated molecules.

2.2. Single-Electron Methods

As the size of the systems increases, lower levels of theory have to be employed. Fortunately, larger systems are somewhat easier to treat theoretically, as the electronic relaxation extends over a larger volume, making choice of basis set and treatment of electron correlation less

Table 1. Experimental and Theoretical excitation energies for ethene through decapentaene in eV.

	C_2H_4	C_4H_6	C_6H_8	C_8H_{10}	$C_{10}H_{12}$
exp.	7.65 [363] est. 8.0	5.92 [311] 5.90 [312]	5.13 [313]	4.41(0–0) [314] 4.38(0–0) [288] 4.61(vert. est.)	4.02 (0–0) [288] 4.21(vert. est.)
CI	7.96 [329] 8.38 [331] 8.06 [335] 8.17 [343]	6.39 [343]	5.15 [340]	4.76 [339]	
MRCI	7.95 [332]	6.2 [337] 6.48 [345]			
CASCI	8.01 [352]				
CASPT2	8.40 [346]	6.23 [346]	5.01 [346]	4.42 [347]	
MRMP		6.21 [351]	5.10 [351]	4.66 [351]	4.05 [351]
CIS [357]		6.22	5.40	4.78	4.33
TDHF	7.56	6.08	5.15	4.53	4.10
TDLSDA		5.62	4.51	3.81	3.23
TDB3LYP		5.70	4.65	4.00	3.51
TDB3LYP [357]		5.59	4.64	3.98	3.52
TDB3P86		5.75	4.68	4.00	3.52
TDB386-30%		5.81	4.77	4.09	3.62

Where no reference is given, the data are obtained for this work, employing the CEP-31G* pseudopotential and basis set throughout.

critical. Time-dependent Hartree-Fock (TDHF) theory, which is also known as the random phase approximation (RPA), is an inexpensive method for calculation of excitation energies. Configuration interaction including single excitations (CIS) [365] is another method that can be used for large systems. TDHF and CIS are not suitable for calculating the energy difference between the 2^1A_g and 1^1B_u states because the 2^1A_g state requires a linear combination of several electronic configurations involving double excitations. However, both methods are well-suited for describing the 1^1B_u state. CIS can actually be considered to be TDHF in the Tamm-Dancoff approximation [366]. Thus, TDHF is more complete, and therefore the 1^1B_u and a number of other excited states are obtained with very good accuracy [348, 367]. In this chapter, TDHF results for butadiene through decapentaene are compared with *ab initio* values and experiments. The geometries of the ground states were optimized with DFT employing Becke's three-parameter hybrid functional (B3) [368], modified to include 30% of Hartree-Fock exchange and Perdew's 1986 correlation functional (P86) [369] (B3P86-30%). DFT will be discussed in more detail in the next paragraph. As basis sets, Stevens, Basch, and

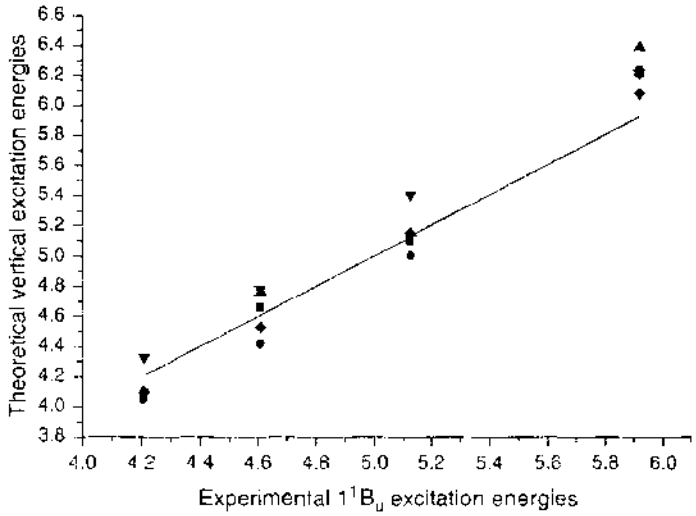


Figure 1. Theoretical excitation energies for polyenes (C_4H_6-$C_{10}H_{12}$) at various levels of theory plotted versus experimental excitation energies measured in the gas phase. All values are given in eV. Squares: MRMP, Ref. [351], circles: CASPT2, Ref. [346], up triangles: CI, Refs. [339], [340], and [343], down triangles: CIS, Ref. [357], diamond: TDHF/CEP-31G*, this work.

Krauss pseudopotentials and split valence basis sets including polarization functions [370] (CEP-31G* keyword in G98 [371]) were used. TDHF/CEP-31g*//B3P86-30%/CEP-31G* data and the CIS results obtained by Hsu et al. [357] are included in Table 1 and in Fig. 1. Both the CIS and TDHF/CEP-31G*//B3P86-30%/CEP-31G* results reproduce experimental excitation energies to the 1^1B_u state with very good accuracy. TDHF values are closer to experiment than the CIS ones. This is in agreement with earlier investigations [348, 367, 372]. In fact, the TDHF/CEP-31G*//B3P86-30%/CEP-31G* results are closer to experiment than the high-level *ab initio* data presented in the last paragraph, with the deviations between theory and experiment being 0.16 eV or less. Moreover, the errors are fairly constant on increasing the chain length. This gives confidence that longer polyenes can be treated reliably. This is an important conclusion from a practical point view, as TDHF is an affordable method that can be used to deal with much longer chains than high-level *ab initio* methods.

To check for possible error cancellations because of the choice of basis set and geometry, TDHF calculations for $C_{12}H_{14}$ with different geometries and basis sets (6-31G*, 6-31+G*, 6-311+G*, 6-311++G**, CC-PVDZ, and CC-PVTZ as provided in G98 [371]) were performed. Vertical excitation energies based on HF ground-state geometries decrease only slightly on increasing the basis set. The values are 4.35, 4.26, 4.25, 4.24, 4.30, and 4.24 eV with the above basis sets. The difference between the smallest (6-31G*) and the largest (CC-PVTZ) basis sets is only 0.11 eV. Including a diffuse function with the 6-31+G* basis set recovers 0.09 eV compared to 6-31G*. Thus, almost no further improvement is achieved in going beyond 6-31+G*. Whereas the choice of basis set is obviously not crucial here, the choice of ground-state geometry is. Using the Hartree-Fock (HF) instead of the B3P86-30% geometry increases the excitation energy of $C_{12}H_{14}$ by 0.52 eV. This can be attributed to the fact that HF overestimates bond-length alternation (BLA). Because the excited states of polyenes have less BLA than the ground states, the B3P86-30% ground-state geometry is closer to the excited-state geometry and the vertical excitation energy is smaller. To make a decision which geometry is more appropriate, calculations at the MP2 level of theory were done. MP2 stands for Møller-Plesset perturbation theory truncated at second order [365]. The MP2/6-31G* level is generally accepted to produce very accurate geometries and is therefore one of the most widely used methods for geometry optimizations. The BLA of $C_{12}H_{14}$ at the MP2/6-31G* level is 0.075 Å, almost identical to that at B3P86-30%/CEP-31G*, 0.074 Å. At the HF/6-31G* level, BLA is too large, 0.123 Å. Thus, the use of B3P86-30% geometries is to be preferred. Finally, the TDHF/6-31+G*//MP2/6-31G* excitation energy of $C_{12}H_{14}$ was computed and compared to that at TDHF/CEP-31G*//B3P86-30%/CEP-31G*. The two values, 3.78 and 3.80 eV, respectively, are almost identical. This shows that the choice of basis set and geometry at the TDHF/CEP-31G*//B3P86-30%/CEP-31G* level are adequate to obtain reliable excitation energies.

2.3. Calculation of Excitation Energies Employing Density-Functional Theory

As an alternative to TDHF, time-dependent density-functional theory (TDDFT) has been developed and tested on a number of molecules [357, 366, 373–385]. Like DFT for ground states, TDDFT were exact for excited states if the exact exchange-correlation functional were known [376]. This includes excited states that have significant two-electron character such as the 2^1A_g state of polyenes. Two significant approximations have to be made for practical calculations. One is the use of approximate exchange-correlation functionals, the other one is the so-called adiabatic approximation that arises when the time-dependent exchange correlation action functional is approximated by a time-independent exchange correlation functional [373]. The adiabatic approximation does not seem to cause severe errors. The use of approximate exchange-correlation functionals is problematic. All of the commonly used functionals lead to an incorrect description of the exchange-correlation potential in the asymptotic region [376]. This results in the underestimation of excitation energies of Rydberg states. In contrast, two-electron excited states are reproduced very well [376]. Overall, TDDFT was shown to be a substantial improvement over TDHF if the whole excitation spectrum is considered [385]. Unfortunately, for theoretical research on conducting polymers, the most problematic excited states for TDDFT are the 1^1B_u states of conjugated

π-systems whose excitation energies are underestimated by several tenths of an eV [357, 385]. Analysis shows that this is a result of the ionic nature (in the valence bond picture) of the 1^1B_u states [384], which cannot be described well with the present approximate functionals. Excitation energies to the covalent 2^1A_g states are reproduced quite accurately despite their multiconfigurational nature. Because the 2^1A_g excitation energies are accurate but the 1^1B_u excitation energies are too low, TDDFT has problems with determining the order of the two most important excited states of polyenes and COPs [385].

Comparison of TDDFT calculations of 1^1B_u excitation energies with a variety of popular functionals are included in Table 1 and plotted versus experiment in Fig. 2. Considered are the local spin density approximation (LSDA) as implemented in Gaussian 98 employing Slater exchange [386–388] and the Vosko-Wilk-Nusair (VWN) [389] correlation functional, Becke's gradient corrected hybrid functional B3 [390] in combination with the Lee-Yang-Parr's (LYP) [391] and Perdew's 86 (P86) [369] correlation functionals. TDLSDA excitation energies are consistently too low. TDB3LYP values of Hsu et al. [357] and the present ones with the CEP-31G* pseudopotential and without diffuse functions are virtually identical. Only for the smallest molecule, butadiene, there is a small difference. The values with hybrid functionals are all practically identical and closer to experiment than with pure DFT, but the excitation energies are still underestimated by about 0.5 eV. Increasing the contribution of HF exchange from 20% to 30% increases the excitation energies by 0.1 eV. Thus, the 1^1B_u excitation energies improve with increasing amount of exact HF exchange. However, at the same time the 2^1A_g excitation energies increase even stronger [384]. Thus, there is no way to adjust hybrid functionals to reproduce the relative energies of both excited states [384]. To aggravate the situation, the underestimation of the 1^1B_u excitation energies increases with increasing chain length. The same holds for the data of Hsu et al. [357]. It was shown that DFT has problems with extended systems [380] and that the excitation energies collapse in the infinite chain length limit to the band gap values [366], which suffer from the band gap problem, as discussed below. Despite its overall success with the calculation of excited states, TDDFT is therefore problematic for the specific application of calculating band gaps of long polyenes and COPs. It will be shown later that DFT also has problems with describing polaron sizes and polarizabilities.

2.4. Frozen Orbital Approximation

2.4.1. Hartree-Fock Orbitals

For infinite systems, excitation energies are usually obtained with band structure calculations [256, 259]. Band gaps of infinite systems obtained with band structure calculations

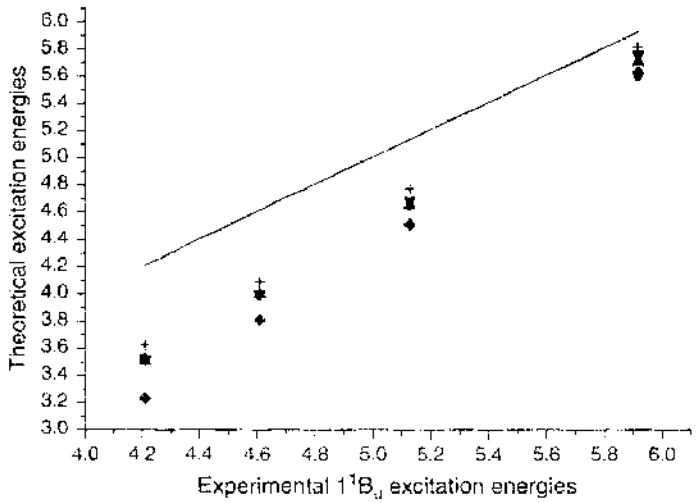


Figure 2. TDDFT excitation energies for polyenes (C_4H_6-$C_{14}H_{16}$) with different exchange-correlation functionals plotted versus experimental excitation energies measured in the gas phase. All values are given in eV. Circles: B3LYP, Ref. [357], up triangles: B3LYP/CEP-31G*, down triangles: B3P86/CEP-31G*, diamonds: LSDA/CEP-31G*, plus signs: BP386-30%/CEP-31G*.

correspond with HOMO-LUMO gaps of finite systems. Thus, band structure calculations use orbital energy differences to estimate excitation energies. Using HOMO-LUMO energy differences as excitation energies is justified at the HF level of theory by Koopmans's theorem [274]. Koopmans's theorem proves that the negative HOMO energy of the neutral molecule corresponds to the difference between the total energies of the neutral molecule and the ion, and therefore the IP, within the frozen orbital approximation. Frozen orbital approximation means that the MOs are optimized for the neutral molecule and that the same MOs are used for the ion, deleting only the orbital from which the electron is removed from the Slater determinant. The energy difference between these two Slater determinants is equal to the negative energy of the removed orbital. Likewise, the LUMO energy corresponds to the EA [365] within the frozen orbital approximation. Using frozen orbitals is an approximation because the MOs of the ions were different, if they were optimized, lowering the total energies of the ions. If electronic excitation is thought to result form removing one electron from the HOMO and adding it to the LUMO, the excitation energy can be approximated by IP minus EA or by the HOMO-LUMO gap within the frozen orbital approximation. Although this procedure is mathematically exact at the HF level, the analogous approach, using DFT orbital energy differences as excitation energies, is highly controversial [392–403]. It is generally agreed on that within DFT, the negative energy of the highest-occupied Kohn Sham orbital corresponds to the IP, provided that the exact exchange-correlation functional is used [393]. Only recently an analog of Koopmans' theorem was proved for DFT (see Ref. [403] and references therein).

The errors in the frozen orbital approximation are well established at the HF level. The results for IPs are usually quite good because of error cancellation. Because HF theory neglects the correlation between electrons of opposite spin, the error is larger for the molecule than for the ion, as a neutral closed-shell molecule has one electron pair more than the corresponding ion. Thus, the energy of the molecule is too high compared to that of the ion because of the correlation error. Because the wave function of the cation is not optimized within the frozen orbital approximation, the energy of the cation is too high compared to that of the optimized ion. Because the correlation error raises the energy of the molecule and the frozen orbital approximation raises the energy of the cation, the energy difference between the two is about correct. For EAs, the correlation error raises the energy of the anion more than that of the neutral molecule, and the frozen orbital approximation raises the energy of the anion additionally. Thus, both errors add up and negative LUMO energies yield very poor EAs. As a result, HOMO-LUMO gaps and band gaps are too large at the HF level [259, 404–407]. Band structure calculations are still useful for obtaining relative band gaps of different systems, as the source of error is well known and the size of the error is quite systematic. Inclusion of correlation at the MP2 [404, 408, 409], MP4 [409], and MRCI [410] levels of theory decreases band gaps, but experimental values are not reached. HOMO-LUMO gaps for polyenes at the HF/CEP-31G* level of theory are given in Table 2 and plotted versus experimental results in Fig. 3, together with DFT results, which will be discussed below. The huge overestimation (5–6 eV) of the excitation energies is visible. Closer inspection shows that the error decreases on chain length increase.

2.4.2. DFT Orbitals

DFT band structure calculations have been applied for a long time, and the errors are established empirically. Within LSDA, HOMOs are very poor approximations for IPs [399].

Table 2. HOMO-LUMO gaps at HF and various DFT levels.

	C_2H_2	C_4H_6	C_6H_8	C_8H_{10}	$C_{10}H_{12}$
HF	14.01	11.65	10.30	9.44	8.87
LSDA	5.44	3.73	2.82	2.28	1.90
B3LYP	7.19	5.31	4.26	3.61	3.17
B3P86	7.28	5.35	4.28	3.62	3.17
B3P86-30% E_{HF} [419]	8.19	6.15	5.03	4.33	3.84

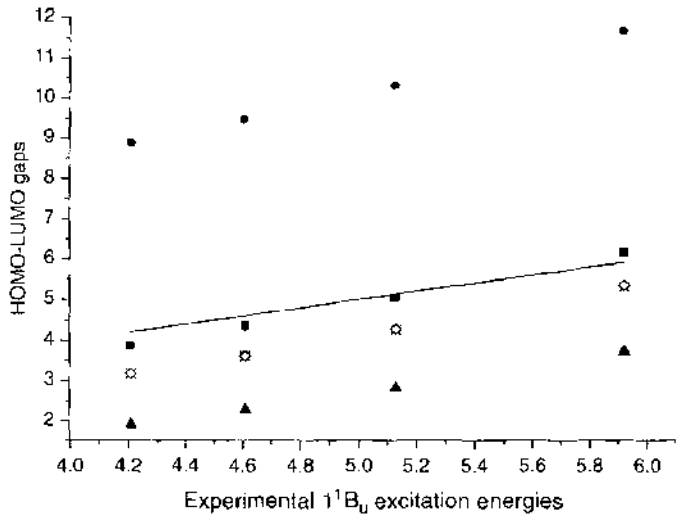


Figure 3. HOMO-LUMO gaps plotted versus experimental excitation energies measured in the gas phase. All values are given in eV. Circles: HF/CEP-31G*, squares: B3P86-30%/CEP-31G*, open diamonds: B3P86/CEP-31G*, crosses: B3LYP/CEP-31G*, up triangles: LSDA/CEP-31G*.

and band gaps are underestimated by about 50% (compare Table 2). This is known as the band gap problem of DFT [392–399, 402, 411–414]. The origin of the band gap problem was heavily debated over the years. The crucial question is whether the band gap problem is caused by the use of approximate exchange-correlation functionals or is intrinsic to DFT. Several authors came to the conclusion that the band gap problem is intrinsic to DFT [394, 397]. An extreme point of view is that DFT orbital energies (with the exception of that of the HOMO) have no physical meaning whatsoever [415]. Nonetheless, evidence is accumulating that the band gap problem can be overcome by improving the functionals and that DFT orbital energies can be used as a good first approximation to excitation energies [398, 403, 416, 417]. A lot of effort has been devoted to overcome the problem with the incomplete cancellation of self-interaction [393] and with the wrong asymptotic behavior of the approximate exchange functionals [401, 412, 418]. Figure 3 shows that DFT HOMO-LUMO gaps are much closer to experimental excitation energies than are HF orbital energy differences.

Mixing some HF exchange into the exchange correlation functional increases the DFT HOMO-LUMO gaps and brings DFT orbital energies with hybrid functionals in close agreement with λ_{max} from UV spectroscopy [399, 419–421]. We tested several functionals and adjusted the amount of HF exchange empirically so that HOMO-LUMO gaps of polyenes are reproduced λ_{max} [399, 419, 420]. HOMO-LUMO gaps for butadiene through decapentaene with LSDA, B3LYP, B3P86, and B3P86-30% functionals are given in Table 2. Figure 3 reveals that DFT HOMO-LUMO gaps obtained at the B3P86-30% level reproduce excitation energies of polyenes with fairly good accuracy. In general, B3P86-30% HOMO-LUMO gaps overestimate the excitation energies for small polyenes and fall off too fast with increasing chain length. This is the same problem that was observed with TDDFT. However, Fig. 3 shows that DFT-hybrid HOMO-LUMO gaps are a huge improvement compared to HOMO-LUMO gaps at HF and LSDA levels. Thus, HOMO-LUMO gaps with hybrid functionals range in quality behind high-level *ab initio* calculations and TDHF but allow for a first estimate of excitation energies.

2.4.3. Semiempirical Methods

Semiempirical methods have a long history in band structure calculations, mainly for infinite one-dimensional systems. Because parameters are used, very good agreement with experiment can be achieved. The focus in this article is on *ab initio* methods and DFT, and therefore semiempirical methods are discussed only briefly. Hückel theory [422–429] in its basic form does not predict a band gap for PA [426]. Extended Hückel [422, 424, 428] or perturbationally corrected [426] versions are very successful in predicting band gaps of

conjugated π-systems. Because the Hückel method is a single-electron theory, the doubly excited, multiconfigurational 2^1A_g state is beyond the scope of Hückel theory [430]. A second widely applied method is the effective valence Hamiltonian (VEH) technique [431–433]. This method is kind of a curiosity. It was parametrized to reproduce HF results without employing experimental parameters. In contrast to HF, however, VEH fails to produce good IPs and succeeds with band gaps. The reason for this was never discovered. Nonetheless, the VEH method was applied successfully to analyze and predict band gaps of COPs [258, 431–434]. The MNDO (modified neglect of differential overlap) method has been developed for cluster-type solid-state calculations [435–437]. MNDO is an approximation to the HF method and suffers from the same overestimation of band gaps as HF. As with HF, the error is systematic so that trends can be established. Another popular semiempirical method is the Pariser-Parr-Pople approximation (PPP) [438–441].

3. EXTRAPOLATION TO INFINITELY LONG CHAINS

The underlying idea of a polymer is that its properties are determined largely by the properties of infinitely long chains, or at least by molecules that are long enough to be close to the infinite limit. Theoretically, either solid-state calculations with periodic boundary conditions or molecular calculations on oligomers with increasing chain length can be performed. On extrapolation, results from molecular calculations should be identical to those from solid-state calculations. With both methods, properties of perfect, defect-free, infinitely long isolated chains are obtained. One interesting question is how long an oligomer has to be to approach the infinite limit. The band model implies that band gaps decrease with each additional repeat unit. The changes per additional repeat unit get smaller with increasing chain length because the effect on a short oligomer is stronger than on a long chain. It is usually assumed that a plot of energy gaps versus inverse chain length is linear over the whole range of chain lengths [263, 283, 426, 427, 442]. Band gaps are an approximation for optical absorption energies. A linear dependence of absorption energies on inverse chain length requires the π-systems of ground and excited states to be completely delocalized. There is evidence, however, that polymers do not obey the prediction of band theory and that ionization as well as excitation produces localized defects [230]. Therefore, analogies have been drawn from comparison with molecular crystals. The issue is, however, whether ideal polymer crystals would behave similarly to molecular crystals or whether the band model would be correct at least in the ideal case. The difficulty in answering this question arises from the fact that polymers do not crystallize well. Therefore, intrinsic properties of ideal polymers cannot be determined experimentally. If the investigated samples are disordered and have short conjugation lengths, their properties must resemble those of molecular crystals. The size of the so-called exciton binding energy is therefore subject to a long-standing controversy [230, 231, 443–449]. If excitation is localized, a saturation of excitation energies with increasing chain length should occur when the chain length exceeds the size of the exciton.

Convergence limits were investigated experimentally for polyenes and thiophenes in solutions and in matrix. For a polydisperse solution of a polymer with 880 double bonds, absorption starts at 1.86 eV, and λ_{max} occurs at around 2.2 eV at 80 K and at 2.44 eV at 300 K [450]. The absorption profile was explained with a distribution of conjugation lengths and a statistical preference for shorter chains. This means that the onset of absorption at 1.86 eV corresponds to λ_{max} of the longest chains. Polyene spectra for medium-sized oligomers recorded for 77 K glasses [451] give a limiting λ_{max} value of 1.77 eV on linear extrapolation. The longest soluble oligomer with more than 100 conjugated double bonds has a λ_{max} value of 1.96 eV. Because the low-temperature spectrum was well resolved, polydispersity was excluded as the reason for the higher λ_{max} compared to the extrapolated value. Instead, the authors suggested that the discrepancy reflects at least in part the inadequacy of the 1/N fits in the long polyene limit. The theoretical results presented below confirm this point of view.

Saturation has also been observed experimentally in solutions of monodisperse samples of alkyl substituted thiophenes with up to 48 rings [94, 281, 452] Although λ_{max} values for monomer through hexamer (see Table 3) extrapolate to a limiting value of 2.27 eV, the excitation energy of long oligomers does not decrease below a value of 2.69 eV. Saturation is

Table 3. Comparison between experimental and theoretical excitation energies for thiophene oligomers.

	2-T	3-T	4-T	5-T	6-T
exp. [299] matrix	3.67	3.08	2.75		
exp. [287, 292] solution	4.05/4.13	3.49/3.50	3.16/3.18	2.99/2.98	2.85/2.86
CASPT2 [465]	3.88	2.86			
TDHF	4.25	3.60	3.26	3.05	2.92
CIS	4.50	3.83	3.47	3.25	3.12
TDLSDA	3.76	2.99	2.55	2.25	2.05
TDB3P86	3.92	3.20	2.80	2.54	2.37
B3P86-30% E_{H-L}	4.81	4.01	3.58	3.32	3.15

All calculations have been carried out with the CEP-31G* basis set and are based on B3P86-30%/CEP-31G* geometries.

reached for a chain length of 20 rings. Because the data were obtained in solution, disorder was thought to be the reason for the saturation. For comparison, TDHF excitation energies of 20-thiophene and 28-thiophene differ by only 0.02 eV, and HOMO-LUMO gaps of 20-thiophene and 40-thiophene differ by only 0.05 eV in the absence of any disorder. Therefore, the experimental data might reflect the intrinsic convergence limit. With the aim of increasing planarity, thiophene vinylene oligomers with 16 thiophene rings and 15 ethylene groups (corresponding to 47 conjugated double bonds) were synthesized [282]. Absorption energies of thienylvinylene oligomers are lower than those of thiophene oligomers. Nonetheless, a deviation from linearity was observed on chain length increase, and the saturation limit was predicted to occur around a 20-mer. For thiophene ethynylene with triple bonds between the thiophene rings instead of double bonds, saturation was observed after the octamer stage (23 conjugated double bonds) [34, 453].

The assumption of a linear dependence of excitation energies on inverse chain length stems from early investigations based on a modified particle in a box model and on semiempirical methods [263]. The approach is applied widely to obtain polymer properties from small numbers (often only four or five points) of oligomer data. A second approach accounts for the experimentally observed saturation. Here the band gap is determined as [454]

$$E(n) = E_\infty + (E_1 - E_\infty)\exp[-a(n-1)]$$

where E_1 is the excitation energy of the monomer, E_∞ that of the polymer, and $E(n)$ corresponds to an oligomer with n repeat units. The parameter a determines how fast $E(n)$ saturates. This equation predicts a deviation of small oligomers from linearity and includes a linear regime for medium-sized oligomers. (The deviation for small molecules from linearity is well known from theory. Usually, the shortest oligomers are left out when the data are plotted.) Theoretical analysis indicates [454] that observed excitation energies depend on the intrinsic size of the exciton and on the size of the oligomer segment. Therefore, excitation energies of short oligomers deviate from linearity because the oligomer is shorter than the exciton. A linear regime is reached for medium-sized oligomers because quadratic intrinsic behavior shows up only at longer chain lengths and because of linear decrease resulting from disorder. The longest oligomers give rise to quadratic saturation. Although disorder enhances the saturation, saturation is also present intrinsically because of the finite exciton size.

To find out whether there is a linear dependence of excitation energies on inverse chain lengths for ideal, planar oligomers in the absence of any disorder, oligomers of acetylene, thiophene, and pyrrole up to $C_{160}H_{162}$, 40-thiophene, and 40-pyrrole were optimized at the B3P86-30%/CEP-31G* level of theory. HOMO-LUMO gaps of these systems and TDHF/CEP-31G* excitation energies for oligomers up to $C_{96}H_{98}$, 24-thiophene, and 24-pyrrole are plotted versus inverse chain lengths in Figs. 4 and 5. As discussed in section 2.2, on the basis of results for shorter oligomers, TDHF is expected to give very accurate results for polyenes. At both theoretical levels, and for all three systems, the dependence of excitation energies on inverse chain length is S-shaped, as is observed experimentally. For HOMO-LUMO gaps this is surprising because saturation was attributed to finite exciton size [454], whereas energy

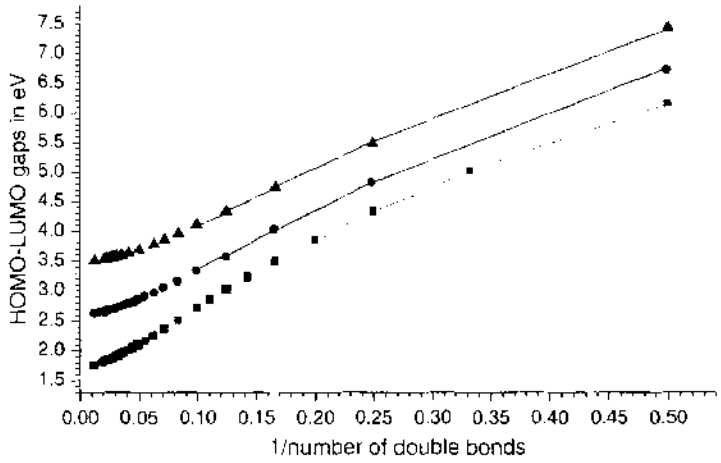


Figure 4. B3P86-30%/CEP-31G* HOMO-LUMO gaps versus inverse chain lengths. Squares: acetylene oligomers, butadiene-$C_{60}H_{62}$, circles: thiophene oligomers, 1-T-40-T, triangles: pyrrole oligomers, 1-Py-40-Py.

gaps obtained for ground-state systems cannot account for excitonic effects in the excited state. As observed before, HOMO-LUMO gaps are too high for small oligomers, and fall off too fast. For acetylene oligomers, this leads to an underestimation of the excitation energy compared to TDHF in the long-chain limit. For thiophene and pyrrole oligomers, HOMO-LUMO gaps overestimate excitation energies of small oligomers more strongly and converge toward TDHF excitation energies for long oligomers. TDHF values converge to 2.2 eV for PA, 2.5 eV for PT, and 3.3 eV for PPy. These values were obtained by extrapolating the last five points with second-degree polynomial fits and comparing the results with those for the longest oligomers. Because extrapolated band gaps increase slightly with every additional data point for a longer oligomer, the extrapolated values are assumed to be lower limits. Convergence is assumed when there is no difference (correct to two significant Figures) between the extrapolated band gap and the calculated value of the longest chain. HOMO-LUMO gaps converge to 1.7, 2.6, and 3.5 eV for PA, PT, and PPy, respectively.

An important difference between HOMO-LUMO gaps and TDHF excitation energies is that TDHF predicts smaller excitation energies for short thiophene than for small acetylene oligomers. This is in agreement with matrix isolation studies [297–299]. Because the smaller slope of the thiophene excitation energies, a crossover occurs near 4-T and octatetraene. The correct description of this crossover confirms the reliability of TDHF excitation energies. The smaller slopes of the thiophene and pyrrole data are associated with an earlier onset of saturation, as predicted by the equation above. This is visible from both HOMO-LUMO gaps and TDHF excitation energies. The different slopes and earlier saturation onset

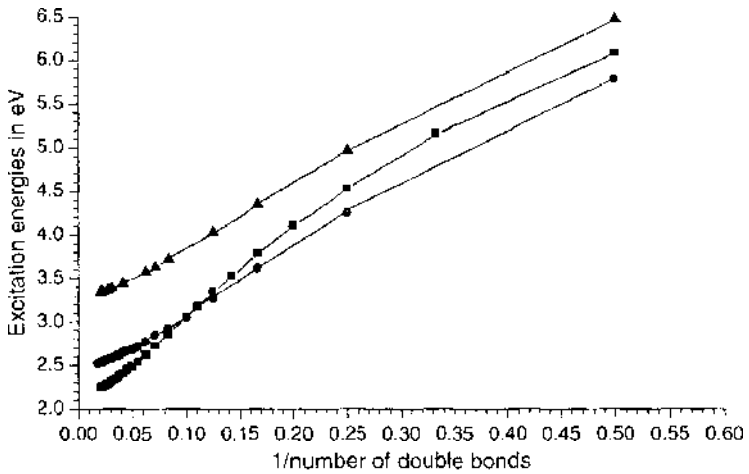
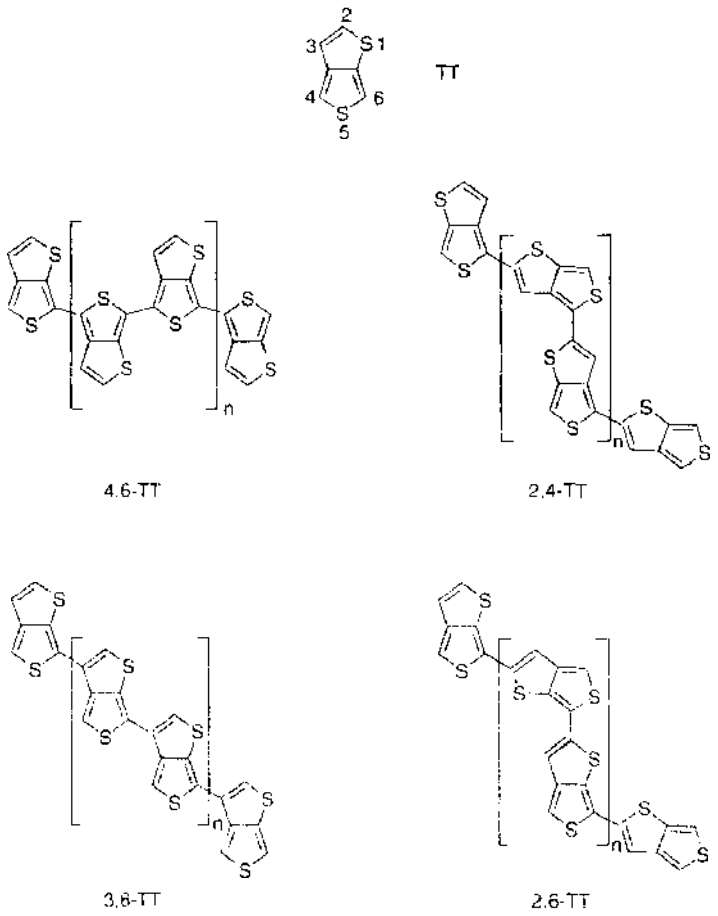


Figure 5. TDHF excitation energies versus inverse chain lengths. Squares: acetylene oligomers, butadiene-$C_{60}H_{62}$, circles: thiophene oligomers, 1-T-24-T, triangles: pyrrole oligomers, 1-Py-24-Py.

for thiophene and pyrrole oligomers indicate that conjugation is less efficient for aromatic polymers than for PA. Therefore, the lower conductivities of polythiophene and polypyrrole might be caused by intrinsic differences in the electronic structure and not only by disorder or mislinking.

As another example for nonlinearity of absorption energies, HOMO-LUMO gaps (at the B3P86-30%/CEP-31g* level) of four kinds of oligomers up to hexamers of thieno[3,4-*b*]-thiophene (TT) that differ only in their connectivity (Scheme 5) are plotted versus inverse chain length in Fig. 6. The HOMO-LUMO gaps of the 2,4-isomer fall off almost linearly with 1/number of double bonds. In contrast, the energy gap decreases faster with chain length for the 4,6-isomers, and those of 2,6- and 3,6-isomers decrease slower. Thus, the chain length dependence and the chain length at saturation depend on the connectivity and reflect the strength of conjugation. For the weakly conjugated 3,6- and 2,6-TT, the band gaps saturate quickly, whereas for the 4,6 conformer, no saturation is seen at the hexamer stage. This is in line with experimental results that show that poly(4,6-TT) has a much smaller band gap than PT and poly(2,4-TT), 0.8 eV [455].

Apart from simulating polymer calculations, it would be desirable to use oligomer calculations and extrapolation in connection with experimental data to determine conjugation lengths in COP samples. The longer the chains are, the more disordered they are assumed to be. Therefore, chain lengths (or molecular weights) do not coincide with conjugation lengths. Calculations that reproduce excitation energies accurately could therefore help to establish conjugation lengths. TDHF calculations show that such an approach is probably futile. Because excitation energies fall off very fast for small systems (up to about 20 double bonds) but changes are very small beyond medium-sized oligomers, the differences in excitation energies in the interesting range are too insensitive to chain length to allow for a precise determination of conjugation lengths. For example, chains with between 14 and 22 thiophene rings have the same λ_{max} value, 2.7 eV. The difference between λ_{max} of oligomers



Scheme 5.

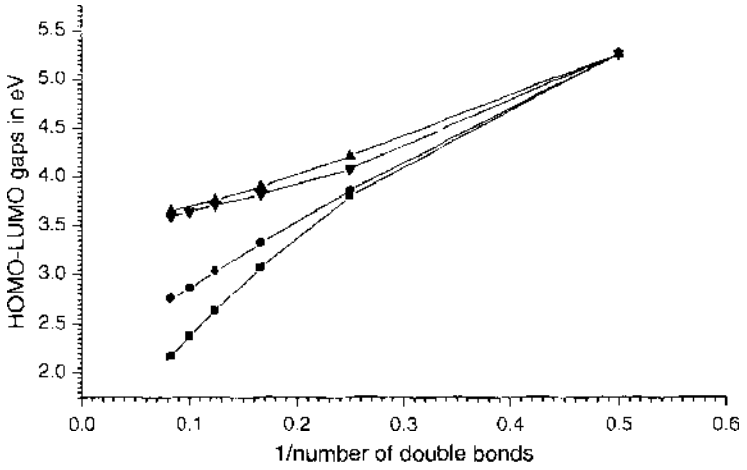


Figure 6. B3P86-3067/CEP-31G* HOMO-LUMO gaps versus inverse chain lengths for thieno[3,4-*b*]-thiophene (TT) monomer through hexamer. Squares: 4,6-TT, circles: 2,4-TT, down triangles: 2,6-TT, up triangles: 3,6-TT.

with 14 rings and the saturation limit is only 0.2 eV. Thus, a huge range of chain lengths leads to very similar λ_{max} values and estimates of conjugation lengths can be ballpark estimates at best.

This paragraph shows that neither experimental nor theoretical excitation energies depend linearly on inverse chain length. Only for a small subset of medium-sized oligomers is there a region in which the dependence is close to linearity. Therefore, linear extrapolation is wrong in any case. The size of the error depends on how many data points are used and where the data points lie on the S-shaped curve. Linear extrapolation underestimates the absorption energies in the infinite limit because saturation is not accounted for. It seems that investigation of the deviations of the dependence from linearity for oligomers can give useful information about strength of conjugation. If polymer properties are needed, calculations on long oligomers are to be preferred over extrapolation. This is feasible nowadays for even for the longest chains presented here, employing single-processor PCs. Because HOMO-LUMO gaps show the same nonlinear dependence on inverse chain length as excitation energies at the TDHF level, exciton formation cannot be the reason for the intrinsic deviation from linearity.

4. EXCITATION ENERGIES OF OLIGOMERS IN LIQUID AND SOLID SOLUTIONS

Theoretical calculations on isolated molecules correspond to gas phase data. Equilibrium structures are valid for the low-temperature limit of 0 K. Before a comparison with experiment can be made, temperature and media effects depending on the type of experiment have to be estimated and removed. Alternatively, temperature and solvent effects may be considered explicitly in the calculations. Differences between gas and condensed phases, however, are not accounted for in the majority of calculations. Therefore, theoretical gas phase calculations should not reproduce experimental values exactly. Any exact agreement between theory and experiment without accounting for media effects can only be coincidental. Additional effects resulting from disorder depend on the system and may cause additional deviations or partial error cancellation.

Ionization and electronic excitation lead to rearrangement of the electron density. Media effects help to distribute the distortion over a larger volume. Thus, in condensed phase, IPs are expected to be lower, EAs to be higher, and excitation energies to be reduced compared with those in the gas phase. In one study, IPs of medium-sized π-conjugated molecules in the gas phase and in the solid state were compared. The so-called solid-state polarization energy was shown to lower IPs of planar polycyclic aromatic hydrocarbons by 1.7 eV, more or less independently of the system [456]. The effect on EAs was not investigated, but the combination of solid-state polarization for IPs of this magnitude plus the unknown effect on EAs should lead to huge effects on band gaps.

4.1. Polyenes

Only for short polyenes are data available for gas phase (as discussed in section 2.1), solutions [288], solid solutions (matrix isolation) [288], and crystals [283]. These data show that excitation energies of unsubstituted polyenes are about 0.4 eV [288] lower in dilute liquid and solid solutions than in the gas phase. Cooling leads to improvement of the resolution without shifting the onset of absorption. Thus, polyenes are planar in the gas phase. Excitation energies of carotenoids [261] in solution are about 0.1 eV higher than those of the unsubstituted polyenes. The close agreement is probably a result of two opposite effects: methyl groups lowering excitation energies through hyperconjugation but increasing nonplanarity. Differences between oxidation and reduction potentials of tert-butyl-capped polyenes [283] are 0.8–0.9-eV smaller than excitation energies of unsubstituted polyenes in the gas phase. The data are plotted in Fig. 7. Because no experimental gas phase data are available for other systems, it is assumed in the remainder of this account that solvent effects are similar in size, about 0.4 eV, for other oligomers and polymers. This is a crude guess, of course.

4.2. Thiophene Oligomers

Thiophene oligomers were studied extensively. Experimental data are available for solutions [281, 286, 287, 291, 292], for isolated molecules in matrix [297–300, 457], for films [458, 459], for self-assembled monolayers (SAMs) [460], and for single crystals [285, 308, 457, 461, 462]. Cooling of solutions of thiophene oligomers leads to a red shift in the absorption of about 0.4 eV [297–300, 457]. This is in contrast to polyenes, where no such effect was observed. The red shift on cooling has been attributed to a transition from a twisted to a planar structure, as thiophene oligomers are known to be nonplanar in the gas phase and in solution [299, 463, 464]. The excited states, however, are planar because they have quinoid structures. Quinoid structures are less prone to out-of-plane distortions, as the interring bonds are double bonds (Scheme 6). Thus, excitation energies decrease on cooling as the matrix forces the oligomers into a geometry closer to that of the excited state. As a result, absorption and emission onset of thiophene oligomers coincide in matrix and are identical to the emission peaks in room-temperature solutions. The common origin of absorption and emission is another important difference between thiophene oligomers and polyenes. As discussed in section 2.1, in longer polyenes, emission occurs from the 2^1A_g rather than from the 1^1B_u state. For short thiophene oligomers (up to hexamers), the 2^1A_g state lies above 1^1B_u [297, 299, 300]. Thus, for thiophene oligomers, absorption, and emission involve the 1^1B_u state. Therefore, the energy of the most intense emission peak can be used to estimate the vertical excitation energy. Note that for thiophene oligomers the estimated solvent effect is of the same size as the effect resulting from nonplanarity, ~0.4 eV. Therefore, the two effects cancel, and

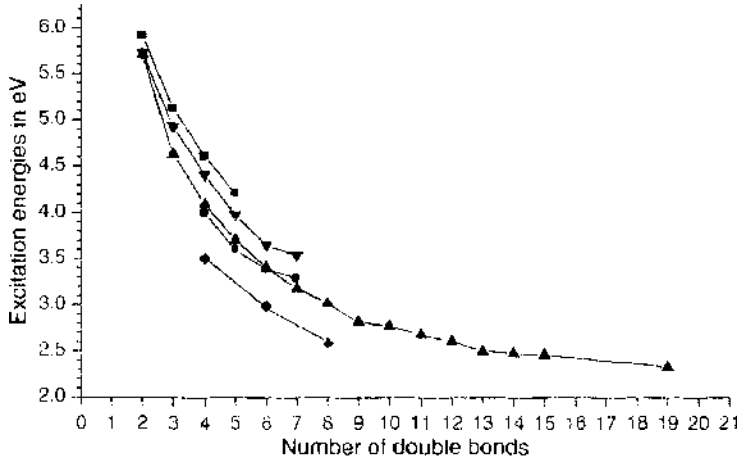
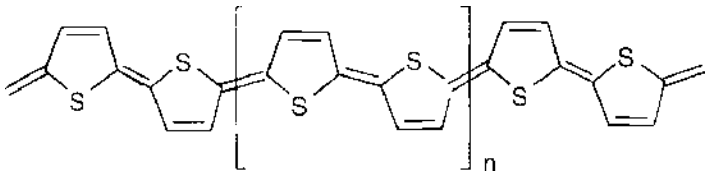


Figure 7. Experimental λ_{max} values in gas and in condensed phases versus the number of double bonds. Squares: gas phase as discussed in Section 2.1, down triangles gas phase 0–0 transitions, up triangles carotenes in solution, Refs. [261] and [426], circles: matrix isolation of polyenes, Ref. [288], diamonds: electrochemical gaps of ter-butyl-capped polyenes, Ref. [283].



Scheme 6.

experimental excitation energies for nonplanar thiophene oligomers in solution, and planar ones in the gas phase, should be about same. As a result of this coincidence, solution data can be compared directly to theoretical gas phase data obtained for planar oligomers. There is no guarantee that such a cancellation holds for other systems.

In Table 3 and Fig. 8, λ_{max} values of thiophene oligomers in room-temperature solutions are compared to gas phase excitation energies at various theoretical levels. The highest level of theory that has been applied is CASPT2 [465]. Excitation energies for 2-T and 3-T are 3.88 and 2.86 eV, respectively. These values are above and below the experimental ones, 3.67 and 3.08 eV, obtained for planar molecules in matrix. Because solvent effects are neglected, the theoretical values should lie above the experimental matrix data and be closer to the RT solution values of about 4.1 and 3.5 eV. That something must be wrong with the CASPT2 values is confirmed by the fact that they are almost identical to the TDLSDA values. TDHF is in better agreement with experiment, overestimating the experimental values by about 0.1 eV. CIS excitation energies are systematically too high. DFT HOMO-LUMO gaps overestimate excitation energies and fall off too fast, approaching the experimental values for long oligomers. TDDFT consistently underestimates excitation energies, and the results deteriorate as the chain length increases. The use of a hybrid functional (B3P86 rather than LSDA) ameliorates the problem somewhat, but the improvement is rather small. Increasing the weight of the HF exchange to 30% in TDDFT calculations has a much smaller effect on excitation energies than on HOMO-LUMO gaps.

4.3. Pyrrole Oligomers

Pyrrole oligomers are less intensively investigated than thiophene oligomers. I am aware of room-temperature solution data involving monomer, dimer, trimer, pentamer, and heptamer [466], and of a matrix isolation study of 2-Py [299]. These data are used for comparison with TDHF and B3P86-30% HOMO-LUMO gaps in Fig. 9. As for polyenes and for thiophene oligomers, HOMO-LUMO gaps are too high for small systems and fall off too rapidly. The error between experiment and TDHF excitation energies is clearly larger than

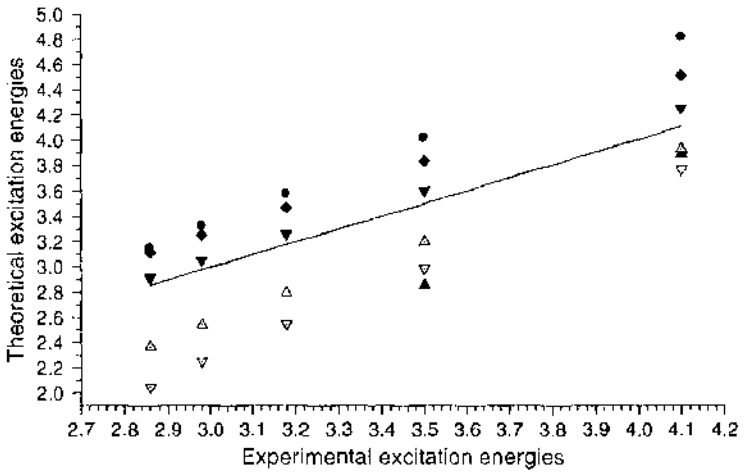


Figure 8. Theoretical excitation energies for thiophene dimer through hexamer at various levels are plotted versus experimental excitation energies measured in solution. All calculations are based on B3P86-30% geometries. Circles: B3P86-30%/CEP-31G* HOMO-LUMO gaps, diamonds: CIS/CEP-31G*, down triangles: TDHF/CEP-31G*, up triangles: CASPT2, Ref. [465], open up triangles: TDB3P86/CEP-31G*, open down triangles: TDLSDA/CEP-31G*.

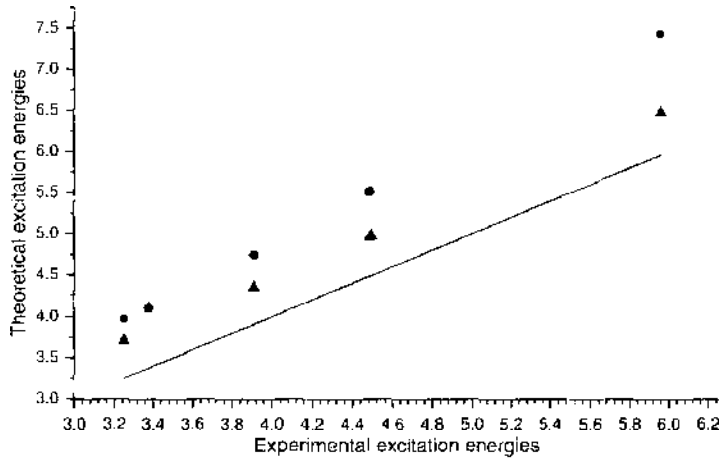


Figure 9. Theoretical excitation energies for pyrrole dimer through hexamer versus experimental excitation energies measured in solution. All calculations are based on B3P86-30% geometries. Circles: B3P86-30%/CEP-31G* HOMO-LUMO gaps, triangles: TDHF/CEP-31G*.

for thiophene oligomers. Excitation energies are overestimated by about 0.4 eV, independent of the chain length. It is not possible at this point to decide whether this is a result of a larger error of TDHF for pyrrole oligomers or of less-efficient cancellation of solvent and geometry effects.

5. INTERCHAIN INTERACTIONS

5.1. Experiment

If COPs were ideal, that is, if the chains were infinitely long and perfectly packed, COPs would behave like huge single crystals. Depending on the distance and the mutual orientation of the chains, the wave-functions overlap, and the electronic states split into as many branches as there are distinct species in the crystal. This so-called Davydov splitting [467] increases with decreasing distance between the chains and depends on the chain length, but it is always smaller than the splitting of the levels caused by intramolecular interactions between the repeat units. Smaller interaction between chains than within chains gives rise to low-dimensionality and anisotropy of COPs. Real samples of COPs are paracrystalline and comprise of polydisperse chains. It is hard to analyze the precise effect of interchain interactions, as too many factors like orientation of crystallites, grain boundaries, and amorphous regions influence the properties as well. In contrast to polymers, oligomers can be crystallized. The longest thiophene oligomer for which single crystals could be grown is octithiophene (8-T) [291]. The best-investigated of all oligomers is sexithiophene (6-T). Spectra of 6-T have been analyzed and compared for single crystals [78, 81, 308, 457–459, 461, 462, 467–469], for polycrystalline films [78, 81, 457, 459, 469], for matrix-isolated species [294, 297–300], and for solutions [287, 292, 470].

As discussed in section 4.2, the lowest excited state of thiophene oligomers is the allowed $1^1A_g \rightarrow 1^1B_u$ transition that is dominated by an electronic configuration in which one electron is removed from the HOMO and added to the LUMO. The same excitation is considered to give rise to the band gap in PT. The following analysis will concentrate, therefore, on the $1^1A_g \rightarrow 1^1B_u$ electronic transition. In crystals, 6-T molecules are planar (C_{2h} symmetry) and are packed in herringbone fashion with space group $P2_1/n$ [285]. A second phase that differs only in a translation of the chains with respect to each other has space group $P2_1/a$ [471]. As a result of the packing structure, the crystals have pronounced two-dimensional character [461]. Thiophene oligomers crystallize with either two or four molecules in the unit cell [457]. In most studies [78, 457], samples with four distinct species were analyzed. In this case, the 1^1B_u excitation splits into four Davydov components of a_g, a_u, b_g, and b_u symmetry. Gerade and ungerade states are nearly degenerate. The a_u and b_u states are one-photon allowed and can be observed in optical spectra [78]. Thus, two peaks designated as lower

and higher Davydov components arising from the same molecular state are expected to be seen in spectra of 6-T crystals.

6-T polycrystalline films and 6-T single crystals have been analyzed with a variety of spectroscopic techniques such as absorption [78, 81, 308, 457, 459, 468], polarized absorption [457, 462, 467], fluorescence excitation [457], emission [457], photoluminescence [461, 469], and electroabsorption spectroscopy [78, 81, 308, 459]. Furthermore, photoluminescence and photocurrent action spectra [461] have been recorded. Absorption starts for films and single crystals with a weak feature at 2.27 eV, followed by several peaks at around 2.4 eV. The λ_{max} occurs at around 2.6–2.7 eV. In thin films, a broad band at around 3.1–3.2 eV is observed that is absent in single-crystal spectra. Spectra of films are not well resolved, and peaks have been assigned with the help of electroabsorption spectroscopy [81, 308]. For single crystals, sharp optical spectra are observed at low temperature [462]. There was some controversy with the detailed peak assignment over the years. It seems that the results are most consistent with the following analysis. The weak feature at 2.27 eV arises from the a_u state and corresponds to the lower Davydov component of the 1^1B_u exciton. Because the exchange coupling for this excitation is weak, it was assumed that the a_u peak is close to the excitation of an isolated molecule [81]. The excitation at around 2.6–2.7 eV is the higher Davydov component of the 1^1B_u exciton. Thus, the Davydov splitting in 6-T is about 0.3–0.4 eV, and the exciton binding energy (the difference between the low Davydov component and the energy to create free charge carriers) is also about 0.3–0.4 eV in 6-T because CT states lie close to the higher Davydov component [308]. The broad peak at around 3.2 eV that is absent for single crystals is a result of coupling of excitons into a polariton mode [78, 81, 459].

The longest oligothiophene that has been crystallized is octithiophene [291]. The data for 8-T and 6-T are very similar. In dimethylformamide, λ_{max} for 8-T occurs at 2.82 eV, slightly below that of sexithiophene at 2.88 eV. The maximum emission is observed at 2.40 eV (compared to 2.45 eV for 6-T). Unoriented films show a broad band peaking at 3.13 eV with satellites at 2.27, 2.50, and 3.44 eV. Fluorescence excitation and emission spectra show the 0–0 transition at 2.19 eV. For single crystals, absorption spectra were recorded with light polarized parallel and perpendicular along the main molecular axis *b*. For light polarized parallel to *b*, absorption onsets at 2.07 eV and peaks at 2.21 and 2.58 eV. Using the same peak assignment as in 6-T, the Davydov splitting is 0.37 eV for 8-T.

The size of the Davydov splitting, depending on chain length, was investigated by comparing polarized absorption spectra of 4-T and 6-T [472]. The splittings of the a_u and b_u states in the *b*-polarized spectrum are 0.36 and 0.32 eV for 4-T and 6-T, respectively. It was therefore concluded that Davydov splitting decreases with increasing chain length. The result was confirmed with INDO/SCI calculations. The Davydov splitting for polythiophene was estimated to be about 0.25 eV. Combining the data for 4-T and 6-T [472] with those for 8-T [291] does not support this conclusion. The peak splitting between the a_u and b_u excitons in the spectrum polarized parallel to *b* is 0.37 eV and is therefore, almost the same as in 4-T and larger than in 6-T. The splitting is smaller (0.23 eV) with polarization perpendicular to *b*. 4-T and 6-T splittings were obtained for polarization parallel to *b* and correspond therefore to the higher value for 8-T. Thus, experiment does not indicate unambiguously decreasing Davydov splitting with increasing chain length.

5.2. Theory

The differences between 6-T spectra in solution and in crystals show that intermolecular interactions are important and that calculations on individual molecules miss major effects. Therefore, oligomers and polymers were investigated theoretically with three-dimensional solid-state calculations. Alternatively, oligomers were studied in clusters of increasing size. The main issues that are addressed with these investigations are the effect of interchain interactions on the band gap, the nature of the lowest excited state, the amount of Davydov splitting, the size of exciton binding energies, and the stability of localized defects such as solitons, polarons, and bipolarons.

Three-dimensional band structure calculations are usually carried out employing DFT with the LDA functional [231, 471, 473–481]. As described in section 2.4.2, LDA HOMO-LUMO gaps underestimate excitation energies, and the error increases with increasing chain

length. Therefore, LDA band gaps for infinite systems are too small. The 1-D band gap of PA, for instance, is 0.74 eV at the LDA level [479]. Inclusion of gradient correction leads to little improvement [479]. The error is the result of the incomplete cancellation of the self-interaction and may be corrected by calculating the self-energy with the Green function (GW) approach, using the LDA wave-function as input [481]. The results are true quasiparticle energies and quasiparticle bands [475]. Quasiparticle excitations or one-particle excitations are the energies required to add or remove one electron to or from the system [480]. For many materials the quasiparticle bands differ from the LDA bands by a k-independent rigid upward shift of the conduction bands [475]. Interactions between quasiparticles correspond to exciton-binding energies. These so-called two-particle excitations can be calculated by solving the electron-hole Bethe-Salpeter equation (BSE) [480] or with the density matrix formalism [482, 483]. The difference between the one-particle and the two-particle energies is the exciton-binding energy [481]. The difference between the one-particle or quasiparticle gap and the exciton-binding energy is the optical gap [478].

At the LDA level, interchain interactions lead to unsymmetric splitting of valence and conduction bands, and therefore to loss of electron hole symmetry [471, 473, 474]. The band gaps of PA and PPV are reduced by about 0.2 eV because of interchain interactions [471, 473, 474, 479]. As DFT for polyenes, LDA/GW does not reproduce the correct order of dark and optically active excited states for PA [477]. Although quasiparticles or polarons are localized when isolated chains are considered, self-trapping is strongly reduced or completely suppressed when interchain interactions are included [471, 473]. The same trend is reported for two-particle excitations. For single isolated PT chains, a binding energy for the ^1B_u exciton of 1.85 eV was calculated [477]. When interchain screening is included via the dielectric constant of neighboring chains, the exciton-binding energy reduces to 0.76 eV [477, 478]. With interchain effects calculated explicitly, the exciton-binding energy in PT is only 0.15 eV [481]. Similar results are obtained for PPV. The exciton-binding energy for an isolated chain amounts to 0.9 eV. Screening reduces the value to 0.54 eV, and explicit calculation of interchain interactions, including wave-function overlap, yields 0.35 eV. For PA, the exciton-binding energy of the lowest-lying singlet state was found to be reduced from 0.6 to 0.05 eV with three-dimensional calculations [231]. Interchain effects were claimed to strongly influence the quasiparticle gap and the exciton-binding energies and to have very little effect on the optical gap [478]. In contrast, the same quasiparticle gap is predicted with one- and three-dimensional calculations for PA [479]. The similar optical gaps obtained with one- and three-dimensional calculations result from the cancellation of one- and two-particle energies.

To obtain the Davydov splitting, the energy difference between excitons that arise from the same molecular state has to be evaluated. In the case of PT and other polymers crystallizing in the herringbone structure, these excitons have a_u and b_u symmetry. With the three-dimensional LDA/GW/BSE approach, very small Davydov splittings are obtained. For PT, Davydov splitting is below 0.01 eV, for PPV it amounts to 0.09 eV [481]. Applying the density matrix formalism, which is equivalent to the BSE method, a Davydov splitting of 0.12 eV was computed for PPV [482] and 0.01 eV and 0.06 eV for the two different crystal structures of PT [484]. Because it was shown that the state ordering for PA and polydiacetylene (PDA) is qualitatively wrong with the LDA/GW/BSE method [480], the Davydov splittings quoted above might be unreliable. This problem reminds us of the wrong ordering of the states obtained with TDDFT for long polyene chains, as discussed in section 2.3. If the energy difference between excitons is not necessarily correct with this approach, the exciton-binding energies themselves might also be subject to error.

An alternative to three-dimensional band structure calculations is the supermolecular approach [442, 485–493]. The calculations are carried out for oligomers of different chain lengths and for clusters of oligomers of different size. Semiempirical methods are applied because the size of the systems is prohibitive for *ab initio* calculations. Geometries may be obtained using the Austin model-1 (AM1) Hamiltonian [435]. Excitation energies are usually calculated on the basis of intermediate neglect of differential overlap (INDO) and subsequent singles configuration interaction (SCI) between the occupied and unoccupied π orbitals to obtain the excitation energies. The INDO/SCI method was parametrized to reproduce optical absorption spectra of organic molecules [487, 489–492].

Fägerström and Stafström tried to model interchain and intrachain electroluminescence processes in oligothiophenes. Because injection of a hole and of an electron creates a positive and a negative polaron, respectively, the geometries of separate cation and anion of 8-T were obtained employing the AM1 method. Then a dimer was formed by bringing the 8-T cation and anion together. Two configurations of the dimer were tested. The first was a model for interchain electron transfer between two separate molecules, the second was a model for intrachain recombination in which the positive and the negative polaron reside on the same chain. For both models, excited states were calculated at the INDO/SDCI level [485]. For two separate parallel 8-T chains at a distance of 3.7 Å, an allowed excited state with no charge separation (according to Mulliken population analysis) was obtained. The excitonic state lies 2.37 eV above the ground state and has a lifetime of 1.6×10^{-9} s. A dark state was found at lower energy, ~1.86 eV. Singlet and triplet states are close in energy. Two separate polarons on the same chain are not stable and can only be obtained when the two chain segments are twisted with respect to each other. For a planar chain, a nonradiative recombination was predicted. With three instead of two parallel chains, the symmetry properties are different. The S_1–S_0 state is dipole allowed and has a very short lifetime, whereas the S_2–S_0 transition has very long lifetime [488].

Cornil et al. [442, 487, 489–493] applied the supermolecular approach to stilbene, polyenes, para-phenylenevinylene, and thiophene oligomers of different chain lengths, in different orientations, and with varying distances. Geometries were obtained with the AM1/CI method, and excited states were calculated at the INDO/SCI level of theory. The results can be summarized as follows. For cofacial dimers or clusters, Davydov splitting results in a symmetry-forbidden first excited singlet state and a symmetry-allowed second excited singlet state. Because luminescence occurs from the lowest excited state, Davydov splitting leads to luminescence quenching. Davydov splitting is strongest for the highly symmetric cofacial arrangement, increases with decreasing interchain distance and increasing cluster size, and decreases with increasing chain lengths of the oligomers (after a peak for medium sized oligomers) [472, 490]. In addition, electronic excitations with different character start mixing at small interchain distances, which leads to a transfer of oscillator strength to CT states and almost complete luminescence quenching [442]. For homooligomers, CT states are higher in energy than the strong Davydov component. For copolymers, the higher Davydov component moves closer to the CT states. This should be desirable when high conductivity is aimed at. In the herringbone arrangement (the crystal structure of polyenes and of thiophene oligomers), the molecules are almost cofacial. The lowest excited state is weakly allowed, and the Davydov splitting is slightly reduced [442]. For sexithiophene, it was shown that the Davydov splitting saturates quickly with cluster size and is predicted to be 0.32 eV for the single crystal [491]. For bisdithienothiophene (BDT), which prefers a different crystal structure with one-dimensional stacks, a Davydov splitting of 0.61 eV and an intense CT band that lies 0.62 eV higher than the high Davydov component were predicted for a 12-unit cluster.

Hartree-Fock as well as density-functional theory do not account for dispersion effects or van der Waals interactions. The lowest level of theory that does is MP2 [408]. MP2 correction to the band gap of PA single chains, dimers, trimers, and two-dimensional arrays was investigated with extended basis sets. The band gap for a single chain is about 3.6 eV. Interchain interactions lower the band gap to 2.98 eV for the dimer, 1.95 eV for the trimer, and 2.59 eV for the two-dimensional array [408]. Thus, interchain interactions are substantially larger with MP2 than with DFT and semiempirical methods.

6. COMPARISON BETWEEN THEORY AND EXPERIMENT FOR POLYMERS

The calculations described in the last paragraph are the most accurate and realistic that are presently possible for COPs in the solid-state. The LDA/GW/BSE approach is the domain of solid-state physicists, and the supermolecular approach is the method of choice for theoretical chemists. In this chapter an attempt is made to assess the accuracy of both approaches compared to experiment. In Fig. 10 the evolution of absorption energies in going from gas phase to crystal is shown for 6-T and PT. The first and the last entries are TDHF excitation

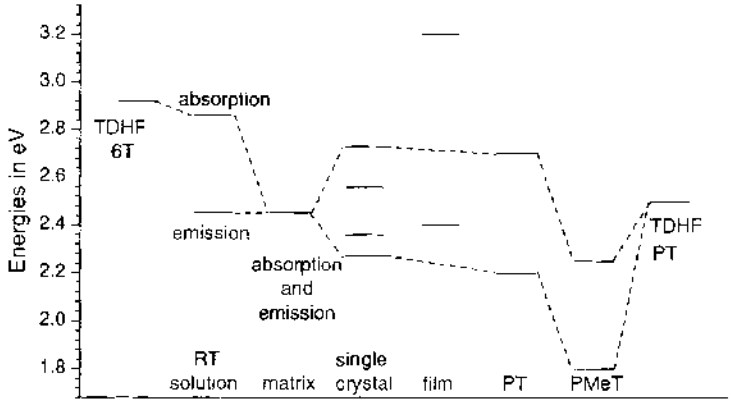


Figure 10. Evolution of excitation energies form gas phase to polymer.

energies of the perfectly planar 6-T and 28-T molecules in the gas phase. At a chain length of 28 repeat units, the excitation energy has converged to 2.5 eV and is considered to be identical to that of an infinite isolated PT chain. For 6-T TDHF yields an excitation energy of 2.92 eV. In room-temperature solution, absorption occurs at 2.86 eV, and emission at 2.45 eV. The difference of 0.4 eV between absorption and emission disappears on cooling and is therefore attributed to a disorder of the ground state in room temperature solution. It was argued in Section 2.2 that TDHF is accurate for gas phase absorption energies. Therefore, the difference between gas phase calculations and matrix absorption energies of planar molecules is attributed to the solvent effect, and the match between room temperature results and theory is a result of the cancellation of the effects of solvent and of nonplanarity, as discussed in Section 4.2. In the crystal, the ^1B_u exciton splits into two optically active Davydov components. The dashed lines connect the ^1B_u absorption energy in matrix with its two Davydov components in the crystal. The lower Davydov component at 2.27 eV lies ~0.2 eV below the absorption and mission peak in matrix; the higher Davydov component at 2.6–2.7 eV lies ~0.2–0.3 eV above. Absorption in polycrystalline films onsets at higher energy and the higher Davydov component is part of the collective polariton state at higher energy. PT shows only a broad absorption feature. The onset of absorption occurs at 2.2 eV, λ_{max} at about 2.7 eV. Comparison with 6-T single crystal spectra indicates that the onset of absorption in PT corresponds to the lower-lying, low-intensity Davydov component, and λ_{max} to the higher-lying, intense Davydov component. Both absorptions of PT occur at slightly lower energies than those of 6-T, and therefore PT should have somewhat increased conjugation lengths compared with 6-T. The last experimental entry is for regioregular polymethylthiophene (PMeT). The lower absorption energy compared with PT was attributed to increased order and not to hyperconjugation [494]. Therefore, PMeT may serve as a model for a well-ordered PT, and the 0.4 eV lowering of the absorption onset reflects increased conjugation length. Thus, ideal three-dimensional PT should have an onset of absorption of about 1.8 eV, λ_{max} of 2.2 eV, Davydov splitting of 0.4 eV, and exciton-binding energy $\geq$ 0.4 eV.

Several conclusion can be drawn from this comparison. First, gas phase absorption energies differ from the lowest absorption in the crystal by a combination of the medium effect plus the state splitting. The difference between TDHF gas phase values and the experimental onset of absorption amounts to about 0.6 eV for 6-T and to 0.7 eV for PT. The similar difference between theoretical gas phase excitation energy and experimental solid-state absorption onset supports the conclusion that the TDHF excitation energy is reliable not only for molecules but also in the infinite chain limit. Second, onset of absorption of PT arises from the lower, and λ_{max} from the higher Davydov component of the same molecular excited state, and not from molecules with different conjugation length. Third, both onset of absorption and λ_{max} of PT are closer to those of 6-T than to those of PMeT. Thus, the conjugation length in PT is probably far from the infinite limit. Fourth, as Davydov splitting is about 0.3–0.4 eV for 6-T, 8-T, PT, and PMeT, there is no indication from experiment that Davydov splitting decreases with increasing conjugation length. There is also no indication that larger separation resulting from the substituents as expected for PMeT decreases Davydov splitting. This is in contrast to predictions obtained with the supramolecular approach at

the INDO/SCI level. Fifth, There is a substantial discrepancy between experimental Davydov splittings and those predicted with the LDA/GW/BSE method for solids. Finally, the optical gap with the LDA/GW/BSE method was found to be 1.74 eV for one- and 1.49 eV for three-dimensional PT. If the analysis above is correct, both values as well as the difference between them are too low.

7. CONDUCTING STATE OF CONJUGATED ORGANIC POLYMERS

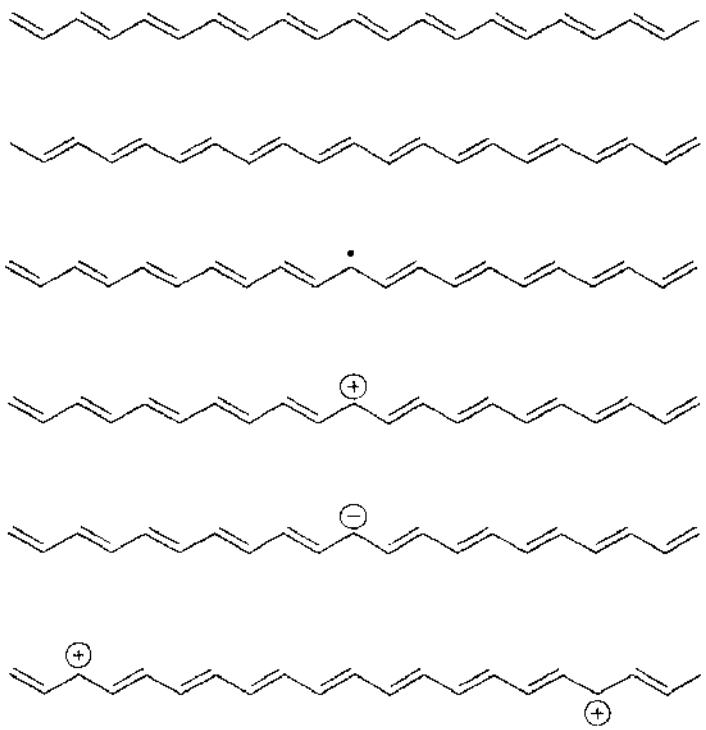
Because COPs have finite band gaps, they do not conduct unless free charge carriers are produced either by excitation of electrons from the valence into the conduction band [495] or by chemical or electrochemical oxidation or reduction [496]. The latter two processes are referred to as p- and n-doping, respectively. In practice, doping is done chemically, by exposing neutral samples to electron-accepting or electron-donating substances, or electrochemically, by charging films at electrodes in electrolyte solution. With both methods, the charges introduced into the polymer are balanced with counter ions that move into the polymer lattice. Thus, all experimental results are obtained in the presence of counter ions.

Although the same terminology is used for organic and inorganic semiconductors, doping of COPs differs substantially from that of inorganic semiconductors. Inorganic semiconductors consist of rigid lattices of individual atoms, whereas COPs consist of more or less crystalline arrays of organic molecules. Inorganic semiconductors are doped by replacing atoms of one element with those of another having either one electron less or one electron more. For instance, silicon may be replaced by boron or phosphorus. This kind of doping does not lead to ionization and does leave the band structure unaltered. If the doping element has one electron less, holes are created in the valence band; if it has one electron more, free electrons are added to the conduction band. In contrast, doping of COPs is achieved by ionization of the polymer chains. This kind of doping leads to ion formation with associated geometry relaxation and alters the band structure [3]. Counter ions are needed to balance the charges. The presence of counter ions influences energy levels, geometries, and crystal structures.

Because doping of COPs alters the band structure, it influences optical spectra [3]. Neutral PA has a strong absorption at 1.9 eV and an onset of absorption of 1.4 eV. On doping, the interband transition at 1.9 eV decreases and a new intense absorption appears in the infrared region at an energy of about half that of the original transition. This behavior is independent of the dopant ion and of whether PA is oxidized or reduced [3]. For PT and PPy, doping leads to two new transitions below the interband transition. The intensity of the original interband transition is reduced, and its energy is blue shifted. The experimental results have been explained successfully in terms of solitons, polarons, and bipolarons [255].

Theoretical investigations indicate that two kinds of COPs have to be distinguished, in principle [255]. The first class comprises systems with degenerate ground states. The most prominent member is PA. Degenerate ground state means that it makes no difference whether the positions of single and double bonds change place. If the chain is infinite, the energy and all other properties remain the same. If such a system develops a defect or is doped, two fragments are obtained on both sides of the defect, in which the positions of double and single bonds are switched with respect to each other (Scheme 7). Such defects are called solitons [246, 247, 255, 497–501]. Individual solitons may be produced on imperfect chains. On excitation or doping, soliton–antisoliton pairs must form [3].

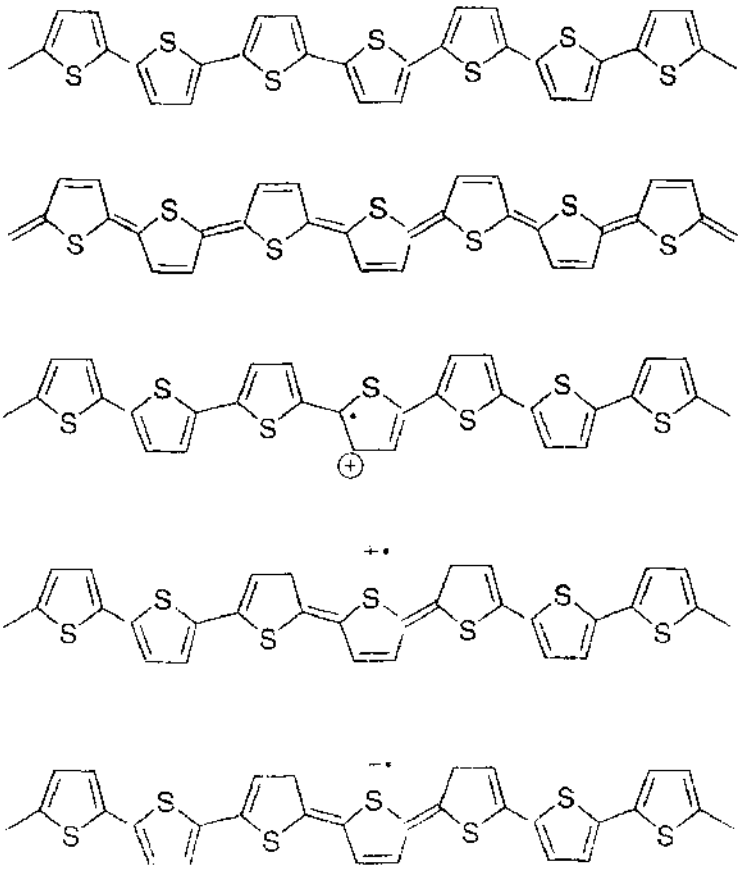
The second class of conducting polymers has a nondegenerate ground state. The majority of COPs fall in this category. Consider polythiophene (PT) (Scheme 8), for instance. Thiophene rings are aromatic. Therefore, the rings contain two double bonds and are connected with single bonds. If the positions of double and single bonds are switched, the thiophene rings are quinoid and the interring bonds are double bonds. Because thiophene is a 6-π-electron aromatic ring, thiophene prefers the aromatic over the quinoid form. Thus, the ground state of PT is nondegenerate, with the aromatic being lower in energy than the quinoid form. Because of this energy difference, thiophene rings next to the defect cannot simply adjust the positions of double and single bonds to bind the lone electron that is left



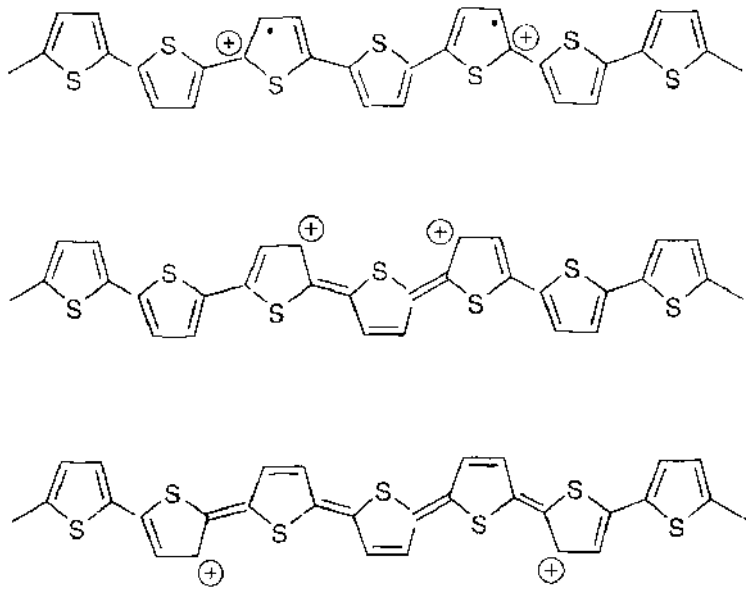
Scheme 7.

behind when its bonding partner is removed. Hence, the cation (or anion) site is coupled to a radical electron. Such a defect is called a polaron [250–252, 502, 503].

Removing a second electron may create a second polaron, or a so-called bipolaron. Scheme 9 shows bipolaron formation in polymers with nondegenerate ground states.



Scheme 8.



Scheme 9.

Because two unpaired electrons are produced, pairing of all electrons can be achieved by switching the positions of single and double bonds between the two charges. This requires a quinoid segment to be formed between the two charges. The size of the bipolaron can be understood to be a compromise between minimizing Coulomb repulsion between the two charges by moving them as far apart as possible while keeping the high-energy quinoid segment as short as possible. This results in charge confinement. The doping processes were shown to be similar in bulk and in dilute solution. Thus, charge storage in the form of bipolarons is not a solid-state effect but a property intrinsic to the molecules [3]. In polymers with a degenerate ground state, there is no energy cost associated with switching the position of single and double bonds. No confinement is required, and a soliton–antisoliton pair may form.

Theoretical treatment of optical spectra of doped bulk COPs is a complicated problem. Macroscopic disorder, actual conjugation lengths, and effect of counter ions on the crystal packing and on the energy levels are hard to determine experimentally, and they are difficult to account for theoretically. Because doping leads to open-shell structures, calculation of optical spectra is even more complicated than for neutral species. In Section 2 important aspects regarding calculation of a much simpler problem, calculation of polyene spectra in the gas phase, are summarized. Because of the size of the systems and the more limited computer recourses from 25 years ago, much simpler approaches had to be developed at the time. The next paragraph describes these efforts in more or less chronological order.

The first theoretical treatment of solitons, polarons, and bipolarons was done employing the Su-Shrieffer-Heeger (SSH) Hamiltonian [246–248, 255, 498]. The SSH method is essentially a Hückel method that takes into account bond length alternation. It does not treat Coulomb repulsion between electrons in a self-consistent way. Counter ions and interchain interactions were not considered. With the SSH method, solitons extend over about 15 carbon atoms in PA. At the center of the soliton, BLA is absent. The localized defect introduces a midgap state in the band gap. On doping, the singly occupied midgap state becomes empty or doubly occupied. P- and n-doping require the same energy, because relaxation effects and Coulomb repulsion are not accounted for. Despite the simplicity of the approach, the midgap state leads to optical transitions in agreement with experimental observations. For polymers with nondegenerate ground states, polarons and bipolarons were found that are associated with two new energy levels in the band gap [3]. Similar to solitons, positive polarons and bipolarons extend over a certain fragment of the chain and lead to reduced BLA or bond-length inversion within that fragment. Bipolarons were found to be more stable than two separate polarons.

Restricted HF calculations on infinitely long Li-doped PA chains [249] were carried out with subminimal Christofferson basis sets. Li atoms were positioned equidistant from three carbon atoms. The model corresponds to a doping level of 33%. Charge transfer was analyzed with Mulliken population analysis (MPA). Even at charge transfer of only 0.1 eV per CH unit, BLA was absent. The charge transfer might actually be underestimated as a result of known problems of MPA with polar bonds, and especially with Li [504]. The presence of Li introduces two 2s bands in the gap, that mix strongly with the π^* band, pulling it down in energy [249]. Using an adiabatic Hückel Hamiltonian with σ-bond compressibility [250], without considering interchain interactions and counter ions, the separation of the radical ion pair in PA was found to be small. Such a closely spaced radical ion pair is a polaron rather than a soliton. Thus, the soliton picture was found to be not unique for PA. At higher doping levels, interaction between the radical electrons of two polarons led to recombination and the formation of two charged solitons carrying no spin. In poly(paraphenylenevinylene) (PPV), defects were predicted to be correlated in pairs as a result of the absence of a degenerate ground state. The lattice distortion was found to involve about five rings. At higher doping levels, uncorrelated solitons form in PA, and bipolarons or correlated charged soliton antisoliton pairs appear in PPV. Overall, PA and PPV were found to be more similar than generally thought. Partial geometry optimizations were carried out with the restricted HF method and STO-3G basis set on quatermers of phenyl, pyrrole, and thiophene with Li and Na counter ions [505]. Band structure calculations were done employing VEH on the RHF geometries. Because about 80% of the charge was found on the two inner rings, it was concluded that bipolarons extend over about four rings in PPP. Two new states at 0.74 eV above the valence band and at 0.88 eV below the conduction band appeared in the gap. For 4-Py only with Na charge transfer observed. Again, two new states appeared in the gap; 4-T behaves very similar to 4-Py. Although individual molecules with only four rings were considered, excellent agreement with experimental data for bulk PT was claimed. Likewise, band structure calculations with a quantum mechanical version of the SSH Hamiltonian on individual PPy chains yielded quantitative agreement with experiment. Bipolaron binding energies with respect to two separate polarons were calculated to be 0.45 eV in PPy and 0.34 eV in PPP [253].

Higher levels of theory lead to much less consistent results. Investigation of the basis set dependence of charge transfer complexes showed that minimal basis sets are not adequate [506]. Correlation effects were shown to be important as well [506]. Polaron sizes increase when going from semiempirical to *an initio* methods and depend strongly on the presence of counter ions [506]. Explicit calculation of the excitation energies of doped thiophene oligomers at the INDO/SCI level [507] led to underestimation of transition energies of 0.6 eV. The second peak from HOMO to the first polaron level was overestimated, and the transition from the first to the second polaron level was underestimated by 1 eV. In contrast, good agreement with experiment was obtained between VEH band energies and experiment. However, bipolarons led to only a single subgap absorption feature. PPV and PT were shown to behave similarly. At the HF/6-31G* [508] level, quinoid structures in dications of thiophene oligomers were predicted to disappear and benzoid structures to develop with an increasing number of rings. Optimization at the CI level [509] of polarons and bipolarons on thiophene oligomers indicated that bipolarons are only stable for short chains. Starting with the hexamer, triplet and singlet states are degenerate, which means that the two polarons do not interact. Two quinoid regions are separated by an aromatic region in the middle. The charge is located near the ends. The same conclusion was drawn from calculations at the HF and ROHF levels [510]. The excitation spectrum of the bithiophene radical cation was investigated at the CASSCF level [511]. The resulting transitions were compared to those predicted by the polaron model, where two new levels in the gap give rise to three intergap transitions. All transitions were found to be multiconfigurational at the CASSCF level. Two transitions were dominated by configurations that correspond to those predicted by the polaron model, but one additional excitation was obtained that has no counterpart in the polaron model.

Three-dimensional band structure calculations on PA employing DFT showed that interchain interactions strongly reduce localization of defects, destabilizing polarons and probably

bipolarons [473]. For reasonable ranges of lattice distortions, no states were found in the band gap [471]. DFT has a tendency to underestimate localization, but the presence of localization on individual chains and the absence of localization in the crystal at the same level of theory shows that localization is overestimated if interchain interaction are neglected. Gradient-corrected DFT calculations [512, 513] result in planar PPy chains in the crystal, although 6-Py has an out-of-plane distortion at the same level of theory. Doping of infinite chains in the crystal with chlorine atoms and molecules leads to out-of-plane distortions and quinoid deformations that extend over two rings. Charge transfer is incomplete, and the interactions are partially covalent. 6-Py cations are almost planar. Positive charges are found at the ends, whereas excess spin density occurs in the middle. Because bipolarons were shown to be unstable with respect to two polarons for thiophene oligomers, π-dimers were invoked to explain conductivity in the absence of an ESR signal [514]. In the gas phase, dimers are unstable or metastable according to DFT calculations. However, stability increases when medium effects are included. Stacks of 4-T cations dimerized on geometry optimization. Thus, a metallic stack becomes semiconducting as a result of Peierls distortion. Three-dimensional band structure calculations employing DFT on PPP in the presence of sodium ions [515] showed that a phase transition to a hexagonal channeled structure occurs in which columns of sodium ions are surrounded by three PPP chains. Doping left the relative energetic positions of the carbon π-bands almost unchanged. However, the Fermi level was shifted upward because Na 3s electrons partially filled previously unoccupied π-bands. A transition to the metallic state was predicted.

DFT calculations, including those employing hybrid functionals, lead to complete delocalization of all defects in the absence of counter ions [516–518]. Only one hybrid functional that includes 50% of Hartree-Fock exchange seems to reproduce localized defects [519]. The highest-level calculations on the longest chains performed so far [516] show that HF theory overestimates localization, predicting a polaron size in PA of 15–16 units. At the MP2 level, the defect extents over 36 CH units. In the presence of counter ions, HF, MP2, and DFT agree that a positive defect extends over nine CH units (counter ion Cl) and a negative defect (counter ion Li) over five CH units [516]. The different size of positive and negative defects may partially be a result of different amounts of charge transfer.

In summary, the nature and the extension of charged defects depends strongly on the theoretical level, the basis set, the presence of interchain interactions, and especially the presence of counter ions. There is experimental evidence for localization, but these results were obtained in the presence of counter ions. There is, therefore, no experimental evidence that localization occurs in the absence of counter ions. The strong localization obtained with semiempirical methods may well be a result of the approximate treatment of Coulomb interactions. Localization decreases from semiempirical to HF to correlated methods. There is very little localization with DFT methods unless a substantial amount of HF exchange is used, although DFT/hybrid methods are usually very successful for geometry optimizations. This problem is probably related to the difficulty of DFT in treating long-range interactions. Because a large dependence of the results on the level of theory quite often indicates very small energy differences, the underlying reason for the theoretical problems might be that energy differences between localized and delocalized defects are small. This possibility is investigated in the following paragraphs.

$C_{102}H_{104}$, $C_{102}H_{104}^{2-}$, 24-T, 24-T^{2+}, 22-Py, and 22-Py^{2+} were optimized at the HF/CEP-31G* and at the B3P86-30%/CEP-31g* levels of theory. Differences in bond lengths were evaluated by comparing bond lengths in neutral chains with those in the charged systems. The results for $C_{102}H_{104}^{2+}$ and for 24-T^{2+} are plotted in Figs. 11–14 at the HF/CEP-31G* and B3P86-30%/CEP-31G* levels of theory. At the HF level, the geometry of the $C_{102}H_{104}^{2-}$ ion corresponds to a soliton–antisoliton pair. Starting from the ends, single bonds decrease and double bonds increase in length, going through a region with no BLA. In the middle part of the molecule, changes in bond lengths are constant and correspond to a complete inversion of single- and double-bond positions. In this region, BLA is exactly equal to that in the neutral chain. With DFT, the trend is similar, but the fragment in which BLA is equal to that of the neutral system but single- and double-bond positions are switched is much shorter because the extension of the solitons is wider.

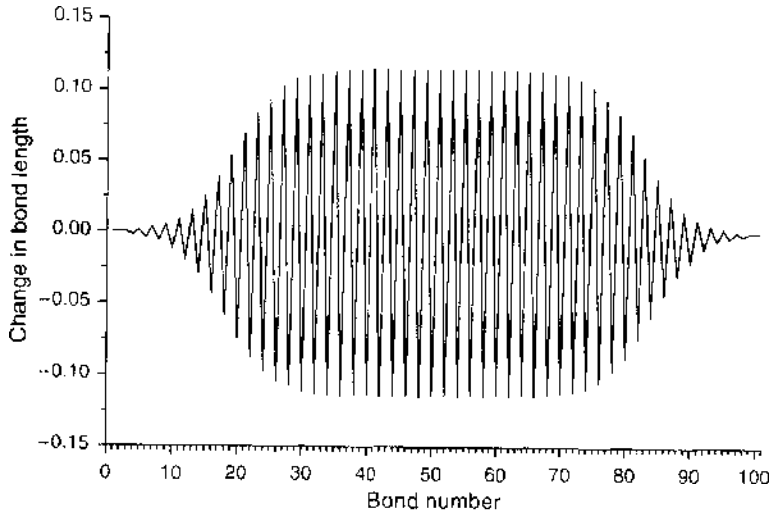


Figure 11. Bond length changes on double ionization in $C_{102}H_{104}$ at the HF level.

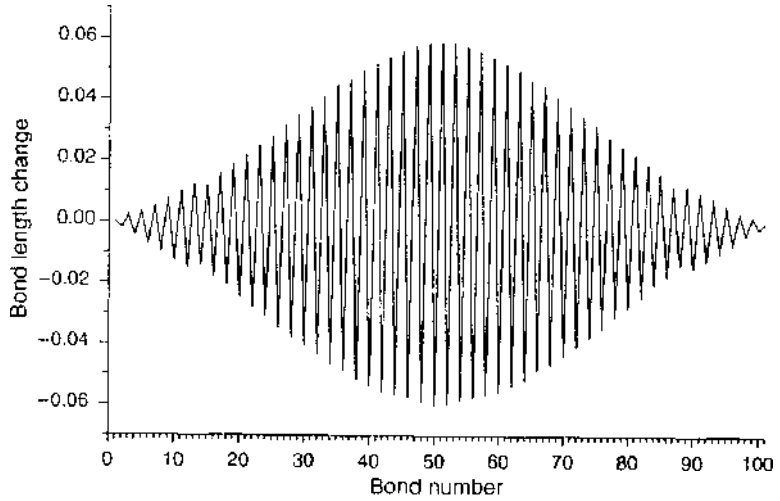


Figure 12. Bond length changes on double ionization in $C_{102}H_{104}$ at the DFT level.

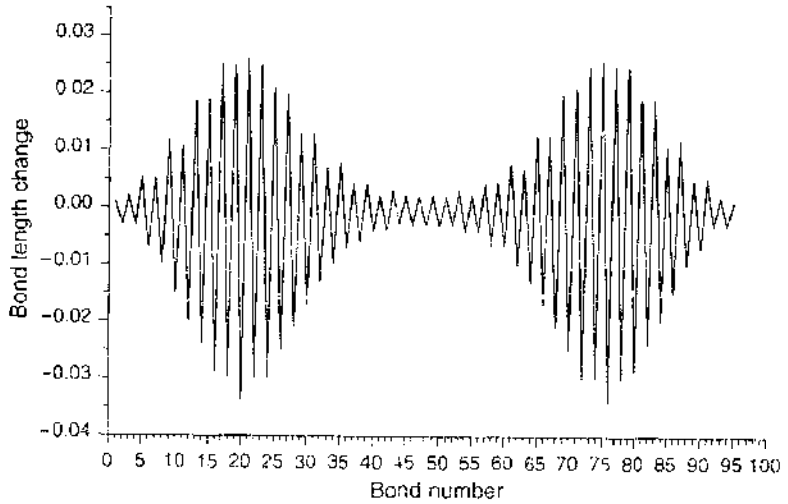


Figure 13. Bond length changes on double ionization in 24-T at the HF level.

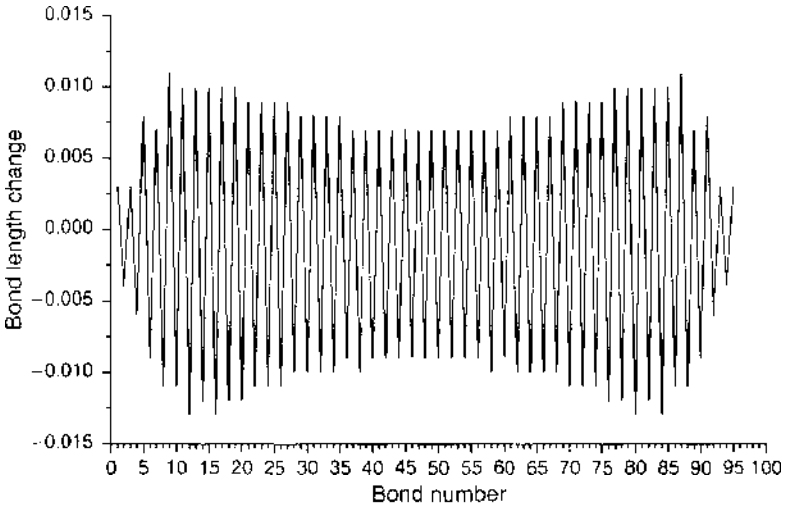


Figure 14. Bond length changes on double ionization in 24-T at the DFT level.

The 24-T^{2+} ion behaves very similar to $C_{102}H_{104}^{2+}$ at the HF level, with two charged defects at the ends of the chain. The difference is that the middle part of the molecule shows no switching of single- and double-bond positions. Thus, 24-T^{2+} develops two polarons on charging. It is noteworthy that this happens although the calculation of a bipolaron was attempted and a closed shell singlet state was defined. The proper treatment of two polarons would require multireference calculations involving at least four electronic configurations [509]. Thus, the two polarons formed, although the calculation is strongly biased toward bipolaron generation. This confirms that bipolarons are unbound in the absence of counter ions at *ab initio* levels, when long enough chains are considered. DFT also produces two polarons, but the extend is again much wider, and the changes in bond lengths in the middle of the molecule are larger.

To examine the energy cost associated with the differences in geometries at HF and DFT levels, single-point calculations were carried out at the HF level on the DFT geometries and at the DFT level on the HF geometries. Because total energy differences increase with chain length, the differences were evaluated per repeat unit. The proper repeat units for PT and PPy are two rings in transconformation. For comparison a segment of similar size, C_8H_8, was considered for PA. The energy differences between the two geometries are 3.02 and 2.44 kcal/mol for PA at HF and DFT levels. 1.30 and 1.10 kcal/mol for PT, and 1.26 and 0.97 kcal/mol for PPy. Especially for PT and PPy, the energy differences are so small that even the highest levels of theory would be challenged. To reach "chemical accuracy" with errors of about 1 kcal/mol, highly correlated calculations with extended basis sets are required. This explains the sensitivity of the geometries to the level of theory and the difficulties encountered during the geometry optimizations, which require 10–20 times as many optimizing steps as with more "well-behaved" molecules.

Theoretical results thus indicate that the geometries of long-conjugated ions can adjust to very different geometries at a small energy cost. When potential energy surfaces are that flat, exact values of bond lengths and angles have much less meaning than for stiff molecules. This is exactly what is required for efficient charge transport. The discussion about exact polaron sizes in the gas phase is thus fairly meaningless. Any external influence such as interchain interactions or counter ions will have a profound influence on the final structure. This explains nicely why all levels of theory agree about the width of polarons when counter ions are considered.

8. POLARIZABILITY

To calculate polarizabilities, molecules or polymers have to be exposed to an electric field. In the finite field approach for molecules, the operator $(-eE\ r)$ is used to represent the potential of the electric field. Calculating polarizabilities of infinite systems in the same way

is problematic because the operator for the potential of the electric field is unbound and destroys the translational symmetry of the polymer [520]. The oligomer approach has been used extensively [407, 521–532], but the problem here is that polarizabilities first increase and then reach an asymptotic limit [533]. Therefore, extrapolation can not be applied using small oligomers. Because the convergence is very slow, required system sizes are often prohibitive.

The first method that has been worked out for treating infinite systems was based on the tight-binding approximation [534, 535]. The chains were assumed to be independent, and infinite, electronic repulsions and geometry relaxations leading to excitons were neglected. Another approach to solving the problem is to average the polarizabilities from various energy-band structure calculations for periodic systems with different but constant potentials [536]. Partitioning the electric-field operator into two components, one describing the polarizing effect and the other one describing the acceleration of the electrons, allows preservation of translational symmetry by taking into account only the first term [407, 520]. The method was implemented at the HF level. More recently it was suggested that the perturbation is only unbound if the interaction with the electronic or nuclear dipole is treated separately [537]. HF calculations with this approach showed good agreement with literature values. An approximated method at the uncoupled HF level was implemented and used to establish trends because polarizabilities were underestimated [533]. The most complete treatment of static and dynamic non-linear optical effects for polymers was developed employing RPA without neglecting any terms [538–540]. The results with this method agree well with those of oligomer calculations.

The uncoupled sum over states (SOS) scheme [533] was compared to RPA polarizabilities of oligomers of increasing size [541]. The SOS approximation underestimated the RPA values by 30–50% but reproduced the trends correctly. Polarizability was shown to increase in the order PPy, polyfuran (PFu), PT, polydiacetylene, PA. For all systems investigated, the highest-occupied π-orbital makes the largest contribution to the polarizability [542]. Inclusion of relaxation effects at the coupled HF level of theory [523] showed similar trends as with RPA, increasing polarizabilities compared to uncoupled HF. Intermolecular interactions were examined employing hexatriene as a model for PA [543]. The polarizability was reduced by 50% and second hyperpolarizability by 95% compared to single-molecule values.

Inclusion of electron correlation effects with MP2-MP4, CCSD, and CCSD(T) decreases the polarizability, α [544, 545]. The first hyperpolarizability β increases at the MP2 level [546]. Basis sets have to be much larger than for HF calculation, or the correlation effect is overestimated. For centrosymmetric acetylene oligomers, β vanishes because of symmetry, α decreases, and the second hyperpolarizability γ increases when correlation effects are included at the MP2, MP4, and CCSD(T) levels of theory. Because correlation was shown to be important, the performance of DFT was investigated [531]. With all common DFT functionals, both α and γ increase substantially for acetylene oligomers. Inclusion of HF exchange ameliorates the problem somewhat, but unreasonably large amounts of HF exchange are needed. As with excitation energies, the problem increases with increasing chain length. Thus, DFT with currently available functionals fails in computation of nonlinear optical effects.

So far, only static polarizabilities have been discussed. Dynamic polarizabilities take into account the frequency dependence [407, 547]. For π-conjugated molecules, the importance of the vibrational part to the first hyperpolarizability depends on the nature of chemical system [530]. The smaller the BLA, the larger the contribution of the vibrational compared to the electronic part. For polymers, dynamic polarizabilities have been calculated, taking interchain interactions into account [548]. Compared to literature values, the frequency dependence tends to lower the polarizability. Dynamic hyperpolarizabilities of PA, OH- and F-substituted PA, PPy, PFu, PPP, PT, and PAn were evaluated [549]. At the HF level, parent PA is more polarizable than the substituted forms. When correlation is included, polyfluoroacetylene is the most polarizable. In the series of polymers with aromatic rings, polyaniline is twice as polarizable as PT, followed by PPP, PPy, and PFu. Calculation of dynamic polarizabilities at the RPA level confirms that vibrational contributions can be very important [538].

9. CONDUCTIVITY

Macroscopic conductivity through bulk COPs is a complicated process that is strongly influenced by heterogeneous disorder [277]. Because COPs are made of very long conjugated molecules, conductivity is largest in chain direction and is limited by the need for hopping between chains. It is therefore interesting to investigate conductivity through small bundles of well-aligned polymer chains [20, 56] or through individual molecules. It is now possible to measure conductivity of individual molecules in self-assembled monolayers with scanning-tunneling microscopy and with atomic force microscopy [49, 90, 91, 97, 99, 105, 106, 550]. Many theoretical studies are done on mesoscopic and nanoscopic devices [38, 50, 98, 100, 101, 103, 104, 107, 280, 550–575] consisting of conjugated molecules connected to jellium or gold contacts, usually via sulfur bridges. Charge is transferred under the influence of an external field. Transport through the molecular wire bridge is a many-body problem that depends on the electronic structure of the molecule, the coupling between the molecule and the contacts (leads), the energy difference between the molecular energy levels and the Fermi energy of the leads, geometry changes resulting from charges entering or leaving the molecule, and the polarizability of the molecular wire. In the operation of the molecular wire device, at least two regimes can be distinguished [280]. If the energy of the tunneling particle coincides with a resonance of the wire, electrode system, the molecule may behave like an ideal macroscopic wire, and the current is independent of the distance between electrodes (neglecting the usual resistivity contributions). In the other regime (at low bias), the particle has to travel through a barrier and the current decreases exponentially with electrode distance.

$$G = G^{\circ} e^{-\gamma L}$$

where G is the elastic conductance, G° is a prefactor that depends on the electronic interactions between the molecular wire and the contacts, γ is a damping factor that depends on the electronic structure of the molecular wire, and L is the length of the molecular wire. The above equation holds for tunneling currents when the bias voltage is lower than the HOMO-LUMO gap. The current through such tunnels is low [38], but the tunnel path is more transparent to electrons than vacuum [554].

Because of the complexity of the problem, approximations have to be made. If the one-electron approximation is used and geometry relaxation is neglected, the current can be described as a scattering process during which the electron that resided initially at one electrode is detected at the other electrode at a different time [280]. The stationary current is the total electron transfer rate times the electronic charge. The current is as highly nonlinear function of the applied voltage. I(V) curves can be calculated with the Landauer formula, which relates the conductance to the electron transmission probability [38, 280, 551, 552]:

$$G = \frac{2e^2}{h} T(E_f)$$

The quantum unit conductance $2e^2/h$ equals about 77 μS [562]. The transmission function $T(E_f)$ can be obtained with Green function techniques [280, 551, 552] or by solving the Schrödinger equation directly for the scattered wave function using the one-electron tight-binding [38] or Hubbard [550] Hamiltonians. Different approaches to calculate transmission functions have been reviewed and compared [556]. The efficiency of a tunnel path is measured in units of transparency. Because transparency controls the intensity of the measured current of the macroscopic circuit outside the junction, this is called electronic conductance G of the molecules, although the major part of the measured dissipation is a result of the macroscopic part of the junction [554]. A major factor in the conductance is the energy of the leads relative to those of the molecule in the junction. Therefore, the I(V) curves show a staircase behavior, each step being associated with the opening of a resonant channel as the bias shifts the energy levels of leads and molecule with respect to one another [280]. If resonance is encountered, the enhancement in conductivity is dramatic [98].

If the transmission function is calculated from the electronic structure of the molecule in the ground state, geometry relaxations that are expected to be important in organic molecules on charging are neglected. The influence of vibrations was examined by calculating the current for different geometries [553]. Geometry distortions that influence the electronic structure have a strong influence on the current. Localized geometry changes on charging, such as polarons, can be explicitly examined with the Marcus theory [563, 571, 576–578].

The influence of the contact geometry was investigated for the p-benzene dithiolate molecule between gold contacts, treating the molecule cluster complex as a single entity in an extended Hückel tight-binding scheme [38]. For a weakly coupled system, the transmission was shown to be dominated by the energy levels of the free molecule. For strong coupling, the energy levels of the molecule are distorted and the transmission reflects this. Transmission was shown to be sensitive to geometry and bonding. The more cluster states interact with the molecule, the greater the transmission. However, calculated conductivities overestimate the experimental values.

The superposition law was examined for tunnel junctions operating in the tunneling transport regime [554]. Superposition law means that the conductance of n molecules is n times the conductance of a single molecule, just like in macroscopic devices. If both ends of the molecules are absorbed independently on the electrodes, as a first approximation, the total junction conductance is the summation of all the conductances in the junction. A correction can be made considering the coupling of the molecules through the surface of the electrodes. If the molecules are connected in the junction, the superposition law becomes quadratic.

A more realistic description of conductance was achieved using the Landauer formula but describing the molecule and the contacts with the HF method [98]. This approach accounts for polarization of the molecule and for the charge transfer caused by contact between molecule and leads. Because charge transfer changes the molecular energy levels, it has a major influence on the conductance. The main conclusion of this investigation is that the Fermi level of the leads does not necessarily occur in the middle of the HOMO-LUMO gap of the molecule, as previously assumed. The injection level is the work function of the metal and lies, for p-benzene dithiol, close to the HOMO level of the molecule. Because the HF method was used, the energy of the HOMO level is reliable, and hole conductivity is well described. In contrast, unoccupied orbitals lie at unrealistically high energies, and electron conductivity cannot be examined.

The opposite problem exists with DFT applying LSDA, as the HOMO-LUMO gap is underestimated. A p-benzenedithiolate molecule was placed between two jellium electrodes in the presence of an external field [555]. The electron wave-functions were computed by solving the Lippman-Schwinger equation (which is equivalent to the Landauer formula [568]). Exchange and correlation were included because LSDA was employed. Geometry relaxation was not accounted for. The differential conductance was computed as the derivative of the current with respect to external bias. The shape of the I(V) curve was the same as the experimental one, but the current was found to be two orders of magnitude too high. The insertion of a gold atom at each metal molecule contact lowered the current by one order of magnitude; replacing the gold atoms with aluminum atoms raises the current by one order of magnitude. In a comment on this work [560] it was suggested that the experimental junction might actually consist of two individual molecules absorbed on each side of the junction, rather than one single molecule attached with both ends to the two leads. Theoretical treatment of such a model explained the experimental results very well [50]. However, the overestimation of the conductance with LSDA may very well be a result of the known problem of DFT with underestimation of band gaps and overestimation of delocalization.

The transmission function was calculated employing Green functions based on the B3PW91 hybrid density functional to compute the conductance of p-benzene dithiol monomer [561], and for monomer through hexamer [559]. In contrast to HF and LSDA, the B3PW91 functional should reproduce roughly the correct energy gap—one requirement for a reliable description of the conduction process. However, the problems detected with DFT for polarizabilities might influence the calculation of conductivities of extended systems. In contrast to what might be expected, HF overestimates conductance compared to DFT by one to two orders of magnitude [561]. This could be because DFT shifts HOMO

and LUMO energies of π-systems compared to HF and experiment by about 1 eV [399, 419]. Therefore, a resonance that is present at the HF level may vanish with DFT and vice versa. A strong correlation was found between the density of states and the transmission function [559]. Delocalized orbitals were shown to allow for better transmission than localized ones. Electronic levels shift nonuniformly through contact with the leads and may switch energetic order [559, 562]. Different MOs of the isolated molecule may combine on interaction with the contacts and open a conduction channel with high transmission function [559, 569]. Conduction was shown to be largest when the molecule was connected to a single gold atom [559].

10. SUMMARY

There are two approaches to deal with extended systems. One is based on using the translational symmetry of infinite polymers. Polymer properties are calculated for repeat units using periodic boundary conditions. It was assumed for a while that this is the only proper treatment of polymers. However, computer power has improved, and algorithms became more efficient, allowing calculations on oligomers with a couple hundred atoms. As experiments showed that polymer properties are quite similar to those of relatively short oligomers, the oligomer approach gained acceptance. Theoretical data presented in Section 3 for thiophene oligomers show that starting with oligomers with 14 rings, the excitation energies are only 0.2 eV higher than at the saturation limit. Beyond 21 rings, the difference is 0.1 eV. These numbers were obtained in the absence of disorder. Thus, the oligomer approach is reliable, employing moderately sized oligomers.

COPs can be used as bulk materials, as nanofibers, or as individual molecules. In all cases, the properties are dominated by those of the polymer chains, but the influence of disorder, medium, and interchain interactions is very different. Bulk materials are usually modeled as idealized cases with three-dimensional band structure calculations, neglecting disorder. Individual polymer chains can be examined more easily and realistically with oligomer calculations.

Theoretical levels may be divided into three categories: semiempirical, *ab initio*, and DFT methods. All three approaches are used extensively in oligomer and polymer calculations. Basis set evaluations showed that the size of basis set becomes less crucial the larger the systems are. Polarized valence double zeta basis sets are adequate for almost any application. Semiempirical calculations are usually successful for the property for which they are parametrized. HF calculations are a substantial improvement over semi-empirical methods, especially over approaches in which Coulomb repulsions are not treated self-consistently. DFT, which is extremely successful in molecular calculations, has to be handled with care when extended π-systems are involved, as currently available exchange-correlation functionals have difficulties with describing long-range interactions. This causes problems with excitation energies, with the extent of solitons and polarons in the absence of counter ions, and with polarizabilities. It is expected that the calculation of conductivities suffers from this deficiency as well.

The relationship between band gaps and excitation energies was evaluated for polyenes in the gas phase, and the performance of theoretical methods was analyzed. It turned out that even the highest-level *ab initio* methods have difficulties with obtaining accuracies better than ± 0.2 eV. They are successful, however, in determining the correct order of the excited states. If only the band gap is required, it is sufficient to calculate the energy of the lowest allowed excited state accurately. For this purpose, TDHF turns out to be more accurate than higher-level approaches. The advantage with this method is that no configuration space has to be selected and that small and large systems can be treated on equal footing. The geometries have to be optimized at a level of theory that produces accurate BLA, as there is a large influence of BLA on the band gap. TDDFT is, in general, superior to TDHF for the calculation of spectra but has difficulties with the lowest allowed excitation of conjugated π-systems, especially for long chain lengths. TDDFT is therefore not reliable for obtaining polymer band gaps.

Because the lowest allowed optical transition in COPs is dominated by a single-electron configuration that is characterized by excitation of one electron from HOMO to

LUMO, HOMO-LUMO gaps can be used as an approximation of the excitation energies. HOMO-LUMO gaps for oligomers are the equivalent of band gaps of polymers. To produce reliable band gaps, orbital energies have to have reasonable values. Hückel theory was parametrized to produce the band gap of PA and is also successful for other π-systems. At the HF level, the negative energy of the HOMO gives a quite accurate estimate of the IP. HF theory overestimates band gaps substantially because the virtual orbitals lie very high. VEH was parametrized to reproduce HF results. Interestingly, it produces good band gaps and bad IPs. Pure DFT methods suffer from underestimating band gaps and do not give reliable estimates for IPs. Hybrid functionals describe HOMO-LUMO energy differences correctly, whereas IPs and EAs are shifted relative to experiment [399, 419]. Moreover, calculations on oligomers show that the energy gaps decrease too fast compared to experiment.

Medium effects have pronounced effects on band gaps. Solvents lower excitation energies, and interchain interactions cause splitting of the excitons into lower and higher lying components. For thiophene oligomers the combined effect is a lowering of the lowest excitation energy of about 0.6 eV compared to the gas phase value. The lower excitation is weakly allowed and corresponds most likely to the onset of absorption of the polymer. The higher-lying component is strongly allowed and gives rise to λ_{max}. Because medium effects lower excitation energies and splitting caused by interchain interactions raises the energy of the higher component, partial cancellation occurs and λ_{max} of polymers lies relatively close ($\pm$0.2 eV) to the gas phase value. This explains why band structure calculation on single chains appears successful for obtaining bulk polymer band gaps.

There is experimental evidence that charged defects in COPs are localized in the presence of counter ions. Crude semiempirical methods reproduced these findings even with gas phase calculations in the absence of counter ions. Higher levels of theory produce defects that are much more delocalized and show no evidence for the existence of bipolarons on long chains in the absence of counter ions. HF theory still overestimates localization compared to MP2, and DFT underestimates it. The energy differences between the structures obtained at different levels of theory are so small that the final conformations are probably determined by the position and the nature of the counter ions.

Polarizabilities of oligomers can be calculated with most quantum chemical program packages. The convergence with increasing chain length is slower than for band gaps. Therefore, several approaches were developed to implement methods to calculate polarizabilities of polymers. This turned out to be difficult because the application of the potential that represents the electric field destroys the translational symmetry of the polymer. The difficulty has finally been overcome, and codes exist that employ coupled perturbed HF theory and TDHF. DFT was tested as well but shown to be unsuitable. Because of the problems with long-range interactions, DFT substantially overestimates polarizabilities and hyperpolarizabilities.

Finally, conductivity through mesoscopic devices is receiving increasing experimental and theoretical attention. Theoretical treatment of conductivities is influenced by practically all of the issues discussed in the preceding chapters. Single molecules are connected to contacts, and a voltage is applied. At low bias, when the Fermi level of the contacts lies in the band gap, charge transport occurs via tunneling, and currents are in the pA range. As the bias is increased, energy levels move with respect to each other, and resonant channels open with a dramatic increase in conductance. A theoretical description requires the exact knowledge of the energy levels and band widths of the oligomer in contact with the leads. Because chemical potentials of different systems tend to equilibrate on contact, electron densities change. Thus, charging occurs with its associated geometrical rearrangements. This situation involves charging in the absence of counter ions. The above description of charged defects shows that the theoretical level has to be chosen carefully. Applying a bias polarizes the molecule. Therefore, polarizability and hyperpolarizabilities should be accounted for. Current implementations use Hückel and Hubbard Hamiltonians, HF and DFT. The shapes of the I(V) curves are usually reproduced well, but currents tend to be overestimated by several orders of magnitude.

In conclusion, there has been intensive theoretical research on describing all aspects of COPs. Theory has contributed a great deal toward understanding the physical processes that govern electrical and optical properties of COPs. Reaching quantitative agreement between

theory and experiment is still not easy. On the one hand, experimental data are influenced by a number of factors that are hard to control. On the other hand, theoretical methods are approximate and biased either toward too much localization (as semiempirical methods and HF theory) or toward too much delocalization (as DFT). Nonetheless, with a clever choice of methods, important insights and reliable comparisons between different systems can be achieved.

ACKNOWLEDGMENTS

The author thanks TÜBITAK, Bilkent University, and the Computer Center at Bilkent for providing computational resources. The help of Ozan Karaltı with the TDHF calculations of thiophene oligomers is acknowledged.

REFERENCES

1. T. A. Skotheim, (Ed.), "Handbook of Conducting Polymers," Marcel Dekker, New York, 1986.
2. T. A. Skotheim, R. L. Elsenbaumer, and J. R. Reynolds, (Eds.), "Handbook of Conducting Polymers," Marcel Dekker, New York, 1997.
3. A. O. Patil, A. J. Heeger, and F. Wudl, *Chem. Rev.* 88, 183 (1988).
4. C. K. Chiang, C. R. Fincher, Y. W. Park, A. J. Heeger, H. Shirakawa, E. J. Louis, S. C. Gau, and A. G. MacDiarmid, *Phys. Rev. Lett.* 39, 1098 (1977).
5. C. H. Naarmann and N. Theophilou, *Synth. Met.* 22, 1 (1987).
6. S. Kivelson and A. J. Heeger, *Synth. Met.* 22, 371 (1988).
7. S. Y. Lee and B. H. Boo, *J. Phys. Chem.* 100, 15073 (1996).
8. A. J. Heeger, *Angew. Chem. Int. Ed. Engl.* 40, 2591 (2001).
9. A. G. MacDiarmid, *Angew. Chem. Int. Ed. Engl.* 40, 2581 (2001).
10. H. Shirakawa, *Angew. Chem. Int. Ed. Engl.* 40, 2575 (2001).
11. J. Roncali, *Chem. Rev.* 97, 173 (1997).
12. J. Roncali, *Chem. Rev.* 92, 711 (1992).
13. Q. Pei, G. Zuccarello, M. Ahlskog, and O. Inganäs, *Polymer* 35, 1347 (1994).
14. A. Aleshin, R. Kiebooms, R. Menon, and A. J. Heeger, *Synth. Met.* 90, 61 (1997).
15. M. C. Morvant and J. R. Reynolds, *Synth. Met.* 92, 57 (1998).
16. L. Groenendaal, F. Lonas, D. Freitag, H. Pielartzik, and J. R. Reynolds, *Adv. Mater.* 12, 481 (2000).
17. G. Liang and T. Cui, *Solid-State Elec.* 48, 87 (2004).
18. F. Wudl, M. Kobayashi, and A. J. Heeger, *J. Org. Chem.* 49, 3382 (1984).
19. M. Kobayashi, N. Colaneri, M. Boysel, F. Wudl, and A. Heeger, *J. Chem. Phys.* 82, 5717 (1985).
20. J. G. Park, B. Kim, S. H. Lee, A. B. Kaiser, S. Roth, and Y. M. Park, *Synth. Met.* 135–136, 299 (2003).
21. Z. Cai and C. R. Martin, *J. Am. Chem. Soc.* 111, 4138 (1989).
22. W. Liang and C. R. Martin, *J. Am. Chem. Soc.* 112, 9666 (1989).
23. L. S. Van Dyke and C. R. Martin, *Langmuir* 6, 1118 (1990).
24. J.-M. Liu and S. C. Yang, *J. Chem. Soc. Chem. Commun.* 1529 (1991).
25. C. R. Martin, R. Parthasarathy, and V. Menon, *Synth. Met.* 55, 1165 (1993).
26. J. Mansouri and R. P. Burford, *J. Membr. Sci* 87, 23 (1994).
27. J. C. Michaelson and A. J. McEvoy, *J. Chem. Soc. Chem. Commun.* 79 (1994).
28. R. V. Parthasarathy and C. R. Martin, *Chem. Mater.* 6, 1627 (1994).
29. M. Dorogi, J. Gomez, R. Osifchin, R. P. Andres, and R. Reifenberger, *Phys. Rev. B* 52, 9071 (1995).
30. M. Granström, J. C. Carlberg, and O. Inganäs, *Polymer* 36, 3191 (1995).
31. M. Granström and O. Inganäs, *Polymer* 36, 2867 (1995).
32. C. R. Martin, *Chem. Mater.* 8, 1739 (1996).
33. D. H. Reneker and I. Chun, *Nanotechnology* 7, 216 (1996).
34. J. M. Tour, *Chem. Rev.* 96, 537 (1996).
35. J. C. Hulteen and C. R. Martin, *J. Mater. Chem.* 7, 1075 (1997).
36. S. Demoustier-Champagne, E. Ferain, C. Jérôme, and R. Legras, *Eur. Polym. J.* 34, 1767 (1998).
37. J. Duchet, R. Legras, and S. Demoustier-Champagne, *Synth. Met.* 98, 113 (1998).
38. E. G. Emberly and G. Kirczenow, *Phys. Rev. B* 58, 10911 (1998).
39. L. Jestin, P. Frère, N. Mercier, E. Levillain, D. Stievenard, and J. Roncali, *J. Am. Chem. Soc.* 120, 8150 (1998).
40. M. Kryszewski and J. K. Jeszka, *Synth. Met.* 94, 99 (1998).
41. Y. Li and D. Haarer, *Synth. Met.* 94, 273 (1998).
42. S. Demoustier-Champagne and P.-L. Stavaux, *Chem. Mater.* 11, 829 (1999).
43. J. Planees, Y. Cheguettine, and Y. Samson, *Synth. Met.* 101, 789 (1999).
44. A. J. G. Zarbin, M.-A. De Paoli, and O. L. Alves, *Synth. Met.* 99, 227 (1999).
45. M. Delvaux, J. Duchet, P.-L. Stavaux, R. Legras, and S. Demoustier-Champagne, *Synth. Met.* 113, 275 (2000).
46. C. Gadermaier, E. J. W. List, P. Markart, W. Graupner, J. Partee, J. Shinar, R. Smith, D. Gin, and G. Leising, *Synth. Met.* 111–112, 523 (2000).

47. J. Roncali, *Acc. Chem. Res.* 33, 147 (2000).
48. Z. Tang, S. Liu, Z. Wang, S. Dong, and E. Wang, *Electrochem. Commun.* 2, 32 (2000).
49. S.-T. Yau, C. Zhang, and P. C. Innis, *J. Chem. Phys.* 112, 6774 (2000).
50. E. G. Emberly and G. Kirczenow, *Phys. Rev. B* 64, 23541 (2001).
51. T. Hassenkamp, D. R. Greve, and T. Bjornholm, *Adv. Mater.* 13, 631 (2001).
52. H. X. He, C. Z. Li, and N. J. Tao, *Appl. Phys. Lett.* 78, 811 (2001).
53. A. B. Kaiser, G. U. Flanagan, D. M. Stewart, and D. Beaglehole, *Synth. Met.* 117 (2001).
54. J. J. Langer, G. Framski, and R. Joachimiak, *Synth. Met.* 121, 1281 (2001).
55. A. G. MacDiarmid, W. E. Jones, Jr., I. D. Norris, J. Gao, and A. T. Johnson, Jr., N. J. Pinto, J. Hone, B. Han, F. K. Ko, H. Okuzaki, and M. Llaguno, *Synth. Met.* 119, 27 (2001).
56. J. G. Park, G. T. Kim, V. Krstic, B. Kim, S. H. Lee, S. Roth, M. Burghard, and Y. W. Park, *Synth. Met.* 119, 53 (2001).
57. T. Piok, C. Brands, P. J. Neyman, A. Erlacher, C. Soman, M. A. Murray, R. Schroeder, W. Graupner, J. R. Heflin, D. Marciu, A. Drake, M. B. Miller, H. Wang, H. Gibson, C. Dorn, G. Leising, M. Guzy, and R. M. Davis, *Synth. Met.* 116, 343 (2001).
58. H. Qiu and M. Wan, *J. Pol. Sci. A* 39, 3485 (2001).
59. C. Wang, Z. Wang, M. Li, and H. Li, *Chem. Phys. Lett.* 341, 431 (2001).
60. G. Dongtao, J. Wang, Z. Wang, and W. Shichang, *Synth. Met.* 132, 93 (2002).
61. E. Frackowiak, K. Jurewicz, K. Szostak, S. Delpeux, and F. Béguin, *Fuel Process. Tech.* 77–78, 213 (2002).
62. L. Huang, Z. Wang, H. Wang, X. Cheng, A. Mitra, and Y. Yan, *J. Mater. Chem.* 12, 388 (2002).
63. A. B. Kaiser and Y. W. Park, *Current Appl. Phys.* 2, 33 (2002).
64. I. Kratochvilova, A. Kocirik, J. Mbindyo, T. E. Mallouk, and T. S. Mayer, *J. Mater. Chem.* 12, 2927 (2002).
65. J. M. Mativetsky and W. R. Datars, *Physica B* 324, 191 (2002).
66. Z. Wei, Z. Zhang, and M. Wan, *Langmuir* 18, 917 (2002).
67. Y. Yang and M. Wan, *J. Mater. Chem.* 12, 897 (2002).
68. D. Allard, S. Allard, M. Brehmer, L. Conrad, R. Zentel, C. Stromberg, and J. W. Schultze, *Electrochim. Acta* 48, 3137 (2003).
69. J. Cao, J. Sun, G. Shi, H. Chen, Q. Zhang, D. Wang, and M. Wang, *Mater. Chem. Phys.* 82, 44 (2003).
70. C. He, C. Yang, and Y. Li, *Synth. Met.* 19, 539 (2003).
71. J. Huang, S. Virji, B. H. Weiller, and R. B. Kaner, *J. Am. Chem. Soc.* 125, 314 (2003).
72. K. Huang and M. Wan, *Synth. Met.* 135–136, 173 (2003).
73. A. B. Kaiser and Y. W. Park, *Synth. Met.* 135–136, 245 (2003).
74. H. Kosonen, S. Valkama, J. Ruokolainen, M. Knaapila, M. Torkkeli, R. Serimaa, A. P. Monkman, G. Ten Brinke, and O. Ikkala, *Synth. Met.* 137, 881 (2003).
75. J. M. G. Laranjeira, H. J. Khoury, W. M. de Azevedo, E. A. de Vasconcelos, and E. F. J. da Silva, *Mater. Char.* 50, 127 (2003).
76. M. Mazur, M. Tagowska, B. Palys, and K. Jackowska, *Electrochem. Commun.* 5, 403 (2003).
77. G. C. McIntosh, D. Tománek, and Y. W. Park, *Synth. Met.* 135–136, 729 (2003).
78. S. Möller, G. Weiser, and C. Taliani, *Chem. Phys.* 295, 11 (2003).
79. J. G. Park, B. Kim, S. H. Lee, and Y. W. Park, *Thin Solid Films* 438–439, 118 (2003).
80. J. G. Park, B. Kim, S. H. Lee, A. B. Kaiser, S. Roth, and Y. W. Park, *Synth. Met.* 135–136, 299 (2003).
81. G. Weiser, *Synth. Met.* 139, 719 (2003).
82. M. J. Winokur and W. Chunwachirasiri, *J. Pol. Sci. B* 41, 2630 (2003).
83. E. Arici, H. Hoppe, F. Schäffler, D. Meissner, M. A. Malik, and N. S. Sariciftci, *Thin Solid Films* 451–452, 612 (2004).
84. C. L. Huisman, A. Huisjer, H. Donker, J. Schoonman, and A. Goossens, *Macromol.* 37, 5557 (2004).
85. C. Jérôme, D. E. Labaye, and R. Jérôme, *Synth. Met.* 142, 207 (2004).
86. P. Kumar, A. Mehta, S. M. Mahurin, S. Dai, M. D. Dadmun, B. G. Sumpter, and M. D. Barnes, *Macromol.* 37, 6132 (2004).
87. S. E. Moulton, P. C. Innis, L. A. P. Kane-Maguire, O. Ngamma, and G. G. Wallace, *Curr. Appl. Phys.* 4, 402 (2004).
88. B. Rajesh, K. Ravindranathan Thampi, J.-M. Bonard, N. Xanthopoulos, H. J. Mathieu, and B. Viswanathan, *J. Phys. Chem. B* 108, 10640 (2004).
89. X. Zhang, R. Chan-Yu-King, A. Jose, and S. K. Manohar, *Synth. Met.* 145, 23 (2004).
90. T. Ishida, W. Mizutani, Y. Aya, H. Ogiso, S. Sasaki, and H. Tokumoto, *J. Phys. Chem. B* 106, 5886 (2002).
91. D. J. Wold, R. Haag, M. A. Rampi, and C. D. Frisbie, *J. Phys. Chem. B* 106, 2813 (2002).
92. V. B. Engelkes, J. M. Beebe, and C. D. Frisbie, *J. Am. Chem. Soc.* 126, 14287 (2004).
93. S. B. Sachs, S. P. Dudek, R. P. Hsung, L. R. Sita, J. F. Smalley, M. D. Newton, S. W. Feldberg, and C. E. D Chidsey, *J. Am. Chem. Soc.* 119, 10563 (1997).
94. H. Nakanishi, N. Sumi, Y. Aso, and T. Otsubo, *J. Org. Chem.* 63, 8632 (1998).
95. S. Creager, C. J. Yu, C. Bamdad, S. O'Connor, T. MacLean, E. Lam, Y. Chong, G. T. Olsen, J. Luo, M. Gozin, and J. F. Kayyem, *J. Am. Chem. Soc.* 121, 1059 (1999).
96. H. Ishii, K. Sugiyama, E. Ito, and K. Seki, *Adv. Mater.* 11, 605 (1999).
97. G. Leatherman, E. N. Durantini, D. Gust, T. A. Moore, A. L. Moore, S. Stone, Z. Zhou, P. Rez, Y. Z. Liu, and S. M. Lindsay, *J. Phys. Chem. B* 103, 4006 (1999).
98. S. N. Yaliraki, A. E. Roitberg, C. Gonzales, V. Mujica, and M. A. Ratner, *J. Chem. Phys.* 111, 6997 (1999).
99. T. Ishida, W. Mizutani, N. Choi, U. Akiba, M. Fujihira, and H. Tokumoto, *J. Phys. Chem. B* 104 (2000).

100. H. B. Weber, J. Reichert, F. Weigend, R. Ochs, D. Beckmann, M. Mayor, R. Ahlrichs, and H. V. Löhneysen, *Chem. Phys.* 281, 113 (2002).
101. G. K. Ramachendran, J. K. Tomfohr, J. Li, O. F. Sankey, X. Zarate, A. Primak, Y. Terazono, T. A. Moore, A. L. Moore, D. Gust, T. Nagahara, and S. M. Lindsay, *J. Phys. Chem. B* 107, 6162 (2003).
102. F. Giacalone, J. L. Segura, N. Martín, and D. M. Guldi, *J. Am. Chem. Soc.* 126, 5340 (2004).
103. Y. Karzazi, X. Crispin, O. Kwon, J. L. Brédas, and J. Cornil, *Chem. Phys. Lett.* 387, 502 (2004).
104. F. Remacle, I. Willner, and R. D. Levine, *J. Phys. Chem. B* 108, 18129 (2004).
105. A. Szuchmacher Blum, J. G. Kushmerick, S. K. Pollack, J. C. Yang, M. Moore, J. Naciri, R. Shashidhar, and B. R. Ratna, *J. Phys. Chem. B* 108, 18124 (2004).
106. X.-Y. Zhu, *J. Phys. Chem. B* 108, 8778 (2004).
107. R. Baer and D. Neuhauser, *Int. J. Quantum Chem.* 91, 524 (2002).
108. T. Tani, W. D. Gill, P. M. Grant, T. C. Clarke, and G. B. Street, *Synth. Met.* 1, 310 (1980).
109. S. Etemad, T. Mitani, M. Ozaki, T. C. Chung, A. J. Heeger, and A. G. MacDiarmid, *Solid State Commun.* 40, 75 (1981).
110. E. M. Genies, P. Hany, M. Lapkowski, C. Santier, and L. Olmedo, *Synth. Met.* 25, 29 (1988).
111. R. P. McCall, M. G. Roe, J. M. Ginder, T. Kusumoto, A. J. Epstein, G. E. Asturias, E. M. Scherr, and A. G. MacDiarmid, *Synth. Met.* 29, 433 (1989).
112. S. D. Phillips, G. Yu, and A. J. Heeger, *Synth. Met.* 28, D669 (1989).
113. H. Antoniadis, B. R. Hsieh, M. A. Abkowitz, S. A. Jenekhe, and M. Stolka, *Synth. Met.* 62, 265 (1994).
114. B. Z. Tang, H. Z. Chen, R. S. Xu, J. W. Y. Lam, K. K. L. Cheuk, H. N. C. Wong, and M. Wang, *Chem. Mater.* 12, 213 (2000).
115. N. Loussaïef, L. Hassine, N. Boutabba, F. Kouki, P. Spearman, F. Garnier, and H. Bouchriha, *Synth. Met.* 128, 283 (2002).
116. G. Dicker, T. J. Savenije, B.-H. Huisman, D. M. de Leeuw, M. P. deHaas, and J. M. Warman, *Synth. Met.* 137, 863 (2003).
117. C. Im, W. Tian, H. Bässler, A. Fechtenkötter, M. D. Watson, and K. Müllen, *J. Chem. Phys.* 119, 3952 (2003).
118. M. Weiter, V. I. Arkhipov, and H. Bässler, *Synth. Met.* 141, 165 (2004).
119. J. Tsukamoto, H. Ohigashi, K. Matsumura, and A. Takahashi, *Synth. Met.* 4, 177 (1982).
120. B. R. Weinberger, M. Akhtar, and S. C. Gau, *Synth. Met.* 4, 187 (1982).
121. S. Glenis, G. Horowitz, G. Tourillon, and F. Garnier, *Thin Solid Films* 111, 93 (1984).
122. B. R. Weinberger, *Solid State Commun.* 51, 167 (1984).
123. G. Horowitz and F. Garnier, *Solar Energy Mater.* 13, 47 (1986).
124. A. Usuki, M. Murase, and T. Kurauchi, *Synth. Met.* 18, 705 (1987).
125. S.-A. Chen and Y. Fang, *Synth. Met.* 60, 215 (1993).
126. S. Karg, W. Riess, M. Meier, and M. Schwoerer, *Synth. Met.* 57, 4186 (1993).
127. M. S. Lee and K. C. Huang, *Solid-State Elec.* 37, 1899 (1994).
128. W. Rieß, S. Karg, V. Dyakonov, M. Meier, and M. Schwoerer, *J. Luminescence* 60–61, 902 (1994).
129. S. B. Lee, P. K. Khabibullaev, A. A. Zakhidov, S. Morita, and K. Yoshino, *Synth. Met.* 71, 2247 (1995).
130. S. Panero, B. Scrosati, M. Baret, B. Cecchini, and F. Masetti, *Solar Energy Mater. Solar Cells* 39, 239 (1995).
131. M. Granström, K. Petritsch, A. C. Arias, A. Lux, M. R. Andersson, and R. H. Friend, *Nature* 395, 257 (1998).
132. K. Murakosji, R. Kogure, Y. Wada, and S. Yanagida, *Solar Energy Mater. Solar Cells* 55, 113 (1998).
133. B. P. Jelle and G. Hagen, *Solar Energy Mater. Solar Cells* 58, 277 (1999).
134. L. Sicot, C. Fiorini, A. Lorin, J.-M. Nunzi, P. Raimond, and C. Sentein, *Synth. Met.* 102, 991 (1999).
135. C. Videlot and D. Fichou, *Synth. Met.* 102, 885 (1999).
136. T. Fromherz, E. Padinger, D. Gebeyehu, C. Brabec, J. C. Hummelen, and N. S. Sariciftei, *Solar Energy Mater. Solar Cells* 63, 61 (2000).
137. M. F. Dustock, B. Taylor, R. J. Spry, L. Chiang, S. Reulbach, K. Heitfeld, and J. W. Baur, *Synth. Met.* 116, 373 (2001).
138. K. Takahashi, T. Iwanaga, T. Yamaguchi, T. Komura, and K. Murata, *Synth. Met.* 123, 91 (2001).
139. C. O. Too, G. G. Wallace, A. K. Burrel, G. E. Collis, D. L. Officer, E. W. Boge, S. G. Brodie, and E. J. Evans, *Synth. Met.* 123, 53 (2001).
140. C. A. Cutler, A. K. Burrel, D. L. Officer, C. O. Too, and G. G. Wallace, *Synth. Met.* 128, 35 (2002).
141. L. Goris, M. A. Loi, A. Cravino, H. Neugebauer, N. S. Sariciftei, J. Polec, L. Lutsen, F. Andries, J. Manca, L. De Schepper, and D. Vanderzande, *Synth. Met.* 138, 249 (2003).
142. G. D. Sharma, S. Sharma, and M. S. Roy, *Solar Energy Mater. Solar Cells* 80, 131 (2003).
143. S. X. Tan, J. Zhai, M. X. Wan, L. Jiang, and D. B. Zhu, *Synth. Met.* 137, 1511 (2003).
144. J. Pan, Y. Xu, L. Sun, V. Sundström, and T. Polívka, *J. Am. Chem. Soc.* 126, 3066 (2004).
145. T. Röder, H.-S. Kitzerow, J. C. Hummelen, *Synth. Met.* 141, 271 (2004).
146. H. Spanggaard and F. C. Krebs, *Solar Energy Mater. Solar Cells* 83, 125 (2004).
147. R. Valaski, R. Lessmann, L. S. Roman, I. A. Hummelgen, R. M. Q. Mello, and L. Micaroni, *Electrochem. Commun.* 6, 357 (2004).
148. J. H. Burroughs, D. D. C. Bradley, A. R. Brown, R. N. Marks, R. H. Friend, P. L. Burns, and A. B. Holmes, *Nature* 347, 539 (1990).
149. D. D. C. Bradley, A. R. Brown, P. L. Burn, J. H. Burroughes, R. H. Friend, A. B. Holmes, K. D. Mackay, and R. N. Marks, *Synth. Met.* 43, 3135 (1991).
150. D. Braun and A. J. Heeger, *J. Appl. Phys.* 58, 1982 (1991).
151. N. C. Greenham, S. C. Moratti, D. D. C. Bradley, R. H. Friend, and A. B. Holmes, *Nature* 365, 628 (1993).

152. R. N. Marks, D. D. C. Bradley, R. W. Jackson, P. L. Burn, and A. B. Holmes, *Synth. Met.* 57, 4128 (1993).
153. D. R. Baigent, R. N. Marks, N. C. Greenham, R. H. Friend, S. C. Moratti, and A. B. Holmes, *Synth. Met.* 71, 2177 (1995).
154. M. Berggren, O. Inganäs, G. Gustafsson, M. R. Andersson, T. Hjertberg, and O. Wennerström, *Synth. Met.* 71, 2185 (1995).
155. J. Cornil, D. Beljonne, D. A. dos Santos, Z. Shuai, and J. L. Brédas, *Synth. Met.* 78, 209 (1996).
156. J. Cornil, D. A. dos Santos, D. Beljonne, and J. L. Brédas, *Synth. Met.* 76, 101 (1996).
157. J. Gao, A. J. Heeger, J. Y. Lee, and C. Y. Kim, *Synth. Met.* 82, 221 (1996).
158. S. Karg, J. C. Scott, J. R. Salem, and M. Angelopoulos, *Synth. Met.* 80, 111 (1996).
159. H. L. Wang, A. G. MacDiarmid, Y. Z. Wang, D. D. Gebler, and A. J. Epstein, *Synth. Met.* 78, 33 (1996).
160. Y. Yang, Q. Pei, and A. J. Heeger, *Synth. Met.* 78, 263 (1996).
161. F. Cacialli, R. H. Friend, N. Haylett, R. Daik, W. J. Feast, D. A. dos Santos, and J. L. Brédas, *Synth. Met.* 84, 643 (1997).
162. F. Cacialli, G. R. Hayes, J. Grüner, R. T. Phillips, and R. H. Friend, *Synth. Met.* 84, 533 (1997).
163. J. C. Scott, S. A. Carter, S. Karg, and M. Angelopoulos, *Synth. Met.* 85, 1197 (1997).
164. T. Yamamoto, H. Suganuma, T. Maruyama, T. Inoue, Y. Muramatsu, M. Arai, D. Komarudin, N. Ooba, S. Tomaru, S. Sasaki, and K. Kubota, *Chem. Mater.* 9, 1217 (1997).
165. D. Chinn, M. DeLong, A. Fujii, S. Frolov, K. Yoshino, and Z. V. Vardeny, *Synth. Met.* 102, 924 (1999).
166. E. J. W. List, L. Holzer, S. Tasch, G. Leising, U. Scherf, K. Müllen, M. Catellani, and S. Luzzati, *Solid State Commun.* 109, 455 (1999).
167. H. Sirringhaus and R. H. Friend, *Synth. Met.* 102, 857 (1999).
168. Y. Z. Wang, R. G. Sun, F. Meghdadi, G. Leising, T. M. Swager, and A. J. Epstein, *Synth. Met.* 102, 889 (1999).
169. T. M. Brown, J. S. Kim, R. H. Friend, F. Cacialli, R. Daik, and W. J. Feast, *Synth. Met.* 111–112, 285 (2000).
170. A. Elschner, F. Bruder, H.-W. Heuer, F. Jonas, A. Karbach, S. Kirchmeyer, S. Thurm, and R. Wehrmann, *Synth. Met.* 111–112, 139 (2000).
171. S. B. Heidenhain, Y. Sakamoto, T. Suzuki, A. Miura, H. Fujikawa, T. Mori, S. Tokito, and Y. Taga, *J. Am. Chem. Soc.* 122, 10240 (2000).
172. T. Kobayashi, H. Yoneyama, and H. Tamura, *J. Electroanal. Chem.* 161, 419 (1984).
173. E. M. Genies, M. Lapkowski, C. Santier, and E. Vieil, *Synth. Met.* 18, 631 (1987).
174. J. C. Gustafsson-Carlberg, O. Inganäs, M. R. Andersson, C. Booth, A. Azens, and C. G. Granqvist, *Electrochim. Acta* 40, 2233 (1995).
175. P. Somani, A. B. Mandale, and S. Radhakrishnan, *Acta Mater.* 48, 2859 (2000).
176. P. Dyreklev, G. Gustafsson, O. Inganäs, and H. Stubb, *Synth. Met.* 57, 4093 (1993).
177. G. Horowitz, F. Deloffre, F. Garnier, R. Hajlaoui, M. Hmyene, and A. Yassar, *Synth. Met.* 54, 435 (1993).
178. W. A. Schoonveld, R. W. Stok, J. W. Weijtmans, J. Vrijmoeth, J. Wildeman, and T. M. Klapwijk, *Synth. Met.* 84, 583 (1997).
179. G. Horowitz, R. Hajlaoui, R. Bourguiga, and M. E. Hajlaoui, *Synth. Met.* 101, 401 (1999).
180. C. K. Chiang, *Solid State Ionics* 9–10, 445 (1983).
181. A. Corradini, M. Mastragostino, A. S. Panero, P. Prosperi, and B. Scrosati, *Synth. Met.* 18, 625 (1987).
182. G. Fumio, K. Abe, K. Ikabayashi, T. Yoshita, and H. Morimoto, *J. Power Sources* 20, 243 (1987).
183. A. G. MacDiarmid, L. S. Yang, W.-S. Huang, and B. D. Humphrey, *Synth. Met.* 18, 393 (1987).
184. T. Nagatomo, C. Ichikawa, and O. Omoto, *Synth. Met.* 18, 649 (1987).
185. K. Naoi and T. Osaka, *J. Electroanal. Chem.* 217, 203 (1987).
186. K. Naoi, H. Sakai, S. Ogano, and T. Osaka, *J. Power Sources* 20, 237 (1987).
187. T. Satoshi and T. Tanaka, *J. Power Sources* 20, 249 (1987).
188. L. W. Shacklette, M. Maxfield, S. Gould, J. F. Wolf, T. R. Jow, and R. H. Baughman, *Synth. Met.* 18, 611 (1987).
189. L. Gi, L. Changzhi, and W. Fosong, *J. Power Sources* 24, 115 (1988).
190. D. Naegele and R. Bitthin, *Solid State Ionics* 28–30, 983 (1988).
191. E. M. Genies, P. Hany, and C. Santier, *Synth. Met.* 28, 647 (1989).
192. M. Mizumoto, M. Namba, H. Miyadera, M. Koseki, and Y. Kobayashi, *Synth. Met.* 28, 639 (1989).
193. T. Nakajima and T. Kawagoe, *Synth. Met.* 28, 629 (1989).
194. L. Changzhi, P. Xinsheng, Z. Borong, and W. Baochen, *J. Power Sources* 39, 255 (1992).
195. M. Shaolin, Y. Jinhai, and W. Yuhua, *J. Power Sources* 45, 153 (1993).
196. K. S. Hwang, J. S. Kim, and M. J. Kong, *Synth. Met.* 71, 2201 (1995).
197. P. Novák, K. Müller, K. S. V. Santhanam, and O. Haas, *Chem. Rev.* 97, 207 (1997).
198. K. Osawa, K. Oshida, T. Nakazawa, T. Nishizawa, T. Ikeda, M. Narita, N. Saito, T. Ito, M. Endo, and S. Bonnamy, *J. Phys. Chem. Solids* 65, 253 (2004).
199. S.-C. Huang, S.-M. Huang, H. Ng, and R. B. Kaner, *Synth. Met.* 57, 4047 (1993).
200. A. Rudge, I. Raistrick, S. Gottesfeld, and J. P. Ferraris, *Electrochim. Acta* 39, 273 (1994).
201. C. Arbizzani, M. Catellani, M. Mastragostino, and C. Mingazzini, *Electrochim. Acta* 40, 1871 (1995).
202. A. Laforgue, P. Simon, C. Sarrazin, and J.-F. Fauvarque, *J. Power Sources* 80, 142 (1999).
203. A. G. Ounnov, V. Z. Mordkovich, and Y. Takeuchi, *Synth. Met.* 121, 1581 (2001).
204. Y. Ikariyama, C. Galiatsatos, W. R. Heineman, and S. Yamauchi, *Sensors Actuators* 12, 455 (1987).
205. H. Shinohara, T. Chiba, and M. Aizawa, *Sensors Actuators* 13, 79 (1988).
206. P. N. Bartlett and S. K. Ling-Chung, *Sensors Actuators* 20, 287 (1989).
207. P. N. Bartlett and S. K. Ling-Chung, *Sensors Actuators* 19, 141 (1989).

208. P. N. Bartlett, P. B. M. Archer, and S. K. Ling-Chung, *Sensors Actuators* 19, 125 (1989).
209. A. Boyle, E. M. Genies, and M. Lapkowski, *Synth. Met.* 28, 769 (1989).
210. V. I. Krinichnyi, O. N. Yeremenko, G. G. Rukhman, Y. A. Letuchii, and V. M. Geskin, *Polym. Sci. USSR* 31, 1819 (1989).
211. G. Fortier, E. Brassard, and D. Bélanger, *Biosensors Bioelectronics* 5, 473 (1990).
212. W. Schuhmann, R. Lammert, B. Uhe, and H.-L. Schmidt, *Sensors Actuators B: Chem.* 1, 537 (1990).
213. R. John, M. Spencer, G. G. Wallace, and M. R. Smyth, *Anal. Chim. Acta* 249, 381 (1991).
214. V. I. Krinichnyi, O. N. Eremenko, G. G. Rukhman, V. M. Geskin, and Y. A. Letuchy, *Synth. Met.* 41, 1137 (1991).
215. I. A. Vinokurov, *Sensors Actuators B: Chem.* 10, 31 (1992).
216. A. Q. Contractor, T. N. Sureshkumar, R. Narayanan, S. Sukeerthi, R. Lal, and R. S. Srinivasa, *Electrochim. Acta* 39, 1321 (1994).
217. H. Korri Youssoufi, A. Yassar, S. Baitoche, M. Hmyene, and F. Garnier, *Synth. Met.* 67, 251 (1994).
218. N. E. Agbor, M. C. Petty, and A. P. Monkman, *Sensors Actuators B: Chem.* 28, 173 (1995).
219. K. Ramanathan, M. K. Ram, B. D. Malhotra, A. Murthy, and S. N., *Mat. Sci. Eng. C* 3, 159 (1995).
220. A. L. Kukla, Y. M. Shirshov, and S. A. Piletsky, *Sensors Actuators B: Chem.* 37, 135 (1996).
221. S. B. Adeljou, S. J. Shaw, and G. G. Wallace, *Anal. Chim. Acta* 341, 155 (1997).
222. J. Feng and A. G. MacDiarmid, *Synth. Met.* 102, 1304 (1999).
223. B. Strehlitz, B. Grundig, and H. Kopinke, *Anal. Chim. Acta* 403, 11 (2000).
224. J. E. G. de Souza, F. L. dos Santos, B. B. Neto, C. G. dos Santos, M. V. C. dos Santos, and C. P. de Melo, *Sensors Actuators B: Chem.* 88, 246 (2003).
225. M. K. Andrews, M. L. Jansen, G. M. Spinks, D. Zhou, and G. G. Wallace, *Sensors Actuators A: Phys.* 114, 65 (2004).
226. M. A. Díaz-García, F. Hide, B. J. Schwartz, M. R. Andersson, A. J. Heeger, and Q. Pei, *Synth. Met.* 84, 455 (1997).
227. G. E. Collins and L. J. Buckley, *Synth. Met.* 78, 93 (1996).
228. D. De Rossi, A. Della Santa, and A. Mazzoldi, *Mater. Sci. Eng.* 7, 31 (1999).
229. J. L. Brédas, (Ed.), "Conjugated Oligomers, Polymers, and Dendrimers: From Polyacetylene to DNA," Proceedings of the Fourth Francqui Colloquium, Brussels, 1998, De Boeck & Larcier, Brussels, 1999.
230. N. S. Sariciftci, (Ed.), "Primary Photoexcitations in Conjugated Polymers: Molecular Exciton versus Semiconductor Band Model," World Scientific, Singapore, 1997.
231. P. Puschnig and C. Ambrosch-Draxl, *Synth. Met.* 135–136, 415 (2003).
232. A. Kumar and J. R. Reynolds, *Macromolecules* 29, 7629 (1996).
233. G. A. Sotzing, J. R. Reynolds, and P. J. Steel, *Chem. Mater.* 8, 882 (1996).
234. B. Sankaran and J. R. Reynolds, *Macromolecules* 30, 2582 (1997).
235. B. C. Thompson, P. Schottland, K. Zong, and J. R. Reynolds, *Chem. Mater.* 12, 1563 (2000).
236. T. Kobayashi, H. Yoneyama, and H. Tamura, *J. Electroanal. Chem.* 177, 281 (1984).
237. H. Yoneyama, K. Wakamoto, and H. Tamura, *Mater. Chem. Phys.* 15, 567 (1986).
238. K. Hyodo and M. Omae, *J. Electroanal. Chem.* 292, 93 (1990).
239. B. P. Jelle, G. Hagen, and R. Ødegård, *Electrochim. Acta* 37, 1377 (1992).
240. B. P. Jelle, G. Hagen, S. Sunde, and R. Ødegård, *Synth. Met.* 54, 315 (1993).
241. M.-A. De Paoli, G. Casalbore-Miceli, E. M. Girotto, and W. A. Gazotti, *Electrochim. Acta* 44, 2983 (1999).
242. J. Yano and S. Yamasaki, *Synth. Met.* 102, 1157 (1999).
243. A. Bessiere, C. Duhamel, J.-C. Badot, V. Lucas, and M.-C. Certiat, *Electrochem. Commun.* 49, 2051 (2004).
244. N. M. Rateliffe, *Synth. Met.* 38, 87 (1990).
245. Y. Cao, G. M. Treacy, P. Smith, and A. J. Heeger, *Synth. Met.* 57, 3526 (1993).
246. W.-P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* 42, 1698 (1979).
247. W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. B* 22, 2099 (1980).
248. W.-P. Su, *Solid State Commun.* 35, 899 (1980).
249. J. L. Brédas and R. R. Chance, R. Silbey, *J. Phys. Chem.* 85, 756 (1981).
250. J. L. Brédas, R. R. Chance, and R. Silbey, *Phys. Rev. B* 26, 5843 (1982).
251. J. L. Brédas, J. C. Scott, K. Yakushi, and G. B. Street, *Phys. Rev. B* 30, 1023 (1984).
252. J. L. Brédas, B. Thémans, J.-M. André, R. R. Chance, and R. Silbey, *Synth. Met.* 9, 265 (1984).
253. J. L. Brédas and G. B. Street, *Acc. Chem. Res.* 18, 309 (1985).
254. J. L. Brédas, F. Wudl, and A. J. Heeger, *Solid State Commun.* 63, 577 (1987).
255. A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W.-P. Su, *Rev. Mod. Phys.* 60, 781 (1988).
256. J.-A. André, J. Delhalle, and J.-L. Brédas, "Quantum Chemistry Aided Design of Organic Polymers." An Introduction to the Quantum Chemistry of Polymers and its Applications, World Scientific, London, 1991.
257. J.-M. André and B. Champagne, in "Conjugated Oligomers, Polymers, and Dendrimers: From Polyacetylene to DNA," De Boeck-Université, Brussels, 1998.
258. J.-L. Brédas, in "Handbook of Conducting Polymers," New York, 1986.
259. J. Ladik "Quantum Theory of Polymers as Solids," New York, Plenum Press, 1988.
260. G. N. Lewis and M. Calvin, *Chem. Rev.* 39, 273 (1939).
261. P. Karrer and C. H. Eugster, *Helv. Chim. Acta* 34, 1805 (1951).
262. H. Kuhn and C. Kuhn, *Chem. Phys. Lett.* 204, 206 (1993).
263. H. Kuhn, *J. Chem. Phys.* 17, 1198 (1949).
264. R. E. Peierls, "Quantum Theory of Solids," Oxford University Press, Oxford, 1955.

265. U. Salzner, *Curr. Organic Chem.* 8, 569 (2004).
266. E. E. Havinga, W. Ten Hoeve, and H. Wynberg, *Synth. Met.* 55–57, 299 (1993).
267. H. Huang and P. G. Pickup, *Chem. Mater.* 10, 2212 (1998).
268. M. Karikomi, C. Kitamura, S. Tanaka, and Y. Yamashita, *J. Am. Chem. Soc.* 117, 6791 (1995).
269. C. Kitamura, S. Tanaka, and Y. Yamashita, *Chem. Mater.* 8, 570 (1996).
270. S. Tanaka and Y. Yamashita, *Synth. Met.* 55–57, 1251 (1993).
271. S. Tanaka and Y. Yamashita, *Synth. Met.* 84, 229 (1997).
272. H. A. M. van Mullekom, J. A. J. M. Vekemans, and E. W. Meijer, *Am. Chem. Soc. Div. Polym. Chem.* 39, 1002 (1998).
273. Q. T. Zhang and J. M. Tour, *J. Am. Chem. Soc.* 120, 5355 (1998).
274. T. A. Koopmans, *Physica* 1, 104 (1934).
275. P. W. Atkins and R. S. Friedman, "Molecular Quantum Mechanics," Oxford University Press, Oxford, 2000.
276. B. Champagne, D. H. Mosley, J. G. Fripiat, and J.-M. André, *Int. J. Quantum. Chem.* 46, 1 (1993).
277. A. B. Kaiser, *Adv. Mater.* 13, 927 (2001).
278. R. H. Baughman and L. W. Shacklette, *Synth. Met.* 17, 173 (1987).
279. V. N. Prigodin, A. N. Samukhin, and A. J. Epstein, *Synth. Met.* 141, 155 (2004).
280. V. Mujica, M. Kemp, A. Roitberg, and M. Ratner, *J. Chem. Phys.* 104, 7296 (1996).
281. P. Bäuerle, T. Fischer, B. Bidlingmeier, A. Stabel, and J. P. Rabe, *Angew. Chem. Int. Ed. Engl.* 34, 303 (1995).
282. I. Jestin, P. Frère, P. Blanchard, and J. Roncali, *Angew. Chem.* 110, 990 (1998).
283. A. Kiehl, A. Eberhardt, M. Adam, V. Enkelmann, and K. Müllen, *Angew. Chem. Int. Ed. Engl.* 31, 1588 (1992).
284. R. E. Martin and F. Diederich, *Angew. Chem. Int. Ed. Engl.* 38, 1351 (1999).
285. G. Horowitz, B. Bachet, A. Yassar, P. Lang, F. Demanze, J.-L. Fave, and F. Garnier, *Chem. Mater.* 7, 1337 (1995).
286. G. Bidan, A. De Nicola, V. Ennée, and S. Guillerez, *Chem. Mater.* 10, 1052 (1998).
287. R. Colditz, D. Grebner, M. Helbig, and S. Rentsch, *Chem. Phys.* 201, 309 (1995).
288. K. L. D'Amico, C. Manos, and R. L. Christensen, *J. Am. Chem. Soc.* 102, 1777 (1980).
289. E. H. Elandaloussi, P. Frére, P. Richomme, J. Orduna, and J. Garin, J. Roncali, *J. Am. Chem. Soc.* 119, 10774 (1997).
290. E. H. Elandaloussi, P. Frére, and J. Roncali, *Chem. Soc. Chem. Commun.* 301 (1997).
291. D. Fichou, M.-P. Teulade-Fichou, G. Horowitz, and F. Demanze, *Adv. Mat.* 9, 75 (1997).
292. F. Martinez, R. Voelkel, D. Naegele, and H. Naarmann, *Mol. Cryst. Liq. Cryst.* 167, 227 (1989).
293. R. Snyder, E. Arvidson, C. Foote, L. Harrigan, and R. L. Christensen, *J. Am. Chem. Soc.* 107, 4117 (1985).
294. A. Yang, S. Hughes, M. Kuroda, Y. Shiraishi, and T. Kobayashi, *Chem. Phys. Lett.* 280, 475 (1997).
295. A. Yassar, J. Roncali, and F. Garnier, *Macromolecules* 22, 804 (1989).
296. R. A. Auerbach, R. L. Christensen, M. F. Granville, and B. E. Kohler, *J. Chem. Phys.* 74, 4 (1981).
297. D. Birnbaum, D. Fichou, and B. E. Kohler, *J. Chem. Phys.* 96, 165 (1992).
298. D. Birnbaum, and B. E. Kohler, *J. Chem. Phys.* 90, 3506 (1989).
299. D. Birnbaum, and B. E. Kohler, *J. Chem. Phys.* 95, 4783 (1991).
300. D. Birnbaum, and B. E. Kohler, *J. Chem. Phys.* 96, 2492 (1992).
301. R. L. Christensen and B. E. Kohler, *J. Chem. Phys.* 63, 1837 (1975).
302. N. DiCésare, M. Belletête, M. Leclerc, and G. Durocher, *Chem. Phys. Lett.* 291, 487 (1998).
303. M. F. Granville, G. R. Holtom, and B. E. Kohler, *J. Chem. Phys.* 72, 4671 (1980).
304. B. E. Kohler, C. Spangler, and C. Westerfield, *J. Chem. Phys.* 89, 5422 (1988).
305. B. E. Kohler and T. A. Spiglanin, *J. Chem. Phys.* 80, 23 (1984).
306. B. E. Kohler and V. Terpougov, *J. Chem. Phys.* 108, 9586 (1998).
307. J. H. Simpson, L. McLaughlin, D. S. Smith, and R. L. Christensen, *J. Chem. Phys.* 87, 3360 (1987).
308. C. Taliani and L. M. Blinov, *Adv. Mater.* 8, 353 (1996).
309. K. Müllen and G. Wegener, (Eds.), "Electronic Materials: The Oligomer Approach," Wiley-VCH, Weinheim, 1997.
310. B. E. Kohler, *Chem. Rev.* 93, 41 (1993).
311. O. A. Mosher, W. M. Flicker, and A. Kuppermann, *J. Chem. Phys.* 59, 6502 (1973).
312. R. McDiarmid, *J. Chem. Phys.* 64, 514 (1976).
313. W. M. Flicker, O. A. Mosher, and A. Kuppermann, *Chem. Phys. Lett.* 45, 492 (1977).
314. J. Gavin, R. M. C. Weisman, J. K. McVey, and S. A. Rice, *J. Chem. Phys.* 68, 522 (1978)
315. L. A. Heimbrook, B. E. Kohler, and I. J. Levy, *J. Chem. Phys.* 81, 1592 (1984).
316. J. Horwitz, T. Itoh, B. E. Kohler, and C. Spangler, *J. Chem. Phys.* 87, 2433 (1987).
317. F. Rohlfing and D. D. C. Bradley, *Chem. Phys. Lett.* 277, 406 (1997).
318. T. A. Moore and P.-S. Song, *Chem. Phys. Lett.* 19, 128 (1973).
319. R. M. J. Gavin, S. Risemberg, and S. A. Rice, *J. Chem. Phys.* 58, 3160 (1973).
320. K. Mandal and T. N. Misra, *Chem. Phys. Lett.* 27, 57 (1974).
321. R. M. J. Gavin and S. A. Rice, *J. Chem. Phys.* 60, 3231 (1974).
322. M. A. C. Nascimento and W. A. Goddard III, *Chem. Phys.* 53, 265 (1980).
323. D. G. Leopold, R. D. Pendley, J. L. Roebber, R. J. Hemley, and V. Vaida, *J. Chem. Phys.* 81, 4218 (1984).
324. M. T. Allen and D. G. Whitten, *Chem. Rev.* 89, 1691 (1989).
325. R. McDiarmid, *Chem. Phys. Lett.* 188, 423 (1992).
326. B. E. Kohler and I. D. W. Samuel, *J. Chem. Phys.* 103, 6248 (1995).
327. R. P. Hosteny, T. H. Dunning, Jr., R. R. Gilman, and A. Pipano, *J. Chem. Phys.* 62, 4764 (1975).

328. R. J. Buenker, S. Shih, and S. D. Peyerimhoff, *Chem. Phys. Lett.* 44, 385 (1976).
329. L. E. McMurchie and E. R. Davidson, *J. Chem. Phys.* 66, 29592971 (1977).
330. D. R. Yarkony and R. Silbey, *Chem. Phys.* 20, 183 (1977).
331. B. R. Brooks and H. F. I. Schaefer, *J. Chem. Phys.* 68, 4839 (1978).
332. R. J. Buenker, S.-K. Shih, and S. D. Peyerimhoff, *Chem. Phys.* 36, 97 (1979).
333. M. A. C. Nascimento and W. A. Goddard III, *Chem. Phys.* 36, 147 (1979).
334. M. A. C. Nascimento and W. A. Goddard III, *Chem. Phys. Lett.* 60, 197 (1979).
335. C. Petrongolo, R. J. Buenker, and S. D. Peyerimhoff, *J. Chem. Phys.* 76, 3655 (1982).
336. M. Said, D. Maynau, and J.-P. Malrieu, *J. Am. Chem. Soc.* 106, 580 (1984).
337. R. J. Cave and E. R. Davidson, *J. Phys. Chem.* 91, 4481 (1987).
338. P. M. Lahti, J. Obrzut, and F. E. Karasz, *Macromol.* 20, 2023 (1987).
339. R. J. Cave and E. R. Davidson, *J. Phys. Chem.* 92, 2173 (1988).
340. R. J. Cave and E. R. Davidson, *J. Phys. Chem.* 92, 614 (1988).
341. R. J. Cave and E. R. Davidson, *Chem. Phys. Lett.* 148, 190 (1988).
342. O. Kitao and H. Nakatsuji, *Chem. Phys. Lett.* 143, 528 (1988).
343. R. J. Cave, *J. Chem. Phys.* 92, 2450 (1990).
344. R. L. Graham and K. F. Freed, *J. Chem. Phys.* 96, 1304 (1992).
345. L. Serrano-Andres, J. Sánchez-Marín, and I. Nebot-Gil, *J. Chem. Phys.* 97, 7499 (1992).
346. L. Serrano-Andrés, M. Merchán, I. Nebot-Gil, R. Lindh, and B. O. Roos, *J. Chem. Phys.* 98, 3151 (1993).
347. L. Serrano-Andrés, R. Lindh, B. O. Roos, and M. Merchán, *J. Phys. Chem.* 97, 9360 (1993).
348. Y. Luo, H. Ågren, and S. Stafström, *J. Chem. Phys.* 98, 7782 (1994).
349. D. Beljonne, E. Meyers, and J. L. Brédas, *Synth. Met.* 80, 211 (1996).
350. D. Beljonne, Z. Shuai, L. Serrano-Andrés, and J. L. Brédas, *Chem. Phys. Lett.* 279, 1 (1997).
351. K. Nakayama, H. Nakano, and K. Hirao, *Int. J. Quantum Chem.* 66, 157 (1998).
352. M. P. Pérez-Casany, I. Nebot-Gil, J. Sánchez-Marín, O. C. Marcos, and J.-P. Malrieu, *Chem. Phys. Lett.* 295, 181 (1998).
353. P. G. Szalay, A. Karpfen, and H. Lischka, *Chem. Phys. Lett.* 130, 219 (1998).
354. R. P. Krawczyk, K. Malsch, G. Hohlneicher, R. C. Gillen, and W. Domcke, *Chem. Phys. Lett.* 320, 535 (2000).
355. J. Lappe and R. J. Cave, *J. Phys. Chem. A* 104, 2294 (2000).
356. P. Cronstrand, O. Christiansen, P. Norman, and H. Ågren, *Phys. Chem. Chem. Phys.* 3, 2567 (2001).
357. C.-P. Hsu, S. Hirata, and M. Head-Gordon, *J. Phys. Chem. A* 105, 451 (2001).
358. F. Blomgren and S. Larsson, *Theor. Chem. Acc.* 110, 165 (2003).
359. J. Li, J. K. Tomfohr, and O. F. Sankey, *Physica E* 19, 133 (2003).
360. C. Nonnenberg, S. Grimm, and I. Frank, *J. Chem. Phys.* 119, 11585 (2003).
361. M. Boggio-Pasqua, M. J. Bearpark, M. Klene, and M. A. Robb, *J. Chem. Phys.* 120, 7849 (2004).
362. J. Catalán and J. L. G. de Paz, *J. Chem. Phys.* 120, 1864 (2004).
363. A. J. Merer and R. S. Mulliken, *Chem. Rev.* 69, 639 (1969).
364. A. M. Mebel, Y. T. Chen, and S.-H. Lin, *Chem. Phys. Lett.* 258, 53 (1996).
365. A. Szabo and N. S. Ostlund, "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory," Macmillan, New York, 1982.
366. S. Hirata, M. Head-Gordon, and R. J. Bartlett, *J. Chem. Phys.* 111, 10774 (1999).
367. Y. Luo, K. Ruud, P. Norman, D. Jonsson, and H. Ågren, *J. Phys. Chem. B* 102, 1710 (1998).
368. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
369. J. P. Perdew, *Phys. Rev. B* 33, 8822 (1986).
370. W. Stevens, H. Basch, and J. Krauss, *J. Chem. Phys.* 81, 6026 (1984).
371. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. J. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, N. Rega, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, "Gaussian 98 Gaussian 98," revision A.11.2:, Gaussian, Pittsburgh, 1998.
372. P. Norman, Y. Luo, D. Jonsson, and H. Ågren, *J. Chem. Phys.* 108, 4358 (1998).
373. R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.* 256, 454 (1996).
374. R. Bauernschmitt, M. Häser, O. Treutler, and R. Ahlrichs, *Chem. Phys. Lett.* 264, 573 (1997).
375. R. Bauernschmitt, R. Ahlrichs, F. H. Henrich, and M. M. Kappes, *J. Am. Chem. Soc.* 120, 5052 (1998).
376. S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.* 302, 375 (1999).
377. D. Sundholm, *Chem. Phys. Lett.* 302, 480 (1999).
378. C. Adamo and V. Barone, *Chem. Phys.* 274, 242 (1997).
379. C. Adamo and V. Barone, *J. Chem. Phys.* 108, 664 (1998).
380. Z.-L. Cai, K. Sendt, and J. R. Reimers, *J. Chem. Phys.* 117, 5543 (2002).
381. W. Hieringer, S. J. A. van Gisbergen, and E. J. Baerends, *J. Phys. Chem. A* 106, 10380 (2002).
382. F. Della Sala, H. H. Heinze, and A. Görling, *Chem. Phys. Lett.* 339, 343 (2001).
383. A. Dreuw, J. L. Weisman, and M. Head-Gordon, *J. Chem. Phys.* 119, 2943 (2003).
384. M. Parac and S. Grimme, *Chem. Phys.* 292, 11 (2003).

385. M. Dierksen and S. Grimme, *J. Chem. Phys.* 120, 3544 (2004).
386. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
387. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
388. J. C. Slater, in "Quantum Theory of Molecular and Solids." McGraw-Hill, New York, 1974.
389. S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* 58, 1200 (1980).
390. A. D. Becke, *J. Chem. Phys.* 98, 5648 (1993).
391. C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* 37, 785 (1988).
392. J. P. Perdew and M. Levy, *Phys. Rev. Lett.* 51, 1884 (1983).
393. A. R. Williams and U. von Barth, In "Theory of the Inhomogeneous Electron Gas." Plenum Press, London, 1983.
394. R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* 37, 10159 (1988).
395. L. Fritsche, *Physica B* 172, 7 (1991).
396. V. G. Malkin, O. L. Malkina, M. E. Casida, and D. R. Salahub, *J. Am. Chem. Soc.* 116, 5898 (1994).
397. M. Levy, *Phys. Rev. A* 52, 50 (1995).
398. E. J. Baerends and O. V. Gritsenko, *J. Phys. Chem. A* 101, 5383 (1997).
399. U. Salzner, J. B. Lagowski, P. G. Pickup, and R. A. Poirier, *J. Comput. Chem.* 18, 1943 (1997).
400. A. Savin, C. J. Umrigar, and X. Gonze, *Chem. Phys. Lett.* 288, 391 (1998).
401. D. J. Tozer and N. C. Handy, *J. Chem. Phys.* 109, 10180 (1998).
402. R. Stowasser and R. Hoffmann, *J. Am. Chem. Soc.* 121, 3414 (1999).
403. O. V. Gritsenko and E. J. Baerends, *J. Chem. Phys.* 117, 9154 (2002).
404. S. Suhai, *Phys. Rev. B* 27, 3506 (1983).
405. A. K. Bakhshi, J. Ladik, and M. Seel, *Phys. Rev. B* 35, 704 (1987).
406. H. O. Villar, P. Otto, M. Dupuis, and J. Ladik, *Synth. Met.* 59, 97 (1993).
407. P. Otto, *Int. J. Quantum Chem.* 52, 353 (1994).
408. R. Pino and G. E. Scuseria, *J. Chem. Phys.* 121, 8113 (2004).
409. S. Suhai, *Phys. Rev. B* 51, 16553 (1995).
410. V. Bezugly and U. Birkenheuer, *Chem. Phys. Lett.* 399, 57 (2004).
411. J. M. Galbraith and H. F. I. Schaefer, *J. Chem. Phys.* 105, 862 (1996).
412. G. E.-L. Chan, D. J. Tozer, and N. C. Handy, *J. Chem. Phys.* 107, 1536 (1997).
413. N. Rösch and S. B. Trickey, *J. Chem. Phys.* 106, 8940 (1997).
414. M. E. Casida, C. Jamorski, K. C. Casidam, and D. R. Salahub, *J. Chem. Phys.* 108, 4439 (1998).
415. P. G. Parr and W. Yang "Density-Functional Theory of Atoms and Molecules." Oxford University Press, New York, 1989.
416. O. Gritsenko, R. V. Leeuwen, and E. J. Baerends, *J. Chem. Phys.* 101, 8955 (1994).
417. D. P. Chong, O. V. Gritsenko, and E. J. Baerends, *J. Chem. Phys.* 116, 1760 (2002).
418. W. H. Green, Jr., D. J. Tozer, and N. C. Handy, *Chem. Phys. Lett.* 290, 465 (1998).
419. U. Salzner, J. B. Lagowski, P. G. Pickup, and R. A. Poirier, *J. Phys. Chem.* 102, 2572 (1998).
420. U. Salzner, J. B. Lagowski, R. A. Poirier, and P. G. Pickup, *Synth. Met.* 96, 177 (1998).
421. C. H. Choi, M. Kertesz, and A. Karpfen, *J. Chem. Phys.* 107, 6712 (1997).
422. R. Hoffmann and W. N. Lipscomb, *J. Chem. Phys.* 37, 2872 (1962).
423. R. Hoffmann and W. N. Lipscomb, *J. Chem. Phys.* 36, 2179 (1962).
424. R. Hoffmann, *J. Chem. Phys.* 39, 1397 (1963).
425. R. Hoffmann, *Tetrahedron* 22, 521 (1966).
426. W. Kutzelnigg, *Theor. Chim. Acta* 4, 417 (1966).
427. W. Kutzelnigg "Einführung in die Theoretische Chemie. Band 2." Velag Chemie, Weinheim, 1978.
428. R. Hoffmann, "Solids and Surfaces: A Chemist's View of Bonding in Extended Structures." Velag Chemie, Weinheim, 1988.
429. R. Hoffmann, C. Janiak, and C. Kollmar, *Macromolecules* 24, 3725 (1991).
430. B. E. Kohler, *J. Chem. Phys.* 93, 5838 (1990).
431. J. L. Brédas, R. R. Chance, R. Silbey, G. Nicolas, and P. Durand, *J. Chem. Phys.* 75, 255 (1981).
432. J.-L. Brédas, R. R. Chance, R. H. Baughman, and R. Silbey, *J. Chem. Phys.* 76, 3673 (1982).
433. J. L. Brédas, R. R. Chance, R. Silbey, G. Nicolas, and P. Durand, *J. Chem. Phys.* 77, 371 (1982).
434. J.-L. Brédas, R. L. Elsenbaumer, R. R. Chance, and R. Silbey, *J. Chem. Phys.* 78, 5656 (1983).
435. M. J. S. Dewar, E. G. Zoebisch, E. F. Healey, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).
436. J. J. P. Stewart, "MOPAC: 6.0." Indiana University, Bloomington, 1985.
437. J. J. P. Stewart, *QCPE Bull.* 8, 62 (1985).
438. R. Pariser and R. G. Parr, *J. Chem. Phys.* 21, 466 (1953).
439. R. Pariser, R. G. Parr, *J. Chem. Phys.* 21, 767 (1953).
440. J. A. Pople, *Trans. Faraday Soc.* 49, 1375 (1953).
441. J. A. Pople, *Proc. Phys. Soc. Sec. A* 68, 81 (1955).
442. J. L. Brédas, J. Cornil, D. Beljonne, D. A. dos Santos, and Z. Shuai, *Acc. Chem. Res.* 32, 267 (1999).
443. H. Bässler, D. D. C. Bradley, and M. Hennecke, *Synth. Met.* 41, 1249 (1991).
444. P. Gomez da Costa and E. M. Conwell, *Phys. Rev. B* 48 (1993).
445. E. M. Conwell, *Synth. Met.* 83, 101 (1996).
446. H. Ågren, F. Gel'Mukhanov, and C. Liegener, *Int. J. Quantum Chem.* 63, 313 (1997).
447. H. Bässler, in "Primary Photoexcitations in Conjugated Polymers: Molecular Exciton versus Semiconductor Band Model." World Scientific, Singapore, 1997.

448. A. J. Heeger, in "Primary Photoexcitations in Conjugated Polymers: Molecular Exciton versus Semiconductor Band Model." World Scientific, Singapore, 1997.
449. H. Bässler, V. I. Arkhipov, E. V. Emelianova, A. Gerhard, A. Hayer, C. Im, and J. Rissler, *Synth. Met.* 135, 377 (2003).
450. P. Wood, I. D. W. Samuel, R. Schrock, and R. L. Christensen, *J. Chem. Phys.* 115, 10955 (2001).
451. R. L. Christensen, A. Faksh, J. A. Meyers, I. D. W. Samuel, P. Wood, R. R. Schrock, and K. C. Hultzsch, *J. Phys. Chem. A* 108, 8229 (2004).
452. T. Otsubo, Y. Aso, K. Takimiya, H. Nakanishi, and N. Sumi, *Synth. Met.* 133–134, 325 (2003).
453. J. M. Tour, *Acc. Chem. Res.* 33, 791 (2000).
454. J. Rissler, *Chem. Phys. Lett.* 395, 92 (2004).
455. G. A. Sotzing and K. Lee, *Macromol.* 2002, 7281 (2002).
456. N. Sato, K. Seki, and H. Inokuchi, *J. Chem. Soc. Faraday II* 77, 1621 (1981).
457. F. Kouki, P. Spearman, P. Valat, G. Horowitz, and F. Garnier, *J. Chem. Phys.* 113, 385 (2000).
458. O. Dippel, V. Brandl, H. Bässler, R. Danieli, R. Zamboni, and C. Taliani, *Chem. Phys. Lett.* 216, 418 (1993).
459. S. Möller and G. Weiser, *Synth. Met.* 122, 41 (2001).
460. F. Garnier, A. Yassar, R. Hajlaoui, G. Horowitz, F. Deloffre, B. Servet, S. Ries, and P. Alno, *J. Am. Chem. Soc.* 155, 8716 (1993).
461. G. Horowitz, S. Romdhane, H. Bouchriha, P. Delannoy, J.-L. Monge, F. Kouki, and P. Valat, *Synth. Met.* 90, 187 (1997).
462. M. Muccini, E. Lunedei, C. Taliani, F. Garnier, and H. Baessler, *Synth. Met.* 84, 863 (1997).
463. C. Alemán and L. Julia, *J. Phys. Chem.* 100, 1524 (1996).
464. H. A. Duarte, D. A. dos Santos, W. R. Rocha, and W. B. de Almeida, *J. Chem. Phys.* 113, 4206 (2000).
465. M. Rubio, M. Merchán, E. Ortí, and B. O. Roos, *Chem. Phys. Lett.* 248, 321 (1996).
466. G. Zotti, S. Martina, G. Wegner, and A.-D. Schlüter, *Adv. Mater.* 4, 798 (1992).
467. M. Muccini, E. Lunedei, C. Taliani, D. Beljonne, J. Cornil, and J. L. Brédas, *J. Chem. Phys.* 109, 10513 (1998).
468. G. Lanzani, M. Nisoli, S. De Silvestri, and F. Abbate, *Chem. Phys. Lett.* 264, 667 (1997).
469. M. Cerminara, F. Meinardi, A. Borghesi, M. Campione, A. Sassella, P. Spearman, G. Bongiovanni, A. Mura, and R. Tubino, *Synth. Met.* 2003, 765 (2003).
470. J. Gierschner, H.-G. Mack, H.-J. Egelhaaf, S. Schweizer, B. Doser, and D. Oelkrug, *Synth. Met.* 138, 311 (2003).
471. P. Vogl and C. K. Campbell, *Phys. Rev. B* 41, 12797 (1990).
472. M. Muccini, M. Schneider, C. Taliani, M. Sokolowski, E. Umbach, D. Beljonne, J. Cornil, and J. L. Brédas, *Phys. Rev. B* 62, 6296 (2000).
473. P. Vogl and C. K. Campbell, *Phys. Rev. Lett.* 62, 2012 (1989).
474. P. Gomes da Costa, R. G. Dandrea, E. M. Conwell, M. Fahlman, M. Lögdlund, S. Stafstrom, W. R. Salaneck, S. C. Graham, R. H. Friend, P. L. Burn, and A. B. Holmes, *Synth. Met.* 55–57, 4320 (1993).
475. P. Puschnig and C. Ambrosch-Draxl, *Phys. Rev. B* 60, 7891 (1999).
476. M. Rohlfing and S. G. Louie, *Phys. Rev. Lett.* 82, 1959 (1999).
477. J.-W. van der Horst, P. A. Bobbert, M. A. J. Michels, G. Brocks, and P. J. Kelly, *Phys. Rev. Lett.* 83, 4413 (1999).
478. J.-W. van der Horst, P. A. Bobbert, M. P. de Jong, M. A. J. Michels, G. Brocks, and P. J. Kelly, *Phys. Rev. B* 61, 15817 (2000).
479. M. Rohlfing, M. L. Tiago, and L. S. G., *Synth. Met.* 116, 101 (2001).
480. J.-W. van der Horst, P. A. Bobbert, M. A. J. Michels, and H. Bässler, *J. Chem. Phys.* 114, 6950 (2001).
481. J.-W. van der Horst, P. A. Bobbert, and M. A. J. Michels, *Synth. Met.* 135–136, 281 (2003).
482. A. Ruini, M. J. Caldas, G. Bussi, and E. Molinari, *Phys. Rev. Lett.* 88, 206403 (2002).
483. A. Ruini, G. Bussi, A. Ferretti, M. J. Caldas, and E. Molinari, *Synth. Met.* 139, 755 (2003).
484. G. Bussi, A. Ruini, E. Molinari, M. J. Caldas, P. Puschnig, and C. Ambrosch-Draxl, *Appl. Phys. Lett.* 80, 4118 (2002).
485. J. Fagerström and S. Stafström, *Phys. Rev. B* 54, 13713 (1996).
486. E. Moore, B. Gherman, and D. Yaron, *J. Chem. Phys.* 106, 4216 (1996).
487. J. Cornil, A. J. Heeger, and J. L. Brédas, *Chem. Phys. Lett.* 272, 463 (1997).
488. J. Fagerström and S. Stafström, *Synth. Met.* 85, 1065 (1997).
489. J. Cornil, D. A. dos Santos, X. Crispin, R. Silbey, and J. L. Brédas, *J. Am. Chem. Soc.* 120, 1289 (1998).
490. D. Beljonne, J. Cornil, R. Silbey, P. Millié, and J. L. Brédas, *J. Chem. Phys.* 112, 4749 (2000).
491. J. Cornil, J. P. Calbert, D. Beljonne, R. Silbey, and J. L. Brédas, *Adv. Mater.* 12, 978 (2000).
492. J. Cornil, D. Beljonne, J. P. Calbert, and J. L. Brédas, *Adv. Mater.* 13, 1053 (2001).
493. J. L. Brédas, D. Beljonne, J. Cornil, J. P. Calbert, Z. Shuai, and R. Silbey, *Synth. Met.* 125, 107 (2002).
494. J. Roncali, Y. Yassar, and F. Garnier, *J. Chem. Soc. Chem. Commun.* 581 (1988).
495. T. Tani, P. M. Grant, W. D. Gill, G. B. Street, and T. C. Clarke, *Solid State Commun.* 33, 499 (1980).
496. J. C. R. Fincher, D. L. Peebles, A. J. Heeger, M. A. Druy, Y. Matsamura, A. G. MacDiarmid, H. Shirakawa, and S. Ikeda, *Solid State Commun.* 27, 489 (1978).
497. A. J. Heeger and J. R. Schrieffer, *Solid State Commun.* 48, 207 (1983).
498. A. J. Heeger, G. Blanchet, T. C. Chung, and C. R. Fincher, *Synth. Met.* 9, 173 (1984).
499. L. M. Tolbert and M. E. Ogle, *J. Am. Chem. Soc.* 112, 9519 (1990).
500. L. M. Tolbert, *Acc. Chem. Res.* 25, 561 (1992).
501. A. J. Epstein and A. G. MacDiarmid, *Synth. Met.* 69, 179 (1995).
502. A. R. Bishop, *Solid State Commun.* 33, 955 (1980).
503. D. A. dos Santos, C. P. de Melo, and H. S. Brandi, *Solid State Commun.* 52, 99 (1984).

504. E. A. Reed, L. A. Curtiss, and F. Weinhold, *Chem. Rev.* 88, 899 (1988).
505. J. L. Brédas, B. Thémans, J. G. Fripiat, J.-M. André, and R. R. Chance, *Phys. Rev. B* 29, 6761 (1984).
506. S. Irle and H. Lischka, *J. Chem. Phys.* 103, 1508 (1995).
507. J. Cornil, D. Beljonne, and J. L. Brédas, *J. Chem. Phys.* 103, 842 (1995).
508. C. Alemán and L. Julia, *J. Phys. Chem.* 100, 14661 (1996).
509. A. J. W. Tol, *Chem. Phys.* 208, 73 (1996).
510. A. D. McFarland and A. V. Fratini, *Am. Chem. Soc. Div. Polym. Chem.* 39, 655 (1998).
511. T. Keszthelyi, M. M.-L. Grage, J. F. Offersgard, R. Wilbrandt, C. Svendsen, O. Sonnich Mortensen, J. K. Pedersen, and H. J. A. Jensen, *J. Phys. Chem. A* 104, 2808 (2000).
512. R. Colle and A. Curioni, *J. Am. Chem. Soc.* 120, 4832 (1998).
513. R. Colle and A. Curioni, *J. Phys. Chem. A* 104, 8546 (2000).
514. G. Brocks, *J. Chem. Phys.* 112, 5353 (2000).
515. C. Ambrosch-Draxl, P. Puschnig, E. Zojer, and G. Leising, *Synth. Met.* 119, 211 (2001).
516. B. Champagne and M. Spassova, *Phys. Chem. Chem. Phys.* 6, 3167 (2004).
517. F. C. Grozema, P. T. van Duijnen, L. D. Siebbeles, A. Goossens, and S. W. Leeuw, *J. Phys. Chem. B* 108, 16139 (2004).
518. A. Dkhissi, D. Beljonne, R. Lazzaroni, F. Louwet, L. Groenendaal, and J. L. Brédas, *Int. J. Quantum Chem.* 91, 517 (2003).
519. V. M. Geskin, A. Dkhissi, and J. L. Brédas, *Int. J. Quantum Chem.* 91, 350 (2003).
520. P. Otto, *Phys. Rev. B* 45, 10876 (1992).
521. G. J. B. Hurst, M. Dupuis, and E. Clementi, *J. Chem. Phys.* 89, 385 (1988).
522. B. Kirtman, *Chem. Phys. Lett.* 143, 81 (1988).
523. B. Champagne, D. H. Mosley, and J.-M. André, *J. Chem. Phys.* 100, 2034 (1994).
524. T. Tangredi Toto, J. L. Toto, C. P. de Melo, M. Hasan, and B. Kirtman, *Chem. Phys. Lett.* 244, 59 (1995).
525. J. L. Toto, T. Tangredi Toto, and C. P. de Melo, *Chem. Phys. Lett.* 245, 660 (1995).
526. J. L. Toto, T. Tangredi Toto, C. P. de Melo, and K. A. Robins, *J. Chem. Phys.* 102, 8048 (1995).
527. B. Champagne, D. Jacquemin, J.-M. André, and B. Kirtman, *J. Phys. Chem. A* 101, 3158 (1997).
528. D. Jacquemin, B. Champagne, and B. Kirtman, *J. Chem. Phys.* 107, 5076 (1997).
529. E. A. Perpète, B. Champagne, and B. Kirtman, *J. Chem. Phys.* 15, 2463 (1997).
530. B. Champagne, E. A. Perpète, T. Legrand, D. Jacquemin, and J.-M. André, *J. Chem. Soc. Faraday Trans.* 94, 1547 (1998).
531. B. Champagne, E. A. Perpète, S. J. A. Gisbergen, E. J. Baerends, J. G. Snijders, C. Soubra-Ghaoui, K. A. Robins, and B. Kirtman, *J. Chem. Phys.* 109, 10489 (1998).
532. B. Champagne, M. Spassova, J.-B. Jadin, and B. Kirtman, *J. Chem. Phys.* 116, 3935 (2002).
533. B. Champagne and J.-M. André, *Int. J. Quantum Chem.* 42, 1009 (1992).
534. C. Cojan, G. P. Agarwal, and C. Flytzanis, *Phys. Rev. B* 15, 909 (1977).
535. G. P. Agarwal, C. Cojan, and C. Flytzanis, *Phys. Rev. B* 17, 776 (1978).
536. J. Ladik, *J. Mol. Struct.* 206, 39 (1991).
537. A. Martinez, *Chem. Phys.* 327, 389 (2000).
538. B. Kirtman, F. L. Gu, and D. M. Bishop, *J. Chem. Phys.* 113, 1294 (2000).
539. D. M. Bishop, F. L. Gu, and B. Kirtman, *J. Chem. Phys.* 114, 7633 (2001).
540. F. L. Gu, D. M. Bishop, and B. Kirtman, *J. Chem. Phys.* 115, 10548 (2001).
541. B. Champagne, D. H. Mosley, J. G. Fripiat, and J.-M. André, *SPIE* 1775, 236 (1992).
542. B. Champagne, D. H. Mosley, J. G. Fripiat, and J.-M. André, *Phys. Rev. B* 54, 2381 (1996).
543. B. Kirtman, C. E. Dykstra, and B. Champagne, *Chem. Phys. Lett.* 305, 132 (1999).
544. B. Champagne, D. H. Mosley, M. Vracko, and J.-M. André, *Phys. Rev. A* 52, 178 (1995).
545. B. Champagne, D. H. Mosley, M. Vracko, and J.-M. André, *Phys. Rev. A* 52, 1039 (1995).
546. D. Jacquemin, B. Champagne, and J.-M. André, *Chem. Phys. Lett.* 284, 24 (1997).
547. B. Champagne, J.-M. André, and Y. Öhrn, *Int. J. Quantum Chem.* 57, 811 (1996).
548. P. Otto, F. L. Gu, and J. Ladik, *Synth. Met.* 92, 161 (1998).
549. P. Otto, M. Piris, A. Martínez, and J. Ladik, *Synth. Met.* 141, 277 (2004).
550. G. Treboux, *J. Phys. Chem. B* 104, 9823 (2000).
551. M. P. Samanta, W. Tian, and S. Datta, *Phys. Rev. B* 53, R7626 (1996).
552. M. Magoga and C. Joachim, *Phys. Rev. B* 56, 4722 (1997).
553. M. Olson, Y. Mao, T. Windus, M. Kemp, M. Ratner, N. Léon, and V. Mujica, *J. Phys. Chem. B* 102, 941 (1998).
554. M. Magoga and C. Joachim, *Phys. Rev. B* 59, 16011 (1999).
555. M. Di Ventra, S. T. Pantelides, and N. D. Lang, *Phys. Rev. Lett.* 84, 979 (2000).
556. L. E. Hall, J. R. Reimers, N. S. Hush, and K. Silverbrook, *J. Chem. Phys.* 112, 1510 (2000).
557. A. Johansson and S. Stafström, *Chem. Phys. Lett.* 322, 301 (2000).
558. V. Mujica, A. E. Roitberg, and M. Ratner, *J. Chem. Phys.* 112, 6834 (2000).
559. P. A. Derosa and J. M. Seminario, *J. Phys. Chem. B* 105, 471 (2001).
560. E. G. Emberly and G. Kirczenow, *Phys. Rev. Lett.* 87, 269701 (2001).
561. C.-K. Wang, Y. Fu, and Y. Luo, *Phys. Chem. Chem. Phys.* 3, 5017 (2001).
562. P. Damle, A. W. Ghosh, and S. Datta, *Chem. Phys.* 281, 171 (2002).
563. J. Filatow and S. Larsson, *Chem. Phys.* 284, 575 (2002).
564. A. Nitzan, M. Galperin, G.-L. Ingold, and H. Grabert, *J. Chem. Phys.* 117, 10837 (2002).
565. M. Ernzerhof and M. Zhuang, *J. Chem. Phys.* 119, 4134 (2003).

566. F. Evers, F. Weigend, and M. Koentopp, *Physica E* 18, 255 (2003).
567. D. S. Kosov, *J. Chem. Phys.* 119, 1 (2003).
568. V. Mujica, A. Nitzan, S. Datta, M. Ratner, and C. P. Kubiak, *J. Phys. Chem. B* 107, 91 (2003).
569. S. Piccinin, A. Selloni, S. Scandolo, R. Car, and G. Scoles, *J. Chem. Phys.* 119, 6729 (2003).
570. H. B. Weber, J. Reichert, R. Ochs, D. Beckmann, M. Mayor, and H. V. Löhneysen, *Physica E* 18, 231 (2003).
571. N. Zimbovskaya, *J. Chem. Phys.* 118, 4 (2003).
572. A. Kopf and P. Saalfrank, *Chem. Phys. Lett.* 386, 17 (2004).
573. J. K. Tomfohr and O. F. Sankey, *J. Chem. Phys.* 120, 1542 (2004).
574. E. A. Weiss, M. J. Ahrens, L. E. Sinks, A. V. Gusev, M. Ratner, and M. R. Wasielewski, *J. Am. Chem. Soc.* 126, 5577 (2004).
575. A. Lahmidi and C. Joachim, *Chem. Phys. Lett.* 381, 335 (2002).
576. L. Rodriguez-Monge and S. Larsson, *J. Phys. Chem.* 100, 6298 (1996).
577. A. Klimkāns and S. Larsson, *Int. J. Quantum Chem.* 77, 211 (2000).
578. E. Johansson and S. Larsson, *Synth. Met.* 144, 183 (2004).

CHAPTER 6

Nanomechanics of Nanoreinforced Polymers

Frank T. Fisher

Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, New Jersey, USA

L. Cate Brinson

Department of Mechanical Engineering, Northwestern University, Evanston, Illinois, USA

CONTENTS

1. Introduction 254
   1.1. Motivation for Nanoreinforced Materials 254
   1.2. Forms of Nanoscale Reinforcement 258
   1.3. Introduction to Viscoelasticity 269
2. Nanoreinforced Polymers—Experimental Results 274
   2.1. Equiaxed Nanoparticulate Reinforcement 274
   2.2. Nanotube Reinforcement 277
   2.3. Nanoplatelet Reinforcement 288
   2.4. The Interphase Region and Viscoelastic Behavior in Nanoreinforced Polymers 292
3. Nanomechanical Modeling 298
   3.1. Bottom-up Approaches 299
   3.2. Top-down Approaches 309
   3.3. Hybrid Modeling Methods 322
4. Current Issues in Nanomechanics of Nanoreinforced Polymers 328
5. Conclusions 332
   Appendix 334
   A.1 Derivation of Mori-Tanaka Method 334
   A.2 Summary of Related Literature 351
   References 352

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (253–360)

1. INTRODUCTION

The past decade has seen an explosion of research at the nanoscale, and polymer nanocomposites have been one of the early material success stories. Research in the area of nanocomposites has progressed based on interdisciplinary interactions between multiple branches of science and engineering as shown in Fig. 1. While the interactions between the branches are sometimes binary (e.g., finite elements represents an intersection of computation and mechanics, often without significant input from materials science and chemistry), for nanocomposites much greater interaction among these different areas is becoming increasingly necessary and, to a large extent, realized. The macroscale behavior of polymer nanocomposites is due to the local, mesoscale, and macroscale interactions and mechanics of polymer chains and the nanoinclusions, requiring the use of sophisticated computational modeling tools at scales from quantum mechanics to structural mechanics, and careful experimental analysis ranging from X-ray photoelectron spectroscopy (XPS) and atomic force microscopy (AFM) to classical fracture mechanics. In this chapter, we will introduce the fundamentals of polymer-based nanocomposites with an emphasis on nanomechanics and the modeling strategies currently used. While we attempt to cover nanocomposites very generally, our frame of reference is grounded within the mechanics and materials science disciplines. In addition, emphasis will be given to nanotube reinforcements over other forms of nanoscale reinforcement due to our more extensive background in this area. While other nanoparticle types will be discussed in general terms and with some limited examples, many of the modeling issues highlighted for nanotubes translate directly to the other nanoparticle forms.

The remainder of this section will review the different types of nanoparticle reinforcements and their specific characteristics, and highlight the unique properties of and potential for nanoreinforced polymer systems. In addition, while an understanding of basic mechanics is presumed throughout this chapter, a brief introduction to viscoelasticity and the mechanical behavior of polymers will be presented in Section 1.3 to assist readers who may be unfamiliar with these areas.

1.1. Motivation for Nanoreinforced Materials

Polymer matrix composites evolved as a structural material in the 1960s, emanating from the pursuit of advanced lightweight, high-strength, corrosion-resistant materials for high-performance defense applications. The advantages of composites are many; they combine light, easily formable polymers with stiff and strong fibers into a highly tailorable structural material. Traditional composites contain a high volume fraction (usually ~60%) of fibers, which can be oriented and constructed to provide materials with directionally-tuned properties. With these advantages, however, comes a complexity of material response significantly beyond that of monolithic materials. The complex microstructure, coupled with the time-and

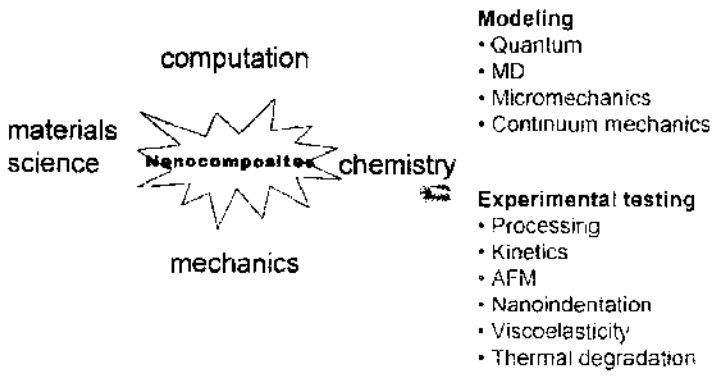


Figure 1. Contributions to the study of polymer nanocomposites hinge on interactions between several fundamental fields of science and engineering.

environmental-dependence of the matrix polymer, makes prediction of long-term performance and assessment of damage evolution challenging [1]. Nevertheless, the performance gains of traditional polymer composites outweigh the uncertainties in the design process and have led to revolutionized structural concepts in aerospace vehicles, sports equipment, medical implants, and automotive parts.

A precursor to nanocomposites, carbon black–filled polymers have been used for decades with primary emphasis on creating electrically conductive polymers with filler particle loadings on the order of 10% [2]. Carbon black particles are agglomerates of amorphous carbon, with typical particle dimensions on the micron scale and have been typically employed in rubbery polymer matrices, where some improvement of mechanical properties was also observed.

The transition to true nanocomposites began in the 1980s with Toyota engineers seeking ways to increase the temperature range of inexpensive plastic materials to replace metal parts in car engines. A layered silicate clay (montmorillonite) was modified via ion exchange, decreasing its hydrophilic nature and swelling the galleries, and then mixed with ε-caprolactam. Ring opening polymerization was then activated which formed nylon-6 reinforced with well-exfoliated clay plates creating a "molecular composite" [3]. At clay loadings of 4%, mechanical properties were significantly improved, and the heat distortion temperature increased by 100°C, enabling the new nanocomposite to be used in the engine compartment for belt covers and other applications. In the United States, General Motors introduced the first commercial automotive exterior application for nanoclay-polymer composites in the 2002 GMC Safari and Astro vans; the use of polymer-clay nanocomposites was later expanded for use on body side molding in the 2004 Chevrolet Impala and the cargo bed of the 2005 Hummer H2 sport utility truck. At this time it is estimated that General Motors is using 660,000 pounds of nanocomposites in its vehicles per year [4].

Around the time that Toyota's work on clay nanocomposites became known, interest in carbon nanotubes exploded just as the technology to create, view, and manipulate nanoscale objects was rapidly ripening [5]. These factors, together with funding availability for exploration of nanomaterials, led to intense development and study of a variety of nanocomposites. Initial results on nanoparticle reinforcement were inconsistent with respect to changes in fundamental thermal and mechanical properties. The assembled work began to indicate the numerous parameters influencing overall nanocomposite performance: dispersion and distribution of inclusions, load transfer from inclusion to matrix, geometric arrangement of nanoparticles, mechanical response of the nanoparticles at the nanoscale, interphase polymer formation and properties, and chemical modification of the nanoinclusion. The most significant factor initially was nanoparticle dispersion: reducing nanoparticle clustering so as to increase the surface area of contact between polymer and nanoparticles. With improved processing methods, modulus increases in nanocomposites are more consistent for well-dispersed systems; however, improvements in strength and toughness remain more elusive.

In addition to mechanical property improvements, nanocomposites offer several strategic multifunctional targets. As with carbon black, some nanoparticles are electrically conductive (e.g., carbon nanotubes), leading to electrically conductive composites. For platelet shaped nanoparticles, thermal and diffusional barrier properties are important, with dramatic decreases in permeability coefficients for water vapor, helium, and oxygen consistent with a model of increased tortuosity of the diffusing vapor particles [6, 7]. These multifunctional considerations are a huge driving force for continued nanocomposite research and development. To summarize, a list of major motivations for using nanoscale reinforcement for nanocomposite applications include:

1. *Mechanical properties of the inclusion materials.* Numerical simulations predict tensile moduli on the order of 1 TPa for nanotubes, making them perhaps the ultimate high-stiffness filler material. Numerical simulations and experimental findings also suggest large elastic (recoverable) strains for nanotubes. Other forms of nanoreinforcement, such as nanoclays and graphite nanoplatelets, also have high modulus values for reinforcement enhancement.
2. *Extremely high strength- and stiffness-to-weight ratios.* Given the exceptional mechanical properties and low densities associated with typical nanoreinforcements,

nanocomposites with high volume fractions of nanoinclusions may result in strength and stiffness weight ratios unachievable with traditional composite materials, offering substantial weight savings for weight-critical applications. Hybrid multiscale composites, where nanoreinforcement is added to the matrix material in traditional micron-sized fibrous composites, have also been proposed.

3. *Multifunctionality*. In addition to their outstanding mechanical properties, nanotubes and other forms of nanoreinforcement have also been shown to have exceptional electrical and heat-related properties, suggesting materials that may be designed to meet mechanical as well as secondary material property specifications. For example, low volume (weight) fractions of nanoreinforcement have been used to enhance the electrical conductivity, increase the working temperature, and (primarily for platelet-shaped reinforcement) improve the barrier and diffusion properties of polymers.

Despite these potential benefits, a number of critical issues must be overcome before the full benefit of such materials can be realized. Such issues include:

1. *Understanding and control of load transfer between the nanoreinforcement and the polymer*. While several examples of excellent load transfer between nanotubes and a polymer have been demonstrated, comprehensive models of load transfer for these and other nanocomposite systems are lacking. Functionalization of the nanotubes (and other nanoparticles) is also being pursued as a way to improve load transfer and between the nanoparticles and the polymer, and models of how functionalization impacts the nanocomposite mechanical response would be desirable.
2. *Dispersion of the nanoreinforcement within the polymer*. Proper dispersion will be necessary for optimal and uniform material properties. Efforts to model the impact of imperfect dispersion on nanocomposite response are ongoing. For nanoplatelet reinforcement, issues related to intercalation and exfoliation of the individual platelets are critical for optimal property enhancement.
3. *Orientation and geometry of the nanoinclusions within the polymer*. While methods have been developed to orient free-standing and as-grown nanotubes, methods to controllably orient nanotubes and other nanoparticles in bulk polymers are still under development. Methods utilizing mechanical load, electrical and magnetic fields, and nano- and micropatterning are being explored to attain morphological control of nanoparticles within the polymer.
4. *Effective properties and behavior of nanocomposites at the nanoscale*. While the properties of nanotubes are known to be dependent on the method of production, reinforcement will depend on the form of the nanotube (single-walled nanotube, multi-walled nanotube, or nanotube bundle). The relationship between these variables and mechanical properties needs to be further elucidated. Further, models of the local nanoscale changes in polymer behavior due to interactions with the embedded nanoparticles also need to be developed.

These points illustrate the need for accurate models of nanocomposite behavior necessary to aid in the interpretation of experimental results: this will lead to a more comprehensive understanding of the key mechanisms underlying the nanocomposite thermomechanical-electrical behavior. Such models in the long term will enable aggressive material and structural design strategies that fully leverage the benefits of such materials. However, modeling of these systems is challenging due to the small length scale of the nanofiller elements. In addition, the need to understand interactions between nanoparticle and polymer chains at the nanoscale, between neighboring nanoparticles, and over regions of many nanoparticles, imply that mechanisms to bridge length scales in modeling are essential.

An example of the need for a multiple-length scale modeling approach is the formation of a nonbulk interphase region in nanocomposites. The interphase is a region of polymer near the nanoparticle that exhibits properties different from the bulk polymer. While experimental work has estimated the interphase thickness for traditional carbon fiber-epoxy composites to be on the order of 1 μm for a 25 μm diameter fiber [8, 9], there is growing evidence that the zone of altered polymer behavior in nanocomposites is much more extensive. From the extensive thin film literature, it is clear that the behavior of polymers near substrates and free

surfaces is dramatically altered and that the influence of the surfaces extends on the order of 10 nm into the polymer [10, 11], suggesting that nanoclay-polymer nanocomposites present a unique opportunity to study the behavior of polymers in confined spaces [12]. Additionally, recent molecular dynamics simulations (discussed further in Section 5.1.1) indicate that the polymer structure and dynamics are affected at least a radius of gyration of the polymer chain away from the nanoparticle surface [13–15].

The fraction of the interphase polymer in nanocomposites may be estimated from simple geometric considerations for a given nanoparticle morphology. For nanotubes (see Fig. 2), the volume fraction of the interphase region V_i is related to the volume fraction of the fiber/nanotube inclusion V_f and the thickness of the interphase region t as [16]

$$V_i = \left[2\left(\frac{t}{r_f}\right) + \left(\frac{t}{r_f}\right)^2\right]V_f = \beta V_f \tag{1}$$

where r_f is the radius of the nanotube (or fiber) and t is the thickness of the interphase region. Likewise, the ratio of the volume fraction of the nonbulk polymer phase (interphase) to the total volume fraction of viscoelastic phases within the composite (interphase and matrix) can be expressed as

$$\frac{V_i}{V_i + V_m} = \frac{\beta V_f}{1 - V_f} \tag{2}$$

These expressions are shown graphically in Fig. 2 as a function of fiber/nanotube volume fraction for various ratios of interphase thickness t to fiber radius r_f. Representative values of (t/r_f) are ~0.05 for carbon fiber composites [8] and ~1.0 for nanotube-reinforced polymers [17], respectively. From Fig. 2, we see that for the case of nanotube-reinforced polymers a significant portion of the nanocomposite can be characterized as the interphase region. In this case, the nonbulk polymer behavior of the interphase region is expected to contribute to the overall thermomechanical and electrical response of the material. For the case for traditional micron-sized fiber polymer composites, the interphase region is much smaller and can often be neglected in micromechanical predictions of effective properties. While fiber sizings are known to significantly enhance the fiber-matrix interface in traditional polymer matrix composites, the sizings in such systems are typically small and often neglected in micromechanical predictions for the effective modulus of these materials. However, for well-dispersed nanocomposites, the volume fraction of interphase is substantial and its effect on overall thermomechanical properties is much larger than that for a traditional micron-sized composite and cannot be neglected. Thus Fig. 2 suggests that the interphase volume fraction will be appreciable for even relatively low loadings of nanotube inclusions; the properties and extent of this interphase is an ongoing topic of research being pursued from both "bottom-up" and "top-down" approaches as discussed further in Section 3.

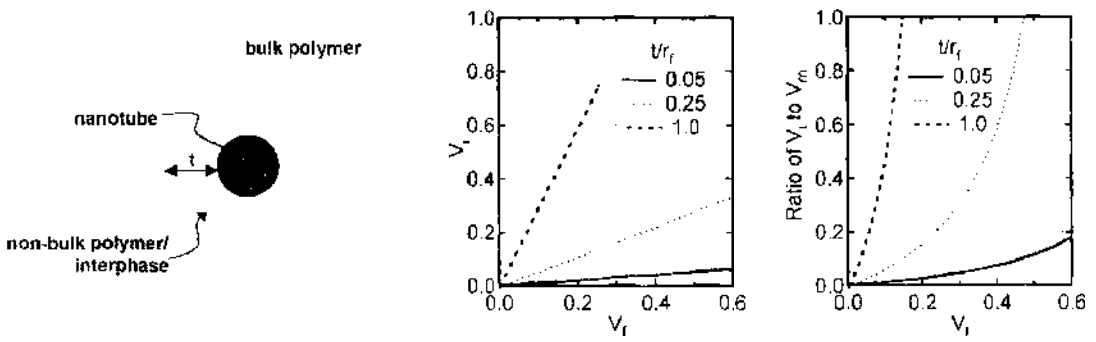


Figure 2. Significance of surface area per unit volume for nanotube-reinforced polymers. (left) Schematic of the interphase region. (center, right) Interphase volume fraction (V_i) and ratio of the interphase (nonbulk) to matrix (bulk) volume fraction (V_m) as a function of fiber volume fraction V_f. Note that t/r_f is ~1 for nanotubes and on the order of ~0.05 for micron-sized fibers. Reprinted with permission from [16], E. T. Fisher and L. C. Brinson, in "Society for Experimental Mechanics Conference Proceedings," Milwaukee, 2002. © 2002, Society for Experimental Mechanics.

1.2. Forms of Nanoscale Reinforcement

For classification purposes it is perhaps easiest to group nanoinclusions based on their geometry, such as the number of dimensions that are nanoscale in length as shown in Fig. 3. In many cases, such a classification is useful as many issues related to processing and modeling are common among nanoinclusions of the same geometry or aspect ratio, L/d, where L and d are the characteristic length and cross-sectional dimension, respectively. *Nanospheres* ($L/d \approx 1$; Section 1.2.1), which we will also refer to as *three-dimensional (3D) nanoparticles or equiaxed nanoparticles*, can be characterized by having all three dimensions of nanoscale length; *nanotubes* or *nanorods* ($L/d \gg 1$; Section 1.2.2) have a cross section on the order of nanometers but are characterized by lengths on the order of microns; and *nanoplatelets* ($L/d \ll 1$; Section 1.2.3) have one dimension (the thickness of the plate) on the order of nanometers while the lateral dimensions may be on the order of microns. Even within these classifications further distinctions can be made, and in many cases definitions can become somewhat arbitrary as the size scale of the inclusion approaches an upper limit where nanoscale effects become negligible. For example, *nanorods* generally refer to solid (as opposed to hollow) cylindrical inclusions, whereas *nanofibers* (or *nanofibrils*) typically denote "larger" cylindrical inclusions with diameters on the order of 100 nm.

1.2.1. Equiaxed Nanoparticulate Reinforcement

A variety of equiaxed nanoparticulate fillers have been added to polymers to enhance the physical properties of the unfilled polymer, dating back to the earliest work with carbon black (see Ref. [2]). Often such composites exploit the size-dependent properties of the nanoparticle while using the polymer as a binder. Examples of equiaxed nanoparticles that have been widely used and studied in nanocomposites include titania (TiO_2), silica (SiO_2), and alumina (Al_2O_3). One of the benefits of this form of inclusion is that they can be added and processed at relatively high weight (volume) loadings, while the surface chemistry of the nanoparticle can be tailored for a particular application. Commercially, for example, equiaxed nanoparticles have been used to drastically alter the optical and refractive properties of commodity plastics and paints (see, for example, Ref. [18]); however, the use of spherically shaped nanoparticles for mechanical property enhancement of polymers is less common (particularly in comparison to the nanotube and nanoplatelet reinforcements described in the next sections). While a brief discussion of relevant experimental data for equiaxed nanoparticulate polymer composites will be presented later in this chapter (see Section 2.1), in many cases mechanical property enhancement is secondary to other property improvements and thus beyond the scope of the present chapter. The reader is referred to the chapter by Schadler for a more complete description of equiaxed nanoparticulate polymer composites [19].

1.2.2. Nanotube Reinforcement

As shown in Fig. 3, nanotube reinforcement refers to a general class of nanoscale-sized cylindrical inclusions with very large aspect ratios (from a mechanics standpoint such aspect

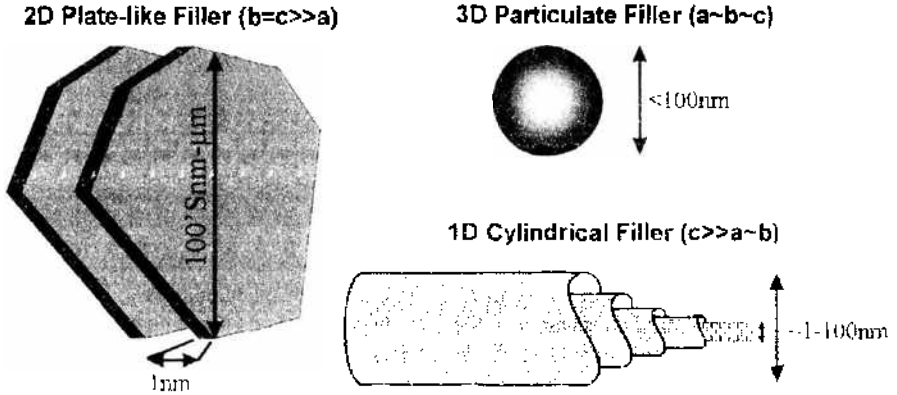


Figure 3. Types of nanoscale reinforcement fillers.

raties can often be treated as "infinitely long"). Carbon nanotubes (discussed in detail in Section 1.2.2.1) are by far the most popular and most studied material to date, motivated by the potential of exploiting the outstanding physical properties of the nanotubes within a host polymer material. Related structures such as bone-shaped carbon nanotubes [20] (see Section 1.2.2.2) and carbon nanocoils [21] (Section 1.2.2.3), which may find particular use for novel applications within an polymer matrix, have been introduced within the literature. In addition there has been a recent renewed interest in the electrospinning process, a versatile method that has been used to create a variety of polymer and composite nanofibers with potential applications as nanoscale reinforcement fillers (Section 1.2.2.4). Other forms of nanofibers, such as boron nitride (BN) [22], which have mechanical properties similar to those of carbon nanotubes but are electrically insulating and have higher oxidation temperatures, and inorganic/ceramic nanofibers electrospun from sol–gel precursors [23], are also being investigated as potential nanoinclusions for nanocomposites applications. Undoubtedly, the list of candidate nanofibers will continue to grow as methods to fabricate unusual nanofibers with potentially interesting properties continue to be developed.

1.2.2.1. Carbon Nanotubes. The outstanding properties that are predicted (and in some cases verified experimentally) for carbon nanotubes are the result of their structure; due to the inherent strength of the carbon–carbon bond and the potential of a defect-free structure, it has been suggested that nanotubes may approach the theoretical limits for many important mechanical properties, including axial stiffness and tensile strength. The desire to exploit these physical properties has resulted in the large research effort currently ongoing within academia and industry to use carbon nanotubes as a reinforcing phase within polymeric materials. A detailed presentation of the state of the art with respect to carbon nanotubes is certainly beyond the scope of the current work; however, a number of excellent reviews can be found in the literature, including general discussions of carbon nanotubes [24–26], physical properties of carbon nanotubes [27], mechanical properties of carbon nanotubes [28], the mechanics of carbon nanotubes [29–31], and potential applications of carbon nanotubes [32].

Carbon nanotubes can be classified into three broad categories: single-walled nanotubes (SWNT), multiwalled nanotubes (MWNT), and nanotube bundles or ropes. Here we will use the abbreviations SWNT and MWNT to refer to single-walled and multiwalled nanotubes, respectively. In some instances in the literature, SWCNT and MWCNT are used as well. Unless otherwise specified, SWNT and MWNT should be taken to refer to carbon nanotubes. SWNTs consist of a single layer of carbon atoms wrapped into a cylindrical shape, which may or may not be capped on each end by one half of a fullerene molecule (see Fig. 4) [33]. Because the graphene hexagonal lattice can be rolled at different angles, the geometry of a particular nanotube is best described in terms chiral vector, which is defined as $C_h = n\hat{a}_1 + m\hat{a}_2$, where $\hat{a}_1$ and $\hat{a}_2$ are unit vectors on the hexagonal lattice and n and m are integers; nanotubes with different chiral vectors (n, m) will have different atomic

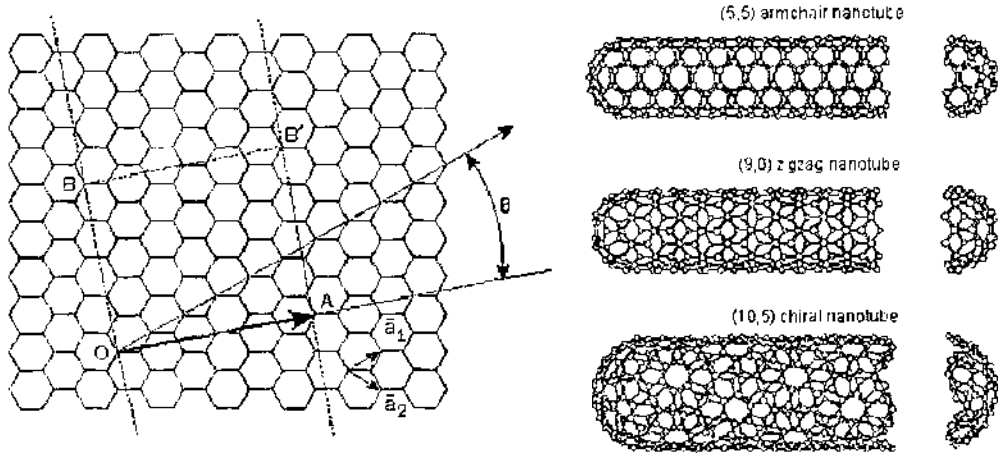


Figure 4. (left) Unit cell and chiral vector for a (4, 2) carbon nanotube. (right) Examples of nanotubes with different chirality. Reprinted with permission from [33], M. Dresselhaus et al., *Phys. World* 11, 33 (1998). © 1998, IOP Publishing.

configurations as shown in Fig. 4. Certain nanotube properties have been found to be strongly dependent on the chirality of the nanotube, including electrical resistivity and fracture behavior. For example, a carbon nanotube will be metallic when the chiral vector satisfies the relationship $n - m = 3q$, where q is an integer, while all other nanotubes will be semi-conductive.

However, other properties, particularly the axial stiffness, have been found to be relatively independent of chirality. While recent work has been able to identify the chirality of an individual nanotube based on STM atomic resolution images [34], at present methods to control nanotube chirality via appropriate processing and to isolate nanotubes of a given chirality are still under development [35]. Thus at this relatively early stage of nanotube-polymer composite mechanical modeling it is premature to consider chirality as a design feature. Typical diameters for SWNTs are on the order of 1 nm, while lengths are often on the order of μm, resulting in very large aspect ratios which in traditional composites theory are desirable from the perspective of load transfer. Both the diameter and the length of the SWNTs are typically dependent on the particular technique used to create the nanotubes.

MWNTs consist of several concentric layers (or shells) of individual carbon nanotubes that are weakly coupled to each other through van der Waals forces. A high-resolution transmission electron microscope (TEM) image of a MWNT is shown in Fig. 5 [25]. The spacing between these individual shells is on the order of 0.34 nm, which is slightly larger than the interlayer spacing between graphene sheets. The diameter and number of shells comprising a MWNT is dependent on the fabrication process, although outer diameters on the order of 15 to 30 nm may be considered as somewhat common [36, 37]. As shown in Fig. 6, there can also be a large variation in diameter within a given batch of MWNT, and only recently has this been taken into account with respect to the interpretation of experimental data [38].

Perhaps more critical is the open question regarding the efficiency with which the inner walls of the MWNT provide reinforcement to the nanocomposite. For example, early Raman spectroscopy measurements found evidence that while all shells of the MWNT are strained under compression, only the outer shell of a MWNT is loaded in tension (due to the weak coupling between the shells) [39]. It has been suggested that such poor load transfer is responsible for the inability to date to experimentally obtain effective mechanical properties for nanotube-polymer composites matching those predicted from standard micromechanics and composites models. In addition, it is becoming increasingly evident that the mechanical properties of MWNTs are dependent on the method of fabrication of the MWNTs; this is directly related to the level of defects in the MWNT atomic structure.

Typically, nanotubes are found to have self-organized into crystalline bundles [40, 41], consisting of several to hundreds of SWNTs or MWNTs arranged in a closest-packed two-dimensional lattice, with adjacent tubes weakly coupled via van der Waals interactions. Optimal reinforcement will likely necessitate complete dispersion of these bundles to maximize the available surface area of the nanotubes within the polymer matrix. From the perspective of structural reinforcement, optimal behavior will be dependent on the proper transfer of load from the matrix to the inclusion (and among the shells or tubes in the case of MWNTs or NT bundles, respectively). While SWNTs are more susceptible to bending due to their extremely small cross sections, for MWNTs and NT bundles interlayer sliding (so-called "sword and

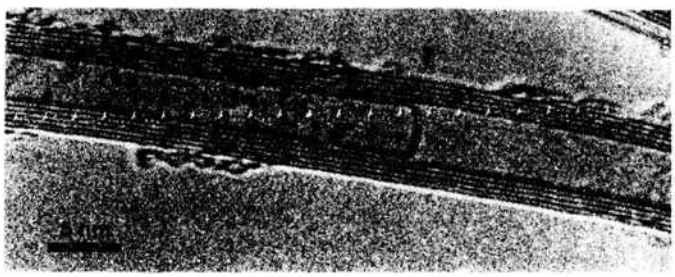


Figure 5. High resolution TEM image of a MWNT with an internal cap highlighted by the arrow. Reprinted with permission from [25], P. J. F. Harris, "Carbon Nanotubes and Related Structures: New Materials for the 21st Century," Cambridge University Press, Cambridge, 1999, © 1999, Cambridge University Press.

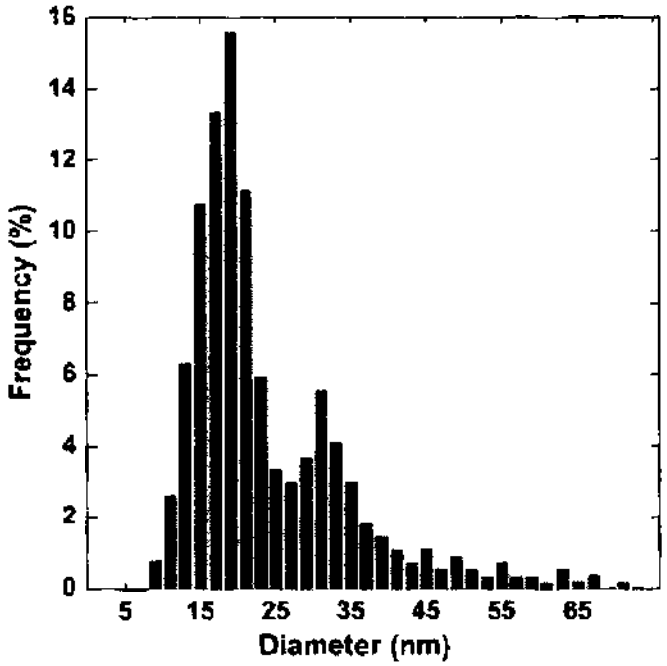


Figure 6. Bimodal histogram of outer diameter for CVD-grown MWCNTs. Reprinted with permission from [38], E. T. Thostenson and T. W. Chou, *J. Phys. D* 36, 573 (2003). © 2003, IOP Publishing Ltd.

sheath" slippage [42]) and weak intertube coupling, respectively, could hinder load transfer. Thus, explicit differences in the structural behavior of these various NT forms will need to ultimately be included in future models of nanotube-related materials.

While a complete discussion of the various processing methods to create carbon nanotubes is beyond the scope of the present chapter, a brief description of some methods is warranted here. For a more complete analysis, the reader is referred to other sources in the literature [19, 25]. In early work, arc discharge [43] and laser vaporization [40] processes were the most common forms of nanotube production, typically resulting in nanotubes with low structural defects and thus excellent physical properties. In these techniques SWNTs are typically formed in the presence of a metal catalyst, which seems to preclude the formation of MWNTs. One difficulty associated with these techniques is the need to process the end product, which is typically found to be quite entangled (see Fig. 7) [44]; in many cases amorphous carbon and other contaminants on the surface of the nanotubes need to be removed via various purifying techniques. An additional problem with these techniques

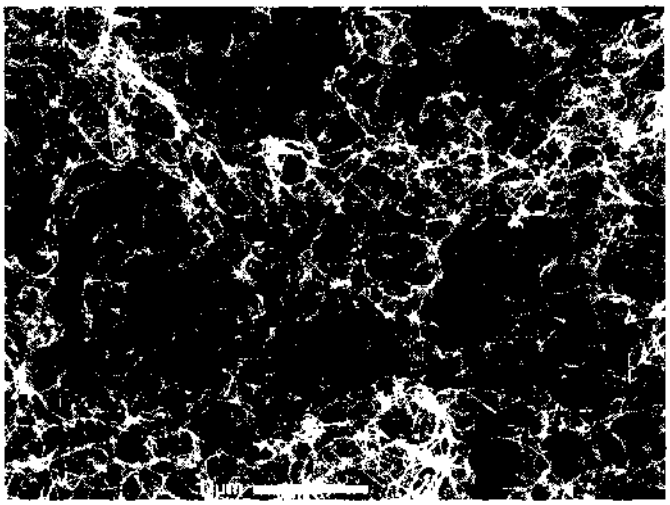


Figure 7. SEM image of SWNT bundles formed via the arc discharge method. Reprinted with permission from [44], C. Journet et al., *Nature* 388, 756 (1997). © 1997, Nature Publishing Group.

is that production yields from these methods are rather limited, and do not seem suited to satisfy the long-term goal of ton-quantity production. A recent advance in the fabrication of high structural quality single-walled nanotubes is the high-pressure CO conversion process (HiPCO) developed at Rice University [45].

In the interest of developing a process that can be scaled for industrial production, a great deal of work has been devoted to techniques that may be classified as chemical vapor deposition (CVD) [46, 47]. While CVD processes have been used to create a wide variety of carbon structures [48], the major drawback of these methods is the reduced structural integrity of the nanotubes. For example, experimental work has suggested that the tensile moduli of CVD nanotubes may be more than an order of magnitude lower than those measured for nanotubes created via other methods due to a greater number of structural defects [49]. However, promising CVD techniques that seem to produce nanotubes with fewer defects (and hence better physical properties) are currently under development [46].

One bottleneck currently hampering nanotube-related research is the high cost and low availability of the raw nanotube material. While significant progress has been made addressing the purity and production yields from nanotube processing techniques, this work is quite challenging because detailed models of nanotube growth are not well developed at present. However, as interest from scientific community continues to expand it is expected that nanotube fabrication techniques will continue to develop, and future advances in nanotube processing are likely to increase the availability of these materials.

The motivation for using carbon nanotubes is based on the ultimate theoretical properties predicted for defect-free nanotubes, as summarized by Collins and Avouris for single-walled carbon nanotubes and presented in Table 1 [26]. Much of the initial work studying the mechanical properties of nanotubes has consisted of computational methods such as molecular dynamics and *ab initio* models. Although primarily used to study SWNTs because of the increase in computational resources necessary to model systems comprised of a larger number of atoms, analyses for larger MWNT systems are increasingly more common. Typically, these computational studies have found nominal values for the axial Young's modulus on the order of 1000 GPa (assuming a shell thickness of 0.34 nm), with values for the Poisson ratio approximately 0.20 to 0.30 [50–52]. *Ab initio* calculations have found that the mechanical properties of nanotubes are similar to those of graphite down to small nanotube radii (on the order of 3 nm), at which point the properties increase due to the enhanced curvature of the tubes [52]. Because strength is closely related to the presence of defects within a material, it has been hypothesized that nanotubes (particularly low defect NTs formed via carbon arc and laser vaporization methods) may approach theoretical limits in terms of strength. For example, in a recent molecular mechanics simulation, NT fracture strains between 10% and 15% were reported, with corresponding tensile stresses on the order of 65 to 93 GPa (compare to the values for other common filler materials listed in Table 2) [53].

In addition, a great deal of progress has been realized in the manipulating and testing of individual nanotubes and nanotube bundles [54–57]. In general, the experimental results

Table 1. Ultimate theoretical properties for carbon nanotubes and comparison with other materials.

Property	SWNTs	By comparison
Density	~1.33g/cm^3	Aluminum has a density of 2.7 g/cm^3
Tensile strength	45 GPa	2 GPa for high strength steels
Resilience	Very high elastic strains; can be bent at large angles without permanent damage	High stiffness materials generally have very low elastic strains
Current carrying capacity	Estimated at 1 billion A/cm^2	Copper wires burn out at 1 million A/cm^2
Field emission	Can activate phosphors at 1–3 V if electrodes are spaced 1 μm apart	Molybdenum tips require fields of 50–100 V/μm and have limited lifetimes
Heat transmission	Predicted to be as high as 6000 W/m K	Pure diamond transmits at 3320 W/m K
Temperature stability	Stable at 2800°C in vacuum, 750°C in air	Metals wires in microchips melt at 600–1000°C

Source: Reprinted with permission from [26], P. G. Collins and P. Avouris, *Sci. Am.* 62 (2000). © 2000, Scientific American.

Table 2. Mechanical properties for filler materials for structural polymers.

Fiber	Diameter (μm)	Density (g/cm^3)	Tensile strength (GPa)	Modulus (GPa)
Carbon	7	1.66	2.4–3.1	120–170
S-glass	7	2.50	3.4–4.6	90
Aramid	12	1.44	2.8	70–170
Boron	100–140	2.50	3.5	400
Quartz	9	2.2	3.4	70
SiC fibers	10–20	2.3	2.8	190
SiC whiskers	0.002	2.3	6.9	—
Carbon NTs	0.001–0.1	~1.33	Up to ~50	Up to ~1000

Source: Reprinted with permission from [69], L. J. Matienzo et al., in "Characterization of Polymers" (N. J. Chou, S. P. Kowalczyk, R. Saraf, and H.-M. Tong, Eds.), Butterworth-Heinnmann, Boston, 1994. © 1994, Elsevier.

have validated the computational predictions. For SWNT bundles the maximum tensile strain was measured to be 5.3%, with the tensile strength of the individual SWNTs estimated to be 13 to 52 GPa [56]. Related tests on MWNTs found that failure occurred via a "sword-in-sheath" mechanism at tensile strains up to 12%, with the tensile strength of the outer shell of the MWNT estimated to be between 11 and 63 GPa [57]. The tensile strength of NTs has been estimated to be 3.6 GPa for CVD-grown MWNTs using a miniature stress-strain puller to test long (~2 mm) NT ropes, with the order of magnitude decrease in strength attributed to an increase in defects [37, 58]. The reader is referred to a recent review of nanomechanical testing of nanotubes and related nanostructures for further discussions of these testing methodologies [59].

In many cases, approaches based on continuum mechanics have been applied successfully for simulating the mechanical responses of individual or isolated carbon nanotubes [60–62]. However, we note that some researchers have suggested that to properly model the bending behavior of nanotubes, values of 5 TPa and 0.067 nm for the Young's modulus and the shell thickness, respectively, of the nanotube should be used [63, 64]. While treating the nanotube as a continuum may be suitable for modeling the linear elastic behavior of the nanotube, other mechanical behavior (such as crack propagation and fracture) will undoubtedly be more dependent on atomic structure and may be ill-suited for such an assumption. In addition, recent theoretical work has developed the appropriate transverse plane (in the plane perpendicular to the axis of the NT) properties for SWNTs [65], MWNTs [66], and SWNT bundles [67], respectively. A set of self-consistent relationships for the physical properties of SWNTs has also been developed [68].

To place the nanotube moduli and strength predictions into proper perspective, representative values for common types of filler materials for structural reinforcement are given in Table 2 [69]. While the predicted properties of carbon nanotubes compare quite favorably to those materials listed here, a greater understanding of the nanotubes themselves, and issues related to their use within a polymer matrix, must be developed to fully utilize the properties of the nanotubes in structural composites.

1.2.2.2. Bone-Shaped Carbon Nanotubes. The template-based synthesis approach was pioneered in the early 1990s as a means to create uniform nanotubes (or, if completely filled, nanorods) [70]. The original approach entails synthesizing the desired materials within the nanochannels of a suitable membrane (e.g., anodic aluminum oxide film, track-etch membrane) by decomposition of a suitable precursor gas. Because the nanochannels in the membranes have monodisperse diameters, analogous monodisperse nanostructures can be obtained, with the smallest carbon nanotube (diameter of 0.4 nm) being synthesized inside porous zeolite [71]. The tunable geometry of the template-synthesized nanostructures are ideal model structures for studying the size-dependent physical, chemical and electronic properties at the nanometer scale, although their use for nanocomposite applications has thus far been limited due to the inherent batch process necessary to fabricate nanotubes in such a manner.

Anodic aluminum oxide (AAO) films have been used previously as templates for CVD growth of carbon nanotubes [72, 73]. One can fabricate AAO films with an extremely uniform nanochannel array using a standard two-step anodization process [74]. Electrolyte, anodization voltage, anodization time, and reaction temperature are four important fabrication factors. For CVD growth of CNTs using AAO as templates, two different synthesis approaches have been developed: catalytic CVD growth [72, 75, 76], and non-catalytic CVD growth [73, 76]. In the latter case, the through-hole AAO film is placed such that nanochannels are running horizontally, in the direction of the gas flow. At elevated temperatures, hydrocarbon gases will be thermally decomposed, resulting in carbon deposition on the wall of the nanochannels. Depending on the deposition time, nanotubes or solid nanorods can be synthesized [76]. The as-synthesized carbon-related nanostructures can be readily freed by etching away the AAO films using HF acid or 6 M NaOH solution, and exhibit a turbostratic structure characteristic of a large number of defects within the nanotube lattice (see Fig. 8). As shown in Fig. 9, by varying the CVD deposition time, template-grown CNTs with different wall thicknesses can be obtained.

A novel four-step extension of this template-based deposition technique has been developed to create the bone-shaped templated carbon nanotubes morphology shown in Fig. 10, [20]. Such a bone-shaped morphology has been used in the short-fiber composites community to promote enhanced load transfer via mechanical interlocking of the enlarged ends [77, 78]; the reader is particularly referred to the comprehensive review article by Zhu and co-workers for a complete description of the state-of-the-art [78]. While yet to be investigated in detail, for nanotube-polymer composites such morphology may provide an alternative method to increase load transfer to the CNTs *without* overly strengthening the interface, providing a means to increase load transfer to the nanotube while maintaining the energy-dissipation mechanisms necessary to maintain (or improve) the toughness of the material.

1.2.2.3. Carbon Nanocoils. Another recent development in the synthesis of carbon nanostructures is the carbon nanocoil [79]. Because of their unique helical structure, carbon nanocoils have been proposed for a variety of applications, including resonating elements [80], nanoscale springs, and novel reinforcement in high-strain composite materials. Using a

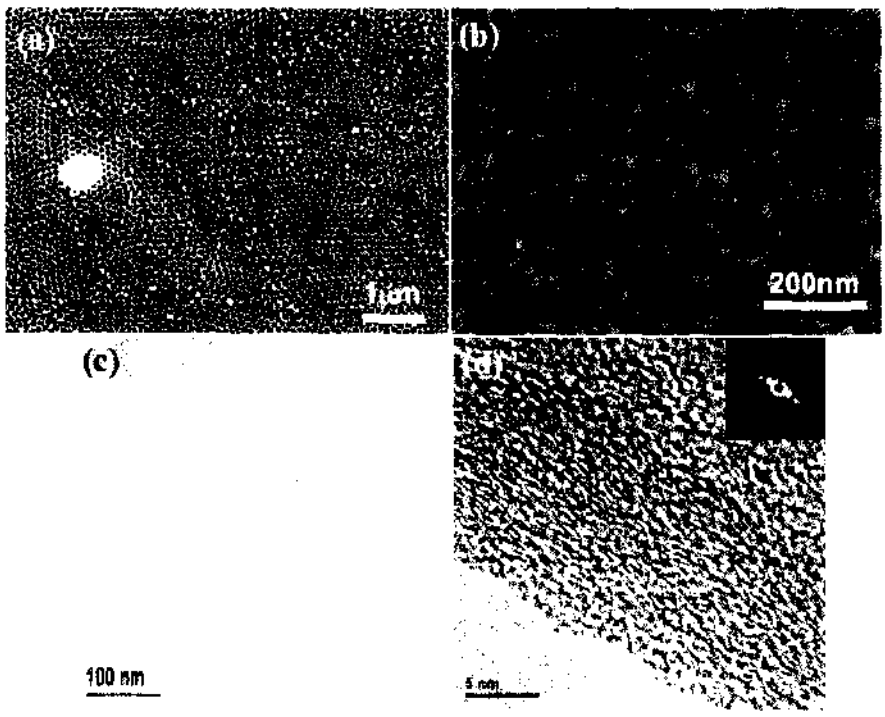


Figure 8. SEM micrographs of (a) top view of as-fabricated AAO film, (b) top view of CNTs grown inside AAO films; TEM micrographs of (c) low magnification of two template-grown CNTs, (d) corresponding HRTEM showing the turbostratic structure standard for CVD grown nanotubes. Images courtesy of Professor T. Xu, University of North Carolina at Charlotte.

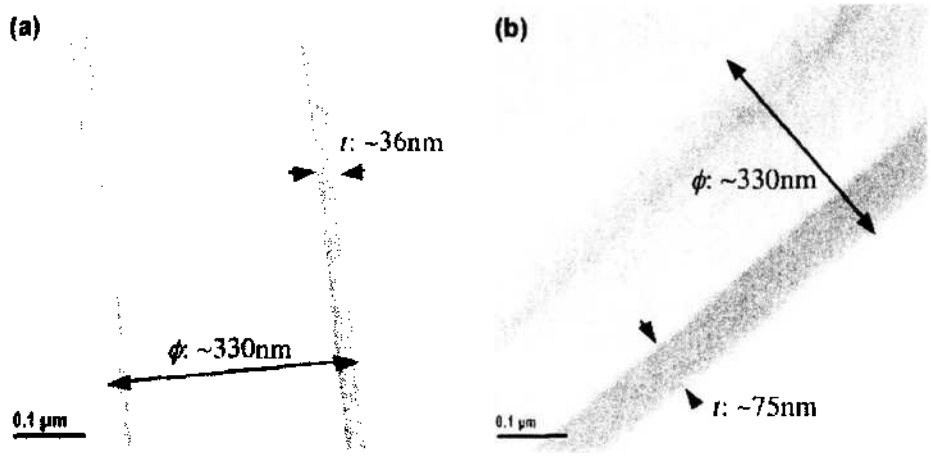


Figure 9. TEM images of turbostratic template-grown CNTs synthesized by (a) 2 hours deposition, and (b) 4 hours deposition. Images courtesy of Professor T. Xu, University of North Carolina at Charlotte.

custom nanomanipulator operated within an SEM, characterization of the mechanical properties of these carbon nanotubes was achieved by direct tensile loading as shown in Fig. 11 [21]. The shear modulus G of the nanocoil was determined based on the spring constant of the nanocoil K and the geometric parameters of the nanocoil. At low-strain levels, the nanocoil behaves as an elastic spring with a spring constant $K = 0.12$ N/m, with a characteristic upturn in the spring constant at higher strain levels as the nanocoil straightens. In addition, it was determined that the nanocoil could be extended up to 42% strain without evidence of plastic deformation, such that the nanocoil returned to its relaxed geometry after loading. The effective shear modulus of the coils was determined to be approximately 2 GPa, which is much lower than the shear modulus estimated for high-quality carbon nanotubes. While we are currently unaware of such work, these unique nanostructures might be of particular interest in high-strain polymeric material applications, where, for instance, the electrical conductivity of the polymer could be increased without adversely compromising the elongation of the host polymer.

1.2.2.4. Electrospun Polymer Nanofibers. A promising emerging technology is the electrospinning of polymer nanofibers (excellent review articles are provided in the literature; see Refs. [81, 82]). Depending on processing conditions, electrospinning has been shown to produce nanofibers with diameters ranging from a few microns down to less than 10 nm. Electrospinning technology enables production of continuous polymer nanofibers (and reinforced polymer nanofibers) from polymer solutions or melts when a high electric field induces an electric force on a polymer liquid strong enough to overcome surface tension. The resulting polymer jet can be collected and dried in what is typically a random mat. There is

Figure 10. Bone-shaped templated carbon nanotubes. Outer diameter of the widened ends and stem are 70 and 40 nm, respectively. The length and wall thickness are ~5 μm and 10 nm. Reprinted with permission from [20], T. T. Xu et al., *Nano Lett.* 3, 1135 (2003). © 2003, American Chemical Society.

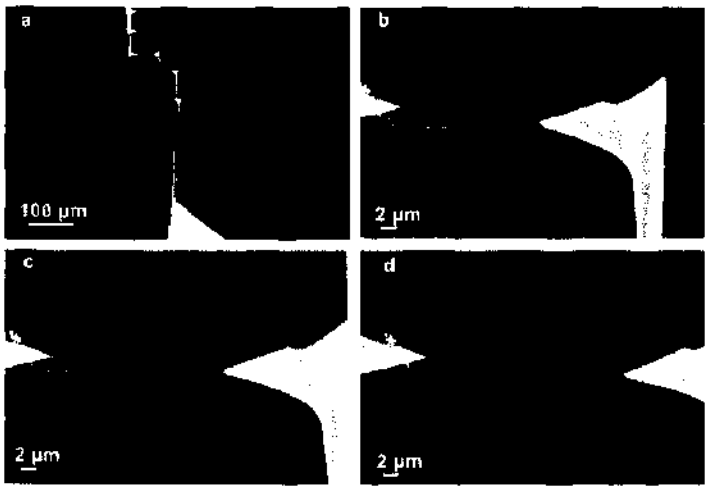


Figure 11. Tensile loading of a carbon nanocoil. (a) Carbon nanocoil clamped between two AFM cantilevers. The left cantilever is stiffer than the right cantilever. (b) Relaxed nanocoil prior to loading. Nanocoil at a relative elongation of (c) 20% and (d) 33%. Reprinted with permission from [21], X. Q. Chen et al., *Nano Lett.* 3, 1299 (2003). © 2003, American Chemical Society.

considerable interest in using these electrospun polymer fibers in a number of applications ranging from protective textiles to sensors to biomedical applications. Because fiber properties typically increase as the diameter decreases, there is also potential in using such nanofibers as a reinforcement material in composites applications.

A review of the recent advances in the electrospinning of polymers is beyond the scope of this chapter. However, these electrospun polymer nanofibers and electrospun nanocomposites are of interest here as appropriate modeling will be necessary to understand and capture:

(1) How the properties of the electrospun nanofibers are dependent on the electrospinning process and resulting nanofiber structure.
(2) How the incorporation of nanoparticles, including montmorillonite [83] (see Fig. 12(a)), graphite nanoplatelet [84, 85], and carbon nanotubes (see Fig. 12(b)) [86–88], change the properties of the nanofibers.
(2) The resultant mechanical properties should these nanometer diameter fibers be incorporated as a reinforcement phase in a matrix material.

The interested reader can pursue this topic beginning with selected references from the literature [81, 82, 89].

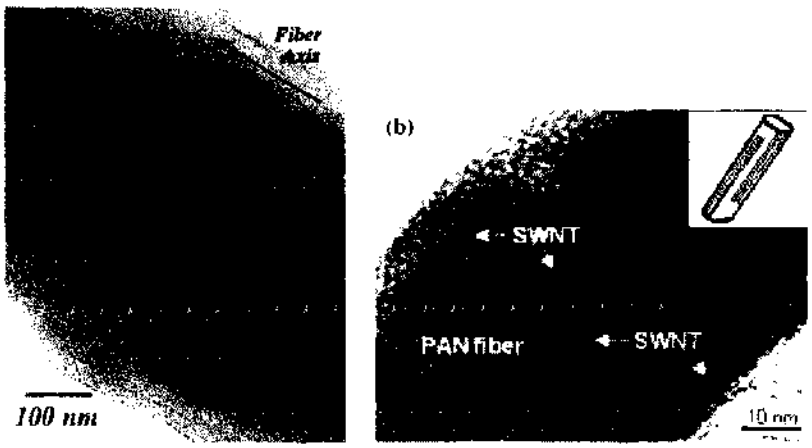


Figure 12. TEM images of electrospun polymer nanocomposite fibers showing alignment of the nanoinclusion along the fiber axis. (a) Exfoliated montmorillonite in nylon 6. Reprinted with permission from [83], H. Fong et al., *Polymer* 43, 775 (2002). © 2002, Elsevier. (b) SWNT in PAN fiber. Reprinted with permission from [86], F. Ko et al., *Adv. Mater.* 15, 1161 (2003). © 2003, Wiley-VCH.

1.2.3. Nanoplatelet Reinforcement

As shown in Fig. 3, nanoplatelet reinforcement refers to a general class of 2D platelike filler material with lateral dimensions on the order of microns with thicknesses of the individual platelets on the order of nanometers. The interest in using nanoplatelet reinforcement in a polymer matrix composite was initiated by the pioneering work with nanoclays by researchers at Toyota [3, 90]. Since that time, research into clay nanocomposites has exploded; the reader is referred to a number of recent review articles for an in-depth discussion of the current state of the art [91–99]. More recently, a number of researchers have looked to extend this work on nanoplatelet reinforcement using graphite nanoplatelets (which, if exfoliated down to individual sheets are referred to as graphene sheets), which have dimensions similar to the nanoclays but in many ways combine the chemistry and properties of carbon nanotubes due to the similarity of the carbon bond network.

Layered silicate clays are an inexpensive material that produced the first commercially successful nanocomposites. The clay particles are easily broken by mechanical forces into micron-sized particles composed of many layers of silica sheets that are bound by van der Waals interactions to one another. The interlayer spacing is on the order of 1.85 nm and contains cations to electrically balance the structure (see Fig. 13) [100]. The individual silica sheets are microns in lateral dimensions and often assigned an in-plane modulus value of 170 GPa (on the basis of experimental measurements made on bulk mica) [101]. However, a recent AFM study found that exfoliated individual clay layers with a thickness of 1 nm were extremely compliant and behaved as "wet tissue paper" in conforming to a porous substrate [102]. Such results call into question the proper mechanical properties that should be assigned to these inclusions in continuum-level models; this is an open topic in the nanocomposites community.

With proper processing the nanoclay can be intercalated with polymer or exfoliated to produce nm thick reinforcing sheets throughout the nanocomposite (see Figs. 14 and 15). In general, the incorporation of the nanoclays into various polymers has been shown to result in: increases in tensile properties (elastic modulus, tensile strength), decreased thermal expansion coefficients and improved thermal stability, increased swelling resistance, decreased gas permeability, and improved flammability properties. A new class of nanoclay-polymer composites, based on the layer-by-layer assembly of individual clay nanoplatelets and appropriate polymers, is also under development [103–105]. Because of the large surface area of the clay

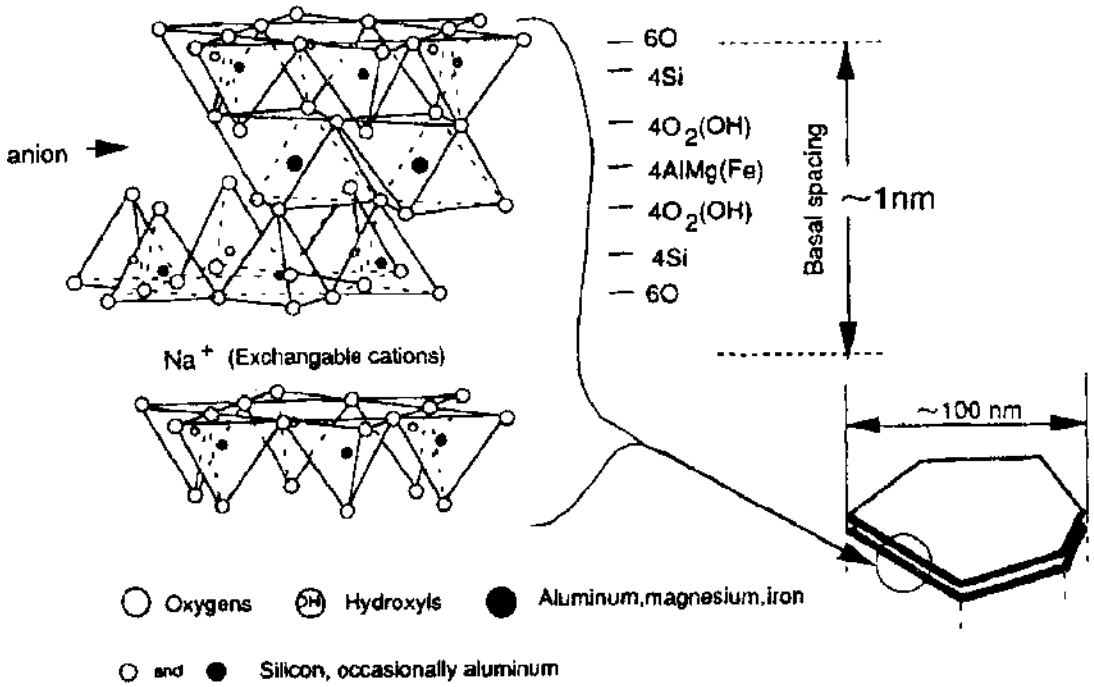


Figure 13. Structure of montmorillonite. Reprinted with permission from [100], M. Kawasumi, *J. Polym. Sci. A* 42, 819 (2004), © 2004, John Wiley and Sons.

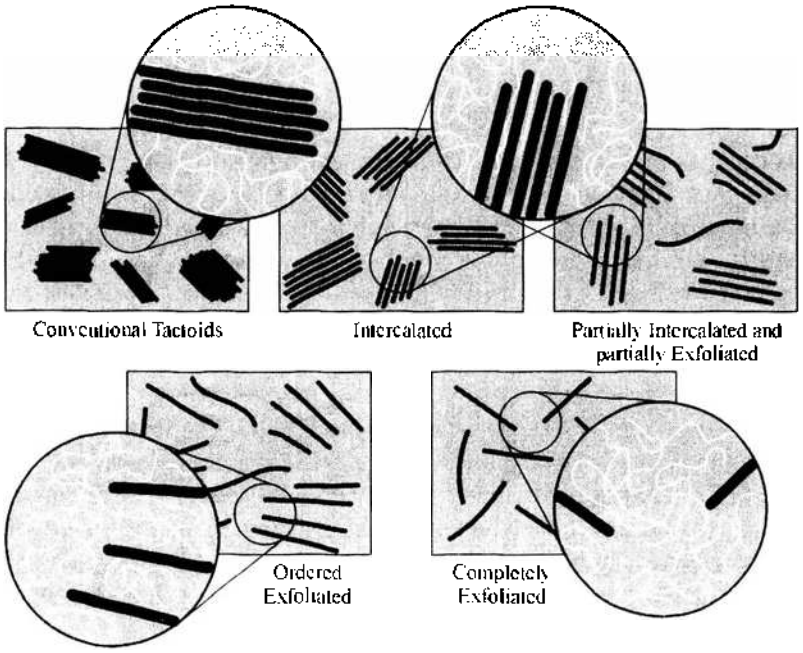


Figure 14. Microstructures of clay-filled polymers from conventional microscale tactoids to fully exfoliated nanocomposite.

nanoplatelets within the nanocomposite, an understanding and control of the clay surface is expected to play an important role in the optimization of these material systems. Models of the mechanical properties of the individual clay nanoplatelets, as input parameters into models of effective clay nanocomposite behavior, will also be critical.

Another interesting platelike nanofiller is exfoliated graphite [106]. Similar to clay, graphite is a layered structure with very strong in-plane covalent bonds within the individual graphene sheets and weak van der Waals bonding between the layers (see Fig. 16(a)). Also similar to clay, graphite is a commodity product with a very low cost of approximately $2/pound. In contrast to clay nanoparticles, the inherent properties of a single nanosheet of graphite approach those of carbon nanotubes, with Young's modulus in the 1 TPa range, and significant electrical and thermal conductivity: resistivity of $\sim 4 \cdot 10^{-5}$ Ohm-cm and thermal conductivity of ~2000 W/m-K (equal to that of type IIa diamond). Such characteristics suggest that exfoliated graphite forms could provide equivalent or superior property enhancements compared to nanotubes but at a fraction of the cost.

As with clay reinforcements, in order for a graphite-based polymer nanocomposite to exhibit superior response, the graphite plates must be properly exfoliated, potentially down to

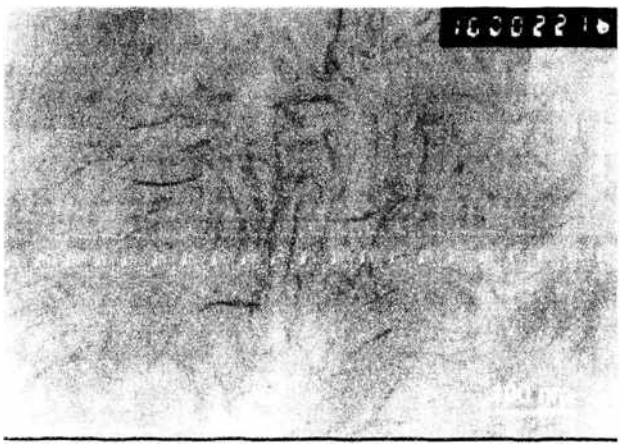


Figure 15. Transmission electron micrograph of 5 wt% well-dispersed, exfoliated aluminosilicate in nylon. Reprinted with permission from [100]. M. Kawasumi, *J. Polym. Sci. A* 42, 819 (2004). © 2004, John Wiley and Sons.

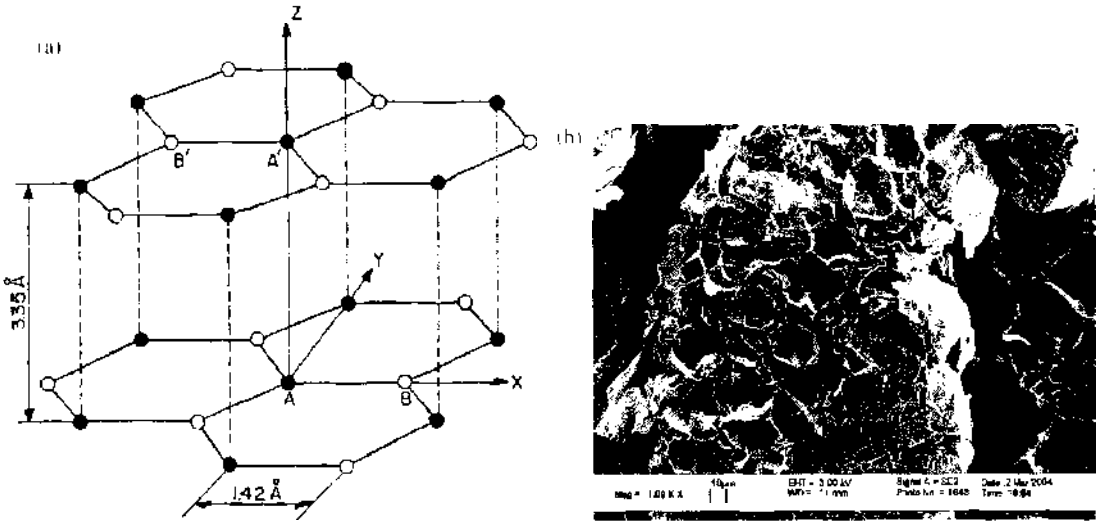


Figure 16. (a) Structure of natural graphite, hexagonal arrangement of covalently bound carbon atoms in-plane with AB stacking of planes held together by van der Waals forces. Reprinted with permission from [107], D. D. L. Chung, *J. Mater. Sci.* 37, 1475 (2002). © 2002, Springer Science and Business Media. (b) the wormlike structure of thermally expanded graphite. Reprinted with permission from [108], A. Yasmin et al., in "Proceedings of the 11th Annual International Conference on Composites/Nano Engineering (ICCE-11)," Hilton Head Island, South Carolina, 2004. © 2004, I. M. Daniel.

individual graphene sheets, to maximize the surface area interactions between the nanosheets and the host polymer matrix. Expansion of acid intercalated graphite is well established, producing characteristic wormlike structures (see Fig. 16(b)) and increasing the volume by several hundred times [107]. Natural graphite has a c-spacing of 3.35 Å (see Fig. 16(a)), which is increased several times by acid intercalation. By heating intercalated graphite to over 900°C, rapid thermal expansion occurs in which pockets of intercalant vaporize, causing shearing of the graphene planes with respect to each other. The resultant wormlike structures of expanded graphite are increased hundreds of times in volume and can be used in polymers to create nanocomposites. Additional processing including pulverization and sonication methods have also been pursued to reduce the nanographite to separated nanosheets [109].

1.3. Introduction to Viscoelasticity

The small size scale and surface to volume ratio of nanoparticles causes their interaction with the polymer chains in nanocomposites to be significant and extensive. As the polymer radius of gyration is also on the nanometer length scale, polymer mobility in the vicinity of the nanoparticles can be fundamentally altered from that of bulk polymer. With good dispersion of nanoparticles, a vast amount of surface area interacts with the polymer chains and at even low volume fractions can cause a dramatic change in the macroscopic manifestations of polymer mobility. Properties such as the glass transition temperature, thermal degradation temperature, time-dependent relaxation modulus, frequency-dependent storage and loss moduli, diffusion coefficient and relaxation times of the nanocomposite all differ from that of the bulk homopolymer. To set the stage to understand and be able to model these properties in the context of nanocomposite response, in this section we outline some fundamental concepts in viscoelasticity theory. The reader is referred to more detailed texts [110–113] for additional background and information.

The classical theory of linear elasticity assumes that the stress and strain within a material are directly proportional in accordance with Hooke's Law, $\sigma = C\varepsilon$, where C is the stiffness tensor of the material and σ and ε are the stress and strain in the material, respectively. The response of an elastic material is independent of strain rate (and thus time and frequency) and temperature (other than thermal expansion effects). For an elastic material, removal of the applied stress (or strain) implies the material will return to its pre-deformed shape, such that the energy required to produce the deformation is recovered. Elasticity theory properly describes the mechanical behavior of a wide range of solids, particularly at low temperatures

and low levels of strain. On the other hand, for a wide range of fluids, the state of stress is proportional to the strain rate of the material but is independent of the strain, and is described for a Newtonian fluid as $\tau = \eta\dot{\varepsilon}$, where η is the viscosity of the fluid and τ and $\dot{\varepsilon}$ are the stress and the rate of strain of the fluid, respectively. The energy required to deform a Newtonian fluid cannot be recovered.

The Hookean law for elastic solids and the Newtonian law for viscous fluids are each idealizations of material behavior, however, and do not represent accurate mechanical models for a wide range of materials. In particular, these models are insufficient to describe the mechanical behavior of most polymer systems, whose behavior can be described as having both elastic and viscous characteristics. Specifically, given an applied deformation some of the energy input into the system is stored within the material (elastic response), while some of the energy is dissipated as heat (viscous response). Materials that demonstrate such behavior are better described by viscoelastic models, which incorporate both elastic-like and viscous-like response characteristics. For such materials, the mechanical response is time-dependent and can be described in terms of an integral equation of the form

$$\sigma_{ij}(t) = \int_{-\infty}^{t} C_{ijkl}(t-\xi)\frac{d\varepsilon_{kl}(\xi)}{d\xi}\,d\xi \tag{3}$$

where σ_{ij} and ε_{kl} are the standard stress and strain tensors and $C_{ijkl}(t)$ is the time-dependent stiffness tensor. The viscoelastic behavior demonstrated by polymers is a direct consequence of the complicated molecular motion that must underlie any mechanical deformation, as is described further in the next section.

1.3.1. Molecular Theory of Polymers and Viscoelasticity

A polymer can be defined as a substance composed of molecules that have long sequences of one or more species of atoms or groups of atoms linked to each other by primary, usually covalent, bonds [114]. Whereas deformation of a solid can be simply thought of as displacements of the atoms from an equilibrium position, polymer deformation requires highly cooperative motion amongst adjacent polymer chains (and perhaps between different segments of the same polymer chain). When subject to a given load (or deformation), instantaneous rearrangements of the polymer chains result in an initial configuration of the local polymer chains that represents the (momentary) minimum free energy of the system. However, if the state of load (or deformation) is maintained over time, long-range cooperative motion of the polymer chains will result in different minimal free energy configurations. While rearrangements on a local scale are relatively rapid, the long-range cooperative motion among the polymer chains can be quite slow; this results in the range of relaxation times that typically characterize viscoelastic behavior. It is this continual rearrangement of the polymer chains that results in the viscoelastic behavior demonstrated by most polymers. The critical parameter describing how the polymer will respond to an applied strain (or stress) is the mobility of the polymer chains. The mobility of the polymer chains is influenced by both the chemical and geometric structure (e.g., the length of the chains, the size of the side groups which are attached to the backbone chain, and entanglements and/or cross-links among the chains) and the available thermal energy. Thus, polymers demonstrate both time-dependent (based on the range of time scales that describe various configurational rearrangements of the chains) and temperature-dependent properties.

The glass transition temperature T_g is a characteristic temperature for a given polymer below which thermal motions of the individual chains are greatly restricted [111]. This glass transition temperature can be considered using the concept of free volume. Free volume is the unoccupied "empty space" on the nanoscale within which the polymer chains can accommodate their configurational rearrangements. At sufficiently high temperatures, enough free volume is present such that the chains can instantaneously achieve their equilibrium volume and thus the material is in thermodynamic equilibrium. Polymers at such temperatures are soft and this state is referred to as the rubbery region of the mechanical response (for thermoplastic polymers the material may be in the melt state). However, as the temperature is reduced, the amount of free volume within the polymer decreases, until eventually the

molecular motion of the chains is impeded due to a lack of available free volume. Because of an insufficient amount of free volume (or alternatively, thermal energy), chain motion is restricted and the viscoelastic properties of the polymer are largely independent of time (or frequency). This is referred to the glassy or pseudo-elastic state of the polymer. The temperature marking the critical free volume at which this transition occurs is called the glass transition temperature, and as shown schematically in Fig. 17 is characterized by a discontinuity in the coefficient of thermal expansion (and also by a discontinuity in the heat capacity) of the material. While it is common to treat T_g as a discrete temperature, in reality slight deviations in the chain configurations at the nanoscale will result in a continuous transition from glassy to rubbery behavior). Polymers below the glass transition temperature are in thermodynamic nonequilibrium and over time will undergo a reversible process known as physical aging to gradually evolve toward the equilibrium state [111, 115].

The difference in mechanical behavior between the glassy and rubbery regimes is most evident when these material properties are measured as a function of temperature (see Fig. 18). The glassy regime of behavior (here for $T < \sim 150°C$) is characterized by a stiff material response and a relatively constant storage modulus (see Section 1.3.2). As the test temperature passes through T_g, the storage modulus quickly decreases. For temperatures greater than T_g, the polymer response is described as rubbery, and displays a storage modulus that is orders of magnitude less than that of the glassy region. The behavior of the storage modulus in this region is strongly dependent on the chemical structure of the polymer. For thermoplastic polymers, the storage modulus will continually decrease as a function of temperature as the polymer softens and ultimately melts at the melting temperature. For a thermoset system, the rubbery region storage modulus will plateau at a relatively constant value until the polymer begins to degrade at sufficiently high temperatures. Also note that the viscoelastic behavior of polymers is frequency-dependent, with testing at higher frequencies providing more elastic-like (short timescale) behavior relative to testing done at lower frequencies.

1.3.2. Mechanical Response of Viscoelastic Materials

The molecular structure of polymers results in their viscoelastic mechanical properties. The viscoelastic time-dependent modulus can be characterized by a Prony series representation of the form

$$E(t) = E_\infty + \sum_{j=1}^{N} E_j e^{-t/\tau_j} \tag{4}$$

where E_∞ is the rubbery asymptotic modulus, E_j are the Prony series coefficients, and τ_j are the relaxation times. Taking the half-sided Fourier transform of Eq. (4) yields

$$\bar{E}(\omega) = \int_0^\infty E(t) e^{-i\omega t}\, dt \tag{5}$$

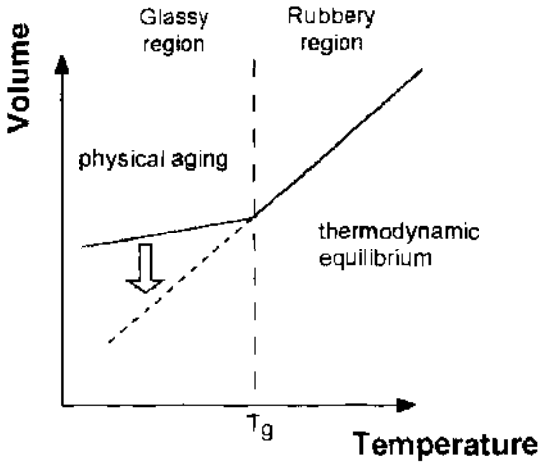


Figure 17. The glass transition temperature and physical aging.

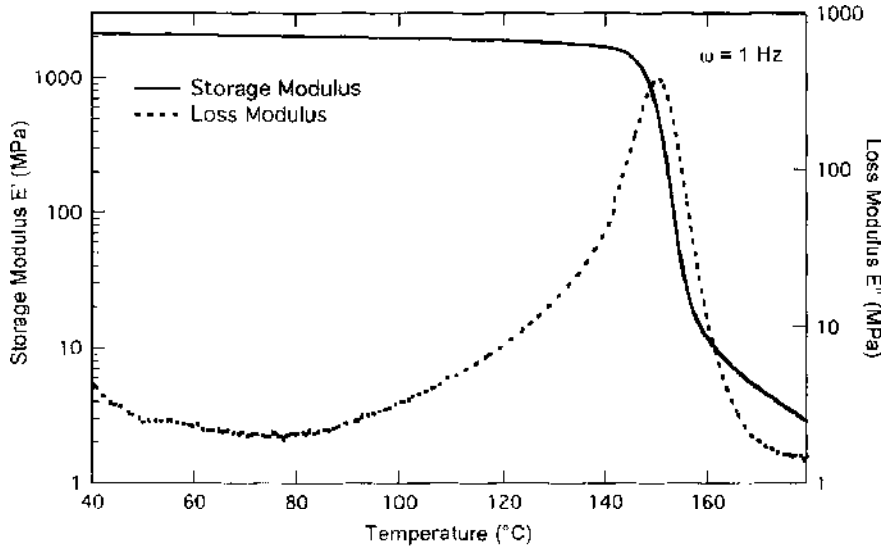


Figure 18. Temperature dependence of the storage and loss modulus for polycarbonate at $\omega = 1$ Hz and a heating rate of 3°C/min.

where i is the imaginary number. The frequency-domain response is described by the complex modulus $\bar{E}^*(\omega)$, which is defined as

$$\bar{E}^*(\omega) = E'(\omega) + iE''(\omega) \tag{6}$$

$$E'(\omega) = E_\infty + \sum_{j=1}^{N} \frac{E_j \omega^2}{1/\tau_j^2 + \omega^2}$$

$$E''(\omega) = \sum_{j=1}^{N} \frac{(E_j/\tau_j)\omega}{1/\tau_j^2 + \omega^2} \tag{7}$$

The terms defined in Eq. (7) are referred to as the storage modulus $E'(\omega)$ and loss modulus $E''(\omega)$, respectively. Example data for the storage and loss modulus as a function of temperature for a polycarbonate sample tested at $\omega = 1$ Hz is shown in Fig. 18. The storage modulus is a measure of the energy stored and recovered by a viscoelastic material per cycle of sinusoidal deformation, while the loss modulus E'' is a measure of the energy dissipated during a similar cycle [112]. Note that as written in Eq. (6) both the storage and loss moduli are real quantities. The ratio of the loss modulus to the storage modulus is referred to as the loss tangent $\tan\delta$, such that

$$\tan\delta = \frac{E''}{E'} \tag{8}$$

The loss tangent is the ratio of energy lost to energy stored in the deformation cycle and is a dimensionless parameter.

Note that as the number of the Prony series elements N representing the time-dependent modulus $E(t)$ in Eq. (4) goes to infinity, one obtains a continuous spectrum of relaxation times describing the viscoelastic response of the polymer. In this case, the time-dependent modulus can be expressed as

$$E(t) = E_\infty + \int_{-\infty}^{\infty} H(\tau) e^{-t/\tau} d(\ln \tau) \tag{9}$$

where $H(\tau)$ is referred to as the relaxation spectrum. The relaxation spectrum represents the infinitesimal contributions to the modulus from relaxation times lying in the range from τ to $\tau + d\tau$ and as such is useful in qualitatively gauging the distribution of relaxation mechanisms at different time scales [111]. Alfrey's first-order approximation of the relaxation spectrum, which is defined as the negative slope of the time-dependent modulus [111], can

be written as:

$$H(\tau) \approx -\frac{dE(t)}{d\ln(t)}\bigg|_{t=\tau} \tag{10}$$

Assuming a Prony series representation of the time-dependent modulus as given in Eq. (4), the relaxation spectrum can be approximated via the analytical expression

$$H(\tau) \approx \sum_{i=1}^{N} \frac{\tau}{\tau_i} E_i e^{-\tau/\tau_i} \tag{11}$$

Equation (11) has been used by the authors to obtain the changes in relaxation spectra for different nanotube-polymer composites as a function of nanotube loading, using the Prony series elements determined from curve-fitting experimental frequency domain data using the procedure described in Section 2.4.1.2 [116, 117].

While experimental measurements of the polymer response can in theory be performed for any length of time, time scales on the order of months or years are typically impractical. To circumvent this difficulty, one can use the principle of *time-temperature superposition*, a method of reduced variables, to extend the time scales (by many orders of magnitude) of the response at a particular temperature of interest. The basic premise of time-temperature superposition is that the material behavior at different temperatures can be superposed via horizontal shifting in log–log space to form a reference (or master) curve at a given temperature. Mathematically, the shifting of material response data collected at different temperatures can be expressed as

$$E(t, T_1) = E\left(\frac{t}{a_T}, T_{\text{ref}}\right) \tag{12}$$

where a_T is the temperature shift factor and T_{ref} is the reference temperature of the master curve. The reader is referred to any classical book on viscoelasticity for a more precise treatment; see for example [110, 111, 113].

Frequency-domain data collected at different temperatures can be shifted in an equivalent manner, providing the material response at a given temperature over a large range of frequencies. An example of the application of time-temperature superposition as applied in the frequency domain is shown in Fig. 19. Here experimental data using dynamic mechanical analysis (DMA) can only be obtained over a rather limited frequency range spanning approximately three orders of magnitude. Using time-temperature superposition, data collected over an experimentally accessible range at constant temperatures can be used to assemble a master reference curve at a single temperature spanning a much larger range of frequencies. Consider a reference temperature T_{ref} for which the frequency domain response of the

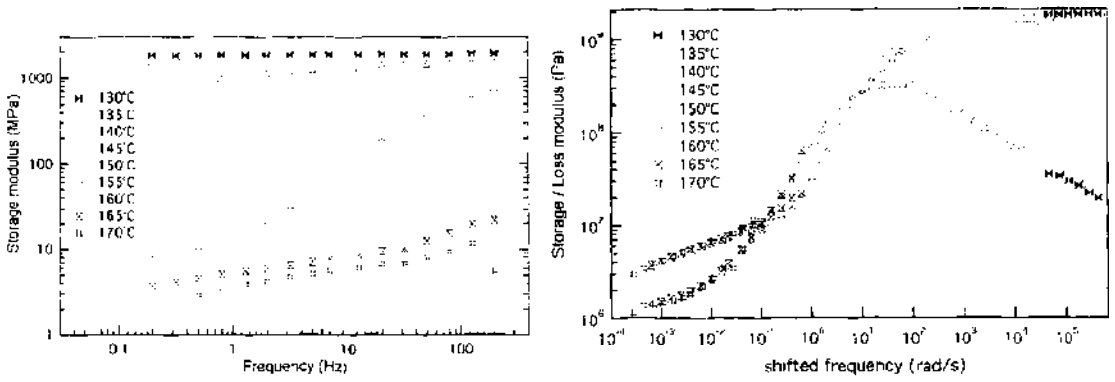


Figure 19. Time-temperature superposition. (left) Raw data (only storage modulus shown) for polycarbonate over a frequency range of 0.1 to 200 Hz limited by the range of a dynamic mechanical analysis machine. (right) Shifted experimental data to provide a master curve of the frequency domain response at a single temperature (here $T_{\text{ref}} = 150°C$).

polymer is desired. At temperatures greater than the reference temperature, the molecular response of the polymer chains will be accelerated as more thermal energy is available for polymer chain reorientation; thus the mechanical response at this elevated temperature represents the behavior of the polymer (at the reference temperature) corresponding to longer timescales (or inversely, smaller frequencies). Likewise, at a temperature below T_{ref}, molecular reorientation is correspondingly slower than that at the reference temperature, thus representing the shorter timescale (or larger frequency) behavior of the polymer (see Fig. 19).

Occasionally slight vertical curve shifting will be required (this is true for both the time and frequency domain experimental data); often the basis of this vertical shifting is the temperature-dependence of the polymer density. Because time-temperature superposition is only discussed here as a means to extend the range of frequencies available for the collection of experimental data, further discussion of topics associated with time-temperature superposition (such as the standard Williams-Landis-Ferry (WLF) representation of the temperature shift factors) is not warranted here. The reader is referred to the literature for further reading in this area [110, 111].

2. NANOREINFORCED POLYMERS—EXPERIMENTAL RESULTS

The focus of the chapter is the modeling of polymer nanocomposites; however, it is important to consider experimental results that guide the modeling efforts. In the following sections we will highlight key experimental work for the different classes of nanoscale inclusions, with a focus on nanotube reinforced polymers. We will also present experimental evidence for the formation of a nonbulk phase in these materials due to interactions between the polymer chains and the nanoinclusion surfaces. An excellent review of experimental results on many types of polymer nanocomposites may be found in a book chapter by Schadler [19]. While only a few results are mentioned explicitly in the text below, the Appendix presents a more complete summary of the polymer nanocomposites literature.

While many of the experimental results on polymer nanocomposites to date rely on methods borrowed from traditional polymer experimental characterization (e.g., DMA, DSC, TGA,...), a number of researchers are developing newer experimental techniques to allow nanoscale mechanical characterization of these materials. In addition to these methods to determine the mechanical behavior of these materials over multiple length scales, a number of experimental techniques from analytical chemistry (XPS, FTIR, NMR,...) and material science (XRD, SEM, TEM,...) are also necessary for complete understanding of nanocomposites. As discussed earlier, there are also a number of issues related to the fabrication of these nanocomposites, such as dispersion and alignment, that must be addressed when fabricating nanocomposite samples. A great deal of progress has been made in this area; the reader is referred to Ref. [19].

2.1. Equiaxed Nanoparticulate Reinforcement

The most commonly used equiaxed nanoparticulate fillers include silica (SiO_2), titania (TiO_2), and alumina (Al_2O_3), although a large number of equiaxed nanoparticles are available [19]. In many applications, these equiaxed nanoparticles are used to impart improvements in thermomechanical and wear properties over those of the unfilled polymer. In other applications, optical and/or electrical properties are the target enhancement. Here, we briefly present selected experimental results published in the literature regarding property enhancements for equiaxed nanoparticle-polymer composites (additional references are provided in Table A.1 in the Appendix). The reader is referred to these articles, and the references contained therein, for additional information and discussion.

As discussed previously, nanoparticle dispersion and surface chemistry are critical issues for these nanoreinforced systems. In one study of nano-SiO_2 (70–110 nm particle size) in PMMA, it was reported that appropriate surface treatment to make the SiO_2 more hydrophobic greatly enhanced the dispersion and led to substantially higher T_g's (as measured by DMA) of the nanocomposite as shown in Fig. 20 [118]. Another study (10 nm silica

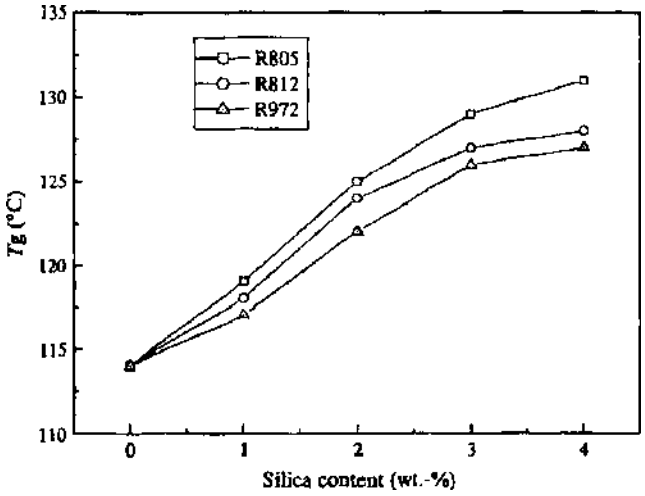


Figure 20. Change in glass transition temperature as a function of nanosilica loading in PMMA (designations in legend refer to slightly different chemistries with strong interactions with the polymer). Reprinted with permission from [118]. Y. H. Hu et al., *Polym. Degrad. Stabil.* 84, 545 (2004). © 2004, Elsevier.

particles in a thermoplastic copolymer) found similar improvements in dispersion and properties with surface treatment, such that a 10 vol% treated nanoparticulate formulation resulted in a forty-fold increase in storage modulus at 170°C [119]. The impact of the size of the nanosilica agglomerate and the amount of bound rubber (controlled via surface chemistry) on the mechanical properties of a nanosilica-filled rubber have also been analyzed [120].

Consistent with the earlier discussion, surface area and interphase formation can cause significant differences in the temperature and mechanical behavior of these nanocomposites. For example, 9 nm SiO_2 particles (with an apparent diameter of 25 nm due to agglomeration), loaded up to 55 wt% in styrene-butadiene, were characterized by a 2 nm bound interphase as characterized by a bimodal glass transition behavior (see Ref. [121] and references therein for an excellent discussion). In another study, 39 nm Al_2O_3 particles in PMMA resulted in an effective T_g decrease of 25°C at only 0.5 wt% loading; however, this T_g depression could be suppressed by coating the spherical nanoparticles to make them compatible with the PMMA matrix [122]. Tsagaropoulos and Eisenberg conducted a comprehensive study of surface interactions and interphase formation for 7 nm diameter nanosilica in a number of polymers [123], identifying a dual glass transition temperature due to the interphase. Huang and co-workers investigated the impact of nano-SiO_2 (particle size 10–15 nm) on the dynamic mechanical properties of linear low density polyethylene (LLDPE) [124]. Other workers have studied the nonlinear viscoelastic behavior of fumed nanosilica in PVA [125, 126] and nano-SiO_2 in a rubber elastomer [127].

The use of nano (as opposed to micro) particulates can in some cases fundamentally change the underlying mechanisms governing thermo-mechanical behavior of the system, although in many cases a detailed understanding of this process is still lacking. As an example, while in traditional composites applications increases in stiffness (modulus) due to the incorporation of stiff fibers often results in losses of ductility and impact resistance due to stress concentrations caused by the fillers; it has been proposed that such stress concentrations can be reduced when nanosized fillers are used [19].

In one case, a large increase in ductility with nano-Al_2O_3 (39 nm) dispersed in PMMA has been reported (see Fig. 21) [128]. As indicated earlier, the interfacial characteristics between the nanoparticles and the matrix largely control the property enhancements; particles that have weak interfaces with the matrix tend to have lower moduli and decreased glass transition temperatures, while stronger interfaces lead to increased moduli and stable or improved thermal properties. Analysis of toughness is more complex, but as with traditional composites a weak interphase can lead to significant improvements in toughness. The mechanisms for equiaxed nanoparticle reinforcement differ however from conventional fiber composites, where toughness arises in the energy dissipated in debonding, frictional sliding

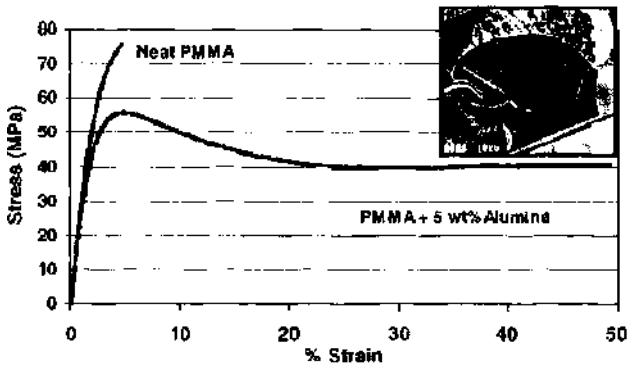


Figure 21. Example of increased nanocomposite toughness due to change of the failure mode for particles with a weak interface. Reprinted with permission from [128], R. Siegel et al., *Scripta Mater.* 44, 2061 (2001). © 2001, Elsevier.

and pullout. Here, there is evidence that a weak interphase leads to cavitation around the equiaxed nanoparticles, relief of local triaxial stress states, and the participation of material in shear yielding rather than crazing. In well-dispersed nanocomposites, this ultimately leads to homogenous shear yielding and can result in an order of magnitude increase in toughness as demonstrated in Fig. 21 [128].

Other studies have also investigated the impact of equiaxed nanoparticulates on nanocomposite thermo-mechanical behavior. The addition of a low concentration of nano-SiO_2 (20 nm) to a rubber-toughened epoxy led to a significant increase in the toughness of the structural adhesive [129], while the addition of nano-TiO_2 was found to improve the creep resistance of PA6,6 [130]. Low loading of surface-modified nanosilica (7 nm) in polypropylene led to an improvement in tensile properties, attributed to a double percolation of yielded zones [131]. Epoxy-silica nanocomposites (particle size 10–20 nm) were fabricated with loadings up to 70 wt%, leading to improved thermal stability but a decrease in the glass transition temperature, attributed to a plasticization effect [132].

Another application of equiaxed nanoparticles in polymers of particular interest is wear-resistant coatings. An excellent review of various experimental results regarding the wear resistance of polymer nanocomposites is given in Ref. [133]. In many cases, equiaxed nanoparticles have been shown to simultaneously increase the wear resistance and decrease the coefficient of friction, whereas for larger-sized particles an increase in wear resistance is typically accompanied by an increase in coefficient of friction [19]. For example, PET filled with 1 to 10 wt% nanosized Al_2O_3 (38 nm average diameter) found the wear resistance doubled compared to that of the pure PET with a decrease in average coefficient of friction (for loadings less than 5 wt% Al_2O_3) [134].

A comparison of micron- and nanosized copper particles found the fundamental wear mechanisms of the composite changed when nanosized particles were used (125 μm vs. 10 nm) [135]. Figure 22 shows that a number of studies have found that the incorporation of equiaxed nanoparticulates within polymers have shown improved wear rates compared to those of the pure polymer (please see the original reference for further discussion and the citations for the work references in the figure [133].) For example, abrasion-resistant nanocomposites based on nano-$CaCO_3$ and PMMA found that the average abrasive weight loss was halved with respect to neat PMMA by adding only 2% nanoparticulate [136]. In another study, the addition of 1–2 vol% of nano-Al_2O_3 to an epoxy was found to simultaneously increase the stiffness, impact energy, and failure strain, in addition to a slight improvement in wear resistance attributed to different wear mechanisms for the composites with nano- and micron-sized inclusions [137]. A recent work reported that an improved wear resistance for a modified epoxy was found by combining nano-TiO_2 with conventional microsized solid lubricants [138].

The sampling of results presented here suggests that the use of equiaxed nanoparticulate reinforcement within a polymer matrix may result in nanocomposite materials with enhanced thermomechanical and wear properties for technological applications. In many cases, the enhanced properties are believed to be due to changes in the fundamental mechanisms

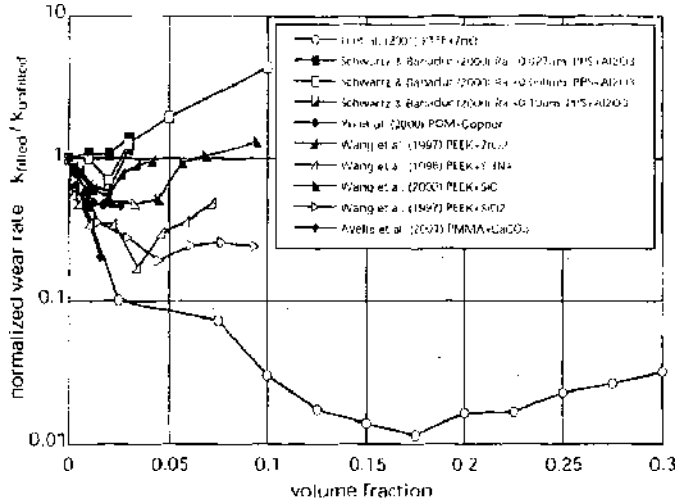


Figure 22. Normalized wear ratio improvements for various equiaxed nanoparticulate composites. Please refer to the original article for citations of work referenced in the figure. Reprinted with permission from [133], W. G. Sawyer et al., *Wear* 254, 573 (2003), © 2003, Elsevier.

governing the response of the nanocomposite compared to those mechanisms controlling the behavior of the pure polymer. Further elucidation of the governing mechanisms at the nanoscale, and their relationship to the exhibited macroscale behavior, are necessary to enable control of these mechanisms and ultimate optimization of these materials and their response. Continued experimental efforts, in conjunction with molecular and multiscale modeling, will enable progress to continue in this direction.

2.2. Nanotube Reinforcement

Research in the area of nanotube-reinforced polymers has exploded in the last several years motivated by the desire to leverage the outstanding physical properties of nanotubes as a multifunctional nanoscale filler in polymeric materials. Despite both the amount and quality of work that has been performed in this area to date, it may be unsettling (although not surprising) to consider that many of the fundamental mechanisms underlying the effective response of these nanocomposites are still the subject of debate. Such a state of affairs can be attributed to the complexity of interpreting experimental results for these materials, which are often dependent on the source of the nanotubes and, more critically, on the processing steps used to produce samples for experimental testing. Differences in polymer chemistry and the complexity of the viscoelastic response of polymers (see Section 1.3) also complicate the slowly emerging picture of nanotube-polymer composite behavior. Thus the authors feel that the timing of the current chapter is appropriate, as we are approaching the stage in nanocomposite development where further advances in this field will be dependent on a stronger coupling of modeling efforts with experimental programs in order to elucidate the basic principles governing the material response.

In this section, we will present a brief sampling of representative experimental work reported in the literature regarding nanotube-polymer composites. Where appropriate, we will focus the discussion on our own experimental efforts. In general, while experimental results achieved to date are promising, the large scatter in data and the inability in many cases to achieve the level of property improvement predicted by straightforward modeling highlight the need for continued development in this field. As one example consider the experimental data shown in Fig. 23 for CVD-grown MWNT-reinforced polystyrene. Here the experimental data is compared with simple predictions for the effective composite modulus E_c obtained using the standard rule of mixtures approximations ($E_c = \sum c_i E_i$ and $E_c^{-1} = \sum c_i / E_i$ for parallel and series models, respectively) and the Mori-Tanaka method (assuming a 3D random orientation of the nanotube inclusions; the Mori-Tanaka method is described in detail in Sections 3.2.1.1 and Appendix A.1). Assuming a nanotube modulus of 450 GPa (to account for the level of defects expected in CVD-grown MWNTs; see Ref. [37]), the modulus enhancement observed experimentally is much less than predicted by this simple

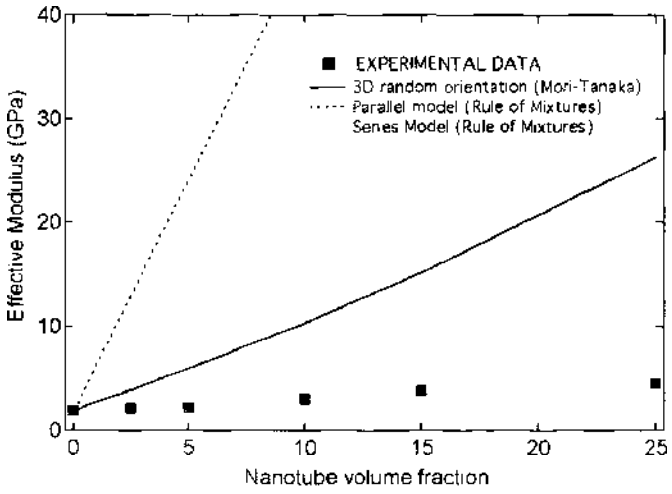


Figure 23. Comparison of experimental data for MWNTs in polystyrene [139] with rule of mixtures and Mori-Tanaka predictions assuming a nanotube modulus of 450 GPa. Reprinted with permission from [140], F. T. Fisher, Ph.D. dissertation, Northwestern University, 2002, © 2002.

modeling. This again emphasizes the need for better mechanical models to assist in the interpretation of nanotube-polymer composite experimental results, as a better understanding of the effective response of these systems will lead to improved fabrication processes and chemistries to further push the observed effective properties towards their optimal values.

The organization of the remainder of this section is as follows: in Section 2.2.1 we will discuss three critical issues related to nanotube-polymer composites: dispersion and alignment of the embedded nanotubes and load-transfer in nanotube-polymer composites. A brief sampling of experimental results for the mechanical properties of these systems will be presented in Section 2.2.2, and a summary of experimental efforts focusing on electrical and thermal properties of nanotube-reinforced polymers will be presented in Section 2.2.3. A discussion of the viscoelastic properties of these nanocomposites, with a focus on the interactions between the nanotubes and the polymer chains, will be delayed until Section 2.4. While not comprehensive, we hope that the references assembled by the authors in the Appendix (see Table A.1) will provide the interested reader a useful resource with which to peruse the literature. Additional information on nanotube-polymer composites can be found in a number of excellent review articles [141–144].

2.2.1. Issues Related to Processing of Nanotube-Reinforced Polymers

While there are a number of difficulties associated with fabricating nanocomposite samples, over the past several years a great deal of progress has been made in this area (see the book chapter by Schadler for a more complete discussion of fabrication and processing of nanocomposites [19]). One issue of critical importance, common for all nanocomposite applications, is obtaining better dispersions of the nanoparticles within the polymer, and progress in this area for nanotube-polymer composites has been realized. Further techniques and chemistries to more thoroughly separate NTs to maximize the nanotube surface area available for interactions with the polymer are continuously under development. In addition, significant efforts are underway to better control the alignment of the nanotubes within the polymer as discussed in Section 2.2.1.2.

Finally, a significant open question is whether, for a particular polymer system, the interface between the embedded nanotubes and the polymer is sufficient for full load transfer to the reinforcement nanotubes; more work from both an experimental as well as a modeling perspective is needed in this area. Thus several groups are attempting to utilize chemical functionalization to covalently tether the nanotubes to the polymer matrix to enhance load transfer between the phases; while such functionalization has been shown to result in enhanced mechanical properties (in comparison to the unfunctionalized nanotube-polymer samples), predictive modeling of the impact of functionalization is at the moment lagging the developments in functionalization chemistry. The impact of functionalization on viscoelastic behavior and nonbulk interphase formation in nanotube-polymer composites will

be discussed in Section 2.4. While the experimental results covered within are promising, continued advances in the areas of fabrication and experimental testing highlight the need for more accurate modeling techniques to better understand the fundamental mechanisms at the nanoscale which govern the behavior of these materials.

2.2.1.1. Nanotube Dispersion within the Polymer. One issue of practical importance for nanotube-reinforced polymers (NRPs) is the separation and dispersion of the nanotubes within the matrix, which is critical as the nanotubes tend to assemble into ropes or bundles due to van der Waals interactions between the individual tubes. While some researchers have been able to separate individual nanotubes from the bundles via ultrasound and polar solvents, maintaining separated nanotubes during the processing of NRPs is still the subject of ongoing work. Some results suggest that the use of a surfactant as a coupling agent may overcome the van der Waals attractive force and allow good dispersion of the nanotubes within the polymer [145]. However, it is unclear whether such processing agents can be employed to promote nanotube dispersion without compromising the nanotube-polymer interface. While the use of harsh acid treatments and chemical functionalization of the nanotube has been shown in several cases to promote nanotube dispersion, such improvements must be balanced with the reduced nanotube properties which result from the creation of defects in the pristine nanotube structure. Because nanotube dispersion is closely related to the van der Waals interactions predominant at the nanoscale, nanotube dispersion is typically more difficult as the diameter decreases.

A number of studies are available in the literature related to the improved dispersion of nanotubes within host polymers. For example, a recent study of the effect of different solvents on the dispersion of SWNTs in epoxy via sonication found solvent effects were most pronounced for, in order, dimethylformamide (DMF) > ethanol > acetone; the effect of solvent choice on the extent of cure reaction was also noted [146]. Noncovalent functionalization of SWNTs with poly(phenyleneethynylene) (PPE) has been shown to drastically improve the dispersion of SWNTs in both polycarbonate and polystyrene [147], whereas plasma coating of Pyrograf III carbon nanofibers (with diameters from 70–200 nm) with pyrrole [148] and polystyrene [149], respectively, has also been used to improve dispersion. Dispersion has also been enhanced by coating SWNTs with poly(vinyl pyrrolidone) (PVP) and sodium dodecyl sulfate (SDS) [150]. Considerable improvement of the dispersion of SWNTs in epoxy based on an acid treatment-flourination treatment has also been reported [151]. A coagulation method has also been found to enhance the dispersion of SWNTs in PMMA [152], while shear mixing using a Haake Polylab bowl mixer has been reported to yield excellent dispersion [139]. As shown in Fig. 24, the level of nanotube dispersion depends strongly on the processing parameters.

One issue complicating the evaluation of dispersion within a polymer is an appropriate method with which to quantify the nanotube dispersion within the polymer. While optical methods are limited to rather large length scales governed by the wavelength of light, SEM techniques are typically limited to fracture surfaces (see Fig. 25). AFM methods, to our knowledge rarely used for dispersion evaluation, are at best only able to probe to very small

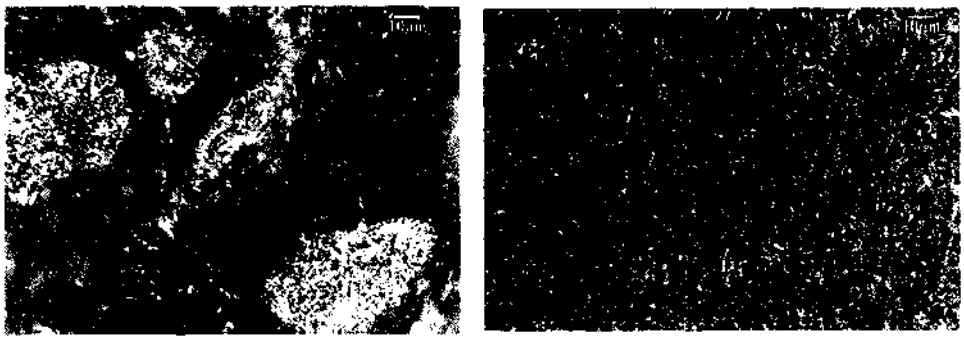


Figure 24. Photomicrographs of (left) poor and (right) good dispersion of CVD-grown multiwalled carbon nanotubes in a polymer. Reprinted with permission from [139]. R. Andrews et al., *Macromol. Mater. Eng.* 287, 395 (2002). © 2002, Wiley-VCH.

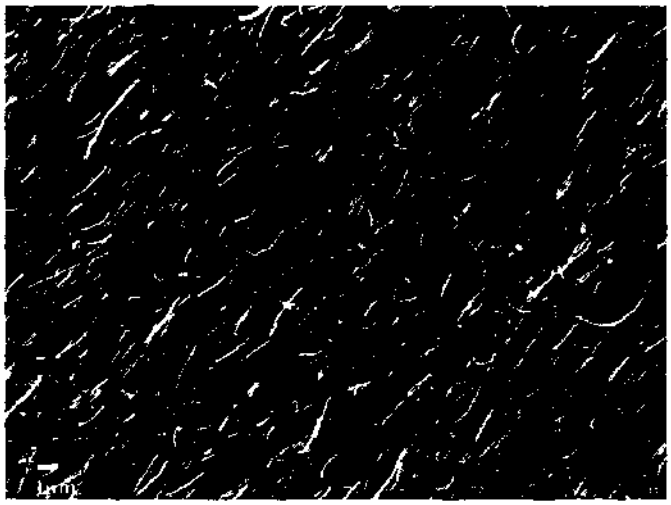


Figure 25. SEM image of MWNT-polycarbonate fracture surface showing excellent nanotube dispersion. Reprinted with permission from [116], A. Eitan et al., *Compos. Sci. Technol.* (2006), © 2006, Elsevier.

sub-surface depths. While TEM methods are most suitable for the highest resolution study of dispersion, such analysis requires tedious TEM specimen preparation to microtome sample thickness thin enough for electron transparency. Even if such an analysis is undertaken, it is necessary to justify that the dispersion obtained over an extremely small area of the sample is representative of dispersion throughout the entire sample. In addition to smaller diameter nanotubes being more difficult to disperse in polymers, the evaluation of their dispersion is likewise more complicated; consider, for example, that SWNTs are typically below the resolution of scanning electron microscopy.

2.2.1.2. Controlled Orientation of the Nanotubes. Material properties may be optimized if the orientation of the nanotubes within the polymer can be controlled, and several techniques have been proposed to address this issue. One group found that cutting thin slices (on the order of 100 nm) of a nanotube-reinforced epoxy film introduced preferential orientation via shear flow [153]. This flow orientation method has also been used to orient small amounts of NTs (0.1% wt) in a urethane acrylate polymer to thicknesses up to 150 μm [154, 155]. An alternative method that may be more suitable for larger samples is tensile loading of the NRP at temperatures above the glass transition temperature of the polymer [156, 157]. A combination of solvent casting and melt mixing was also found to produce a high degree of nanotube alignment [158], while spin casting [159] and extrusion have also been used to induce nanotube alignment within the nanocomposite [160, 161]. While individual SWNTs and SWNT ropes have been aligned in the presence of electric [162] and magnetic [163] fields only recently has this work been extended to nanotube-reinforced polymers; see, for example, [164] and [165, 166], respectively.

Recent work has used a melt-spinning process to create PMMA/nanotube fibers in which a high degree of alignment of the nanotubes is obtained; larger samples are then made from hot pressing a sheaf of aligned PMMA/NT fibers [158, 167]. While one might anticipate higher stiffnesses for aligned samples, this work shows a decrease in the low frequency storage modulus for these samples with the explanation that alignment leads to decreased tube-tube interaction and consequently decreased effectiveness of the nanotube network to stiffen the composite. Additional work, described in more detail in Section 1.2.2.4, is seeking to exploit the electrospinning process to create polymer nanocomposite nanofibers with alignment along the axial direction induced during the electrospinning process; see, for example, work with carbon nanotubes [86] and montmorillonite [83]. Alignment within a polymer via stretching has also been used as a preliminary step to create aligned fibers of SWNTs (postalignment the polymer is removed by annealing of the nanocomposite at 1000°C in hydrogen) [168]. Based on these results, one could foresee using such polymer/aligned NT fibers to weave a composite with controlled, but non-uniaxial NT alignment, which could them mimic traditional composites in terms of ply layup to achieve desired isotropic to directionally preferred property values.

These efforts at alignment emphasize that models of nanotube-reinforced polymers must account for the orientation distribution of the nanotubes within the polymer for proper comparison between the model and experimental results. It is unlikely that nanocomposites will ever see the extreme control over orientation obtained in traditional continuous fiber composites, where pre-preg lamina or unidirectional aligned fibers are selectively stacked for a given application. However, the differences between random orientation, partial alignment or controlled network structure of nanotubes within a polymer nanocomposite will dramatically affect the overall composite thermal, mechanical and electrical properties and must be appropriately reflected in modeling strategies.

2.2.1.3. Load Transfer Across the Nanotube-Polymer Interface. Another area of considerable interest is the NT-polymer interface and load transfer between the polymer and the nanotubes [36, 39, 157, 169–173]. Poor load transfer for MWNTs and SWNT ropes embedded in a polymer has been attributed to the relative slipping of individual tubes within the MWNT [39] and the SWNT rope [171], respectively. However, other researchers have found evidence of promising nanotube-polymer interactions in composite materials. For example, a strong interface between MWNTs and polystyrene (PS) [36] and polyhydroxyaminoether (PHAE) [156] has been reported. Analysis of SWNT bundle-PMMA thin films found that PMMA was able to intercalate within the bundles, which would likely enhance the interface between the nanotube and polymer phases [174]. Significant wetting and interfacial adhesion for SWNT bundles embedded in an epoxy resin has also been reported [175]. In another system, SEM images of fracture surfaces show that the nanotubes are wrapped with PMMA layers, believed to indicate a strong interface between the phases and attributed to the initiator used within the polymerization process [170]. Other researchers found that NTs aligned parallel to the direction of crack propagation tended to break between the crack faces (rather than pullout from the matrix), leading to the conclusion that a relatively strong interface exists between the two phases [36].

A value of 500 MPa has been obtained experimentally for the interfacial shear strength for a polyurethane-NT system based on a fragmentation experiment, which is an order of magnitude larger than typically measured in conventional fiber-based composites [176]. (Compare with the value for the interfacial shear stress for a polystyrene-nanotube composite of 160 MPa calculated using molecular mechanics [177].) Because of this demonstrated adhesion within the urethane matrix, the use of NTs as nanoscale strain sensing devices has been suggested, where low fractions of nanotubes (0.1% wt) make the host polymer Raman-active, allowing changes in the Raman spectrum to be related to strain within the material [154, 178].

In addition, recent work by the Wagner group has demonstrated a novel AFM-based testing methodology to probe the interfacial shear strength of nanotube-polymer composites [179, 180]. In these experiments, a nanotube attached to an AFM tip was first placed into a heated (softened) polymer that was then cooled below the T_g as shown in Fig. 26. Subsequently, the AFM was used to pull the nanotube out of the polymer layer and the force-displacement monitored, providing an estimate of the shear strength of nearly 50 MPa. In addition, as the nanotube interface sustains stresses greatly exceeding the bulk polymer yield stress during the experiments, it was concluded that the properties of local polymer near the nanotube differ substantially from the bulk properties.

In situ Raman spectroscopy measurements of a nanotube-polymer sample under load were used to qualitatively measure the strain in the nanotubes as the samples were loaded, and it was reported that the Raman peak position only shifted significantly in compression. It has been suggested that this is evidence that only the outer layer of the MWNT was loaded in tension, whereas all layers of the MWNT are loaded in compression [39]. More recently, Raman spectroscopy has been used to quantify the impact of functionalization on load transfer in a MWNT-polycarbonate system [116]. As shown in Fig. 27, the epoxide-functionalization increases the load transfer from the PC to the MWNT, resulting in a greater shift in the Raman peak as a function of strain applied to the composite. The differences in the Raman response for different sources of SWNTs and MWNTs dispersed in the same epoxy has also been reported [181].

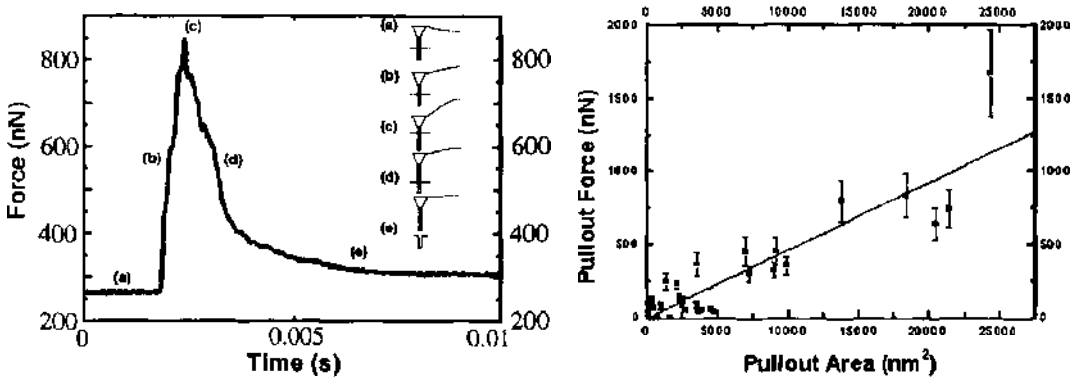


Figure 26. AFM-based pullout test for measurement of interfacial shear stress in nanotube-polymer composites. A nanotube pre-attached to an AFM tip is pushed into a heated (soft) polymer which is then rapidly cooled. Pullout force is measured based on the deflection of the AFM cantilever. (left) Representative plot of pullout force versus time. (right) Pullout force versus the embedded area of the nanotube within the polymer. The slope of the curve was used to calculate an interfacial separation stress of 47 MPa. Reprinted with permission from [180], A. H. Barber et al., *Appl. Phys. Lett.* 82, 4140 (2003). © 2003, American Institute of Physics.

To our knowledge the results shown in Fig. 27 represent the first report of a semi-direct experimental method to characterize the load transfer enhancement due to nanotube functionalization. More typically, load transfer enhancements are inferred based on improvements in bulk macroscopic experimental data; see, for example, an amide functionalization technique resulting in dramatic property improvement for 1 wt% loading of SWNTs in PMMA [182] and the *in situ* polymerization of chemically functionalized MWNTs in MMA monomer [183]. Characterization of the load transfer behavior in nanotube-polymer composites, and how it can be altered via functionalization strategies, is perhaps one of the most important questions presently unanswered in this area. We believe that the convergence of the expansion of the experimental methods described here with continued advances in computational modeling techniques such as those discussed in Section 3 (see, for example, Refs. [177, 184]) will result in significant advances in our understanding of the interface and load transfer in these systems. Such an understanding will be necessary to tailor the interface in these systems for optimal effective properties.

2.2.2. Mechanical Properties of Nanotube-Reinforced Polymers

As discussed in the previous sections, despite considerable recent progress there are still a number of issues related to the processing of nanotube-polymer composites that need

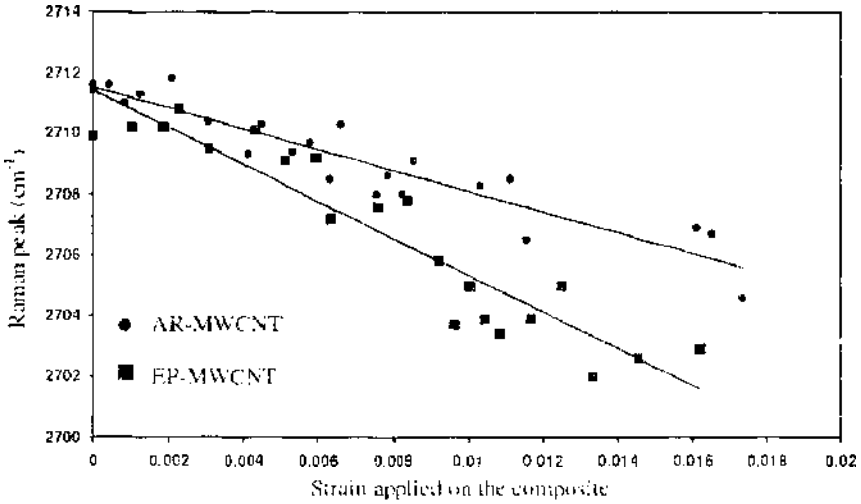


Figure 27. Shift in Raman peak as a function of composite applied strain. AR = as-received MWNTs. EP = epoxide-modified MWNTs. Reprinted with permission from [116], A. Eitan et al., *Compos. Sci. Technol.* (2006). © 2006, Elsevier.

to be addressed. Additional complications hampering experimental interpretation include differences in properties and behavior between various sources of nanotubes, the complexity of the chemistries of the host polymers (often making it difficult to generalize results obtained for different polymers), differences in the chemistry of the fabrication technique, and differences in sample processing (which may impart various degrees of alignment in the sample). Such a state of affairs again emphasizes the need for the development of comprehensive nanomechanical models to go hand-in-hand with these experimental programs. Both the shear quantity and continually increasing number of papers presented in the literature prevents a thorough review of all experimental results here; thus the goal of the current section is to simply provide a summary of the current level of mechanical property enhancements that have been achieved to date using nanotubes as a reinforcing phase within a polymer matrix. The interested reader is referred to the collection of references provided in Table A.1, as well as a number of recent reviews [141–144], for additional work in this area.

Representative experimental values of the glassy, elastic-like modulus (typically elastic testing at room temperature; in all cases data was collected at temperatures well below the glass transition temperature of the host polymer) for *multiwalled* carbon nanotube-reinforced polymers are shown in Fig. 28. All data was normalized with respect to the modulus of the polymer. Also shown in Fig. 28 are 3D Mori-Tanaka micromechanical model predictions for the nanocomposite effective modulus assuming isotropic properties for each of the constituent phases with modulus ratios $E_{rat} = E_{NT}/E_{matrix}$ of 500 and 100 (see Section 3.2.1.1 and Appendix A.1 for further description of the Mori-Tanaka method). For clarity only data up to a volume fraction of 5% are shown on the plot; thus for some of the references additional results for larger volume fractions of MWNTs were truncated. In Fig. 28 data points that are characterized as "aligned" are done so based on claims made within the original reference; typically in such work the alignment is processing-induced and the degree of alignment is relative (compared to other samples processed using a different technique) rather than absolute. To date, true unidirectional alignment of nanotubes in a composite has not been achieved.

Two conclusions are readily apparent from analysis of Fig. 28: there is a large degree of scatter amongst the data, and most modulus enhancements are significantly less than those predicted using a modulus ratio of 500. Assuming a modulus of 1000 GPa for the nanotube, and given that structural polymer moduli are on the order of 1 to 3 GPa, a modulus ratio $E_{rat} = 500$ is considered representative of nanotube-polymer composites, whereas $E_{rat} = 100$

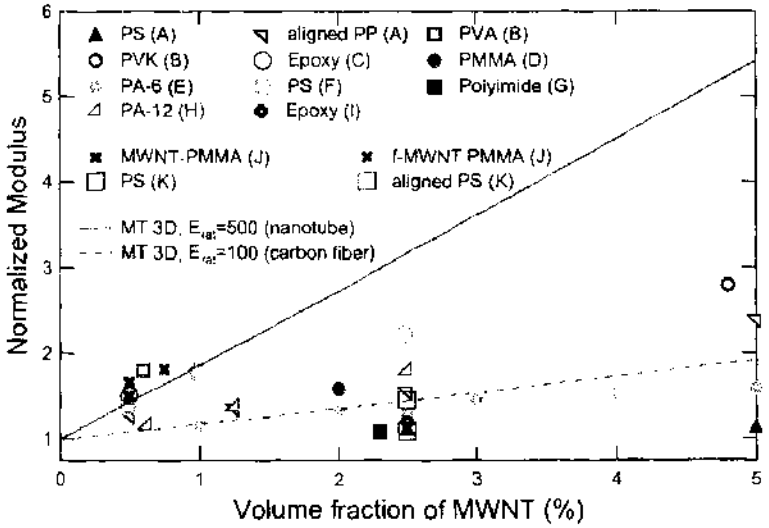


Figure 28. Representative experimental data for glassy modulus of MWNT-reinforced polymers normalized with respect to the modulus of the polymer. Also shown are Mori-Tanaka predictions assuming a 3D random orientation of nanotubes for two values of $E_{rat} = E_{NT}/E_{matrix}$. Key: (A) Ref. [139]. (B) Ref. [185]. (C) Ref. [145]. (D) Ref. [186]. (E) Ref. [187]. (F) Ref. [159]. (G) Ref. [188]. (H) Ref. [189]. (I) Ref. [39]. (J) Ref. [183]; f-MWNT refers to functionalized MWNTs. (K) Ref. [160].

is shown for purposes of comparison. Note that in most cases the experimental data is below that of the model predictions even though the model applied here assumes 3D random orientation of the nanotubes and thus could be considered conservative.

While the data included in Fig. 28 are but a small sampling of published experimental values, the following generalizations can be made which are consistent with additional interpretations found in the literature. Data points from (A) in Fig. 28 show that partial alignment of the embedded nanotubes (accomplished through spin-coating the nanotube-polymer samples) led to comparatively higher modulus enhancements [139]; see also data points from reference (K) in the figure [160]. Data points labeled (B) in Fig. 28 show that the enhanced crystallization of PVA (as opposed to PVK) due to the presence of the nanotubes led to higher modulus enhancement [185]; such increases are consistent with the hypothesis of an additional reinforcement mechanism provided by the embedded nanotubes, namely the nonbulk properties of polymer in close proximity to the nanotubes. Such an effect is also apparent from data point (C), which shows a large increase in modulus slightly above the 3D MT prediction for $E_{rat} = 500$; while this result is much higher than many of the other published data for MWNT composites, those researchers also saw a very significant increase in the T_g of the nanocomposite [145], consistent with the formation of a nonbulk interphase polymer region which, as above, has been attributed to additional reinforcement enhancement of nanotube-polymer composites [116]. This effect will be discussed in much greater detail in Section 2.4.

As discussed previously, another question is whether the inner walls of the embedded multiwalled carbon nanotubes actually provide reinforcement to the nanocomposite. Such a point was raised by analysis of data points from references (G) [188] and (I) [39] as presented in the original references; see also Ref. [38]. Lastly, an *in situ* polymerization process for MWNT-PMMA composites showed large levels of modulus improvement, an effect that was further enhanced by functionalization of the MWNTs as demonstrated by data points (J) in Fig. 28 [183]. That these points lie above the 3D MT predictions is consistent with the existence of an interphase region of altered polymer properties and higher modulus, which accounted for in the simple two-phase MT predictions considered in Fig. 28.

As shown in Fig. 29, those researchers also found that the functionalization also results in increases in toughness of the nanocomposite (defined as area under the stress-strain curve) [183]. Other researchers point out the complexity for toughening enhancements of polymers via the incorporation of nanotubes; while some researchers have found significant toughening of the nanocomposite in comparison to the base polymer, for example with PMMA (see Fig. 30) [190] and UHMWPE [191], recent work with vapor-grown carbon nanofibers found an absence of toughening attributed to enhanced crystallization of the polypropylene matrix [172]. Clearly, the details of the polymer matrix chemistry and structure, and its interaction

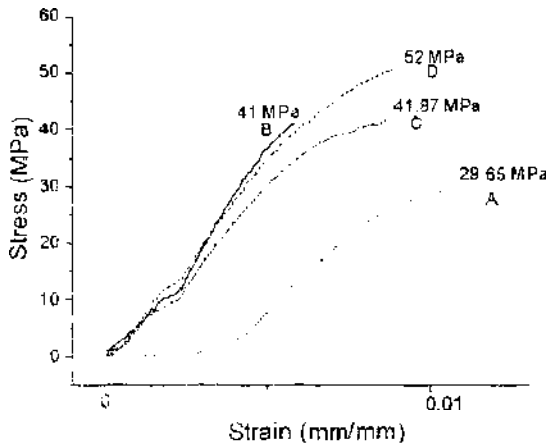


Figure 29. Stress-strain curves for *in situ* polymerized carbon nanotube-reinforced PMMA. A = PMMA; B = [illegible] MWNTs; C = 1 wt% functionalized MWNTs; D = 1.5 wt% functionalized MWNTs. Approximate areas under the curves as a measure of toughness: A = 0.15 MPa; B = 0.116 MPa; C = 0.219 MPa; D = 0.262 MPa. Reprinted with permission from [183], C. Velasco-Santos et al., *Chem. Mater.* 15, 4470 (2003). © 2003, American Chemical Society.

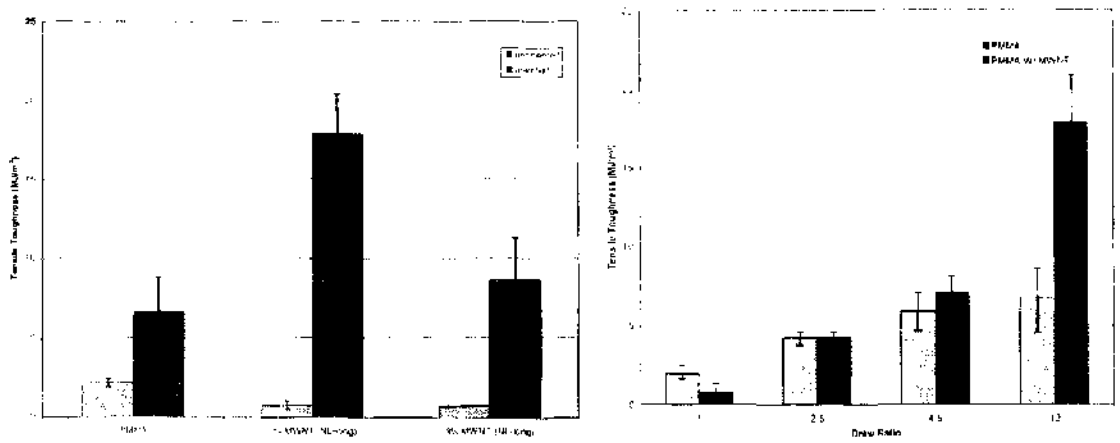


Figure 30. Toughening of MWNT-PMMA nanocomposites. (left) Tensile toughness as a function of MWNT orientation (relative alignment achieved via melt drawing at a 10:1 draw ratio). (right) Tensile toughness as a function of draw ratio for 1 wt% MWNT samples. Reprinted with permission from [190], R. E. Gorga and R. E. Cohen, *J. Poly. Sci.* 42, 2690 (2004), © 2004, John Wiley and Sons.

with the nanoparticles, can lead to improvements or reductions in toughness depending on the different mechanisms activated by the reinforcement. This topic represents a rich area for future research efforts.

While less work has been published on single-walled carbon nanotube composites, a picture similar to that presented above for MWNTs is emerging. The importance of dispersing the SWNTs within the polymer, more difficult than encountered for MWNTs due to the increased surface area of the SWNTs, cannot be overemphasized. For example, recent work found that while the SWNTs affect the crystallization rates of the polypropylene matrix, little change in mechanical properties was attributed to micron-sized aggregates of the SWNTs [192]. On the other hand, increases in modulus up to 30% based on ultrasmall loadings (≤0.25 wt%) of SWNTs in polycarbonate suggest that the entanglement of the SWNTs may enhance the reinforcement effect [193]. A sampling of representative modulus data for various SWNT-polymer systems is shown in Fig. 31. As was the case for MWNTs, results reported with chemical functionalization tend to result in better modulus enhancements (see also Fig. 32). While fewer data points are available for SWNT reinforcement, it appears that in comparison to the micromechanical predictions assuming a 3D random orientation of inclusions, the results for SWNT reinforcement are higher. We stress that such an observation

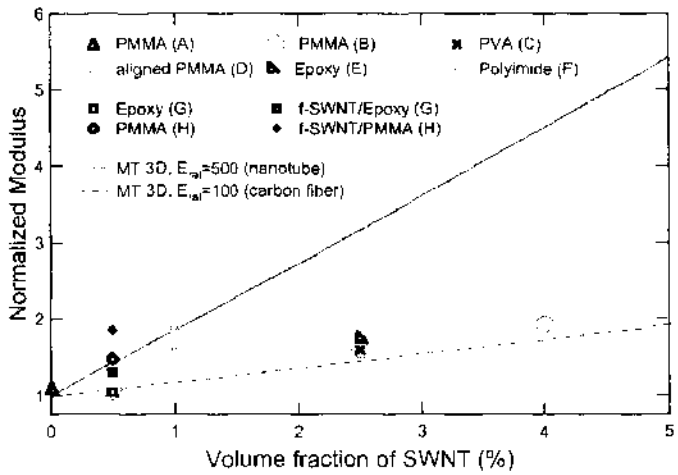


Figure 31. Representative experimental data for glassy (elastic) modulus of SWNT-reinforced polymers normalized with respect to the modulus of the polymer. Also shown are Mori-Tanaka predictions assuming a 3D random orientation of nanotubes for two values of $E_{rat} = E_{CT}/E_{mat}$. f-SWNT refers to functionalized SWNTs. Key: (A) Ref. [194]. (B) Ref. [158]. (C) Ref. [150]. (D) Ref. [152]. (E) Ref. [195]. (F) Ref. [196]. (G) Ref. [151]. (H) Ref. [182].

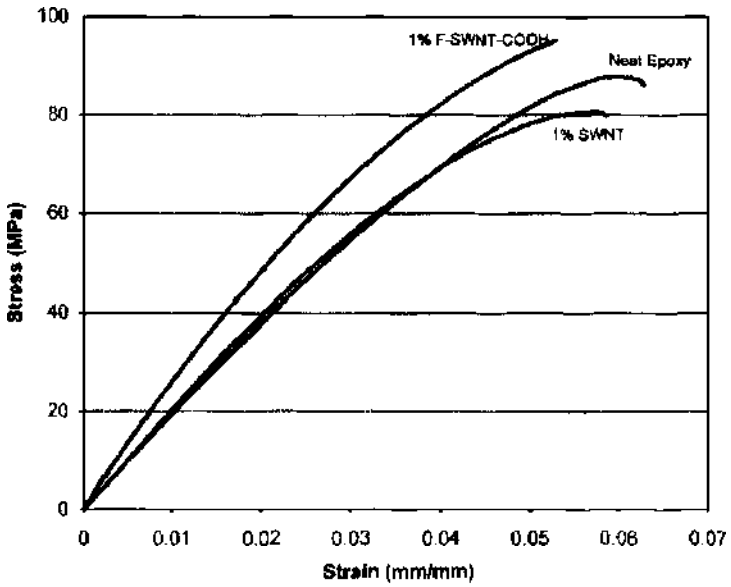


Figure 32. Tensile stress versus strain curves for SWNT-epoxy nanocomposites. F-SWNT-COOH refers to a carboxylation functionalization of the SWNTs. Reprinted with permission from [151], J. Zhu et al., *Nano Lett.* 3, 1107 (2003). © 2003, American Chemical Society.

must be considered in the context of several critical issues: (1) work with SWNTs has in general appeared later in the literature than work for MWNTs, and thus may indirectly benefit from previous work done in this area; (2) the inherent modulus of the SWNTs may be much greater than that of the MWNTs (particularly when considering CVD-grown MWNTs); and (3) the much larger surface area available within the nanocomposite for the case of the SWNT will enhance the effects and reinforcement provided by the interphase (see Section 2.4).

We end this brief section by noting that recent results have also shown that soaking SWNT buckypaper with various polymers enhances the mechanical properties of the buckypaper [197], and that improvements in toughness with the incorporation of SWNTs in PMMA have been observed [161]. The mechanical characterization of SWNTs incorporated within a thermoelastic elastomer have also been reported [198].

2.2.3. Electrical and Thermal Properties of Nanotube-Reinforced Polymers

For many applications, the addition of low volume fractions of NTs is being pursued as a means to increase the conductivity of insulating polymers to levels required to provide electrostatic discharge and electromagnetic radio frequency interference protection [196, 199–201]. Consider the experimental data for CVD grown MWNTs in an epoxy shown in Fig. 33, [199]. It is clear that the nanotubes are much more efficient than the carbon black filler which is currently added to improve polymer conductivity. This is attributed to a percolation-type process, which is achievable for relatively low volume (weight) fractions of the nanotubes due to their high aspect ratio and outstanding conducting properties. Similar results have been reported for arc grown MWNTs in poly(m-phenylenevinylene-co-2,5-dioctoxy-p-phenyleneviylene) (PmPV) [202], SWNTs in PMMA [152, 203], MWNTs in a polycarbonate masterbatch processed via melt-mixing [204] and diluted with PE [205], MWNTs in a rubbery epoxy resin [206], MWNTs in polyaniline (PANI) [207], and SWNTs in a polyimide investigated for electrostatic charge mitigation [200]. These results demonstrate that electrical percolation can be obtained at extremely load nanotube concentrations within the polymer. For example, percolation was obtained for 0.05 vol% MWNTs in poly(propylene) [139], 0.0025 wt% MWNTs in epoxy [208], and 0.1 vol% loading of SWNTs in a polyimide, resulting in a 10 order of magnitude increase in electrical conductivity [196] in agreement with a percolation model for those systems [209]. However, for PmPV and PVA polymer

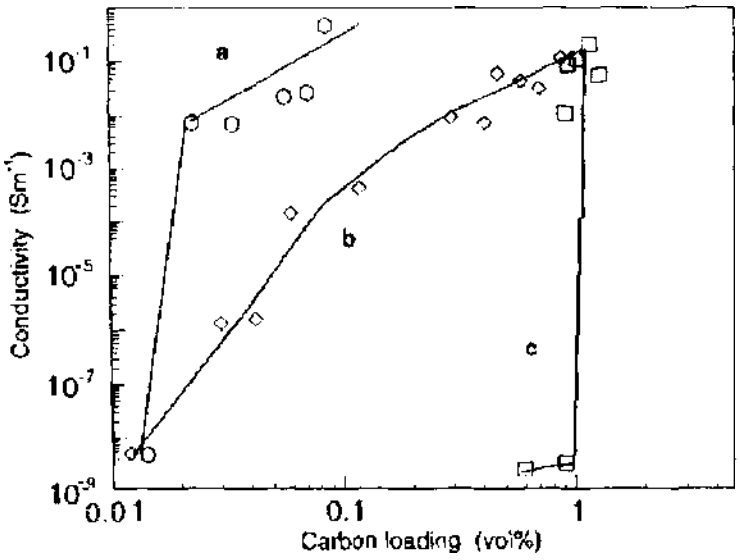


Figure 33. Electrical conductivity of CVD grown NTs in an epoxy. (a) CVD grown NTs. (b) Carbon black with copper-chloride. (c) Carbon black only. Reprinted with permission from [199], J. Sandler et al. *Polymer* 40, 5967 (1999). © 1999, Elsevier.

systems, the presence of a thick polymer coating surrounding the nanotubes may limit the electrical conductivity obtainable for ultra-high loadings of nanotubes [210].

Similar to electrical conductivity, significant increases in thermal conductivity using small amounts of nanotubes have also been reported. For example, increases in both thermal and electrical conductivity were found for an industrial epoxy reinforced with up to 1 wt% SWNTs, with the improvements in thermal conductivity observed over a wide range of temperatures (see Fig. 34) [211]. A loading of 1 wt% SWNT in epoxy increased the thermal conductivity of an epoxy by 120%, a much larger increase than exhibited for vapor-grown carbon fibers [212]. The thermal conductivity enhancement of a silicone elastomer was found to be greater with CVD-grown MWNTs than a corresponding weight loading of carbon black [213]. Recent work for CVD-grown MWNTs in a rubber found thermal conductivity enhancement for unmodified nanotube filler, and that while moderate chemical functionalization improves the thermal conductivity of the composite, excessive chemical treatment decreases the effective conductivity [213, 214]. Modeling efforts have suggested that the effective thermal conductivity of the nanocomposite is strongly influenced by the thermal interface resistance of the system [215–217].

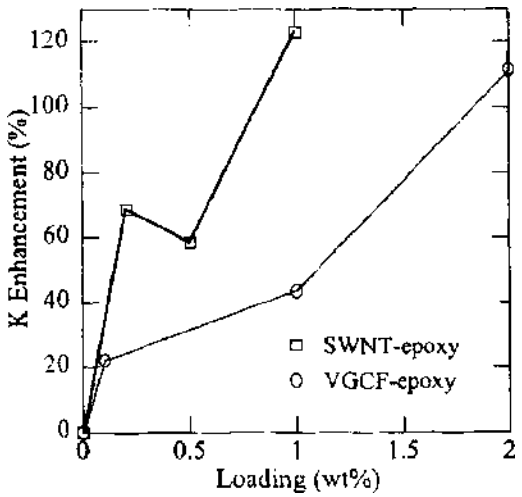


Figure 34. Thermal conductivity enhancement for epoxy as a function of SWNT and vapor-grown carbon fiber weight loadings. Reprinted with permission from [211], M. J. Biercuk et al. *Appl. Phys. Lett.* 80, 2767 (2002). © 2002, American Institute of Physics.

We note one factor that distinguishes the use of nanofillers for mechanical property enhancement is the existence of a "critical threshold" (i.e., percolation) for electrical and thermal properties at which substantial improvements in these properties are observed. Often these improvements result in satisfactory performance improvement being achieved at an extremely small nanofiller loading (see, for example, a recent Monte Carlo study of the level of nanotube loading necessary for percolation in nanotube-polymer composites [218]). For mechanical property enhancement, however, desired mechanical properties should continue to increase as nanofiller loading is increased, and concerted efforts are underway to continue to push the level of nanofiller that can be loaded into polymer nanocomposites for these applications. Readers interested in the use of nanotube fillers for electrical and thermal property enhancement are directed to the references provided above, and references contained therein, for further discussion of research in this area.

2.3. Nanoplatelet Reinforcement

As discussed in Section 1.2.3, interest in the use of nanoclays as nanoscale reinforcement materials has greatly increased since the initial published work by researchers at Toyota in the early 1990s [3, 90]. A comprehensive review of nanoclay-polymer composites work is beyond the scope of the current section; the reader is referred to a number of review articles for focused reading in this area [91–99]. Our goal here is to simply highlight some of the experimental work that has been published in this area, beginning with nanoclays and later focusing on graphite nanoplatelets (alternatively referred to as graphite, expanded graphite, and graphene reinforcement), to set the stage for the discussion of the nanomechanical modeling of such nanocomposites beginning in Section 3. Discussion of changes in the viscoelastic behavior of clay-nanocomposites (in comparison to the behavior of the unreinforced polymer) will be postponed until Section 2.4. Additional papers describing work in this area are also provided in the Appendix.

As discussed in Section 1.2.3, the mechanical properties of nanoclay-polymer composites are strongly dependent on the degree of dispersion and intercalation/exfoliation of the nanoclay within the polymer, and processing methods have been predominantly focused in these areas. In general, increases in elastic modulus and yield strength are accompanied by decreases in ultimate strain (i.e., strain to failure). Representative examples of experimentally measured nanocomposite mechanical properties are shown in Figs. 35 [219] and 36 [220]. The blending of polyimide with organically modified montmorillonite found increases in tensile modulus up to 110% in comparison to the pure polyimide for 2 wt% loading of nanoclay, with decreases in the thermal expansion coefficient and increases in the glass transition temperature and decomposition temperature also noted [221]. Other results show a

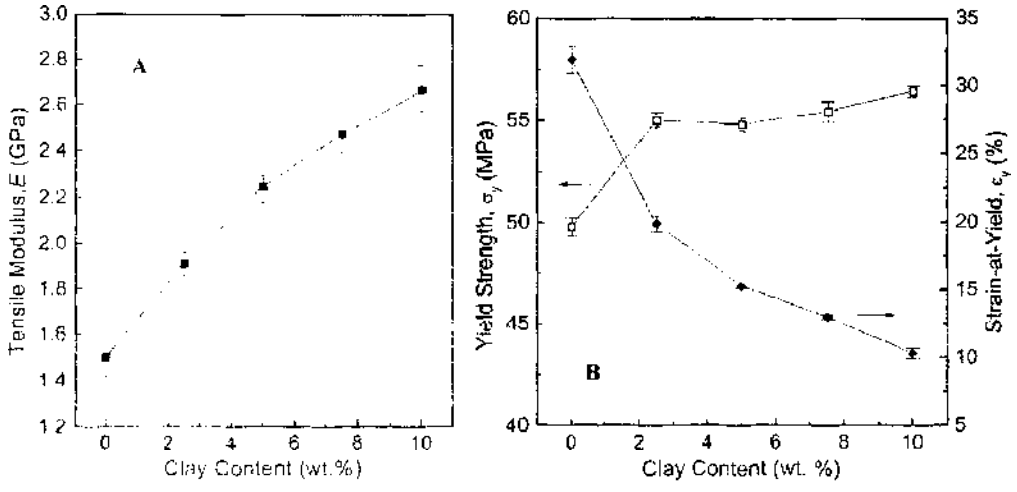


Figure 35. (A) Tensile modulus and (B) yield strength and strain at yield as a function of clay loading for a polyamide-nanoclay composite. Reprinted with permission from [219], T. X. Liu et al., *Compos. Sci. Technol.* 63, 331 (2003), © 2003, Elsevier.

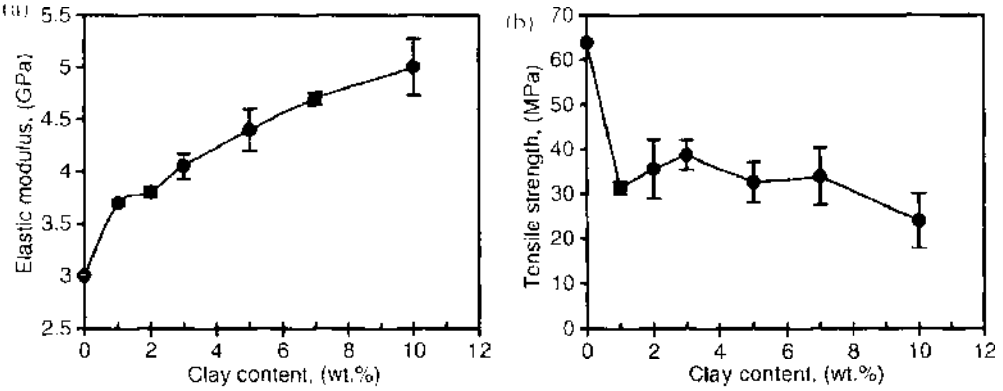


Figure 36. Effect of clay content on the mechanical properties of a three-roll mill compounded epoxy-nanoclay composite. Reprinted with permission from [220], A. Yasmin et al., *Scripta Mater.* 49, 81 (2003). © 2003, Elsevier.

factor of three increase in fracture toughness at 1 wt% nanoclay loading without decreasing strength [222].

However, it is difficult to make sweeping generalizations with regards to the effective mechanical properties of nanoclay-polymer composites, as the structure and hence properties can be sensitive to the processing of the nanocomposite. For example, a collapse of the exfoliated structure of the nanoclay based on mechanical processing of the nanocomposite has been noted [223], and decreases in cross-linking of a cross-linked polyester matrix have been attributed to greater levels of exfoliation of the nanoclay [224].

Issues related to the impact of processing on nanoclay composite properties have also been studied. Nanoclay exfoliation in nylon 6 has been found to be more effective using a twin-screw (as opposed to a single screw) extruder; the resulting nanocomposites were found to have mechanical properties similar to those obtained for similar *in situ* polymerized systems [225]. The impact of molecular weight on the melt mixing of a masterbatch nylon 6—nanoclay composites has also been studied, with higher molecular weights generally leading to higher levels of nanoclay exfoliation [226, 227]. While the difference in moduli between the low molecular weight (LMW) and high molecular weight (HMW) nylon 6 was small, the moduli of nanocomposites based on HMW nylon 6 are 10–15% higher than those based on the LMW nylon 6. Differences in the thermomechanical properties of the skin and core of injection-molded nylon 6—nanoclay composites have also been observed [228]. A non-solvent compounding using a three-roll mill found enhanced elastic moduli up to 80% in comparison to that of the pure epoxy at 10 wt% nanoclay loading, although the rate of enhancement decreased at higher clay contents attributed to the aggregation of clay particles at higher loadings [220]. Epoxy-aided dispersion of nanoclay in PMMA has been found to result a three-phase nanocomposite structure with significant improvements in tensile modulus and impact strength compared to those of the pure PMMA and the PMMA-nanoclay composite [229]. Such results emphasize the need for continued development of the processing-structure-property relationships for this class of nanocomposites.

Recently, a new class of nanoclay-polymer composites has been explored using the layer-by-layer assembly method to create thin nanocomposite films with much higher loadings of nanoclay than can be achieved using other methods [104, 105]. Such processing is based on the sequential absorption of polyelectrolytes, producing a nanocomposite structure that has been referred to as "artificial nacre." While typical film thicknesses were found to be on the order of 1 μm for 50 polyelectrolyte/nanoclay bilayers, modulus values on the order of 10 GPa have been measured experimentally [104, 105].

In addition to mechanical property enhancement, the use of nanoclay reinforcement has been proposed to enhance the barrier properties of polymers. The reduced permeability of nanoclay-filled polymers to both liquid and gas diffusion is of interest in areas such as membranes and packaging applications. An example is shown in Fig. 37, where water vapor permeability is significantly reduced as a function of nanoclay content; similar results were also reported for the diffusion of oxygen and helium gas [6]. Decreases in diffusivity have

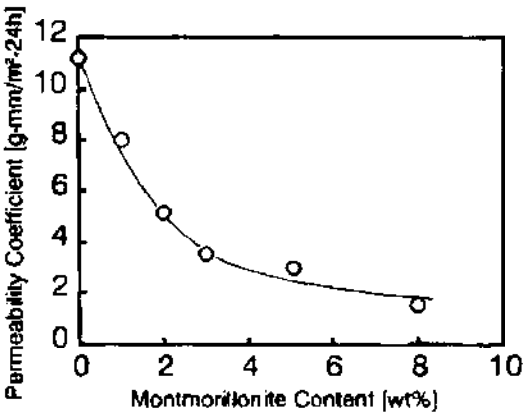


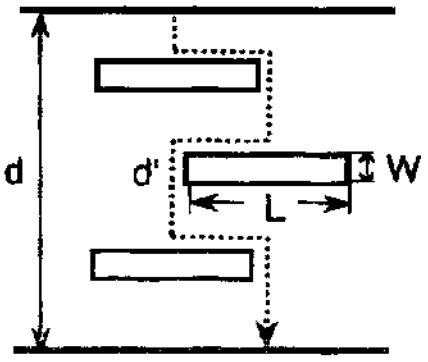
Figure 37. Permeability coefficient of water vapor in a polyimide-clay hybrid. Reprinted with permission from [6], K. Yano et al., *J. Polym. Sci.* 31, 2493 (1993). © 1993, John Wiley and Sons.

been attributed to the increased of tortuosity of molecules diffusing through the polymer [230]. As shown in Fig. 38, the ratio of the permeability coefficient of the composite P_c to that of the pure polymer P_p can be expressed as

$$\frac{P_c}{P_p} = \frac{1}{1 + (L/2W)V_f} \tag{13}$$

emphasizing the importance of exfoliation (to increase the value of $L/2W$) in decreasing the permeability. Interestingly, such changes in permeability may impact the processing of nanoclay composites; for example, the importance of adequate degassing to reduce nano- and microvoids in an epoxy-nanoclay nanocomposite has been identified (see Fig. 39) [220]. Other property enhancements, such as decreased flammability [231, 232] and increased thermal and dimensional stability, have also been observed [19, 92, 99].

While carbon black has been used in polymers for decades as a filler to impart electrical conductivity, and research into nanoclay as a filler material exploded in the early 1990s, more recent work has examined the use of expanded graphite as a nanoscale filler. While natural graphite flakes have a surface area of ~4 m^2/g, expanded graphite ranges from 40 m^2/g to 1500 m^2/g depending on the intercalant chemistry and processing conditions [107, 233]. Consequently, if proper dispersion of the graphite nanoplates in the polymer is obtained with the higher surface area material, the potential for improved properties in the



Total path of a diffusing gas

d' = d + d·L·Vf / 2W
d : thickness of a film
L : length of a clay
W: width of a clay
Vf: volume fraction of a clay

Tortuousity factor

τ = d' / d
= 1 + L·Vf / 2W

Equation for a permeability coefficient

Pc = Pp / τ
= Pp / (1+L·Vf / 2W)
Pp : permeability coefficient of a matrix polymer

Figure 38. Tortuous path of diffusing molecules through a platelet filled material. Reprinted with permission from [6], K. Yano et al., *J. Polym. Sci.* 31, 2493 (1993). © 1993, John Wiley and Sons.

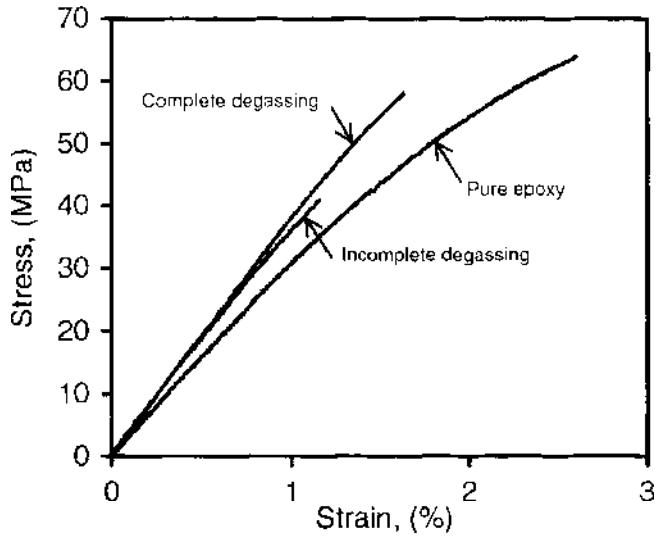


Figure 39. Effect of degassing on the tensile strength of a 3 wt% clay nanocomposite. Reprinted with permission from [220], A. Yasmin et al., *Scripta Mater.* 49, 81 (2003). © 2003, Elsevier.

composites is significant. The first steps in this area were focused on electrical properties, with several sources noting that the percolation threshold for expanded graphite (~1 wt%) is significantly less than that for graphite (~3 wt%) or carbon black (~8 wt%) (see Fig. 40) [234, 235].

Enhancements of thermo-mechanical properties have also been reported for graphite nanocomposites. Elastic modulus or storage modulus increases of 25% at 2 wt% have been reported [236]. In addition, several methods for exfoliating the expanded graphite have been examined, each resulting in different platelet dimensions. It was found that the larger plate diameter tended to correspond to better nanocomposite properties [109]. Changes in glass transition temperature have also been noted, with increasing T_g for the thinner and wider nanoplatelet forms.

A recent paper has compared the impact of graphite, expanded graphite (EG), and graphite nanoplatelet (GNP) on nanocomposite properties under identical processing conditions for loadings from 1 to 5 wt% [237]. The graphite nanoplatelet material was created via sonication of expanded graphite before being incorporated into the host polymer (PMMA) via a solution based processing method. The sonication increased surface area from 20 m^2/g for EG to 30 m^2/g for the GNP. This work demonstrated substantial increases in glassy elastic stiffness, 60% at 1 wt% and 135% at 5 wt% for the GNP reinforcement, exceeding that

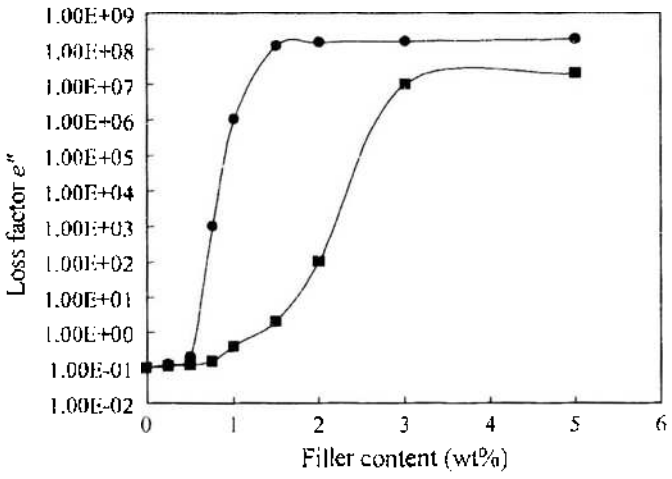


Figure 40. Electrical conductivity of PMMA/graphite (■) and PMMA/expanded graphite (•) as a function of weight fraction. Reprinted with permission from [235], W. Zheng and S. C. Wong, *Compos. Sci. Technol.* 63, 225 (2003). © 2003, Elsevier.

of both the as-received graphite and the EG forms. All cases also showed a change in the T_g of the material, with a maximum increase of 38°C for GNP at 5 wt%. In addition to the increased surface area, the GNP exhibited better adhesion with the PMMA matrix as seen in Fig. 41. While these results are promising, some effects could not yet be fully explained, including a decrease in the T_g for the EG composites at 5 wt%. Modeling of the interaction of the surface chemistries of the different graphite forms with the polymer matrix and the morphologies of the *in situ* platelet forms will be helpful to untangle the complex behavior of these systems.

While the experimental results highlighted here demonstrate the successful use of nanoinclusions (nanoparticles, nanotubes, and nanosheets, respectively) within polymers, in many cases interpretation of these experimental results is hampered by lack of a detailed modeling approach suitable to relate these experimental results to the fundamental principles governing the mechanical response. As modeling efforts progress and a more complete picture of the nanoscale mechanisms governing the effective behavior of the material is developed, this information will feed back into the processing and chemistry of these nanocomposites to enable even greater property enhancements to be realized. In addition, as discussed in the next section, one characteristic that is common to all nanoinclusion forms is the impact of the embedded nanoinclusions on local polymer in the vicinity of the nanoinclusion.

2.4. The Interphase Region and Viscoelastic Behavior in Nanoreinforced Polymers

Because the nanoinclusions are on the same length scale as the radius of gyration of the polymer chains, local polymer morphology will differ from that of the bulk polymer. We refer to this region of nonbulk polymer behavior as the interphase, borrowing a term used in the composites community that refers to the region separating the fiber and matrix phases. While in traditional composites research the interface region is generally attributed to a host of factors (such as the use of fiber sizings, mechanical imperfections, unreacted polymer components, etc.), as discussed in Section 1.1 such regions are generally much smaller (in terms of volume fraction of the composite) than the interphases present in nanocomposites (see Fig. 2). Here we limit our discussion specifically to the change in molecular mobility of the polymer chains in this region due to the presence of, and interactions with, the nanoscale inclusions. We note that in addition to the experimental work discussed in this section, similar findings have been presented based on computational studies of nanocomposites (see Section 3.1.1).

2.4.1.1. Experimental Evidence of Nonbulk Interphase Polymer. A number of experimental studies have shown changes in the viscoelastic behavior of polymer nanocomposites compared to the behavior of the pure polymer. In most early reports such changes were noted as shifts in the glass transition temperature of the nanocomposite, however recently

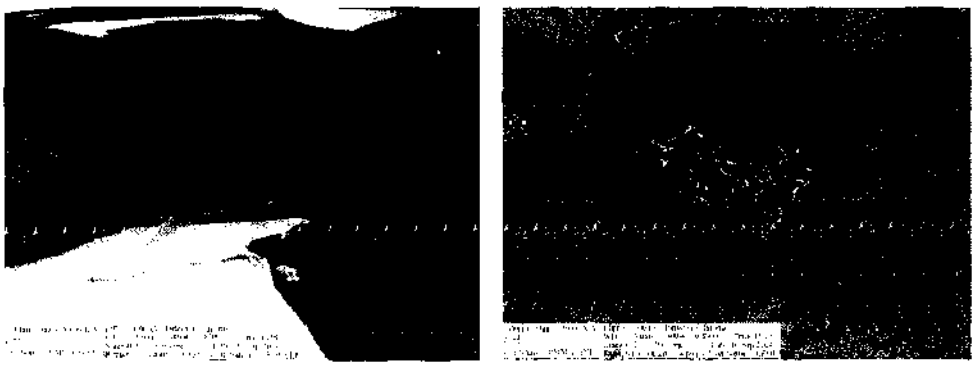


Figure 41. Plates protruding from a fracture surface of EG/PMMA (left) and GNP/PMMA (right) nanocomposites at 5 wt% of particle loading. Clear evidence of better polymer adhesion to the nanoparticles is seen for the case of GNP. Reprinted with permission from [237], T. Ramanathan et al., submitted to Polymer (2006). © 2006, Elsevier.

more complete analyses of the property changes are common, with the interphase generally recognized to be a primary mechanism causing such changes in viscoelastic behavior. As discussed previously, generalizations of the interphase phenomena in nanocomposites is complicated due to the number of parameters (e.g., nanoinclusion type, nanoinclusion-polymer interactions, and polymer chemistry) that are expected to influence the size and properties of the interphase. Here we will present a few examples from the literature demonstrating the significance of the interphase in polymer nanocomposites, with a focus on our own in-depth study of the interphase in MWNT-reinforced polycarbonate. Many of the results focus on viscoelastic measurements to probe the nanoparticle-polymer interactions as these techniques are especially sensitive to changes in polymer mobility [238]. Additional literature references can be found in the Appendix.

A number of studies have looked at the viscoelastic behavior of nanosilica-reinforced polymers. In one particularly extensive study, the effective mechanical behaviors of various nanosilica-filled polymers were characterized as demonstrating two glass transition temperatures [123]. Such behavior was attributed to the morphological changes in the nanocomposite systems as the loading of nanoparticles increased in terms of tightly and loosely-bound polymer. The change in the loss tangent of nanosilica in SBR was also attributed to an interphase formation, despite the agglomeration of the 9 nm equiaxed nanoparticles into clusters with an apparent diameter of 25 nm [121]. However, the effective behavior of the interphase and its impact on the overall nanocomposite properties can be complex due to the sensitivity of the interphase on the chemistry of the system; for example, despite each system showing improved thermal stability, the glass transition temperature has been found to increase for nanosilica-reinforced PMMA [118] and decrease for an epoxy system [132]. A recent study has examined the change in dynamic mechanical properties of nanosilica-filled LLDPE as a function of surface treatment [124]. Other modes of viscoelastic response, such as creep resistance, have been shown to change upon the addition of nanoparticles; for example, by the addition of 21 nm diameter nano-TiO_2 to nylon 6 [130].

Similar changes in viscoelastic behavior have also been observed in nanoclay-polymer composites. Here changes in the glass transition temperature seem to be particularly sensitive to the interaction chemistry, as both increases and decreases in T_g are often reported. For example, even for nanoclay-reinforced epoxy, both lower [239] and higher [240] glass transitions have been observed. Similar results have been observed for polymer thin films, where the nature of the chemical interactions between the polymer chains and the substrate dictate whether increases or decreases in T_g occur [241]. Changes in viscoelastic behavior due to nanoclayreinforcement have been reported for a number of polymer systems, including: polyimide [221], polysiloxane [242], nylon [243], PEO [244], PMPS [245], and diblock copolymers [246]. Changes in the creep behavior of nanoclay-filled nylon have also been reported [247]. Similar alterations in the viscoelastic behavior have also been reported for expanded graphite-reinforced nanocomposites [235, 248, 249].

Early experimental testing of the viscoelastic behavior of nanotube-reinforced polymers has also found that the addition of NTs to the polymer can result in changes in viscoelastic behavior of the nanocomposite. For example, a 25°C shift in the glass transition temperature (from 63 to 88°C), as measured by the peak of the loss tangent curve, has been measured for an epoxy reinforced with 1 wt% MWNTs [145]. However, the changes were only achieved when surfactant was used within the processing of the nanocomposite; non-surfactant samples showed smaller increases in T_g, attributed to poorer dispersion of the NTs within the sample, highlighting the importance of surface area on the changes in viscoelastic behavior. Shifts in T_g have also been observed for polyimides [196], ABS [250], and PMMA [186], while broadening of the loss peak has been reported for PVOH [169]. Interphase formation may also be related to changes in polymerization and crystallization due to the incorporation of the nanotubes within the polymer [170, 185, 192], and has been described as acting as effective crosslinks in a polyimide system [188].

We have recently conducted a comprehensive characterization of interphase formation in MWNT-reinforced polycarbonate, the details of which are available elsewhere [116]. Figure 42 shows the temperature-dependent viscoelastic response of the polycarbonate (PC) control sample as well as samples created with as-received (AR) and epoxide-functionalized

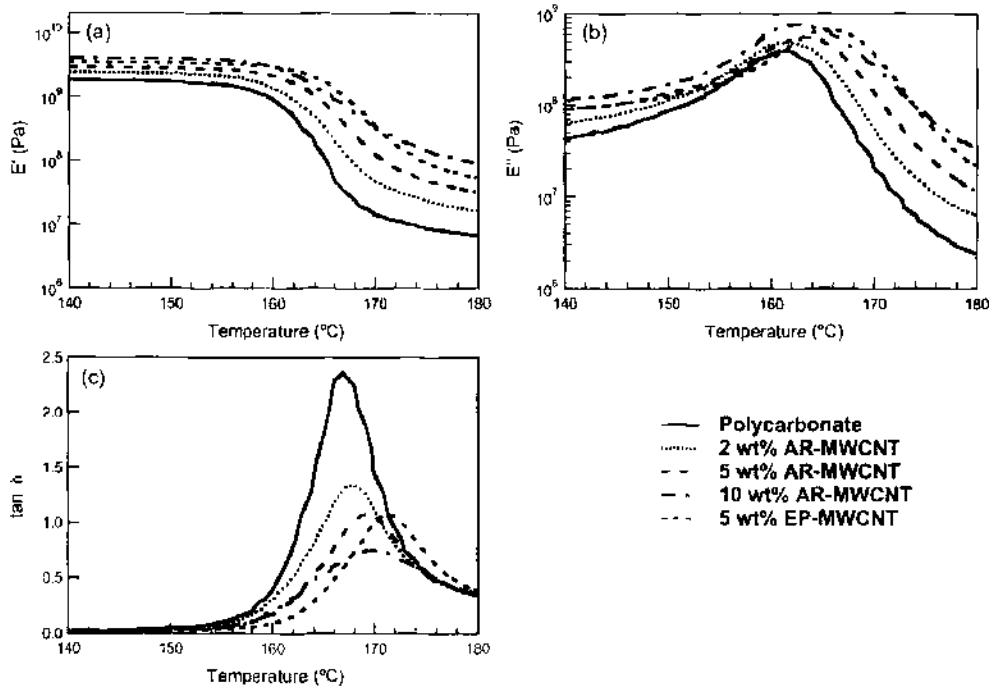


Figure 42. Temperature-dependent viscoelastic properties of MWNT-polycarbonate samples. (a) Storage modulus. (b) Loss modulus. (c) Loss tangent. AR = as-received MWNTs, EP = epoxide-modified MWNTs. Reprinted with permission from [116], A. Eitan et al., *Compos. Sci. Technol.* (to be published, 2006). © 2006, Elsevier.

(EP) MWNTs. For the unfunctionalized samples we see slight increases in the glassy storage modulus (analogous to the elastic-like improvements demonstrated during elastic testing), with corresponding shifting and broadening of the loss modulus and loss tangent peaks to higher temperatures with increasing nanotube loading. However, the functionalized samples exhibited more dramatic shifting of the loss peak for the same weight percent loading, indicating a greater influence of the interphase on the effective response for this system.

Figure 43 schematically illustrates the difference in the extent of the interphase in the unfunctionalized and functionalized cases. The formation of an adhered layer of polymer for the case of the unfunctionalized samples can be enhanced by controlling the processing of the MWNT-polycarbonate samples via solvent selection [116]. Functionalization and subsequent tethering of short chain epoxide modules to the MWNT enable covalent bonding of the nanotubes to the polymer matrix, which further enhances interphase formation, such that eventually the viscoelastic properties of the interphase dominate those of the bulk polymer matrix. Experimental characterization of the relaxation spectra of the nanocomposite

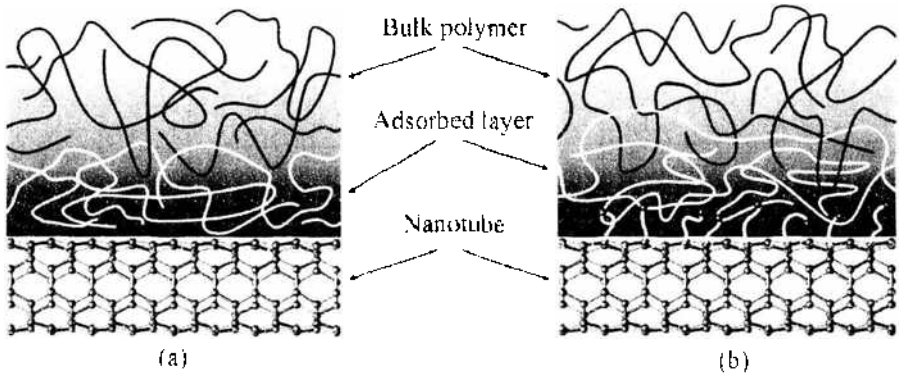


Figure 43. Proposed mechanisms for influence of absorbed layer on interphase formation in nanotube-polymer composites. (a) Unfunctionalized nanotube and absorbed layer formation. (b) Functionalized nanotube with short-chain polymer tether linking nanotube to matrix. Reprinted with permission from [116], A. Eitan et al., *Compos. Sci. Technol.* (to be published, 2006). © 2006, Elsevier.

samples (see Section 2.4.1.2) confirm the decreased mobility of the interphase polymer, and are consistent with the significant enhancement of the interphase effect in the case of the functionalized samples. Such results suggest that some tailorability of the interphase and its properties may be achievable upon careful control of the nanocomposite processing conditions. Extensions of this work to investigate the effect of the interphase region on other mechanical properties, such as strength and toughness, will be the subject of future efforts.

Further evidence of the interphase region for this system was found upon examination of high-resolution SEM imaging of the fracture surface of unfunctionalized MWNT samples [251]. As shown in Fig. 44, protruding from the fracture surface are nanotubes coated (or sheathed) by a layer of polymer. The presence of this polymer sheathing is consistent with models of a nonbulk polymer interphase region proposed for nanotube-polymer composites [16, 117, 140, 169, 252]. In the process of probing these structures with an AFM tip, the polymer sheathing suddenly contracted and balled up, exposing a much thinner structure that may be either a bare nanotube or a MWNT coated with a much thinner layer of polymer (see Fig. 45). While the MWNT shown in Fig. 45 fractured, either due to contact with the AFM tip or the subsequent balling of the polymer coating, in other cases the MWNT did not fracture during this experiment. Of the 26 structures protruding from the same MWNT-PC fracture surface examined, 22 demonstrated a similar "balling up" response [251]. From video recording of these *in situ* SEM experiments, the apparent outer diameters of the coated MWNTs before and after this balling up phenomenon were measured; the volume of the newly formed globules matched the apparent decrease in volume of the coated MWNT within experimental error. Estimates for the thickness of the polymer sheathing based on this balling up observation were similar to the sheathing thicknesses directly measured from SEM images similar to those shown in Fig. 44.

For those cases where the balling up of the polymer sheath did not occur, the coated tubes were clamped to the AFM tips and nanopullout tests from the polycarbonate matrix conducted. At the end where the coated MWNT was pulled from the fracture surface, high-resolution SEM images reveal what appear to be two distinct polymer layers coating the fractured MWNT [251]. This suggests that here there may be two interfaces that influence the effective mechanical properties of the system: the MWNT-inner polymer layer and inner polymer layer-outer polymer layer interfaces. Efforts to further elucidate this polymer sheathing phenomena, and the impact of this sheathing on the mechanical behavior of nanotube-polymer systems, are ongoing.

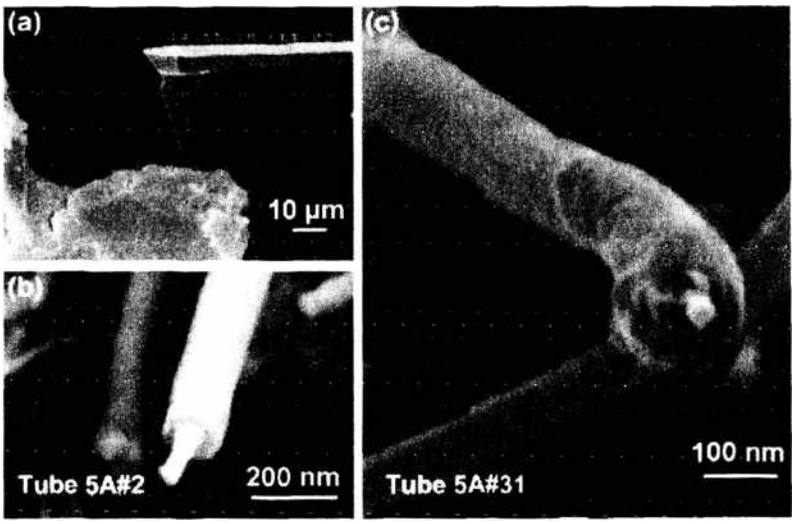


Figure 44. High-resolution SEM images of AR-MWNT fracture surface. (a) Far-field image of the AFM tip approach for nanotube pullout experiment. (b,c) Nanotube structures coated with a polymer sheath protruding from the fracture surface. Reprinted with permission from [251], W. Ding et al., *Nano Lett.* 3, 1593 (2003). © 2003, American Chemical Society.

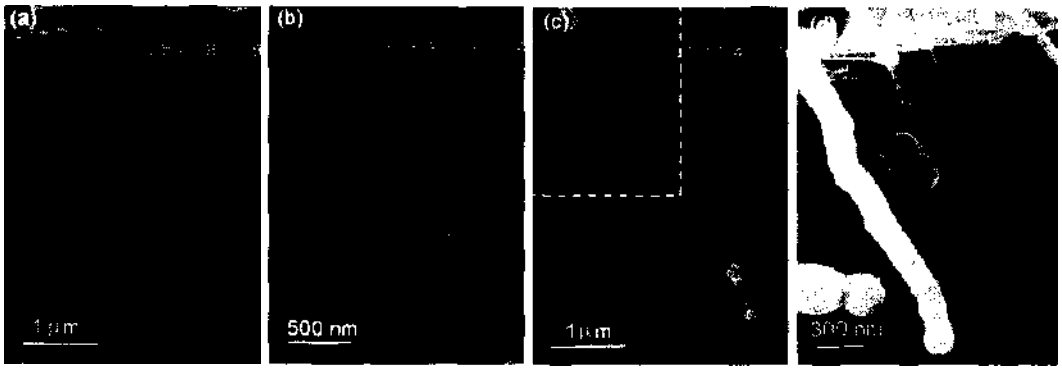


Figure 45. SEM time-lapsed images of balling up of the polymer sheath after contact with the AFM tip. (a) The AFM tip approaches a coated MWNT. (b) The AFM tip is brought into contact with the coated MWNT. (c) After contact, the MWNT fractures and the polymer coating balls up on each side of the contact point as highlighted by the arrows. (d) Higher magnification image of the upper polymer globule identified in (c). Reprinted with permission from [251], W. Ding et al., *Nano Lett.* 3, 1593 (2003). © 2003, American Chemical Society.

We complete this section by highlighting our recent work studying the effective viscoelastic behavior of single-walled carbon nanotube—PMMA nanocomposites [182]. Because of the smaller diameters (and hence greater surface area available) for the single-walled nanotubes, interphase effects are expected to be enhanced. This is indeed the case as shown in Fig. 46 and Table 3, where the storage and loss modulus of the pure PMMA and the nanotube-reinforced samples are shown as a function of temperature. In addition to increases in the storage modulus at low temperatures, 1 wt% of SWNTs is shown to result in a broadening of the loss tangent response around the glass transition temperature. Additionally, amino-functionalization of the SWNTs results in further property enhancements, with a notable shift in the loss tangent of almost 40°C. Similar results for other SWNT-reinforced polymer samples have also been reported [194, 196, 250], including ~50 μm solution spun PAN fibers with embedded SWNTs as shown in Fig. 47 [253]. As discussed in Section 3.1, recent computational models support these findings (see, for example, Refs. [254, 255]).

Results such as those presented in this section suggest that interphase formation may provide an additional reinforcement mechanism within nanocomposites [116, 256]. For example, recent results imply that the polymer matrix in close vicinity to the carbon nanotube is able to withstand stresses that would otherwise cause considerable yield in a bulk polymer specimen [180], while for PVA it was found that the reinforcement provided by the nanotubes scaled with the nanotube surface area [257]. As discussed in Section 3.2.1.1, we are currently developing a complementary experimental and modeling program to assess changes in the viscoelastic behavior of the interphase based on macroscale experimental data. Such characterization of the interphase properties will also be supported by continued development of bottom-up and hybrid modeling approaches as described in Section 3. Development of

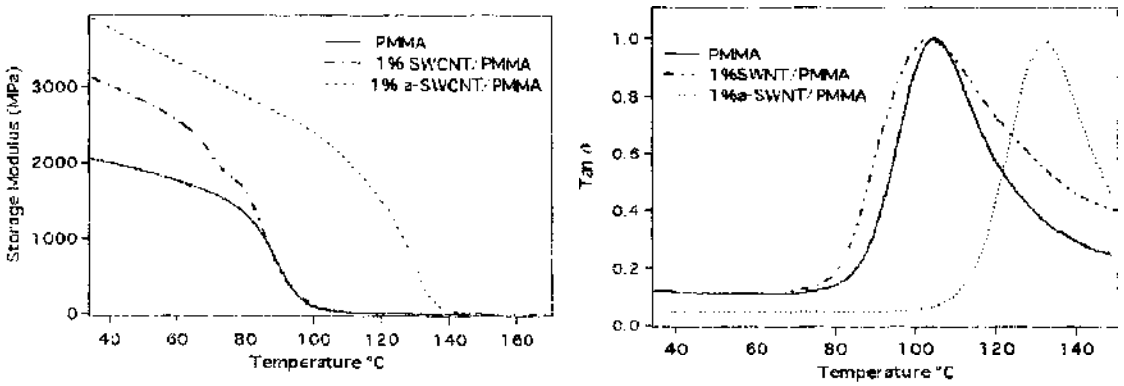


Figure 46. Dynamic mechanical analysis results for (left) storage modulus and (right) loss tangent as a function of temperature for PMMA, SWNT-PMMA, and amino-functionalized SWNT-PMMA composites. During the test the temperature sweep was conducted at a constant rate of 3°C/min. Reprinted with permission from [182], T. Ramanathan et al., *J. Polym. Sci. B* 43, 2269 (2005). © 2005, John Wiley and Sons.

Table 3. Experimental results for 1 wt% PMMA reinforced with as-received and amide-functionalized (a-SWNT) nanotubes.

Sample	Tg by DMA (°C)	Storage Modulus (GPa) at RT
PMMA	105 ± 2	2.1 ± 0.1
SWNT-PMMA	105 ± 3	3.1 ± 0.3
a-SWNT-PMMA	138 ± 2	3.9 ± 0.6

Source: Reprinted with permission from [182], T. Ramanathan et al., *J. Polym. Sci. B* 43, 2269 (2005), © 2005, John Wiley and Sons.

accurate models across a number of length scales will be necessary to better understand, and ultimately control and optimize, interphase formation in polymer nanocomposites.

2.4.1.2. Effective Relaxation Spectra of Nanotube-Polymer Composites. Although the observations noted in the previous section are qualitatively sensible based on the hypothesis of a reduced mobility polymer phase surrounding the nanotubes, quantitative predictive models of how the mobility of the polymer chains influences the glass transition temperature are not available. Thus we were led to investigate the frequency-domain behavior and relaxation spectra of MWNT-reinforced polycarbonate using dynamic mechanical analysis [117].

Starting with a reference curve obtained using time-temperature superposition (see Fig. 19), the storage and loss moduli data were fit to a 30-term Prony series using a linear least squares solver DYNAMFIT [258]. DYNAMFIT assumes that the relaxation times τ_j are equally spaced in log time and then calculates the Prony coefficients E_∞ and E_j that best fit the data, equally weighing storage and loss moduli contributions to the root mean square (rms) error of the solution. An extra constraint imposed within the code forces all Prony coefficients E_j to be greater than zero; while such a constraint does slightly increase the rms error describing the fit, negative Prony coefficients are not physically reasonable and thus were not considered.

Once the Prony series terms for the nanocomposite samples have been found, the relaxation spectra can be approximated using Alfrey's expression given in Eq. (10). Examples of relaxation spectra for pure and MWNT-reinforced polycarbonate samples are shown in Fig. 48. Two sets of data are shown, with the data in Fig. 48(a) for unfunctionalized MWNT-PC and the data in Fig. 48(b) including functionalized MWNT-PC (same specimens as shown in Fig. 42). In Fig. 48(a), the change in relaxation spectra as a function of unfunctionalized nanotube loading shows an increase in the relaxation spectrum at longer times with increasing nanotube volume fraction. Also apparent is a slight broadening of the primary relaxation peak towards longer times. The broadening of the primary relaxation peak and the increase in the relaxation spectrum at longer times are indicative of an increased number of relaxation modes and the introduction of longer time relaxation processes within the reinforced samples, respectively. These changes are consistent with the hypothesis of a reduction in molecular mobility within the interphase. However, the location of the primary relaxation

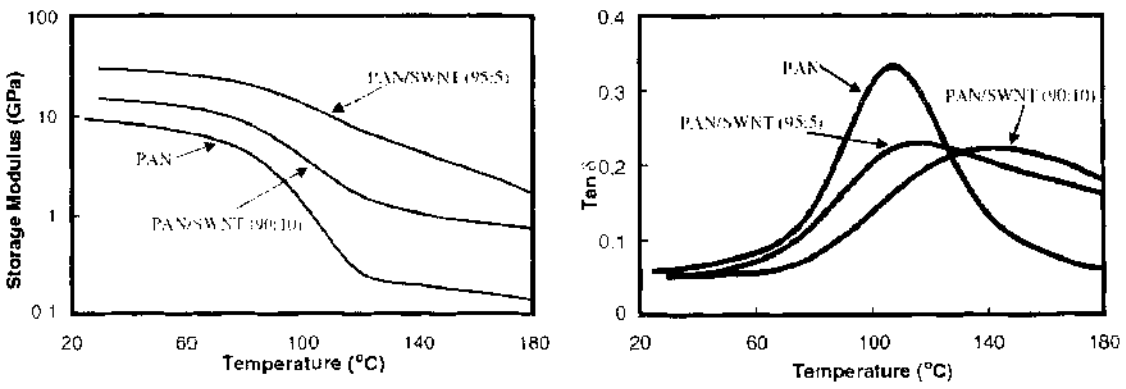


Figure 47. Viscoelastic behavior of ~50 μm diameter solution spun PAN-based nanocomposite fibers. (left) Storage modulus. (right) Loss tangent. Reprinted with permission from [253], T. V. Sreekumar et al. *Adv. Mater.* 16, 58 (2004), © 2004, Wiley-VCH.

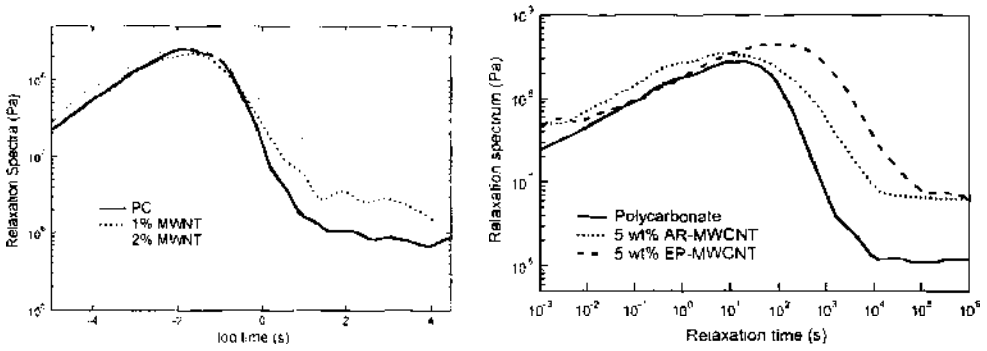


Figure 48. Relaxation spectra for MWNT-reinforced PC samples. (left) Change in relaxation spectra for unfunctionalized samples. Reprinted with permission from [117]. F. T. Fisher et al., *Adv. Compos. Lett.* 13, 105 (2004). © 2004, Advanced Composites Letters. (right) Change in relaxation spectra for epoxide-functionalized (EP-MWCNT) samples. Reprinted with permission from [116]. A. Eitan et al. *Compos. Sci. Technol.* in press (2006). © 2006, Elsevier.

peak does not appreciably change with the addition of the nanotubes, suggesting that the primary relaxation mechanism within the nanocomposite is the same as that within the bulk polymer sample. This suggests that the nanotubes reduce the molecular mobility of the interphase region, and that polymer chains well separated from the nanotubes are not affected by the presence of the nanotubes and retain the mobility of the bulk polymer sample.

More recent work, which used the relaxation spectra as a means to characterize the impact of functionalization on MWNT-polycarbonate nanocomposites, is shown in Fig. 48(b). The slight shift in the location of the main peak of the relaxation spectra for the data shown in Fig. 48(a) versus 48(b) is due to changes in the solvent used for sample processing. Similar to the earlier discussion, an increase in the longer time relaxation processes for the unfunctionalized MWNTs is observed. Noteworthy in this case, however, is that upon functionalization the peak of the effective relaxation spectra has shifted an order of magnitude, suggesting that in this case functionalization has had a much more dramatic impact on the primary relaxation mechanisms of the nanocomposite. Such results indicate that for the functionalized sample the entire matrix behavior can now be characterized as different from the behavior of the pure polycarbonate matrix polymer. Additional methods for more quantitative characterization of the change in mobility of the nonbulk polymer interphase currently under development are described in Section 3.2.1.1. These results again highlight the need for further modeling developments across (and between) multiple length scales in order to better understand the nature and extent of the changes in polymer mobility in nanocomposites as a function of synthesis and composition.

3. NANOMECHANICAL MODELING

From the standpoint of modeling, polymer nanocomposites are challenging because of the vast range of length scales that are relevant and need to be considered for such materials. Modeling at atomistic scales using the principles of quantum mechanics and molecular dynamics has been developed to model collections of atoms, but these methods remain computationally expensive and to date have been limited to looking at the behavior of individual nanoparticles and collections of short polymer chains. For larger length scales, the fields of micromechanics and continuum mechanics are well established but cannot directly capture details at the atomistic level that are important for nanostructured materials. The area of nanomechanics is the bridge between these length scales, which is necessary to develop accurate models of nanocomposites and related materials. The development of nanomechanics into a well-grounded area of study will require contributions from these length scales as shown in Fig. 49.

At present, there are two major approaches that have been used to describe material response of nanocomposites at the nanometer length scale as indicated in Fig. 49. The "bottom-up" approach is based on quantum and molecular dynamics with full atomistic detail and typically addresses either individual nanostructures or a small number of short polymer

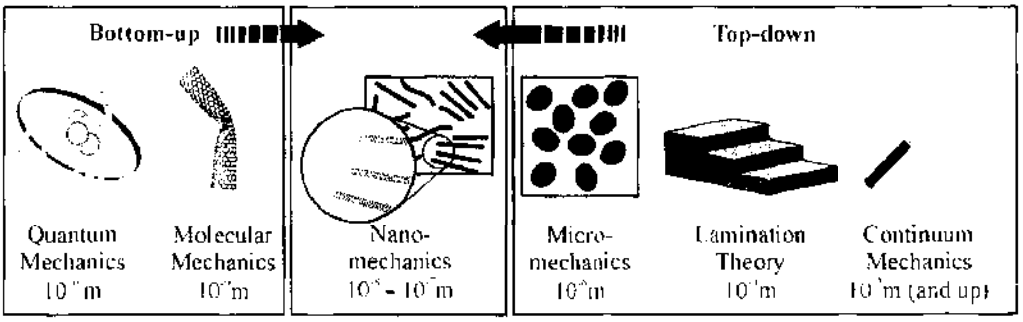


Figure 49. Nanomechanics and other modeling length scales.

chains interacting with a nanoinclusion. The "top-down" approach is based on continuum mechanics, treating the nanoinclusions as continuum elements in methods that attempt to account for the morphology of the nanocomposite. Each of these areas and the various modeling strategies within will be described in detail in the sections below. What is missing in the existing literature is a seamless connection between the molecular level simulations of a single nanoinclusion with a few short polymer chains to the large domain/long time scale simulations of a realistic nanocomposite. Multiscale modeling is a topic of intense current interest and is needed to properly address the changes in properties of the polymer chains near nanoinclusions, including the chemistry between NT and polymer, and its impact on macroscopic properties over many such interacting nanoparticle-polymer domains.

In the following sections, we will describe the major current approaches to modeling of nanocomposites and mention a few specific papers and their results as concrete illustrations of the accomplishments, complications and issues that arise (see also the recent review in Ref. [259]). Space limitations preclude a detailed summary of all the work in this area. However, in Table A.1 in the Appendix provides a more exhaustive summary of the existing nanocomposite models as of the time of this writing.

3.1. Bottom-up Approaches

The bottom-up approaches all start with the consideration of the atomistic description of the nanostructure and surrounding polymer chains. A reader interested in a thorough description of the underlying fundamental principles of these atomistic simulations is referred to a number of articles and books in this area [260–263]. Because polymers are long-chain molecules of high molecular weight, the number of degrees of freedom required to model realistic systems is daunting. For example, a 10 nm radius sphere of low molecular weight polystyrene (15,000 MW) contains about 160 PS chains, each chain consisting of about 2500 atoms. A 10–10 carbon nanotube (~1.3 nm diameter) has ~150 carbon atoms per nm in length and a typical SWNT can be 1 micron (1000 nm) long. To begin to model polymer-nanotube interaction therefore requires at least one nanotube and a large set of sufficiently long polymer chains to capture the local density changes and conformational alterations in the polymer chains near the nanoparticle. As the typical radius of gyration (R_g) of polymers is on the order of 10 nm, even simulations using short nanotubes and sufficient polymer chains to achieve a model size to encompass a realistic R_g can involve millions of atoms. Consequently, most simulations using molecular dynamics consider very short nanotubes (on the order of 10 nm long) and short polymer chains (~10–50 atoms long). Additionally, a united atom model is often used for the polymer chains in which hydrogen atoms are not explicitly included. Another approach uses polymer bead-necklaces, where each bead is representative of a small number of monomer units as shown in Fig. 50.

3.1.1. Molecular Dynamic Modeling

While molecular simulations studying the mechanical behavior of individual nanotubes have appeared in the literature [50, 52, 53, 264–266] and atomistic modeling methods have been used since the mid 1980s for polymers [261], atomistic simulations of nanotube and nanoparticulate-reinforced polymers are much less developed [13, 173, 254, 267–270].

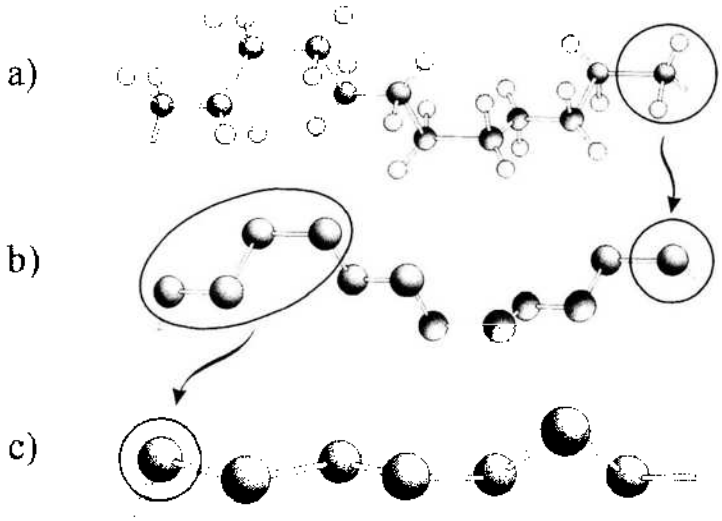


Figure 50. Several approximations to represent a chain of polyethylene in a molecular simulation: (a) full atomistic detail with hydrogen atoms explicitly included; (b) united atom model where hydrogen atoms are implicitly included in each carbon backbone atom; (c) bead-necklace representation where each bead represents four CH_2 monomer units.

As previously mentioned, because of the high molecular weight of polymers, the size of simulations required to truly represent bulk polymer response is prohibitively large. At present, research uses small nanoparticle segments and modest collections of short polymer chains together with coarse graining strategies to generate atomistic simulations for nanocomposites. However, shorter polymer chains may lose some detail for configurational possibilities and entanglements critical for accurate modeling of nanocomposites.

Several papers have used classical molecular dynamics methods to examine changes in the mobility of polymer chains near nanoparticles. The results of these studies help to elucidate the type and extent of polymer modification and can be used in a hierarchical modeling scheme to incorporate an interphase region or graded properties of the polymer near the vicinity of the nanoparticles. While elastic properties may not be greatly altered, the viscoelastic mechanical properties, diffusion characteristics, and thermal properties of the polymer chains within a few radii of gyration of the nanoparticle can be significantly affected. For example, recent work examined polyethylene chains near nanotubes using a multibody Tersoff-Brenner potential for the C–C bonds within the nanotubes and a truncated Lennard-Jones (LJ) potential for the interactions between the polymer chains and the NTs [254]. A united atom model was used for the polymer chains, where the CH_2 and CH_3 groups are the individual units, and interactions between chains are considered with an intra-polymer potential. Small model cells were equilibrated then cooled to obtain density-temperature plots and the glass transition temperature of the polymer was inferred from the slope discontinuity. The results demonstrated a 20°–60°C increase in the T_g for the cell containing the nanotube (see Fig. 51), depending on the size of the simulation and related to the molecular weight of the polymer. It is noteworthy that this work also accurately captured the increase in T_g with increasing molecular weights of the polymer chains. In addition, the coefficient of thermal expansion (CTE) for the cell with the nanotube was significantly higher than that of the polymer alone due to the additional excluded volume introduced by the nanotube [254]. This latter result is in contrast to data for traditional fiber composites in which for typical laminates, the CTE of the composite is reduced over that of the neat polymer, and zero CTE configurations have been designed [271].

Smith et al. have examined interactions of polymer with spherical nanoparticles via molecular dynamic simulations. In their studies, polymer and nanoparticle interactions were examined using a coarse-grained bead-necklace representation of the polymer chains and a truncated LJ potential for polymer-polymer interaction [15, 272]. Nanoparticles created of a rigidly bonded bead network and having a neutral, repulsive or attractive interaction with the polymer were studied by appropriately assigned potentials, with neutral potentials identical

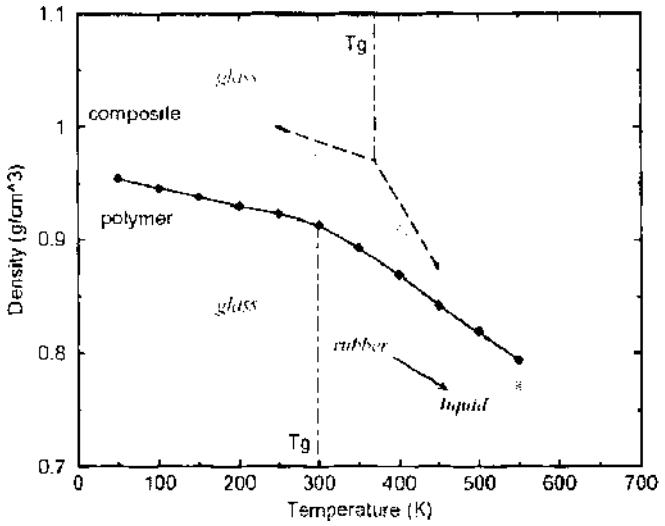


Figure 51. Molecular dynamics calculation of density as a function of temperature and increase in glass transition temperature (T_g). Reprinted with permission from [254], C. Wei et al., *Nano Lett.* 2, 647 (2002). © 2002, American Chemical Society.

to the polymer bead–polymer bead potential. The results revealed that the shear modulus and viscosity of the polymer increased for both neutral and attractive systems and decreased in repulsive systems. Repulsive systems led to faster polymer dynamics, while neutral and attractive systems resulted in slower polymer dynamics. The difference between neutral and attractive systems could be related to an effective corregation of the nanoparticle surface due to the significant barriers for polymer motion along the attractive surface in the interfacial polymer. It was also shown that weak polymer-nanoparticle interactions promoted aggregation of the nanoparticles, while stronger interactions promoted adsorbed polymer layers on the particles and better dispersion [272]. These results are consistent with recent experimental work showing an increased thickness of adsorbed polymer sheathing for functionalized nanotubes [116, 251].

In similar work, molecular dynamics were used to simulate bead-spring unentangled polymer chains near equiaxed nanoparticles approximately 10 nm in size [14]. The simulation consisted of 400, 200, or 100 polymer chains, each 20 monomer units long. For attractive interactions, T_g increases near the particle, while for non-attractive interactions, T_g decreases near the particle. For all cases, the polymer chains are flattened and elongated near the nanoparticle surface, persisting for a distance R_g radially from the particle, compared to bulk polymer chain configurations. Since the flattening does not depend on interaction type, it is suggested that the change in morphology is due to a geometric constraint of the nanoparticle on the polymer chains. A characteristic relaxation time of the melt is defined based on the radially-averaged intermediate scattering function. This relaxation time is shown to increase near the nanoparticle for attractive systems and to decrease slightly for non-attractive systems.

Nanoparticle clustering was examined by MD of bead-spring polymer chains in a melt with polyhedral nanoparticles [273]. Consistent with [272], this study showed that the nanoparticle-matrix interaction strength controlled clustering, with higher interaction (between the polymer and nanoparticle) leading to more dispersed states. Weaker interactions, increased particle concentration, and decreased temperature led to clustering. Although not conclusive, results also indicated that clustering was not a first order phase transition behavior. Rheology simulations showed that viscosity increased for dispersed systems, which is inconsistent with colloidal suspension behavior, but consistent with the concept of an altered polymer region surrounding the nanoparticles affecting the dynamics of the polymer chains more substantially in a dispersed system. These results are useful to help understand the origins and effects of nanoparticle clustering, however, there could be additional geometric effects for platelike and tubelike nanoparticles that were not addressed in this study.

Kumar and co-workers investigated the solidification of a polymer melt in the presence of nanoplatelets [274]. They showed that the unusual relaxation behavior of these composites could be traced to the attractive interactions between polymer chains and filler which create a percolating polymer-filler transient network. The lifetime of this network then controls the mechanical properties of the resulting nanocomposite, especially at low frequencies. More recent work has considered polymer-spherical filler composites and attempted to address the role of this transient network on the mechanical behavior of the composite [275]. The work suggests that the presence of the interface zone leads to an increase in the local segmental relaxation times of the chains, but that the long time behavior of these composites is directly attributable to the presence of this transient network.

While not specifically for nanocomposites, some recent molecular simulation work has focused on the properties of nanosized polymer structures. For example, Boehm and De Pablo have examined nanosized thin films and beams with reference to the types of structures formed in lithography for features in semiconductor fabrication [276]. Their results on the dynamics and mechanics of confined polymer are very relevant for confined polymer regions in nanocomposites. While most other molecular simulations reviewed here use Lennard-Jones potentials, this work uses discontinuous molecular dynamics (DMD), in which the particle-particle interaction is a discontinuous square well form based on distance between interaction sites. The DMD form saves an order of magnitude of computation time compared to continuous interaction potentials such as the Lennard-Jones potential.

The results for the DMD calculations on polymer nanostructures indicate that the T_g decreases for freestanding films due to enhanced segmental mobility near free surfaces, consistent with experimental data [10]. However, T_g increased for all simulations of films on a substrate, of which there were several interaction potentials, including covalent grafting to the substrate. Uniquely, their simulations include explicit calculation of the mechanical response of the nanostructures based on results for bending of small-scale beams. The simulations show that the Young's modulus decreases for thinner cantilever beams, which is consistent with the result of a decreased T_g and increased polymer mobility for the situation of a freestanding film. The results are attributed to the rapid stress relaxation that can occur in the vicinity of the polymer-air interface. Unfortunately no mechanical results for the films on substrates were reported. Magnitudes of molecular strains are calculated in the nanobeam bending scenario and compared to classical continuum calculations. It is seen that the more disperse, localized nature of the plastic strains seen in the molecular model is lost in the continuum model that predicts continuous high strains at the root of the cantilever (see Fig. 52). In contrast, the molecular simulation shows plastic strains much more diffuse and associated near the free surface of the nanobeam. It is pointed out that continuum analyses that account for the density and stress fluctuations might be able to better capture the results for nanoscale structures. The implication of these simulations for nanocomposites is clearly that the constraint of the nanoparticle surfaces within the polymer

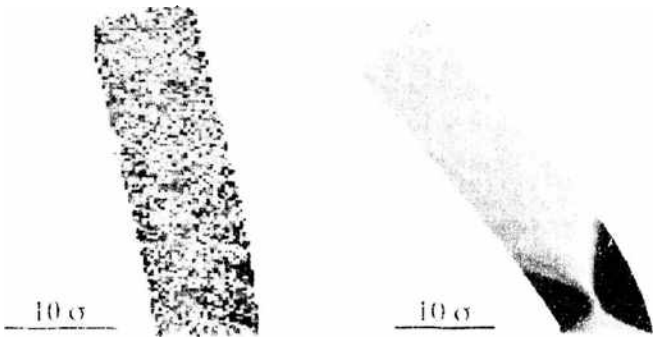


Figure 52. Residual molecular strain (left) and residual continuum plastic strain (right) in a nanocantilever beam bending simulation. Note that the residual strain is more pronounced and concentrated in the continuum simulation that does not account for local fluctuations in density and stresses. σ is a dimensionless parameter related to the potential cutoff distance used in the simulation. Reprinted with permission from [276]. T. R. Bohme and J. J. de Pablo, *J. Chem. Phys.* 116, 9939 (2002). © 2002, American Institute of Physics.

will significantly alter not only the T_g of the material, but also the mechanical properties of the nanocomposite, with results depending on the nature of the interactions between the particle and the polymer chains.

Frankland et al. [184, 277] used molecular dynamic simulations of a nanotube in a polyethylene matrix to study the shear strength and pullout of a nanotube in a polymer matrix. The model was a single nanotube in either a crystalline or amorphous PE matrix, with aligned chains of PE taken as the initial state for the crystalline matrix before equilibration. To simulate functionalization, a low density of crosslinks were added between the NT and the polymer chains. All nonbonded interactions were treated with LJ potentials. Results for interfacial shear strength indicate little change in the initial τ_C for an amorphous matrix with cross-linking, a doubling for the crystalline matrix case, and an order of magnitude increase in the effective τ_C due to restricted sliding of linked polymer chains for the crosslinked cases. Simulation of complete pullout was done for the nonbonded case only [277], where a periodicity in the velocity profile scales with the width of the carbon ring in the nanotube. A simple frictional model was developed in which an effective viscosity can relate the pullout force to the velocity.

Molecular mechanics simulations and elasticity calculations were used to quantify some of the important interfacial characteristics that control the performance of a nanocomposite material. In the absence of chemical bonding between CNT and the matrix, it was found that the nonbond interactions, consisting of electrostatic and van der Waals forces, result in CNT-polymer interfacial shear stresses (at 0 K) of about 138 and 186 MPa, respectively, for CNT/epoxy and CNT/PS (see Fig. 53) [278]. The high interfacial shear stress calculated, about an order of magnitude higher than traditional micron-sized fiber reinforced composites, was attributed to intimate contact between the two solid phases at the molecular scale. Simulations and calculations also showed that local nonuniformity of the CNT and mismatch of the coefficients of thermal expansions between CNT and polymer matrix also promote the stress transfer ability between the two [278]. Other work found that morphological irregularities in the NT structure, such as kinks due to surface defects and changes in outer diameter due to the interior NT structure, in addition to chemical functionalities at defect sites on the NT surface, could lead to increased interaction with the host polymer without explicit chemical functionalization and covalent bonding. This type of NT surface roughness has been shown to increase the energy required for nanotube pullout [278]. Additionally, a fundamental elasticity model was used to illustrate that thermal mismatch between nanotube and host polymer could lead to residual stresses in the polymer inhibiting pullout.

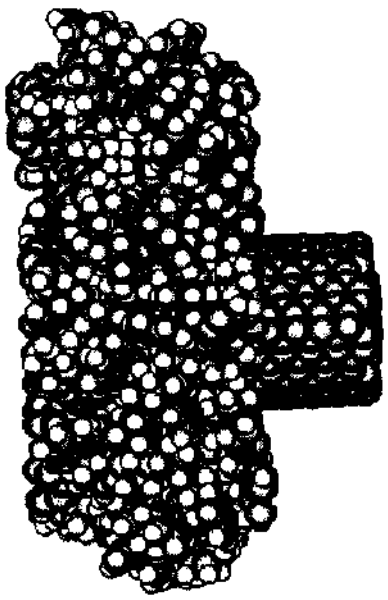


Figure 53. Molecular simulation of a single-walled carbon nanotube being pulled from a polystyrene matrix. Reprinted with permission from [278]. M. Wong et al., *Polymer* 44, 7757 (2003). © 2003, Elsevier.

These various molecular dynamic simulations to date show the extraordinary impact that nanoparticles can have on polymer mobility and intrinsic properties and that such effects permeate for several R_g away from each nanoparticle surface. The changes in mobility depend on the nanoparticle surface, geometry, and chemical interaction with the surrounding polymer, and correspond to changes in glass transition temperature, viscosity, modulus and strength. These changes in turn will impact the associated macroscopic behaviors as well as global properties such as damage tolerance and fracture toughness. While such results provide great insight, many simplifications were made in each case to enable realistic computation times. Notably, polymer chain length, nanoparticle size and shape and interactions with other nanoparticles all need further attention. In addition, although one paper has demonstrated use poly(m-phenylenevinylene) as a coupling agent to enhance the interfacial characteristics of nanocomposites due to helical wrapping around the nanotube [173], the details of bound functional groups on the nanoparticles and their influence on surrounding polymer behavior have yet to be addressed. Some of these issues will be best tackled with multiscale simulation strategies as described in Section 4.

3.1.2. Monte Carlo Methods

In contrast to molecular dynamics, Monte Carlo (MC) approaches have been less utilized to examine polymer and nanocomposite properties. A detailed description of the fundamental principles of the Monte Carlo method can be found in a number of sources in the literature; see, for example, Ref. [260]. The fundamental limitation to MC is that it is inherently a static calculation, and thus evolution or dynamics cannot occur. In an MC simulation, all atoms are moved randomly in each trial step, and a move is accepted if the resulting energies are within certain defined ranges. Many such random moves are made in order to probe the parameter space, but configurations are not walked through time, whereas molecular dynamics solves Newton's equations explicitly, so true dynamics are calculated. With MD one can obtain relaxation times, observe stress relaxation, change temperature and obtain density changes—these are all dynamic problems that MC approaches cannot address. Nevertheless, information on static response can be obtained with MC, such as equilibrium configurations and response to static load. In addition, while MD simulations typically examine short chains ($N \sim 20$) and a single small particle (average particle diameter nearly three times that of the polymer chain bond length), MC simulations can examine long chains and many larger particles, thus approaching more realistic polymer systems.

One example of a Monte Carlo–based approach examined the chain conformations in polymers filled with nanoparticles based on the influence of the excluded volume and the distribution (random versus periodic) of the nanoparticles in the polymer [279]. The study looked at $N = 50{,}000$ polymer chains with the nanoparticle volume fraction varying from 0.2% to 60%. With increase in volume fraction or decrease in inclusion size, there is a decrease in the free volume available for the chains, leading to increased effective stiffness.

A more in-depth study used Monte Carlo simulations of polymer chains and nanosized particles to elucidate polymer-particle interactions, where an offsite lattice MC method is used to account for polymer-particle interactions [280]. Those authors discuss two components to the interaction: a steric repulsion effect that dominates with high density and an attractive bridging effect that dominates under more dilute conditions. This effect has been seen in the case of polymer chains confined between two parallel walls, and the concepts carry over to nanocomposites with the exception that the repulsion effects are less intense in the nanocomposite due to particle curvature. Particle aggregation is observed, though particles remain protected by a layer of interacting polymer. In the polymer, a layering effect was observed due to an oscillatory component of the polymer-particle interaction force. This oscillating component results in layers of dense chains near nanoparticles, with decreasing density moving away from particle.

From the standpoint of modeling the mechanical properties of polymer nanocomposites, it appears at this time that molecular dynamics is favored due to its versatility and the direct incorporation of interatomic forces that simplifies the interpretation of mechanical properties. By comparison, Monte Carlo methods are more suitable for the

determination of equilibrium structures: for example, atomic scale structure of intercalated PEO/montmorillonite nanocomposites [281] and the structure of polymers in the presence of impenetrable surfaces [282]. However, hybrid techniques are currently under development, where MC-type approaches are used to prevent solutions from becoming trapped in local energy minima are complemented with fully atomistic models of the polymer chain to enable modeling of relevant length and time scales: see, for example, Refs. [262, 283–285] and references therein. Such methods may increasingly find use in the modeling and simulation of polymer nanocomposites.

3.1.3. Equivalent Continuum Models

An approach initiated by Odegard and co-workers called the "equivalent continuum model" can be considered to be a pseudomolecular mechanics approach [268, 286–288]. While all the final calculations are done in the continuum domain, close examination of molecular mechanics is used in the development of the model to account for nanostructures and nanostructural features including nanotubes and polymer chains, classifying it in the "bottom-up" category rather than the "top-down" side. Similar pseudo-MD continuum approaches have been developed by other researchers, most notably a "structural molecular mechanics" method by Li and Chou [289]. Other papers have utilized aspects of this method in the context of nancomposites [290, 291]. In this section, we will briefly describe the approach based on Odegard's work in order to place the method and its contributions relative to other bottom-up and top-down approaches. For the interested reader, a separate chapter in this handbook describes the equivalent continuum method in more detail [288].

The equivalent continuum method is developed based on equating energies of molecular mechanics, truss mechanics, and continuum mechanics for the nanostructure(s) of interest. For molecular mechanics, the most general form of the expression for the total molecular potential energy, U^m, is

$$U^m = \sum U^r + \sum U^\theta + \sum U^\tau + \sum U^\omega + \sum U^{vdW} + \sum U^{es} \tag{14}$$

where U^r, U^θ, U^τ, U^ω, U^{vdW}, U^{es} are the energies for bond stretching, bending, torsion, inversion, van der Waals interactions and electrostatic interactions, respectively. In practice, only the most dominant of these terms are included, with Lennard-Jones potentials typically used for the van der Waals interaction terms. For the nanostructure of interest, appropriate force fields are used and the total molecular potential energy determined by summing over all interacting atoms as indicated.

The equivalent continuum method thus relates the molecular model to an equivalent truss system. Figure 54 shows an example for a unit cell of a graphene sheet: the chemical bonds on the left are represented in molecular mechanics by the equation above. An equivalent

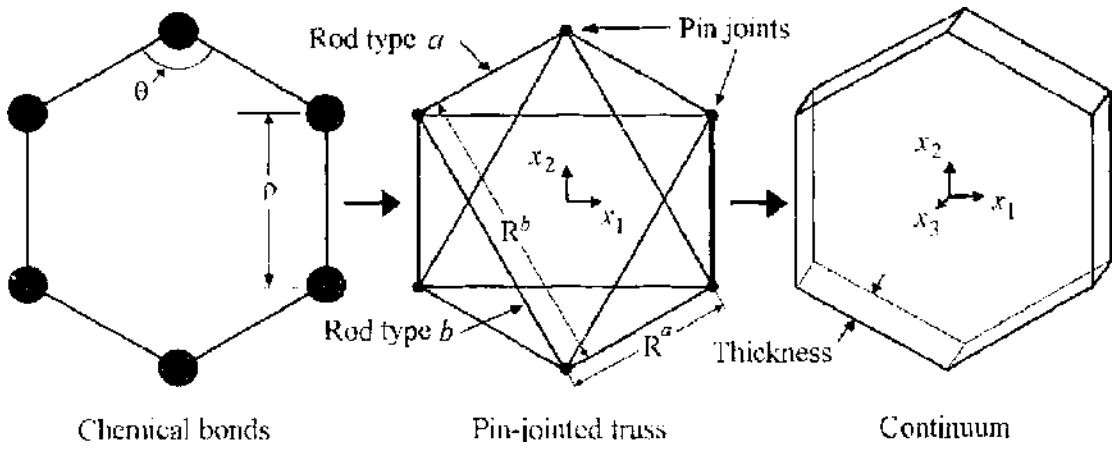


Figure 54. Equivalent continuum approach for the unit cell of a graphene sheet. Reprinted with permission from [268], G. M. Odegard et al., *Compos. Sci. Technol.* 62, 1869 (2002). © 2002, Elsevier.

truss (center image) is formed, in this case with a rods representing the direct bond stretching between atoms and b rods representing the bond angle distortion (torsional springs could alternatively be used). The strain energy of the truss can be easily calculated based on the elastic moduli of the rods and the geometry and deformation of the truss [268]. The energies of the truss and the molecular mechanics models are equated, resulting in expressions for the Young's moduli of the rods based on the force constants from the molecular model. Alternatively, the strain energy of the truss is expressed in terms of geometry and the molecular force constants.

In the last step of the method, the truss model is replaced by an equivalent continuum model by equating the elastic strain energies of the truss and the continuum element under identical loading conditions. Elastic strain energies for continuum structures can be found by consulting any traditional elasticity book in terms of the continuum elastic constants of the material [292]. These can be rewritten to include geometric measures of the structure in question, and such is done in this example in order to determine the relevant continuum thickness for a graphene sheet. The result from the analysis of the graphene sheet demonstrated that to model a graphene sheet as a continuum, the appropriate thickness to use is between 0.57 and 0.69 nm, which is double the commonly used interlayer graphene spacing of 0.34 nm. A similar result is obtained in the equivalent continuum analysis of a SWNT as illustrated in Fig. 55: again the relevant thickness to use for a continuum description of the SWNT (0.65 nm) is approximately double that typically assigned in the literature.

The equivalent continuum method can be applied also to nanoparticles in the presence of polymers as has been done in several papers [286, 287]. For example, in Ref. [286] an MD simulation was used to determine the equilibrium positions of polymer chains surrounding a SWNT. An equivalent truss model was built for the polymer and nanotube, and then an equivalent continuum model was built for a cylindrical continuum domain to represent the NT and local polymer (see Fig. 56). The transversely isotropic moduli of the continuum were then found by equating strain energies of the systems as before. In this manner, the moduli of a unit cell consisting of a nanotube and polymer were created. From this point, a traditional micromechanics approach (see Section 3.3) was used to develop effective properties of a nanocomposite with many interacting inclusions by using the moduli of the polymer-NT unit cell (referred to as an effective fiber) together with additional bulk matrix material.

A recent application of the equivalent continuum model for nanoparticulate-polymer composites is highlighted in Fig. 57 [287]. Initially, similar to the work of Smith, Glotzer, and others [13–15, 272, 273], a molecular dynamics simulation involving coarse-grained simulations and a reverse-mapping technique is used to determine the molecular structure of the

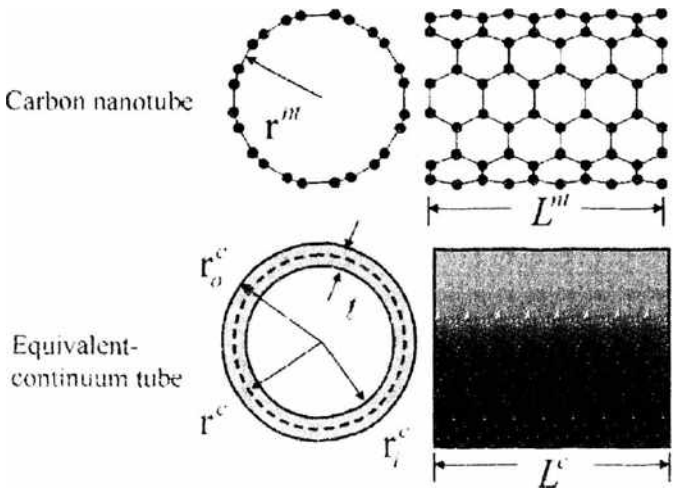


Figure 55. Equivalent continuum representation of a carbon nanotube. Reprinted with permission from [268], G. M. Odegard et al., *Compos. Sci. Technol.* 62, 1869 (2002), © 2002, Elsevier

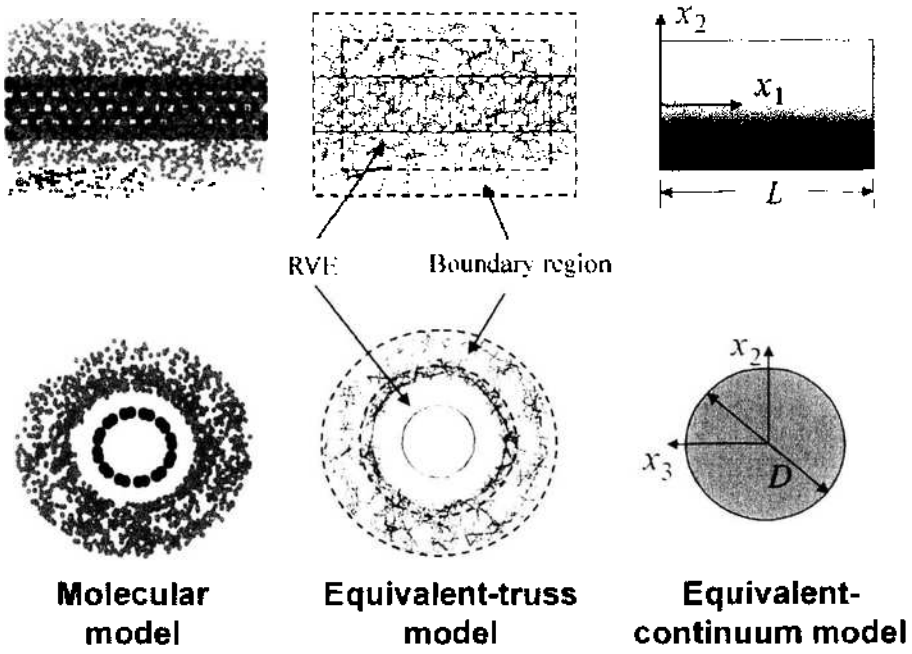


Figure 56. Equivalent-continuum modeling of an effective fiber applicable for nanotube-polymer composites. Reprinted with permission from [286], G. M. Odegard et al., *Compos. Sci. Technol.* 63, 1671 (2003). © 2003, Elsevier.

polymer and interphase region surrounding the equiaxed nanoparticle. In this work, a variety of specific silica-polyimide interfacial conditions were modeled, including (a) no surface treatment, (b) hydroxyl groups bonded to the silicon atoms, (c) phenoxybenzene ($-C_6H_4-O-C_6H_5$) chemically bonded to the silica, and (d) a hydroxylated surface as in (b) covalently

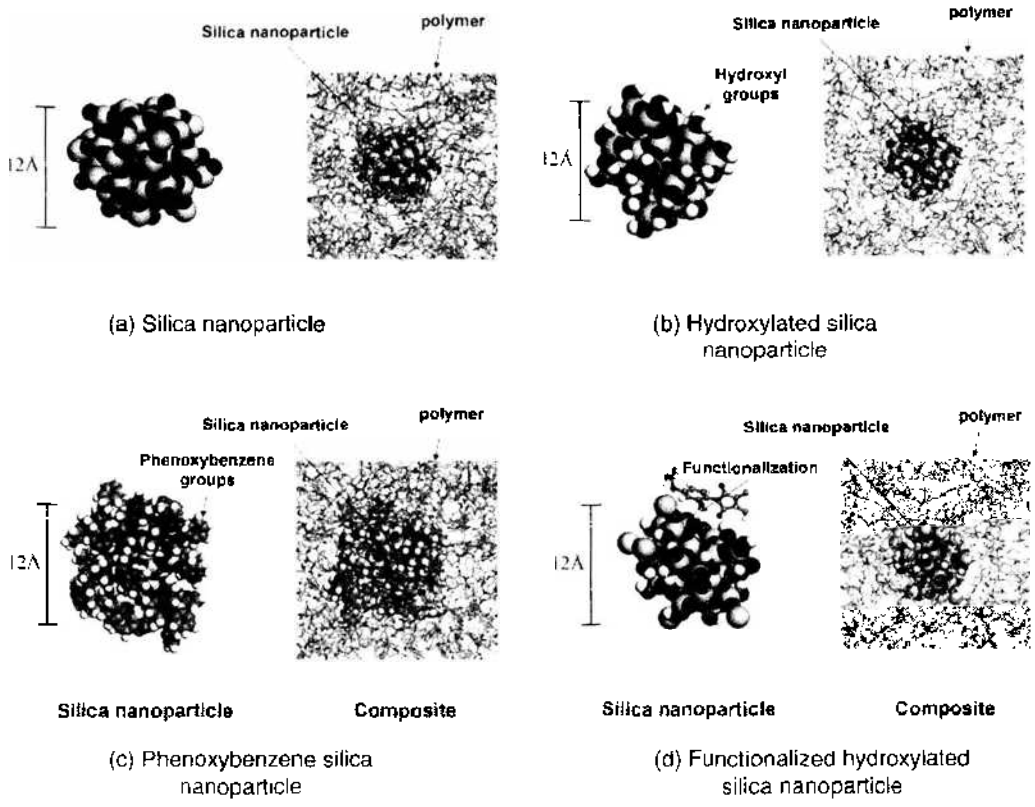


Figure 57. Representative volume element (RVE) of the molecular structure of four types of silica nanoparticles and their incorporation into a polyimide matrix. Reprinted with permission from [287], G. M. Odegard et al., *Polymer* 46, 553 (2005). © 2005, Elsevier.

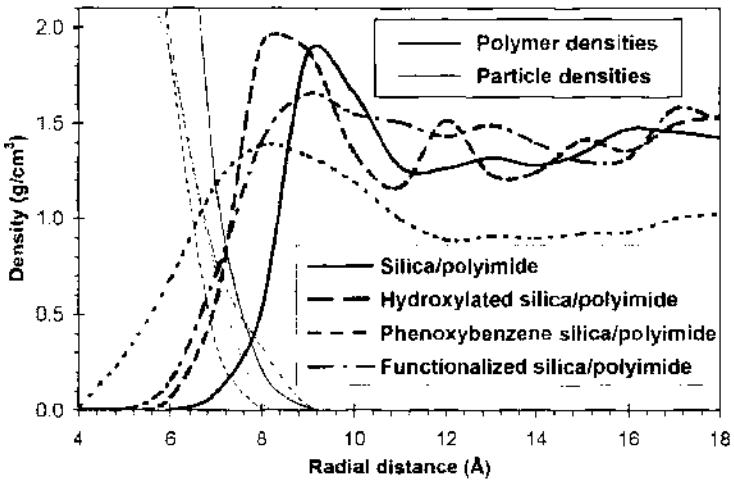


Figure 58. Radial density profiles for silica-polyimide nanocomposites. The density of the core structure of the silica remained constant (2.65 g/cm^3). The density of the pure bulk polyimide was 1.33 g/cm^3. Reprinted with permission from [287], G. M. Odegard et al., *Polymer* 46, 553 (2005). © 2005, Elsevier.

bonded to the surrounding polyimide. Only in the last case (d) was the nanoparticle explicitly bonded to the surrounding polyimide molecules.

The impact of the surface chemistry of the embedded silica nanoparticle on the surrounding polyimide matrix is evident upon inspection of the radial density profiles shown in Fig. 58. For the non-functionalized composites (see Fig. 57) an approximately 33% increase in the maximum polymer density compared with the "far-field" radial density at 1.8 nm is noted; a corresponding value of only 6% is observed for the functionalized sample (Fig. 57).

The equivalent continuum method as described above was used based on the final molecular equilibrium configurations of each case. Results showed a decrease in effective composite modulus in all cases that was attributed to the decreased polymer densities near the nanoparticle (see Table 4). However, the differences in local polymer density/structure as a function of surface treatment also resulted in changes of the overall properties, with the functionalized system resulting in the smallest decrease in stiffness due to the direct bonding with the matrix polymer. This effect was found to be more important to the stiffness than the local polymer density.

From a continuum sense, this *decrease* in moduli for the nanocomposites as a result of silica nanoparticle "reinforcement" was attributed to the formation of an effective third phase (interphase) of polymer surrounding the nanoparticle with mechanical properties *less* than that of the pure polymer. The properties of this interphase polymer were determined based on the solution of the inverse three-phase micromechanical problem; given the properties of the silica nanoparticles and the effective properties of the composite (determined from the molecular simulations), the volume fractions of the phases 1.7%, 45.2%, and 53.1% for the silica nanoparticle, interphase, and polyimide matrix), and assuming that the interphase

Table 4. Effective isotropic properties for the overall composite RVE and the non-bulk interphase for silica-based nanocomposites with various surface treatments. The effective volume fraction of the silica nanoparticle within the RVE was 1.7%. E = Young's modulus, G = shear modulus. (Data from [287].)

		Pure silica	Pure polyimide	Silica NPC (Fig. 58a)	Hydroxylated silica NPC (Fig. 58b)	Phenoxybenzene silica NPC (Fig. 58c)	Functionalized silica NPC (Fig. 58d)
Composite RVE	E(GPa)	88.7	4.2	3.4	3.3	2.2	4.0
	G(GPa)	41.0	1.5	1.2	1.2	0.8	1.5
Interphase modulus	E(GPa)	—	—	2.4	2.2	0.3	3.5
	G(GPa)	—	—	0.9	0.8	0.1	1.3

material is isotropic with a Poisson ratio equal to that of the pure polyimide. The effective properties of this interphase polymer based on this model are given in Table 4, which again highlights how surface chemistry and interactions at the nanoscale can alter local mechanical properties of the host polymer matrix, leading to perhaps unanticipated changes in mechanical properties at the macroscale continuum level. Note that while the interphase properties in the case of the functionalized silica nanoparticle are still less than those of the bulk polyimide, these properties are greater than the interphase properties for the un-functionalized case, suggesting that here functionalization was successful (in a relative sense) in increasing the mechanical properties of the nanocomposite. Such examples illustrate the utility of the equivalent continuum approach to begin to effectively bridge between molecular simulations and higher order continuum approaches. Further discussion of this approach is discussed elsewhere in this handbook [288].

3.2. Top-down Approaches

In contrast to bottom-up approaches that directly account for atomistic information for the nanoparticles and polymer chains, classical continuum mechanics offers methods to study these materials without atomistic level discretization. A number of papers have assessed nanocomposite properties from the continuum side, including methods involving structural mechanics such as beam theory, micromechanics, finite elements, and density functional and field methods used in complex fluids [38, 61, 268, 289, 294–305]. Most common have been estimates of nanocomposite elastic moduli based on simple rule of mixtures and micromechanics approaches such as the Mori-Tanaka or Halpin-Tsai methods. Initial work typically greatly overestimated modulus due to an inherent assumption of straight, aligned inclusions, which was not representative of the experimental materials. Recent papers have illustrated better agreement with experimental data on elastic moduli by using the classical Mori-Tanaka micromechanics approach with randomly aligned inclusions [298–300]. As will be shown in this section, these and other methods are being expanded to encompass more critical features of nanocomposite microstructure to improve predictions and understanding of material behavior.

For all of the top-down approaches mentioned below, a critical assumption is that the components of nanocomposites all behave as a continuum and can be treated with traditional elasticity, viscoelasticity and other constitutive relations developed for continua. Since the nanoparticles have a dimension on the order of nanometers, such an assumption must be carefully considered. For example, a 10–10 single-wall carbon nanotube has only 20 atoms around its circumference, and its behavior is best treated by atomistic simulations that can include quantum effects and interatomic potentials. In addition, as described earlier the confined polymer chains in the vicinity of the nanoparticles have different molecular configurations from polymer chains in the bulk as interactions with the nanoparticles affect their behavior. Nevertheless, continuum approaches have been used quite successfully to explain some larger scale mechanical behavior of nanoparticles, and interpretation of experimental results for nanotube properties have in many cases used basic continuum theories since the initial work in this area [60, 306]. For example, it was found that a bending analysis modeling a MWNT as a continuum was satisfactory if the wall-thickness of each layer of the MWNT was considered [61].

With the understanding that the explicit molecular-level interactions between nanoparticle and local matrix are not well represented, continuum approaches offer a clear way to systematically examine macroscale nanocomposites and their properties as a function of microstructure. The most useful methods are in the micromechanics domain where either analytical or numerical approaches allow discretization of a unit cell of matrix material and nanoparticle(s) that can then be used to obtain global properties for nanocomposites of many inclusions via homogenization methods. In this section, we will discuss micromechanics approaches first, with an emphasis on the Mori-Tanaka method. This is followed by a section on the use of the Halpin-Tsai equations and a section on finite element approaches. Then hybrid methods which couple two or more approaches as a start to multiscale modeling to enable more accurate incorporation of local behavior are discussed. Perhaps the

most promising methods under current investigation are multiscale methods which couple molecular dynamics approaches locally with continuum methods globally.

3.2.1. Micromechanics modeling

Micromechanics methods have been applied with great success to traditional composite materials to understand and predict material thermomechanical response as a function of microstructural features. Theories are typically based on classical solutions for a single inclusion in an infinite medium [307], with a variety of assumptions introduced to create predictive models for composites with dilute to high volume concentration of interacting inclusions. Such methods include the self-consistent and generalized self-consistent methods [308], the Mori-Tanaka method [309–311], the Halpin-Tsai method [312, 313], and method of cells [314]. Very rudimentary methods are the Voigt and Reuss upper and lower bounds, respectively, which assume two phases arranged in a parallel or series fashion, respectively. A critical evaluation of these and other micromechanics approaches are discussed by Christensen [315] and more recently by Tucker [316].

All of the micromechanics methods predict response of the composite based on volume fraction of inclusion plus individual properties of the matrix and inclusion phases and can typically account for inclusion shape and orientation distribution within the composite. In general, most established micromechanics approaches provide similar results for modulus predictions at relatively low volume fractions, with these differences becoming more pronounced at higher volume fractions or as the modulus of the inclusion decreases towards zero (for the case of voids) [315, 317]. These methods all predict the effective modulus of heterogenous materials, but cannot provide results for stress or strain localization due to the effective medium averaging approach.

Given the success and utility of micromechanics models for traditional composites, many have been applied to the study of nanocomposites to interpret and predict mechanical response for these complex materials. Micromechanics are based on a continuum approach, and thus the results are subject to the accuracies of that assumption for the particular case of consideration. Early attempts for nanocomposites often used elementary micromechanics approaches based on the assumption of unidirectional alignment of fibers and thus predictions were often in great discrepancy with experimental data. However, more recent applications have taken into account random orientation of nanoparticles, nanoparticle aspect ratio, nanoparticle clustering, and effects of an interphase, with the resulting predictions now in closer agreement with experimental data and becoming increasingly useful for interpretation of microstructural influences. In the sections below, we will outline how micromechanics methods have been applied to nanotube and nanoparticle composites, with special emphasis on how these methods have allowed better insight into nanocomposite structure-property relationships.

3.2.1.1. Mori-Tanaka Method. Many micromechanics methods are based on the classical Eshelby inclusion analysis for a single inclusion in an infinite matrix. A key concept is that of the dilute strain concentration tensor, A_r^{dil}, which relates the average strain in inclusion r, $\bar{\varepsilon}_r$, to the far field strain tensor $\bar{\varepsilon}$:

$$\bar{\varepsilon}_r = A_r^{\text{dil}}\bar{\varepsilon} \tag{15}$$

Using the concept of eigenstrains [310, 311, 314, 318, 319], it can be shown that

$$A_r^{\text{dil}} = [\mathbf{I} + S_r C_0^{-1}(C_r - C_0)]^{-1} \tag{16}$$

where S_r is the Eshelby tensor accounting for inclusion shape, C_0 and C_r are the stiffness tensors of the matrix and inclusion materials, and $\mathbf{I}$ is the fourth order identity tensor. Use of the Eshelby approach and Eq. (16) generalizes the methods for any ellipsoidal inclusion shape, ranging from spheres to cylinders to needle- and plate-shaped inclusions. For the "dilute" case were the volume fraction of inclusions is small enough that interaction between

inclusions is negligible, the dilute approximation for the effective composite modulus is given by

$$C = C_0 + f_r(C_r - C_0)A_r^{\text{dil}} \tag{17}$$

where f_r is the volume fraction of the inclusion.

To consider non-dilute concentrations of inclusions, the various micromechanics methods diverge in their analytical approach to account for the interaction of the inclusions. The self-consistent method assumes that a single inclusion is embedded into an infinite effective medium of as yet unknown properties and an algorithm is developed to yield implicit formulae for the effective moduli that can be solved numerically. On the other hand, the Mori-Tanaka method assumes that each inclusion feels an average matrix strain $\bar{\varepsilon}_0$ as opposed to the far field applied strain,

$$\bar{\varepsilon}_r = A_r^{\text{dil}}\bar{\varepsilon}_0 \tag{18}$$

which results in closed form explicit expressions for the effective moduli,

$$C = \left(f_0 C_0 + \sum_{r=1}^{N-1} f_r C_r A_r^{\text{dil}}\right)\left(f_0 \mathbf{I} + \sum_{r=1}^{N-1} f_r A_r^{\text{dil}}\right)^{-1} \tag{19}$$

where there can be N different types of inclusions, with volume fractions f_r and each with distinct moduli, C_r. For additional derivation details for the Mori-Tanaka method, see the Appendix or a basic reference [310, 311, 314, 318]. The expression in Eq. (19) is valid for unidirectionally aligned inclusion phases. In order to account for other orientation distributions, orientational averaging must be performed. For a random orientation of inclusions, Eq. (19) becomes

$$C = \left(f_0 C_0 + \sum_{r=1}^{N-1} f_r \{C_r A_r^{\text{dil}}\}\right)\left(f_0 \mathbf{I} + \sum_{r=1}^{N-1} f_r \{A_r^{\text{dil}}\}\right)^{-1} \tag{20}$$

where brackets $\{\}$ represent the average of a quantity over all possible inclusion orientations.

Figure 59 shows traditional Mori-Tanaka predictions for the enhancements to elastic stiffness for a composite with nanotube (fiber) or graphite plate reinforcement. The predictions assume the inclusion modulus is 500 times that of the matrix material, reasonable for nanotubes or graphite platelets. Note the large difference between aligned versus randomly

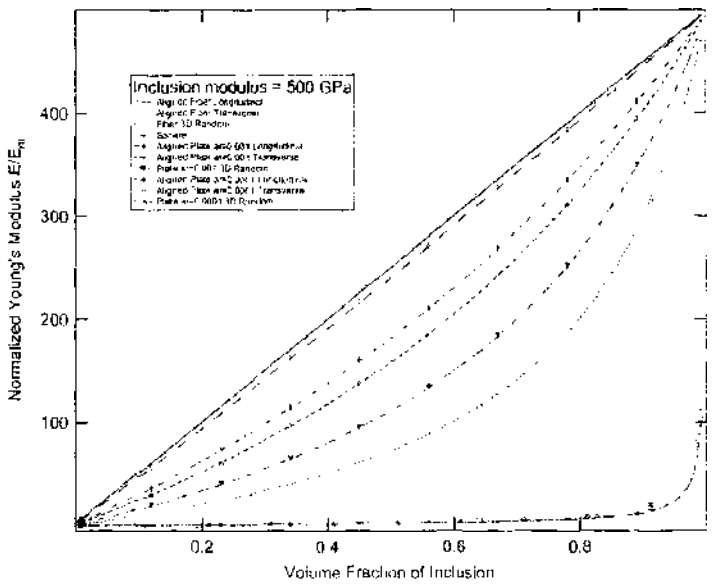


Figure 59. Illustration of baseline Mori-Tanaka modulus predictions for varying inclusion volume fraction with an inclusion modulus 500 times the matrix modulus, E_m (ar = aspect ratio).

oriented inclusions. As most nanocomposites have randomly oriented nanoparticles at relatively low volume fraction, Fig. 60 shows the same data for less than 15% volume fraction inclusion. Also in Fig. 60 is the prediction for inclusion stiffness 100 times that of the matrix for comparison. Note the large difference in reinforcing effect due to the change in the fiber modulus alone. While uniaxially aligned fibers have a slightly higher stiffness than those for high aspect ratio plates, the 3D random orientation of plates is more effective at increasing stiffness for the same volume fraction. Note also that spherical particles do not provide much stiffness enhancement. The predictions shown here assume perfect dispersion of nanoparticles, straight/flat nanoparticles, no clustering, elastic matrix properties, a perfectly bonded interface, and no alteration of matrix properties near the nanoparticle. Thus, they represent the simplest possible assessment of composite moduli. As will be demonstrated in the sections below, many modifications to this simple picture have emerged to address some of these individual issues and have shed light on the underlying mechanisms of nanoparticle reinforcement in composites.

While the MT predictions in Figs. 59 and 60 illustrate the effect of different shapes of nanoinclusions on the elastic properties of nanocomposites, investigation of viscoelastic behavior can provide a richer set of information due to its sensitivity to the impact of nanoparticles. Many viscoelastic characterizations are performed in frequency or temperature space, resulting in the storage and loss moduli as shown in Sections 1.3.2 and 2.4. The MWNT reinforced polycarbonate system considered earlier (see temperature domain data in Fig. 42) was also tested in the frequency domain at multiple temperatures and the results shifted via time-temperature superposition (TTSP; see 1.3.2 and Fig. 19) to obtain master curves for the storage and loss moduli as a function of frequency. As shown in Fig. 61, the addition of small weight fractions of nanotubes to the polycarbonate produces significant broadening of the loss moduli and increase of the loss response at low frequencies, in addition to the enhancements in elastic stiffness demonstrated in the storage modulus.

One of the advantages of analyzing the frequency domain response of nanotube-reinforced polymers is that micromechanical models are available for the analysis and interpretation of results. Micromechanics methods developed for elasticity can be applied to linear viscoelastic behavior in a straightforward manner using the Dynamic Correspondence Principle [113]. In this approach, a direct analogy between elasticity solutions and viscoelasticity solutions in the frequency domain is used in which the elastic moduli are replaced by their corresponding complex viscoelastic moduli of the form $\bar{E}^* = E' + iE''$, and associated field quantities are allowed to be complex [320]. Thus the Mori-Tanaka solution for a multiphase composite with

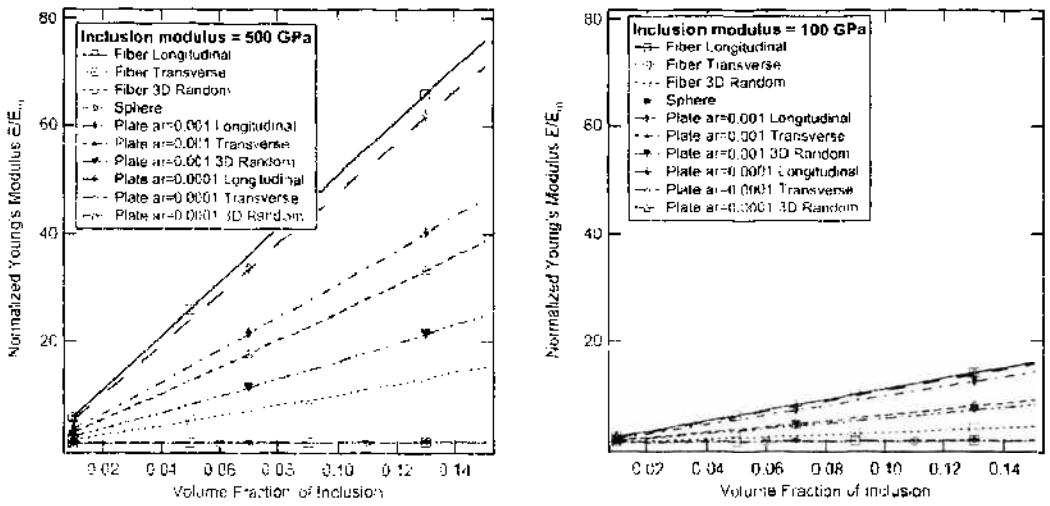


Figure 60. Comparison of low volume fraction reinforcing effect for inclusions with $E_{\text{inc}} = 500$ (left) and $E_{\text{inc}} = 100$ (right). Vertical scales are identical in both plots for easy comparison of impact of inclusion modulus alone (ar = aspect ratio).

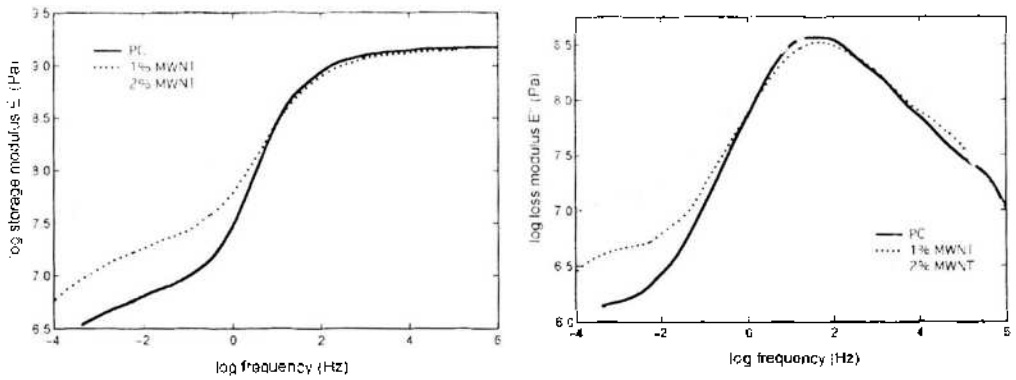


Figure 61. Frequency domain data obtained using dynamic mechanical response for MWNT-reinforced polycarbonate. The pure PC response was obtained using the TTSP data shown in Fig. 19. Reprinted with permission from [140], E. T. Fisher, Ph.D. dissertation, Northwestern University, 2002. © 2002.

viscoelastic phase materials can be written as

$$\bar{C}^* = \left(\sum_{r=1}^{N} c_r \{ \bar{C}_r^* \bar{A}_r^* \} + c_0 \bar{C}_0^* \right) \left(\sum_{r=0}^{N} c_r \{ \bar{A}_r^* \} \right)^{-1} \tag{21}$$

In this form, the Mori-Tanaka method has been employed to investigate viscoelastic and interphase effects in microcomposites, nanocomposites and block copolymers [252, 311, 319, 321, 322]. Results for two and three phase viscoelastic composites showed that the Mori Tanaka method captures the essential features of the storage and loss moduli as a function of frequency when compared to a finite element simulation, though some discrepancies in the magnitudes of the moduli values exist, especially with extreme property differences between the phases [252, 321].

Using this approach, the transverse viscoelastic response of a two phase composite with a viscoelastic matrix (using the Prony series elements for the polycarbonate found from the analysis of Fig. 19) and 1 and 5 vol% of cylindrical isotropic inclusions with $E = 500$ GPa and $\nu = 0.3$ (appropriate for MWNTs) can be obtained as shown in Fig. 62. Here the transverse modulus is considered because it is a matrix-dominated property. While the model predictions show an increase in the storage modulus of the composite (analogous to the increase in elastic modulus that would be obtained for an elastic micromechanics analysis), there is almost no change to the transverse loss modulus response of the system. Since the loss modulus is a measure of the viscous behavior of the composite, the presence of an elastic inclusion (nanotube) should not affect E'' significantly. However, such a result is not the case for NRPs as is clearly seen in Fig. 61 above, in which the loss modulus is profoundly affected by just 1 and 2 weight percent (0.5 and 1 volume percent) of MWNT inclusions. The obvious discrepancy between Fig. 62 and Fig. 61 is due to the existence of a substantial interphase region with altered polymer properties in nanocomposites. For a micromechanics model prediction to match the experimentally observed changes in the loss moduli, a viscoelastic interphase region with properties differing from the bulk matrix

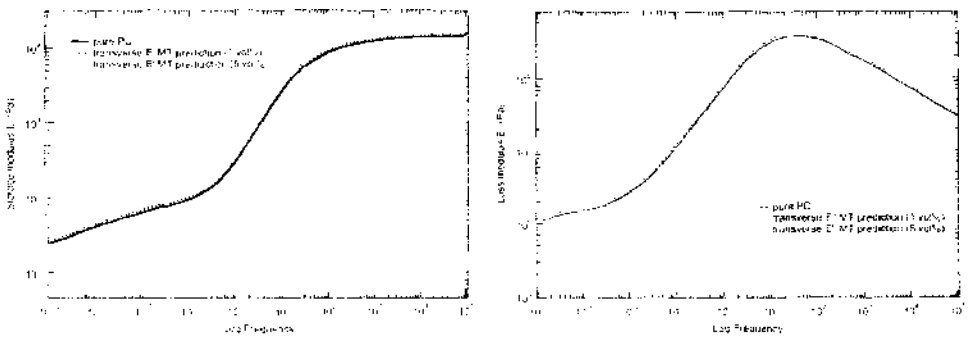


Figure 62. Two-phase frequency domain Mori-Tanaka prediction for transverse modulus assuming aligned cylindrical elastic inclusions and a viscoelastic matrix.

properties must be included. The interacting effects of two viscoelastic phases of differing properties will influence the effective viscoelastic response of the composite, resulting in frequency behavior that differs from that of the bulk matrix material alone.

Application: partially exfoliated nanoplatelets. While the ideal situation for a clay nanocomposite is individually dispersed nm thick clay sheets, current processing methods typically lead to composites that are considered partially intercalated and partially exfoliated as shown in Fig. 14. Thus for nanoclay-reinforced polymers, the idealized model of a two phase composite is clearly inappropriate. Recent work by Luo and Daniel [296] has addressed this issue by developing a three-phase model to account for the intercalated and exfoliated nature of many clay nanocomposites. In their work, the model accounted for the microstructure in a successive fashion as shown in Fig. 63. First, the anisotropic modulus properties of the multilayer intercalated clay were calculated based on a parallel layered plate theory using basic elasticity theory (a similar approach for the effective particle was developed by Boyce and co-workers [301]). The properties of the silicate plates were taken to be 170 GPa and the polymer nanolayers were assigned the matrix modulus multiplied by a factor to account for the confinement effect. Secondly, the effective property of the exfoliated region was calculated for a specific volume fraction of individual silicate layers randomly dispersed in the polymer matrix using the Mori-Tanaka method. Finally, the exfoliated effective modulus was used as the matrix phase with inclusions of the intercalated multilayer clay in a subsequent Mori-Tanaka calculation to yield the overall effective properties of the nanocomposite with inclusions at dual length scales.

Luo and Daniel's work compares their three-phase model to experimental data from two different epoxy clay nanocomposites at several different volume fractions with good results [296]. They then examine the effect of the degree of exfoliation, clay cluster aspect ratio, intragallery polymer modulus and poisson ratio on the effective modulus (see Fig. 64). They found that the fraction of exfoliated clay had the most influence on overall modulus, and that intercalated clusters with high aspect ratios were also effective at providing modulus enhancement to the nanocomposite.

Application: inverse problem for the interphase behavior. The viscoelastic Mori-Tanaka approach has been used recently to explore the nonbulk polymer phase near nanotubes in a nanocomposite [140, 323]. The procedure solved an inverse problem to determine appropriate properties of an interphase region given experimental data on the polymer and nanocomposite behavior and a micromechanics model including the interphase. In that work, the nonbulk polymer phase is assumed to be a distinct interphase region, within which interactions between the nanotubes and the polymer chains alter the characteristic relaxation times of the material. Recalling the Prony series form for a viscoelastic modulus discussed previously, the change in relaxation times of the interphase region can be modeled via the introduction of a mobility parameter α, which relates the mobility (and hence the mechanical properties) of the interphase to that of the pure polymer matrix [16, 140], such that

$$E(t) = E_\infty + \sum_{j=1}^{N} E_j e^{-t/\alpha\tau_j} \tag{22}$$

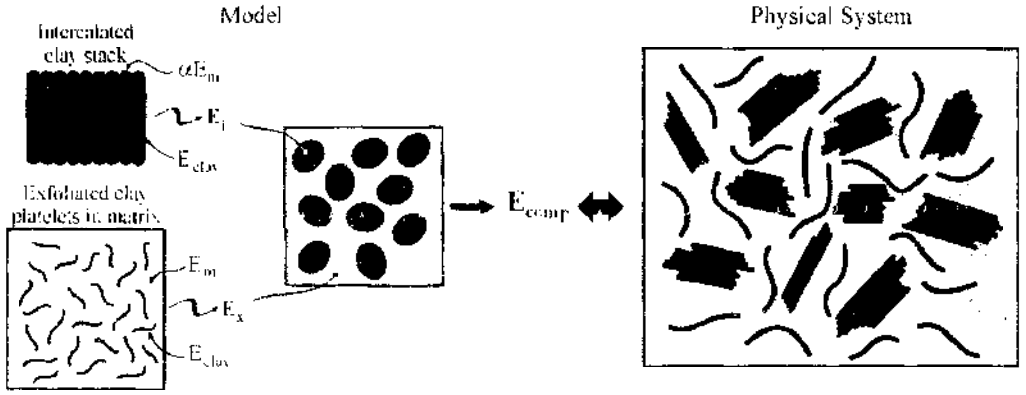


Figure 63. Modeling schematic for Luo and Daniel approach to account for partially intercalated and partially exfoliated clay nanocomposites. Current processing methods typically result in microstructure similar to that shown here.

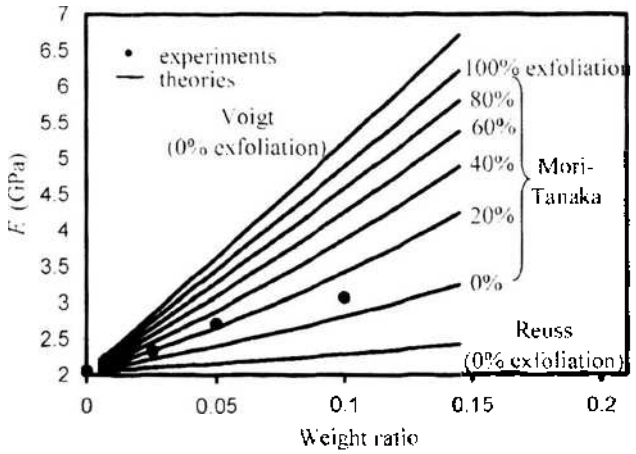


Figure 64. Young's modulus of epoxy/clay nanocomposite as a function of clay concentration for different exfoliation ratios. Reprinted with permission from [296]. J. J. Luo and I. M. Daniel, *Compos. Sci. Technol.* 63, 1607 (2003). © 2003, Elsevier.

It is also likely that interactions between the polymer chains and the nanotubes will result in changes in the magnitude of the time and frequency domain moduli of the interphase region with respect to that of the pure polymer. While not included in Eq. (22), such changes could be incorporated into this model via multiplication of the Prony coefficients E_i.

The effect of the mobility parameter α on the time domain modulus of the material is shown schematically in Fig. 65. In this manner the mobility parameter α can be used to qualitatively characterize the change in mechanical properties of the nonbulk polymer phase. For $\alpha < 1$ the polymer chains are more mobile, representative of a repulsive interaction between nanotube and polymer, which shifts the transition region of the time-dependent response to shorter times. For $\alpha > 1$, the polymer chains are less mobile, representative of an attractive interaction between nanotube and polymer, and the response of the nonbulk phase is stiffer than the bulk polymer. By definition, $\alpha = 1$ describes the bulk polymer response. Assuming such a form for the mechanical properties of the nonbulk polymer interphase region, one can use various micromechanical models to interpret experimental data for carbon nanotube-reinforced polymers.

Viscoelastic data in a form similar to that shown in Fig. 42 for a MWNT reinforced polycarbonate were examined, where the broadening of the loss peak was postulated to be due to decreased mobility polymer interphase near the nanotubes. Mori-Tanaka modeling results for the effective frequency domain response for a 2 wt% nanotube-reinforced polycarbonate sample are shown in Fig. 66. It is noted that in the three-phase model, the interphase is

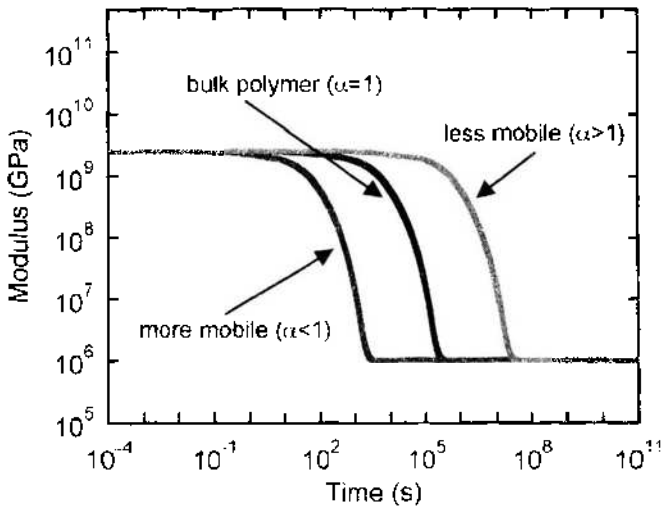


Figure 65. Time-dependent modulus as a function of the mobility parameter α.

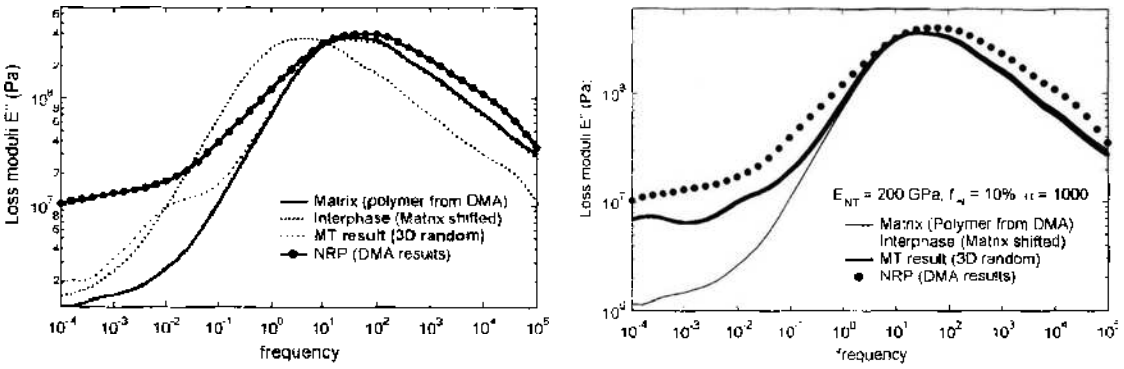


Figure 66. Experimental data and viscoelastic Mori-Tanaka predictions for 2 wt% MWNT sample loss modulus, assuming an interphase volume fraction f_{int} of 10%. (left) Mobility parameter α of the interphase = 100. (right) Mobility parameter α of the interphase = 1000. Reprinted with permission from [140], E. T. Fisher, Ph.D. dissertation, Northwestern University, 2002, © 2002, E. T. Fisher.

a separate inclusion phase and is not geometrically constrained to surround the nanotubes. Extensions of this approach to account for arbitrary inclusion shape using a hybrid modeling approach are discussed in Section 3.3.

For the nanocomposite model, the viscoelastic matrix properties were equal to those obtained for the bulk polycarbonate samples tested, and a 3D random orientation of NTs was assumed. A nanotube (elastic) modulus of 200 GPa was chosen by fitting the high frequency portion of the Mori-Tanaka effective storage modulus to the experimental storage modulus of the nanocomposite. The weight fraction of the nanotubes from the experimental data was converted to volume fraction for the model based on a 2:1 relationship, and the interphase volume fraction was chosen as 10%. As described, the interphase viscoelastic properties were modeled as the bulk matrix material with a simple shift in relaxation times using the mobility parameter approach. Using an iterative approach, a mobility parameter α = 1000, corresponding to a three decade shift in interphase relaxation times was determined. It should be noted that the prediction of nanocomposite response with no interphase (all bulk matrix properties) is very close to the pure matrix response (see Fig. 62) for the loss modulus, as the nanotube does not contribute to the damping response. These results agree with molecular dynamics simulations showing a three order of magnitude change in chain mobility [324] and relaxation times [325] near physioabsorbing surfaces [326].

Figure 67 compares the Mori-Tanaka prediction and the experimental data for the effective loss modulus for the 1 wt% MWNT-PC sample. Using the same values for the interphase volume fraction (10%) and the mobility parameter (α = 1000), we again see very good qualitative agreement between the Mori-Tanaka model and the experimental data.

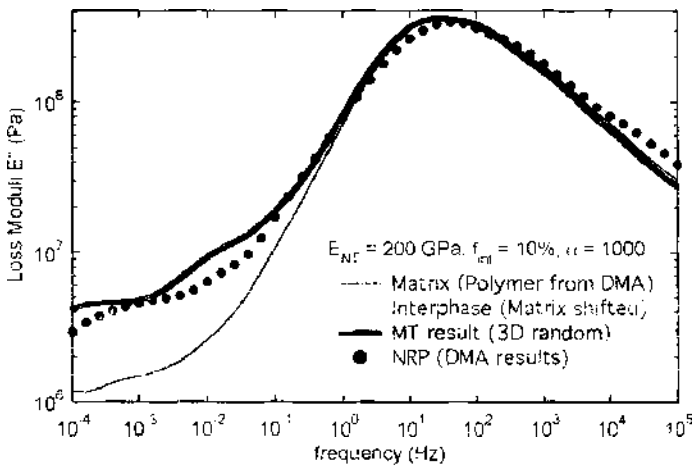


Figure 67. Experimental data and Mori-Tanaka prediction for 1 wt% MWNT sample loss modulus, assuming f_{int} = 10% and α = 1000. Reprinted with permission from [140], E. T. Fisher, Ph.D. dissertation, Northwestern University, 2002, © 2002, E. T. Fisher.

While many assumptions are made in this initial modeling effort to capture interphase properties, the results show that presence of a substantial interphase zone is essential to capture the nanocomposite properties. Additionally, the results predict a 10^3 decrease in the mobility of the interphase that is consistent with an increase of long time-scale relaxation modes for polymer chains in the vicinity of an attractive nanoparticle interface.

The work described here represents our initial efforts to adapt a "top-down" approach using established micromechanics analyses to model the mechanical response of nanotube-reinforced polymers. Eventually, we foresee a more multiscale approach, where the viscoelastic properties of the interphase are based on molecular dynamics simulations of the NT-polymer system and/or nanoscale experimental data. The extension of such modeling efforts for other nanoparticle-reinforced polymer systems is underway. These initial modeling efforts are illustrative of the type of future research directions necessary to better understand the fundamental nanomechanics of nanoreinforced polymer systems as discussed further in Section 4.

3.2.1.2. Halpin-Tsai Methods. Another micromechanical method that has been used to model the mechanical response of nanoreinforced polymer composites is the Halpin-Tsai approach [312, 313]; see, for example, work with nanoclay-reinforcement [297, 327] and nanotube-reinforcement [38, 194, 257]. The expressions below are taken from the Halpin-Tsai description provided by Tucker [316]; for a more complete discussion, the reader is directed to an original review of the method by Halpin [313].

It is to be emphasized that the Halpin-Tsai equations are derived to represent composites consisting of *continuous, aligned fibers*. As nanocomposites do not contain continuous fibers (continuous from one end to another of a part), nor are the fibers typically aligned, the HT expressions should be used with extreme caution. Three of the five transversely isotropic properties of the composite are of the form

$$\frac{P}{P_m} = \frac{1 + \zeta\eta V_f}{1 - \eta V_f}, \quad \text{with } \eta = \frac{(P_f/P_m) - 1}{(P_f/P_m) + 1} \tag{23}$$

where P is one of the properties given in Table 5, P_f and P_m are the corresponding fiber and matrix property, V_f and V_m are the volume fractions, and ζ is a parameter that is a function of the matrix Poisson ratio and is dependent on the properties under consideration (see Table 5). For a continuous, aligned fiber composite, two additional properties are needed to provide the five independent material properties and are given below as

$$\begin{aligned} E_{11} &= V_f E_f + V_m E_m - \underline{4\left[\frac{\upsilon_f - \upsilon_m}{1/k_f - 1/k_m}\right]^2 \left(\frac{1}{k_{23}} - \frac{V_f}{k_f} - \frac{V_m}{k_m}\right)} \\ \upsilon_{12} &= V_f \upsilon_f + V_m \upsilon_m + \underline{\left[\frac{\upsilon_f - \upsilon_m}{1/k_f - 1/k_m}\right]\left(\frac{1}{k_{23}} - \frac{V_f}{k_f} - \frac{V_m}{k_m}\right)} \end{aligned} \tag{24}$$

where the underlined terms in Eq. (24) are often assumed negligible.

Table 5. Parameters for the Halpin-Tsai relations.

P	P_f	P_m	ζ	
k_{23}	k_f	k_m	$\frac{1 - \upsilon_m - 2\upsilon_m^2}{1 + \upsilon_m}$	Plane-strain bulk modulus, aligned fibers
G_{12}	G_f	G_m	$\frac{1 + \upsilon_m}{3 - \upsilon_m - 4\upsilon_m^2}$	Transverse shear modulus, aligned fibers
G_{23}	G_f	G_m	1	Longitudinal shear modulus, aligned fibers
K	K_f	K_m	$\frac{2(1 - \upsilon_m)}{1 + \upsilon_m}$	Bulk modulus, particulate (isotropic)
G	G_f	G_m	$\frac{7 - 5\upsilon_m}{8 - 10\upsilon_m}$	Shear modulus, particulate (isotropic)

The original Halpin-Tsai expressions for continuous aligned fibers given in Eq. (23) were then modified to account for finite fiber lengths by an empirical approach [313]. Here it was noted that as ζ goes to infinity Eq. (23) reduces to the standard rule of mixtures (Voigt) upper bound, whereas for ζ approaching zero the Reuss lower bound is obtained. Thus ζ was treated as a geometric parameter describing the shape of the fiber, and via comparison of the Halpin-Tsai relations with the results from a 2D finite element analysis, appropriate values of ζ selected for various properties for short fiber composites determined as shown in Table 6. In addition to the three properties listed in Table 6, standard Voigt approximations are used to determine the effective Poisson ratio ν_{12} and the transverse shear modulus G_{23} using the parameters for continuous fibers in Table 5, providing the five independent properties necessary to describe transversely isotropic behavior. For spherical reinforcement, the composite response is isotropic and is completely specified by the bulk and shear modulus expressions given in Eq. (23) using Table 5.

While the Halpin-Tsai approach may be appealing due to the simplicity of the expressions, we note that a recent comparison of the Halpin-Tsai and Mori-Tanaka methods found that Mori-Tanaka provided better predictions in comparison to a finite element-based solution even for aligned short-fiber composites [316]. As described in Appendix A.1, the Mori-Tanaka method can readily be derived for multiple inclusion phases and accounts for particle–particle interaction. In addition, the Mori-Tanaka method can directly account for random orientation of the inclusions via appropriate orientational averaging inherent within the derivation, whereas Halpin-Tsai relations for randomly oriented fiber composites are derived by averaging the Halpin-Tsai relations for the unidirectional composite properties to account for orientational distribution of inclusions as described below (see Fig. 68).

For example, for a two-dimensional random orientation of inclusions the in-plane composite tensile and shear moduli E_c and G_c based on the Halpin-Tsai approach can be written as [328]

$$E_c = \left[\frac{3}{8}\frac{1+2(l/d)\eta_L V_f}{1-\eta_L V_f} + \frac{5}{8}\frac{1+2\eta_T V_f}{1-\eta_T V_f}\right]E_m$$
$$G_c = \left[\frac{1}{8}\frac{1+2(l/d)\eta_L V_f}{1-\eta_L V_f} + \frac{1}{4}\frac{1+2\eta_T V_f}{1-\eta_T V_f}\right]E_m \tag{25}$$

where

$$\eta_L = \frac{(E_f/E_m)-1}{(E_f/E_m)+2(l/d)}, \quad \eta_T = \frac{(E_f/E_m)-1}{(E_f/E_m)+2} \tag{26}$$

Such an expression for the tensile modulus has been used previously to model nanotube-polymer composites [36]. However, such a procedure to account for a distribution of fiber orientations is a simplification that in actuality models a composite with randomly oriented small domains of aligned fibers as shown in Fig. 68(a). A more physically representative model of random orientation, such as that shown in Fig. 68(b), can be accounted for within a Mori-Tanaka approach [311] as shown in Appendix A.1. Further discussion of the differences between these methods of accounting for the orientation of the nanotubes can be found elsewhere [140].

It is for these reasons that the authors believe the Mori-Tanaka approach to be a more suitable micromechanics analysis for nanocomposite materials.

Table 6. Parameters for the Halpin-Tsai relations for aligned short fibers of length l and diameter d.

P	P	P_m	ζ	
E_{11}	E_c	E_m	$2(l/d)$	Longitudinal modulus
E_{22}	E_c	E_m	2	Transverse modulus
G_{12}	G_c	G_m	1	Longitudinal shear modulus

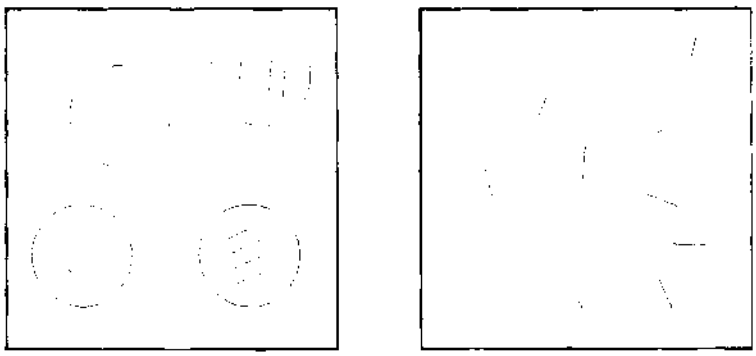


Figure 68. Models to account for random orientation of embedded inclusions. (a) Random orientation of unidirectional composite properties (Halpin-Tsai). (b) Physically appropriate model of random orientation of inclusions (Mori-Tanaka). Reprinted with permission from [140], E. T. Fisher, Ph.D. dissertation, Northwestern University, 2002. © 2002, E. T. Fisher.

We end this section by noting that in earlier work other semi-empirical expressions have been borrowed from the composites community and applied to nanotube-polymer systems [169]. For example, the Krenchel expression for the effective composite modulus [329], developed for short-fiber composites, can be written as

$$E = \eta_0 \eta_1 V_f E_f + (1 - V_f) E_m \tag{27}$$

where η_0 and η_1 are efficiency factors related to the fiber orientation and length, respectively. Using this expression, researchers have attempted to characterize the embedded nanotube orientation by fitting η_0 to experimental data [169, 330].

3.2.2. Finite Element Approaches

Given the power of finite element simulations to simulate structural response, the method is often used to examine the response of materials, explicitly considering the impact of various aspects of their microstructure, such as inclusions, voids or grain structure. With the emerging interest in nanocomposites, a number of researchers have begun to apply this approach to characterize nanocomposite properties. In the simplest cases, a single unit cell of nanotube and polymer is considered [289, 302, 303], whereas in other cases larger representative volume elements (RVEs) are considered in which many nanoparticles interact [301, 304]. To date, most analyses focus on linear elastic properties and simple interface conditions, ultimately providing predictions for the elastic stiffnesses of nanocomposites as a function of loading fraction, nanotube properties, and in some cases nanotube orientation. The results obtained to date in the finite element simulations are thus not much different than those obtained using the more sophisticated micromechanics methods such as the Mori-Tanaka technique discussed earlier. However, even in the simple elastic analyses, the finite element simulations allow one to examine local shear stresses near the nanoparticles and the effects of strain shielding in multi-inclusion cases, aspects that cannot be studied with homogenization methods. In the future, finite element simulations will enable more sophisticated examination of the role of the interphase, particle morphology, load transfer and even damage and failure mechanisms within the nanocomposite. In the discussion below, a few examples of current work using finite element approaches are summarized.

In Liu and Chen [302], the finite element method is used with traditional elasticity mechanics to determine the response of the smallest representative volume element (RVE): a unit cell consisting of one nanotube and its surrounding polymer. Using a simple micromechanics approach in which that unit cell is considered to be part of a periodic array, the effective properties of a composite constructed of an aligned arrangement of nanotubes is developed. Results for longitudinal and transverse elastic properties including poisson ratio are presented. The method has recently been extended to consider a unit cell of square cross section [303]. In both cases the nanotube and matrix phases are assumed to be perfect linear

elastic materials, homogeneous and isotropic, and perfectly bonded to one another. It is shown that due to the inherent nanotube stiffnesses, appreciable increases in nanocomposite stiffness can be obtained at relatively low volume fractions. As will be presented later in the hybrid modeling section, this simple approach can lead to significant errors with respect to experimental data.

Another finite element approach to study nanotube-polymer composites employed conventional three dimensional finite element models to model large RVEs of a polymer matrix containing 50 nanotubes with a range of aspect ratios [304] (see Fig. 69). Both aligned and randomly oriented configurations were considered at a low loading of 0.5 vol% to predict the effective stiffness and coefficient of thermal expansion for the nanocomposites. Model comparisons were made with traditional carbon fibers assigned appropriate anisotropic properties. Additionally, the CNTs were assumed to be solid cylinders, whereas their hollow nature may impact the results, especially for random orientations. Results for aligned nanotubes were compared directly to Halpin-Tsai (HT) and composite cylinders micromechanics methods, and it was shown that the HT approach has significant errors. As discussed earlier in this chapter, as the HT method is a semi-empirical method designed for short fibers and thus the inadequacy is not surprising. Improvements in properties are demonstrated for CNT composites over conventional fiber composites, however the improvement is based solely on the inherent properties of the nanotubes themselves. Thus, the critical issue of the alteration of the surrounding matrix polymer due to the interaction of the nanoparticle at the molecular chain level is not considered.

The finite element method has also been used in an interesting approach that infuses some aspects of molecular level modeling of the nanotube to an overall continuum finite element approach [289]. Here a molecular structural mechanics procedure is developed equating the energies for stretching, bending and torsion modes of chemical bonds to equivalent beam structures for the nanotube. The interfacial condition between the nanotube, created out of the equivalent structural beam elements, and the polymer matrix, created from traditional isoparametric elements, is taken to have either weak van der Waals interactions or strong covalent bonding. The former is simulated by connecting matrix and nanotube with truss elements that are governed by a force-displacement curve matching a Lennard-Jones potential. The latter is accounted for by direct connection of the matrix and nanotube nodes. Similar to other work [302, 303], the model has been used to study the response of a single nanotube surrounded by a cylindrical shell of matrix material. In this manner, the effective Young's modulus of a unidirectionally oriented nanocomposite is predicted and found to agree well with simple rule of mixture calculations. Analysis of the stress distributions reveal that stress

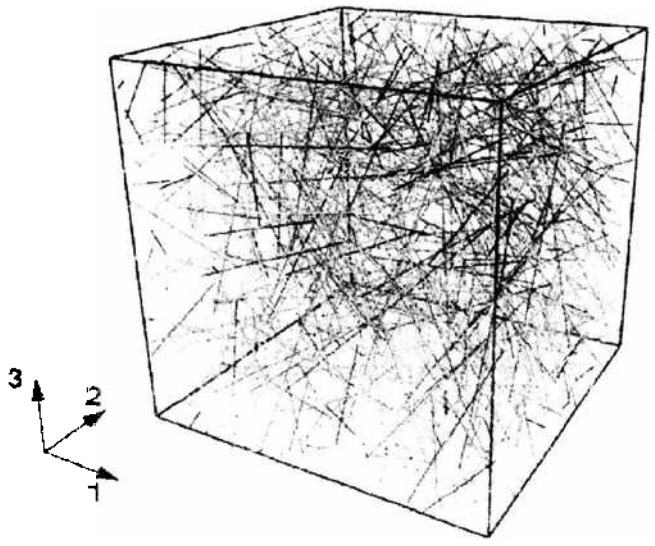


Figure 69. Sample unit cell of many nanotube inclusions, randomly oriented in a polymer matrix material. Reprinted with permission from [304], H. R. Lusti and A. A. Gusev, *Mod. Simul. Mater. Sci. Eng.* 12, S107 (2004), © 2004, IOP Publishing Ltd.

concentrations occur near the ends of the nanotubes and that the van der Waals bonding reduces the shear stresses in the matrix near the nanoparticle.

Another example of the finite element approach has modeled clay-based nanocomposites, taking into account the interphase region of polymer in close contact with the nanoparticle [305]. Here it was recognized that both structural and chemical properties of the material can be altered by the nanofiller. Structural enhancements are due to the filler and its orientation, while chemical enhancements are the altered characteristics of the interphase zone near the nanoparticles or intercalated into the galleries. Another unique feature considered in this work is the influence of the degree of overlap of the platelets in the composite. The interphase was explicitly considered, though modulus enhancement of the region was a modest factor of 2 based on previous MD results [262]. Interphase thickness was correlated to the radius of gyration (R_g) of the polymers, but was taken such that the interphase volume fraction was only six times the filler volume fraction. The model is developed initially for unidirectionally oriented nanoparticles, though the importance of off-axis orientations is recognized and addressed by a second level of modeling. In this additional consideration, upper and lower bounds for the effect of orientation are evaluated by considering a series of cells with a Gaussian orientation distribution that are mechanically connected in either a series or parallel fashion. Only static elastic modulus of the overall nanocomposite is considered, and the modeling illustrates that clay volume fraction, degree of exfoliation, and orientation were the most critical parameters. Influence of the interphase was also found to be important, but at the modest quantity and altered properties for the interphase used here, its role was secondary in determining overall response.

A finite element approach has also been employed by Boyce and co-workers to examine the local properties of nanoclay polymers [301]. This is an extensive work, which also initially examines the comparison between traditional micromechanics approaches and the numerical approach for stiffness of conventional particle composites as a function of volume fraction, aspect ratio, and particle/matrix stiffness ratio (see Fig. 70). The advantages of the numerical approach are evident at that point, as examination of the local stress and

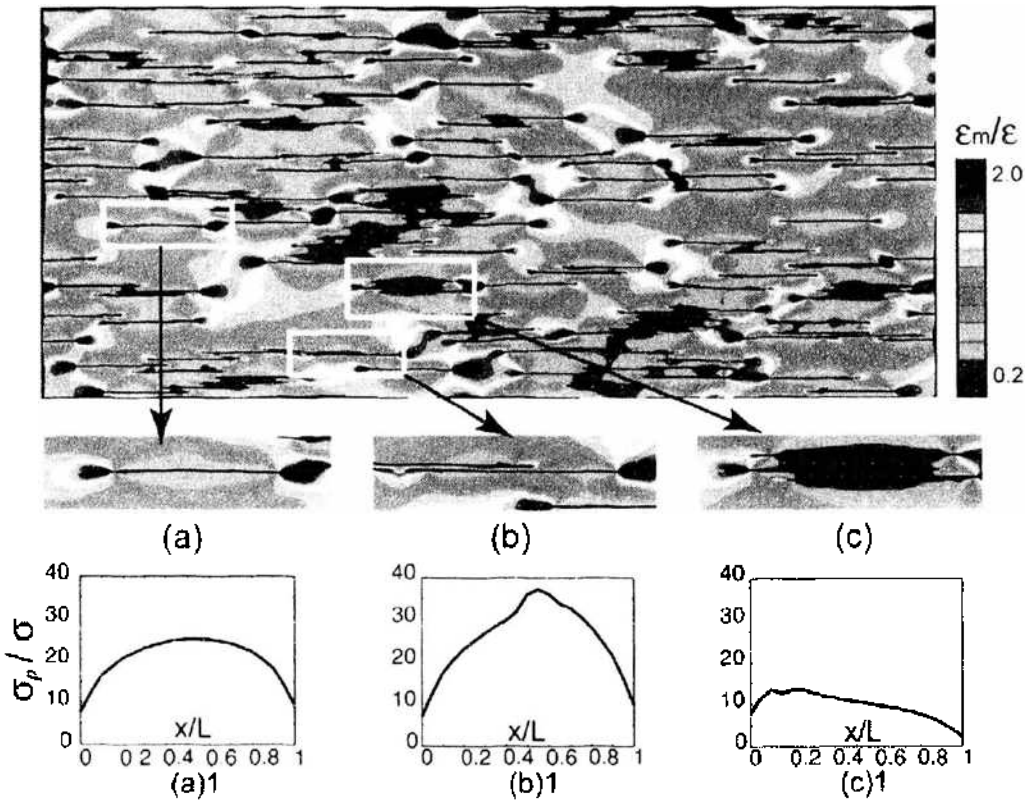


Figure 70. Effect of strain shielding caused by overlapping platelet reinforcements in a polymer composite: (a) isolated, (b) partly overlapped, and (c) completely overlapped particles and their associated stress distributions. Reprinted with permission from [301]. N. Sheng et al., *Polymer* 45, 487 (2004). © 2004, Elsevier.

strain fields provides insight into the strain shielding of high aspect ratio particles that leads to a nonlinear reinforcing effect with increasing aspect ratio. The clay nanocomposites are considered to be composed of multilayer intercalated clay particles that are aligned in a polymer matrix. The presence of fully exfoliated particles interdispersed with the intercalated particles as studied by Luo and Daniel (see Fig. 63). A representative unit cell for a random microstructure of aligned clay particles is developed in finite elements to study the local and global response of the nanocomposite. Several cases are considered to model the clay nanoplatelets, ultimately resulting in an elasticity approach in which the fully anisotropic moduli of parallel multilayered plates are used. The properties of the silicate are obtained from molecular dynamic simulations and the properties for the intercalated polymer are similar to the matrix for tensile modulus, but weaker in shear. These anisotropic properties for the intercalated clay particle are then used in the RVE to determine nanocomposite response and compared to experimental data for nylon-clay systems. The effect of semi-crystalline and transcrystalline matrix material was also thoroughly investigated. Their work demonstrates that models must capture the intrinsically hierarchical microstructure of the nanoclay reinforcements to adequately represent the nanocomposite response. They note that to predict the out-of-plane response for aligned composites, replacing the discrete stack with an effective pseudoparticle is not accurate due to the low shear modulus of interlayer galleries.

There is still an enormous potential for finite element simulations to provide understanding and guidance to nanocomposite properties and research. While a few of the papers to date have considered some aspects of interfacial interaction [289, 301, 305], none to date have considered the critical effect of the extended interphase region of altered polymer mobility and altered polymer properties surrounding the nanoparticle. Nearly all consider only uniaxially aligned nanoparticles, which is not representative of the configuration of current nanocomposites. For platelet-based nanocomposites, some consideration has been given to account for the intercalated polymer layer, but *in situ* morphologies including nanotube or nanoplatelet curvature have not been explicitly addressed. Additionally, at present all the finite element studies have examined only linear elastic response and not yet addressed viscoelastic, local plasticity, or fracture and damage. As these issues are critical for further progress in nanocomposites, they are ripe for new modeling simulations to assist in intelligent design of these materials. In addition, the coupling of finite elements to molecular and micromechanics approaches is an area of much-needed development. The hybrid model mentioned next is one step in that direction, accounting for interacting particles, interphase, viscoelasticity and through use of finite element unit cells and a micromechanics approach.

3.3. Hybrid Modeling Methods

While traditional micromechanics techniques have been used extensively to model the mechanical behavior of polymer nanocomposites as discussed in Section 3.2.1, there are a number of differences between polymer nanocomposites and their microscale counterparts that are not accounted for in traditional micromechanics approaches. Recognizing this limitation, a number of efforts are underway to adapt these techniques to make them more suitable for nanocomposites studies. In this section we will define hybrid techniques as those that couple traditional micromechanics with other approaches to more accurately address issues specific for polymer nanocomposites. We mention briefly first a method by which the equivalent continuum model is coupled to micromechanics, and then the remainder of the section is devoted to a hybrid finite element-micromechanics approach.

In Odegard et al. [286], molecular dynamics and the equivalent continuum model [268] (see Section 3.1.3) are used to determine the transversely isotropic properties of an effective fiber that is then used within a Mori-Tanaka model to describe bulk nanocomposite behavior. Significant differences are predicted for random versus aligned effective fibers and for different nanotube lengths. Specifically, it is shown that increasing the aspect ratio of the nanotube results in increased moduli, though very small gains are achieved after about 200 nm in length. While the use of the effective fiber approach here accounts in some sense for an interphase region around the nanoparticle, additional work can be done to more fully explore the role of the interface and interphase by this method. Due to the significant

computational time-savings over MD simulations, this type of hybrid method is indicative of the type of multiscale simulations necessary to advance our understanding and design of nanocomposites.

In a series of papers by the authors and co-workers [298–300, 331], a hybrid micromechanics-finite element approach was developed to adapt micromechanics analyses to account for inherent nanoparticle curvature and the extensive interphase region. We will first discuss two methods developed to account for nanotube waviness in the elastic regime and then mention more recent work on incorporating the interphase and full viscoelastic representation. As shown in the SEM images in Fig. 71, nanotubes in polymer composites appear to be wavy (not straight), a feature not typically associated with traditional fiber reinforced composites. Note that this waviness is inherently distinct from the uniform and controlled waviness of the yarns in traditional textile composites, as the curvature distributions are random and the volume fraction of the fibers is quite low in nanocomposites. Note that wavy cylindrical inclusions cannot be accounted for directly in a traditional micromechanics approach, as the Eshelby tensor (see Eq. (16)) is not available for such non-ellipsoidal inclusions. In order to better understand the influence of the nanotube curvature distributions on overall nanocomposite response, hybrid finite element-micromechanics models were developed. As discussed below, using experimental data, we demonstrate that nanotube waviness can limit the modulus enhancement of the nanotube-reinforced polymer, suggesting that this waviness may be one reason why the modulus enhancements for NRP systems measured to date are typically less than anticipated using standard micromechanics models.

Two related methods have been developed to utilize the solution of a finite element analysis to incorporate the effects of this waviness into subsequent micromechanics models [298–300]. Each method uses the finite element unit cell shown in Fig. 72 with appropriate boundary conditions. The waviness of the nanotube is assumed to be sinusoidal in shape with an embedded geometry of the form $y = a\cos(2\pi z/\lambda)$, where λ is the sinusoidal wavelength and z is the fiber axial direction (see Fig. 72). The sinusoidal nanotube is then embedded within an infinite matrix material such that a convergent solution for the dilute approximation case is obtained. (A proof that this infinite matrix condition is satisfied is presented elsewhere [299]; typically, the finite element cells shown in Fig. 72 contained greater than 99.5% matrix.) The response of such a finite element cell, assuming that the Poisson ratios of the inclusion and matrix phases are equal, is dependent on three dimensionless ratios: the waviness ratio a/λ, the wavelength ratio λ/d, and the ratio of the phase moduli $E_{rat} = E_{NT}/E_{mat}$. Further details of the finite element modeling are given elsewhere [300].

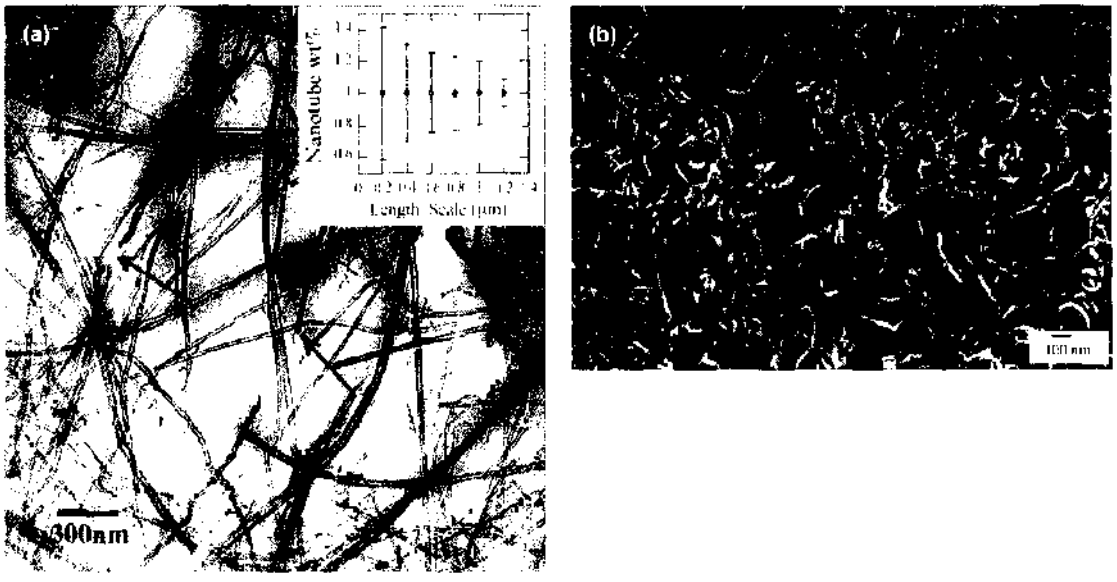


Figure 71. Micrographs showing the waviness of nanotubes embedded in polymers. (a) TEM image of 1 wt% MWNTs in polystyrene. Reprinted with permission from [36], D. Qian et al., *Appl. Phys. Lett.* 76, 2868 (2000). © 2000, American Institute of Physics. (b) SEM image of 50 wt% MWNTs in poly(vinyl alcohol). Reprinted with permission from [169], M. S. P. Shaffer and A. H. Windle, *Adv. Mater.* 11, 937 (1999). © 1999, Wiley-VCH.

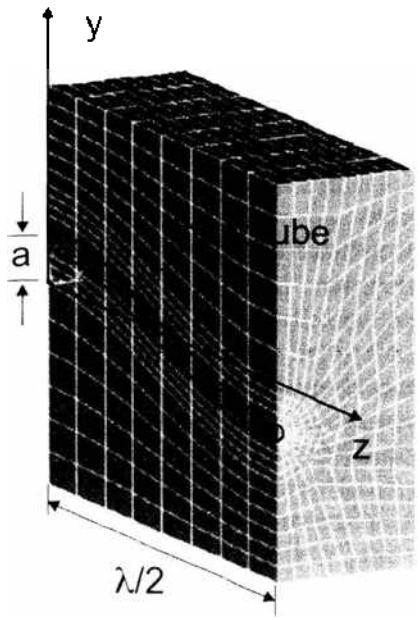


Figure 72. Finite element cell model of an embedded wavy nanotube. For the model shown, $w = a/\lambda = 0.1$ and $\lambda/d = 35$. Reprinted with permission from [299], F. T. Fisher et al., *Compos. Sci. Technol.* 63, 1689 (2003). © 2003, Elsevier.

Such a model implicitly introduces two simplifications into the analysis: the treatment of the nanotube inclusion as a solid cylinder neglects the hollow nature of the nanotubes, and modeling the nanotube as a continuum disregards the specific form of the nanotube (SWNT, MWNT, or bundle) and any possible relative motion between individual shells or tubes in a MWNT and an NT bundle, respectively. In each case the reduction in modulus due to waviness estimated from such an analysis can be viewed as conservative, in that accounting for the hollow nature of the NTs or modeling relative sliding of the tubes or shells would further reduce the effective stiffness of a wavy nanotube. In addition, in this work the individual phase materials were modeled as linear elastic and isotropic, and perfect bonding between the phases was assumed.

In the first model, which we refer to as the effective reinforcing modulus (ERM) model, the unit cell shown in Fig. 72 was constrained in the y direction at a single point to prevent free body translation, and an infinitesimally small axial displacement, Δ, prescribed to all nodes on the plane $z = \lambda/2$. The effective modulus of the finite element cell is defined as

$$E_{\text{cell}}^{\text{FEA}} = \frac{F_{\text{tot}}\lambda}{2A\Delta} \tag{28}$$

where F_{tot} is the sum of all nodal resultant forces on the displaced plane and A is the cross-sectional area of the cell. To extract the effective reinforcing moduli of the *embedded* wavy inclusion (as it exists in the matrix) from Eq. (28), a parallel model of the effective cell response, independent of the previous analysis, was assumed of the form

$$E_{\text{cell}}^{\text{parallel}} = c_{\text{NT}} E_{\text{ERM}} + (1 - c_{\text{NT}}) E_{\text{matrix}} \tag{29}$$

where c_{NT} is the nanotube volume fraction within the finite element cell and E_{matrix} is the matrix modulus. From Eqs. (28) and (29), the E_{ERM} of the embedded inclusion can be calculated as

$$E_{\text{ERM}} = \frac{E_{\text{cell}}^{\text{FEA}} - (1 - c_{\text{NT}}) E_{\text{matrix}}}{c_{\text{NT}}} \tag{30}$$

Thus E_{ERM} represents the modulus of a straight inclusion that, under identical loading conditions, would yield the same effective finite element cell response as that obtained with the wavy inclusion.

Figure 73 shows the E_{ERM} as a function of waviness for several different values of E_{ratio} and a wavelength ratio of 100. For all simulations a matrix modulus of 1 GPa was used. As expected, for zero waviness we obtain the straight nanotube results $E_{ERM} = E_{NT}$. We note that E_{ERM} is strongly dependent on the waviness and quickly decreases with increasing nanotube curvature. This drop in E_{ERM} is less pronounced for smaller E_{ratio} values because the mechanical constraint of the surrounding matrix material in this case is more significant.

During the development of the E_{ERM} model, an alternative solution was conceived to incorporate inclusion waviness into micromechanical predictions of effective stiffness. In this alternative model, which we call the Numerical Strain Concentration Tensor (NSCT) method, the complete dilute strain concentration tensor A_r^{dil} (see Section 3.2.1.1) for a wavy cylindrical inclusion is found via the solution of six separate finite element models with appropriate boundary conditions, a procedure that is described in detail in the literature [300]. Once the dilute strain concentration tensor A_r^{dil} has been determined, it can then be used directly in the Mori-Tanaka solution (see Eq. [20]) to predict the effective modulus of the nanocomposite. The major difference between these two models is that the ERM model solves a single finite element model, analogous to a numerical tensile test, and then treats the wavy nanotube as an isotropic ellipsoidal inclusion with a reduced modulus E_{ERM}. For the NSCT model, the solution to the six independent finite element models (each with the geometry of the embedded wavy nanotube) yields an orthogonal effective response such that the isotropic simplification used in the ERM model is unnecessary. A comparison of the unidirectional (in the direction parallel to the long axis of the nanotube) effective response of the unit cell by each of the two models as a function of nanotubes waviness is illustrated in Fig. 74.

For both the ERM and NSCT models, the distribution of nanotube waviness within a nanocomposite can be accounted for by using a multiphase micromechanics approach as shown in Fig. 75, where a waviness distribution is chosen based on qualitative examination of SEM micrographs, resulting in the percentage of nanotubes exhibiting different curvature magnitudes in the composite. Each curvature magnitude is then treated as a separate phase in an N-phase Mori-Tanaka analysis to account for the coexistence of the various nanotube curvatures in the composite. For the ERM model, given the effective reinforcing modulus and Poisson ratios of the phases, the Mori-Tanaka method described in Section 3.2.1.1 was used where the nanotube is treated as a continuum and the standard Eshelby tensor for a cylindrical inclusion used for each NT phase (as the waviness is factored in via the reduction in modulus via the ERM). For the NSCT model, as the dilute strain concentration tensor A_r^{dil} is found explicitly based on the finite element model, it is substituted directly into Eq. (20) for each value of nanotube waviness.

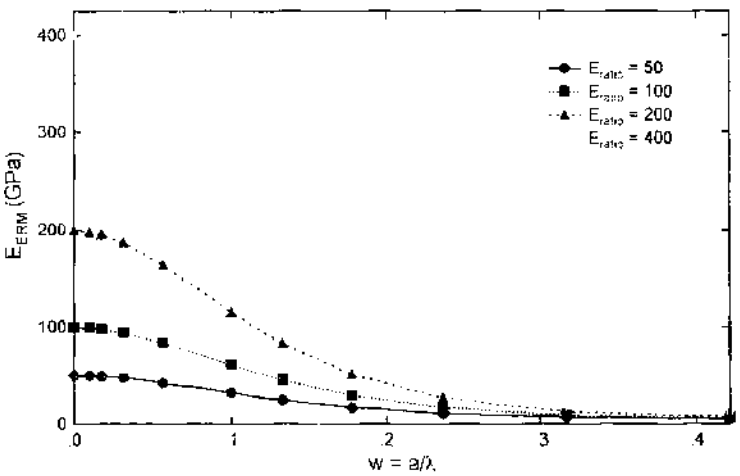


Figure 73. E_{ERM} as a function of nanotube waviness ratio (a/λ) for different ratios of phase moduli with wavelength ratio $\lambda/d = 100$. $E_{matrix} = 1$ GPa. Reprinted with permission from [140]. F. T. Fisher, Ph.D. dissertation, Northwestern University, 2002. © 2002, F. T. Fisher.

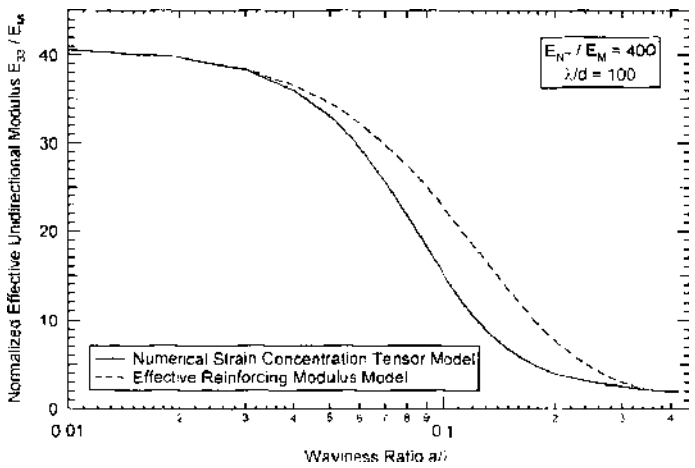


Figure 74. Effective composite modulus E_{33} (in the z direction of Fig. 72, parallel to the NT axis) with increasing waviness ratio (a/λ) for the ERM and NSCT models. Reprinted with permission from [300]. R. D. Bradshaw et al., *Compos. Sci. Technol.* 63, 1705 (2003). © 2003, Elsevier.

Given such an approach, the effective nanocomposite modulus can be determined if the distribution of nanotube waviness shown in Fig. 75 is known. To simply illustrate the procedure, two waviness distributions were assumed (see Ref. [299]), where distributions 1 and 2 represent minimal and more significant level of nanotube waviness within the polymer. The ERM model predictions obtained using these two assumed waviness distributions, for various orientations of the nanotubes, are shown in Fig. 76 and compared with experimental data for MWNTs in polystyrene [139] and epoxy [39], respectively.

What is most striking about the results presented in Fig. 76 are the large discrepancies between the Mori-Tanaka predictions assuming *straight* nanotubes and the experimentally measured moduli. While the experimental modulus has been enhanced with the addition of the NTs, the realized improvements in modulus are significantly less than what the micromechanics predictions with straight nanotubes would indicate. Integrating moderate nanotube waviness into the effective moduli predictions is shown to significantly decrease the moduli predictions, suggesting that NT waviness may be one factor limiting the modulus enhancement of nanotube-reinforced polymers.

The ERM and NSCT predictions for a two-phase nanocomposite with a 3D random orientation of wavy nanotubes are shown in Fig. 77. Here isotropic constituent phases, a 10% NT volume fraction (where all NTs have the same waviness), and $E_{\text{ratio}} = 400$ were assumed. We see that for shorter wavelength ratios ($\lambda/d = 10$), the difference between the models is minimal until very large values of the waviness ratio a/λ are considered, at which point the NSCT model predicts a stiffer effective response as it more appropriately accounts for stiffening in the direction perpendicular to the long axis of the nanotube (the y-direction in Fig. 72)

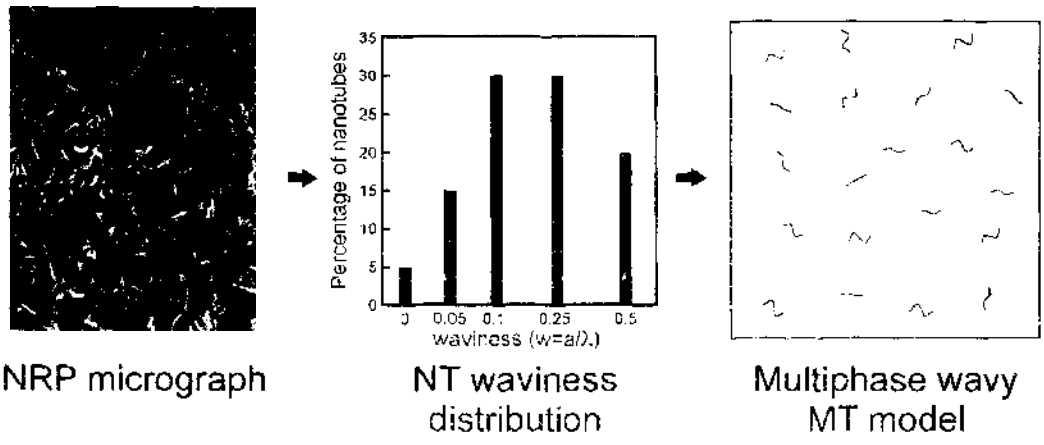


Figure 75. Model of a nanotube-reinforced polymer using a multiphase composite analysis to account for a given waviness distribution function compatible with both hybrid micromechanics-finite element methods described. Reprinted with permission from [299]. F. T. Fisher et al., *Compos. Sci. Technol.* 63, 1689 (2003). © 2003, Elsevier.

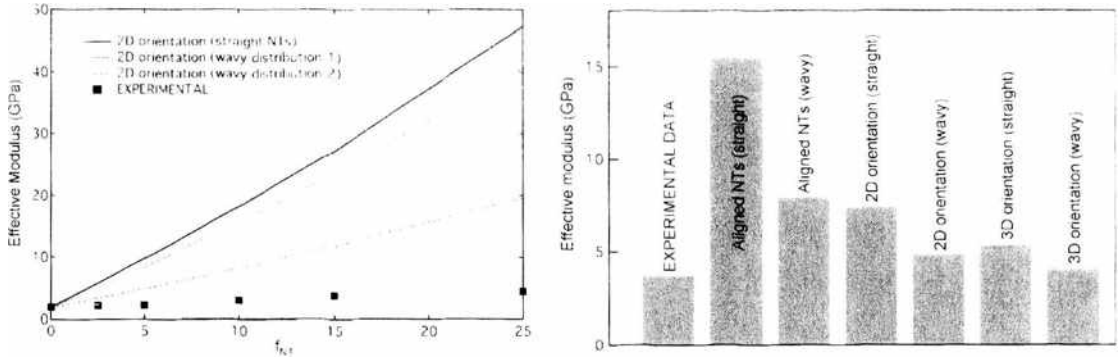


Figure 76. (left) Experimental data for MWNTs in polystyrene [139] and (right) 5 wt% MWNTs in epoxy [39]. Also shown are micromechanical predictions of the NRP effective modulus assuming different random orientations of straight and wavy (ERM method) nanotubes. Reprinted with permission from [299], E. T. Fisher et al., *Compos. Sci. Technol.* 63, 1689 (2003), © 2003, Elsevier.

within the 3D randomly oriented composite. For larger wavelength ratios ($\lambda/d = 100$), the difference between the two models is more significant, although it should be noted that this difference is exaggerated here given the large NT volume fraction modeled. We also note that Fig. 77 assumes that all of the nanotubes have identical values of a/λ. As discussed previously, there is likely to be a distribution of NT waviness within the material, such that only a fraction of the nanotubes would be characterized by a/λ and λ/d parameters for which the difference between the ERM and NSCT results is significant.

Thus motivated by micrographs showing that nanotubes embedded within polymers often exhibit significant curvature, a hybrid model was developed that incorporates this curvature into traditional micromechanical methods via a multiphase composite approach. Using material properties representative of nanotube-reinforced composites, it was shown that nanotube waviness can reduce the predicted effective moduli of these materials, and may be one reason why the modulus enhancement of nanotube-reinforced polymers, while significant, is somewhat less than predicted using standard micromechanical techniques. For samples with moderate waviness, the differences in moduli predictions given by the two models may be minimal and likely masked by other factors, such as a poor NT-polymer interface, inadequate NT dispersion, and nanotube degradation due to the nanocomposite processing; in this case the ERM model may be preferable due to its simplicity. For cases where significant nanotube waviness is expected or has been observed, the NSCT model is preferred because it more accurately models the full impact of the wavy nanotube on the effective moduli of the nanocomposite. While for some applications and property targets (such as impact resistance, energy absorption and fracture toughness) nanotube waviness may be beneficial, the

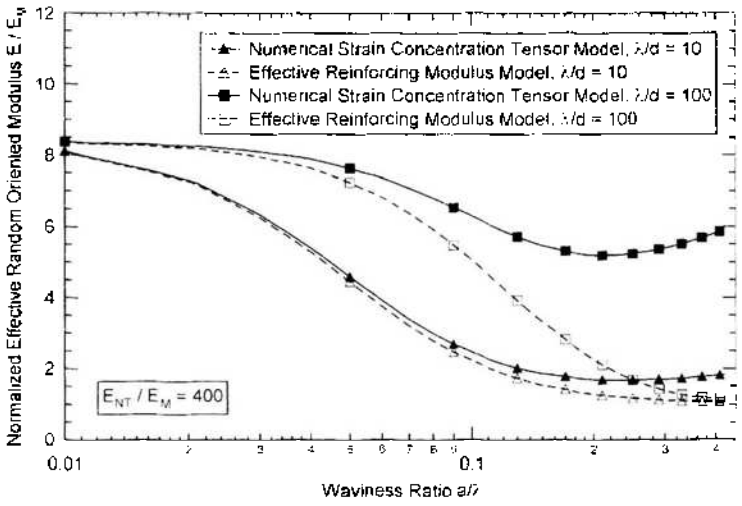


Figure 77. Young's modulus predictions for a nanotube-reinforced polymer with 3D randomly oriented wavy NTs using the ERM and the NSCT models for $E_{NT}/E_M = 400$. Reprinted with permission from [300], R. D. Bradshaw et al., *Compos. Sci. Technol.* 63, 1705 (2003), © 2003, Elsevier.

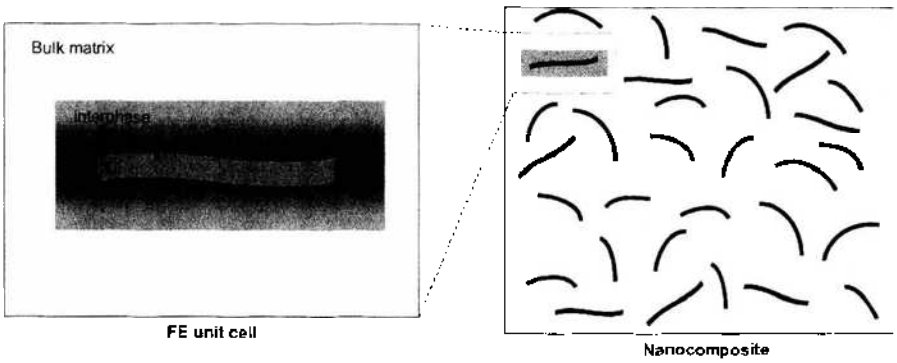


Figure 78. Hybrid FE-micromechanics model. Left unit cell enables strain concentration tensor for non-ellipsoidal inclusion geometries and nonbulk interphase properties to be calculated.

deleterious effect on modulus enhancement provided by the nanotube inclusions must be considered.

While such results demonstrate how waviness of the embedded nanotubes may reduce the effective reinforcement provided by the nanotubes within the nanocomposite, at the moment it is impossible to differentiate the impact of nanotube waviness from competing reinforcement-limiting mechanisms when examining experimental data. Complementary "bottom-up" modeling techniques such as those described in Section 3.1 will in the future be useful to clarify these effects. While the work reported here is an application of a micromechanics method to a nanostructured material, the integration of atomic scale modeling could readily be adapted into such an analysis, including atomic scale information such as the transversely isotropic nature of the NT mechanical properties and inter-layer (MWNTs) and intertube (NT bundles) sliding. In the future a fusion of true nanoscale and microscale modeling will provide even more insight, and quantitatively accurate predictions, of this material behavior.

In a more recent application of the hybrid micromechanics-numerical approach, both viscoelasticity and the effect of the interphase have been considered (see Fig. 78) [331]. In this work, a finite element unit cell was constructed containing a nanotube or nanoplatelet that was surrounded by a discrete interphase domain of polymer which was then embedded in the bulk polymer matrix material. The properties of the interphase is assumed to be related to the properties of the matrix material by a time shift to correspond to altered mobility of the polymer in the vicinity of the nanoparticle. After determination of the dilute strain concentr ·ion matrices (via the NSCT approach) for the inclusion and the interphase separately, the Mori-Tanaka method is employed to provide overall viscoelastic property predictions. Although the standard three phase Mori-Tanaka method can account for an interphase domain as described earlier and be performed in the viscoelastic domain, the annular morphology of the interphase surrounding the inclusion is lost in that approach, adding a significant level of approximation. While methods for coated inclusions [332] could be employed to obtain effective moduli, they cannot also consider the effects of nanoparticle curvature. The hybrid finite element-micromechanics approach is also amenable to approaches to incorporate gradient interphases with spatially varying properties. To date the hybrid method has been implemented for the case of aligned inclusions for transverse and longitudinal properties and results compare qualitatively well with experimental data. In particular the influence of the interphase is observed in a broadening of the loss peak and it was illustrated that curvature of nanoparticles can reduce the anisotropy of aligned nanocomposites [331].

4. CURRENT ISSUES IN NANOMECHANICS OF NANOREINFORCED POLYMERS

While the previous section has reviewed the contributions of a number of approaches to nanomechanical modeling for nanoreinforced polymers, it is clear that there are several fundamental issues that need to be addressed to enable more accurate modeling and deeper

understanding of nanocomposites. In this section, we will discuss these major issues, many of which are under intense investigation by research teams across the world. While the widespread interest and efforts invested in nanomechanics spur rapid progress, at the same time many ongoing investigations also unveil new phenomena that must be further explored. Thus the topics presented here represent the cutting edge of nanomechanics research in polymer nanocomposites today. While some of these issues will require a number of years of continued study, others will either evolve or be eclipsed by new insight over time.

One issue with continued uncertainty is the *inherent properties of the nanoparticles* themselves and how to best represent these at either the molecular or continuum level under various conditions. On the molecular side, while the bonding potentials used for the carbon-carbon bonds in the nanotubes are considered accurate at present, the nature of how to model, for instance, the reinforcement effect provided by the inner shells of MWNTs and the interactions between layers in clay and graphite nanoplatelets are less mature. In addition, for the platelet forms, the intercalants used in processing result in interlayer species where the bonding potentials can be less standardized. For all nanoparticle forms, the role of defects needs to be elucidated [53, 333–335]. The type and density of defects needs to be quantified as a function of processing methods and the impact on inherent nanoparticle properties described. In addition, as chemical functionalization has been demonstrated to be a successful strategy to improve load transfer between particle and matrix, the influence of functionalization on nanoparticle properties needs further investigation. The magnitude of the changes in inherent nanoparticle properties with different types and quantities of functionalization is not currently well understood [53].

Ideally, these issues can be addressed at the molecular level with MD and MC simulations that can provide information on continuum properties to be used in higher length scale simulations. New nanomechanical experiments could also provide insight into nanoparticle properties under these different conditions. Currently, many micromechanics and finite element approaches do not account for the nanotube geometry explicitly, ignoring the hollow nature of nanotubes and the intricacies of the inner shells in MWNTs. Additionally, ideal isotropic properties are often used for the nanofiller given the uncertainty in the changes caused by defects or functionalization and the lack of knowledge of out-of-plane properties. For example, new work has sought to develop relationships for the transversely isotropic properties that would be more appropriate to represent elastic properties of SWNTs, MWNTs, and SWNT bundles [65–68].

Beyond the inherent nanoparticle properties, much more work needs to be done on the *nature of the interfacial and interphase regions* in polymer nanocomposites. The molecular modeling to date described in Section 3.1 has focused on the influence of idealized nanoparticles on the dynamics of relatively small numbers of short polymer chains. While this information has provided great insight, the extent of the changes in polymer properties both spatially and quantitatively is still unclear. Here again, the role of functionalization needs further examination. The optimal density of functional groups and the ideal length and properties of the functional group all remain to be explored. To improve modeling at the molecular level, additional enhancements in atomistic modeling and multiscale methods need to be addressed as discussed below. For continuum approaches to modeling, information either from atomistic modeling or from new experimental data on the interphase region could be incorporated. Currently, experimental efforts are underway to probe the nature of the interfacial bonding through nano-pullout experiments. In addition, methods akin to those used in the polymer thin film community to study changes in polymer dynamics near substrates and free surfaces could be used to probe changes in the glass transition and polymer mobility near nanoparticles [10, 11]. Results of these types of experiments and molecular modeling should yield information on the property gradient in the polymer near nanoparticles that could be used in continuum modeling. The manifestation of these property changes in the polymer according to various functionalization strategies will be critical.

Another area of emerging interest is that of *strength and toughness modeling* for polymer nanocomposites. To date, the modeling in this area has been quite limited. Several MD papers mentioned in Section 3.1.1 address some aspects of nanotube pullout from a polymer. However, a number of simplifying assumptions have been made that need to be addressed.

On the continuum side, recent work has modified the classic shear lag model to account for the hollow nature of a nanotube and obtained expressions for the interfacial shear strength as a function of nanotube diameter [336]. The results are in good agreement with experimental pullout data from the same group. In different work, a local density approximation model and classical elastic shell theory are used to describe nanotube pullout for multiwall nanotubes [337], and the distribution of the interfacial shear stress along the length of the nanotube obtained. In another approach [290], the molecular structural mechanics method of Li and Chou [289] (see Section 3.2.2) is used to create a continuum fiber equivalent to a nanotube. This fiber is then subjected to a classical shear lag analysis to determine the interfacial shear stress and normal stress in the fiber. While these first efforts to address pullout are notable, a vast amount of work remains to be done to understand the origins of toughness and strength in nanocomposites. Issues to be addressed include the roles of nanoparticle morphology (e.g., curved nanoparticles, interaction and entanglement with other nanoparticles), the properties of the interphase zone surrounding the nanoparticle, pullout for platelet forms, and extending information from force-displacement response on a single nanoparticle to address fracture toughness for a multi-nanoparticle composite.

The limitations of atomistic simulation and continuum mechanics, along with practical needs arising from the heterogeneous nature of many materials, have motivated research on *multiscale simulations* that bridge atomistic simulations and continuum modeling. Multiscale modeling is a critical area for current and future study in polymer nanocomposites. To date, the multiscale approaches that have been applied to nanocomposites still reside ultimately on the continuum side and are merely informed by some smaller length scale modeling. This is true for the equivalent continuum approaches in which molecular models are used to develop the energies from which continuum elements are derived for the nanostructures of interest. It is also true for the hybrid micromechanic-finite element approaches, which use some information from either experiments or modeling to define interphase or nanoparticle properties, but the entire simulation takes place on the continuum realm. In one sense, these approaches may be considered as hierarchical multiscale modeling.

In contrast to these efforts, the leading edge of multiscale modeling is in concurrent modeling approaches which link modeling from quantum or atomistic levels to the continuum level simultaneously [338–340]. Typically, the smaller length scale modeling is performed only in local areas of interest or activity, and these are linked directly to a surrounding continuum modeling of the entire structure. It has been found in these methods that a critical issue is the nature of the overlap between atomistic and continuum regions. In particular, sharp boundaries between the domains lead to spurious nonphysical results, and thus an overlap region which can accommodate the transition is essential.

These approaches have been developed for study of fracture events in materials, where atomistic modeling is present only at the crack tips. Concurrent modeling has also been used to study the deformation of nanotubes, where areas of locally high deformation or defect sites are modeled with atomistic algorithms, while areas away from these sites use traditional finite elements [341, 342]. This scheme has been used with MD only in regions of flaws to determine the effect of missing atoms on NT strength. Further work has employed molecular mechanics, molecular dynamics and interlayer potentials to study the deformation of single- and multiwall NTs [341]. The results from these studies suggest that intertube corrugation and surface tension are two factors that contribute to load transfer in NT-polymer composites.

While multiscale modeling is an emerging area of research and has been applied to nanotubes, this approach has not yet been extended to polymer nanocomposites. As mentioned in Section 3.1.1, molecular modeling of polymers alone often uses coarse graining approaches, which are a form of multiscale modeling in that monomer groups are lumped together in order to create longer effective chains. Ideally, future research will focus on connecting these types of coarse graining approaches for polymers near nanoparticles to traditional finite elements to enable study of a larger scale polymer nanocomposite but retaining atomistic detail in the vicinity of localized deformation events such as microcracking, nanoparticle buckling, or pullout. Improved multiscale simulations will also allow for consideration of larger numbers of effectively longer polymer chains at the molecular level, interacting with realistically sized nanoinclusions and will allow for bond breaking and reformation. Such simulations

will provide more information on the polymer dynamics and properties near nanoparticles with different surface chemistries and the impact of these interaction zones around many nanoparticles on macroscopic material response.

An example of the role of the different length scales in nanocomposites is shown in Fig. 79. Here a hierarchical strategy is shown where the MD simulations are performed for a single nanoinclusion and its surrounding polymer, providing input on the characteristics and extent of the interphase zone in a micromechanics model. The micromechanics model then accounts for the network of interacting nanostructures and feeds into a fracture mechanics-type model to evaluate strength and toughness of a bulk nanocomposite. In contrast, in a concurrent strategy, the simulation for fracture toughness would consist of simultaneous calculations of fracture events at different length scales. While the bulk nanocomposite would be a continuum mesh, increasing level of detail would replace the continuum analysis near the crack tip and nanotube pullout regions.

An area of modeling not covered in detail in this chapter and also not yet addressed substantially in the nanocomposite literature is *modeling of multifunctional properties* of nanocomposites. While the improvements in thermo-mechanical properties of nanocomposites show great promise, the real advantage to these systems is in being able to take advantage of improvements in thermal conductivity, electrical conductivity, and diffusion barrier properties in addition to thermomechanical properties. A few modeling efforts have suggested that despite the intrinsic thermal conductivity and aspect ratio of carbon nanotubes, thermal conductivity enhancement for nanotube-polymer systems is limited by the large thermal interface resistance in these systems [215]. Molecular dynamics simulations have suggested that this thermal interface resistance can be reduced by chemical functionalization and subsequent bonding of the nanotubes to the polymer, although the functionalization leads to a decrease in the NT conductivity [216, 217].

In a different approach, the percolation threshold of polymer nanocomposites has been studied using a Monte Carlo statistical method [218]. MC theory was used to explain the low percolation threshold of NTs in a low conductivity medium by showing that the critical volume fraction is inversely proportional to the aspect ratio of the filler. Thus for nanoparticles with aspect ratios on the order of 1000 or more, the percolation threshold can occur at volume fractions as low as 0.01%. Both thermal and electrical conductivity were examined and compared to experimental data. It was found that in polymer nanocomposites, the resistance of the contacts of the percolating network can be quite large and can significantly reduce the expected nanocomposite conductivity. These results indicate a need for functionalization strategies to reduce this contact resistance in nanocomposites [218]. These limited modeling efforts to date indicate the potential to tune nanocomposite composition and morphology to achieve better multifunctional properties. When performed hand-in-hand with thermomechanical modeling strategies, it will be possible to optimize nanocomposite design to achieve the balance of properties desired for a given application.

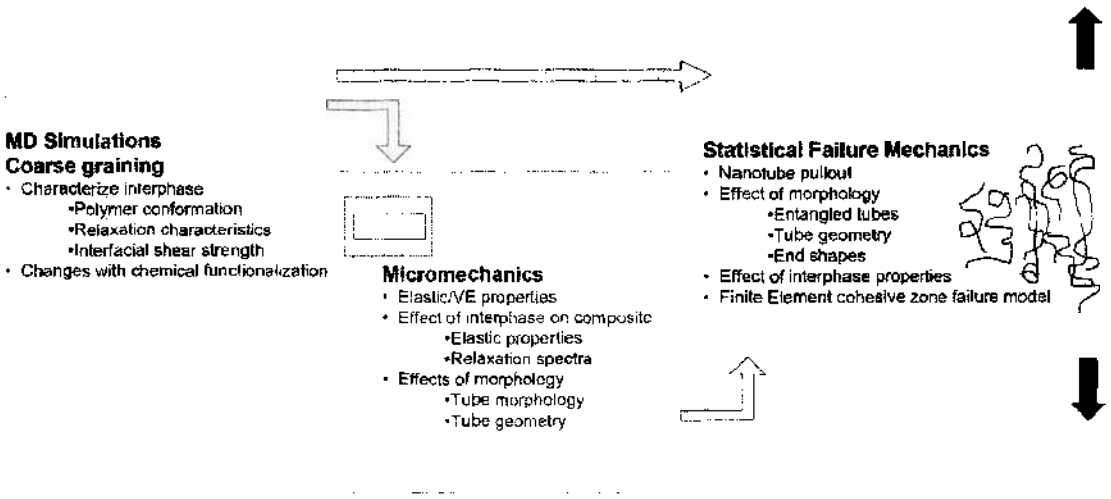


Figure 79. Overview of a multiscale modeling strategy for polymer nanocomposites.

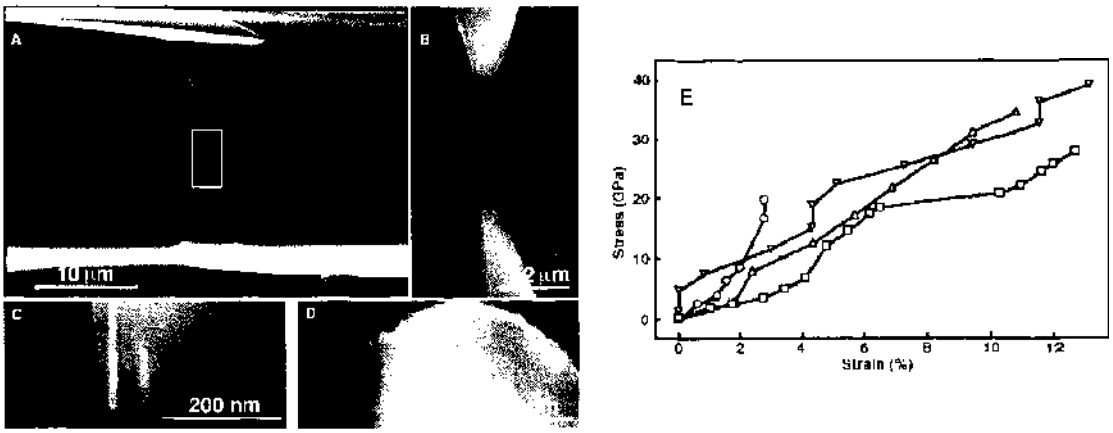


Figure 80. Tensile loading of an individual MWNT. (A,B) SEM images of two AFM tips holding a MWNT. (C, D) High magnification SEM images showing the remnants of the MWNT on both AFM tips after fracture. (E) Plot of stress versus strain curves for individual MWNTs. Reprinted with permission from [57], M.-F. Yu et al., *Science* 287, 637 (2000). © 2000, AAAS.

To complement these modeling developments, continued advances in the ability to perform *quantitative mechanical experiments at the nanoscale* are necessary. For example, while work has been reported using various types of scanning probe microscopy (SPMs) such as atomic force microscopy (AFM) to probe the mechanical properties of nanostructures [49, 55, 60, 343], extensions of these techniques for polymer nanocomposites are at a much earlier stage. One success to date, however, has been the AFM-based nanoscale measurement of the interfacial shear strength of a nanotube-polymer composite (see Fig. 26) [180]. MEMS-based testing systems to probe the mechanical properties of nanomaterials and nanocomposites are also under development [344].

Finally, within the last several years there has been a push to develop 3D nanomanipulation systems which can operate within an electron microscope [345]. Attachments and clamping of nanostructures for mechanical loading can be accomplished via methods such as focused ion beam (FIB) deposition and electron beam induced decomposition [346]. Using such nanomanipulators, researchers have been able to subject individual MWNTs (see Fig. 80) [57] and single-walled carbon nanotube ropes [56] to tensile loading, realize the sliding between nested shells of a MWNTs [42], and load carbon nanocoils in tension (see Fig. 11) [21]. Such nanomanipulation systems are particularly well-suited to probe the mechanical properties of nanocomposites on the nanoscale, and the extension of nanoscale tests that have been developed for individual nanostructures to characterize polymer nanocomposites is envisioned.

5. CONCLUSIONS

In this chapter we have discussed the current state of modeling and future directions for polymer nanocomposites. The inherent properties of various nanoparticles and their size scale make them attractive candidates as a reinforcement filler material in polymer-based structural composites. The size scale is crucial for two reasons: (1) the radius of gyration of typical polymers is on the nanometer length scale and (2) the nanometer size increases the surface to volume ratio of the nanoparticles orders of magnitude over that of conventional polymer filler materials. Together, these two facts indicate that with good dispersion of the nanoparticles in a polymer, tremendous impact on overall polymer nanocomposite properties can be achieved at very low volume fractions. With good dispersion nearly all of the polymer will be within a R_g or two of a nanoparticle such that the conformation, mobility and properties of the polymer is altered. In addition to changing mechanical properties, the thermal stability, electrical and thermal conductivity, and diffusion properties are of the nanocomposite can also be greatly altered. The potential of multifunctional materials with controllable electrical and thermal properties, in addition to large enhancements in the mechanical behavior, has driven the tremendous amount of work dedicated to these material systems that continues today.

We have reviewed the experimental data in the field for polymer nanocomposites and showed evidence of the impact of equiaxed nanoparticles, nanotubes, and nanoplatelet reinforcements on the mechanical, thermal, electrical and diffusion properties. The work has also highlighted evidence for the existence of an extensive interphase region of nonbulk polymer behavior in nanocomposites even at very low (less than 1 vol%) loading. The characterization and extent of this interphase region is still a challenge, and a number of efforts are being directed toward nanotube pullout and other experiments to provide more detailed information on the interfacial load transfer.

Because of the complexity of nanocomposites and the difficulty of experimentation and data interpretation on the nanometer length scale, there is a critical need for accurate models of these systems. Modeling can be used not only to predict properties of a particular composite composition, but perhaps more importantly to assist in the analysis and interpretation of experimental data. However, modeling the behavior of nanoparticle reinforced polymers is inherently difficult because of complexities related to: the structure and properties of the nanoparticles, the orientation and dispersion of the nanoparticles within the polymer, the characteristics of the interface and load transfer between the nanoparticles and the polymer, a lack of understanding of the impact of the nanoparticles on the molecular mobility of the polymer chains, and the range of length scales characteristic in these materials.

We have presented the current range of models, divided into the bottom-up approaches that begin with atomistic detail, and the top-down approaches that are based on continuum assumptions. In each of these areas, significant strides have been made identifying key issues in nanocomposites and explaining the impact of specific structural features on properties. The molecular modeling approaches have demonstrated how the structural, thermal, and mechanical behavior of the polymer in the vicinity of nanoparticles is affected by the nanoinclusion and how the chemical interactions between the polymer and the nanoparticle play a key role in these areas. The continuum approaches have been able to account for microstructural issues such as intercalated polymer in nanoplatelet systems or the *in situ* geometry of nanoparticles to illustrate their impact on overall behavior of systems of many such inclusions. However, there are significant limitations with these approaches: pure atomistic simulations are still for small unit cells (small nanoparticles and a limited number of very short polymer chains) that neglect the details of realistically sized systems, while pure continuum approaches lack information regarding the nanoscale behavior of the nanoparticles and do not account for the gradients of properties present due to nanoparticle-polymer interactions.

Future modeling efforts should be focused toward multiscale approaches in which different length scales, from atomistic simulations to continuum theories, work in concert to accurately model the response of nanocomposites. Some strides have been made in this direction, with the development of some initial multiscale or hybrid modeling approaches. These methods have led to better understanding of the influence of the interphase and the morphology and arrangement of the nanoparticles on nanocomposite properties. However, much work remains to be done. More explicit coupling across the length scales and application of concurrent modeling approaches to examine large-scale nanocomposites with localized molecular and atomistic detail will bring needed accuracy and ensuing insight. In addition to modeling small strain elastic and viscoelastic properties, even more important is progress in modeling damage propagation and fracture, as well as the multifunctional properties of conductivity and diffusion. Currently, increases in toughness come at the expense of stiffness and strength. However, with better understanding of the interaction of nanoparticles and polymer chains across multiple length scales, it should be possible to greatly improve damage tolerance in systems that exhibit increased stiffness and strength. Many permutations to the design of nanocomposites are possible by targeted functionalization of nanoparticles to improve matrix interaction, use of secondary synergistic particles, and morphology control, all of which will be able to fundamentally change local and global nanocomposites behavior. By development of improved modeling strategies, we will be able to tailor the nanocomposite design to simultaneously optimize the key multifunctional properties of interest for given applications.

ACKNOWLEDGMENTS

We would like to thank our collaborators over the last few years, and in particular Linda Schadler (RPI), Ami Eitan (Intel), Roger Bradshaw (Louisville), Terry Xu (University of North Carolina–Charlotte), and Rod Ruoff (Northwestern). Additional thanks to Peter Golovin and Hua Liu (Northwestern) and Kon Chol Lee (Stevens) for creating some of the figures used in this chapter and to Dr. Debbie Burton for assistance in the final stages of manuscript preparation. Portions of our work described in this chapter have been supported by the NASA Langley Research Center Computational Materials: Nanotechnology Modeling and Simulation Program, and the NASA University Research, Engineering and Technology Institute on Bio Inspired Materials (BIMat) under award no. NCC-1-02037.

APPENDIX

A.1. Derivation of Mori-Tanaka Method

Assume that the composite is comprised of N phases; the matrix will be denoted as phase 0 with a corresponding stiffness C_0 and volume fraction f_0, while an arbitrary rth inclusion phase (where $r = 1$ to $N - 1$) has a stiffness of C_r and a volume fraction f_r. Each phase is assumed to be linearly elastic and isotropic, and perfect bonding between the inclusions and the matrix is assumed. The inclusions are further assumed to be ellipsoidal with a circular cross section ($a_1 = a_2$), an aspect ratio α_r (ratio of length to diameter), and aligned along the 3-axis as shown in Fig. A.1.

The two models shown in Fig. A.1 represent the composite model and a "comparison material" with properties identical to those of the matrix. Unless required, explicit tensor notation will be omitted for clarity. Displacements are now prescribed on the boundary of each material to give rise to a uniform strain ε_a in each material. The stresses required to produced this uniform strain in each material are

$$\bar{\sigma} = C\varepsilon_a, \quad \sigma_0 = C_0\varepsilon_a \tag{A1}$$

where $\bar{\sigma}$ and σ_0 are the average stress of the composite and comparison materials, respectively.

The strain field within the matrix material of the composite will not be uniform due the presence of the inclusions (and hence the average matrix strain $\bar{\varepsilon}_0$ will not equal ε_a), but rather will be perturbed by an amount $\bar{\varepsilon}_0^{\text{pt}}$ such that

$$\bar{\varepsilon}_0 = \varepsilon_a + \bar{\varepsilon}_0^{\text{pt}} \tag{A2}$$

where an overscore represents the volume average of the stated quantity. The average strain in the rth inclusion is further perturbed from that of the matrix,

$$\bar{\varepsilon}_r = \bar{\varepsilon}_0 + \varepsilon_r^{\text{pt}} \tag{A3}$$

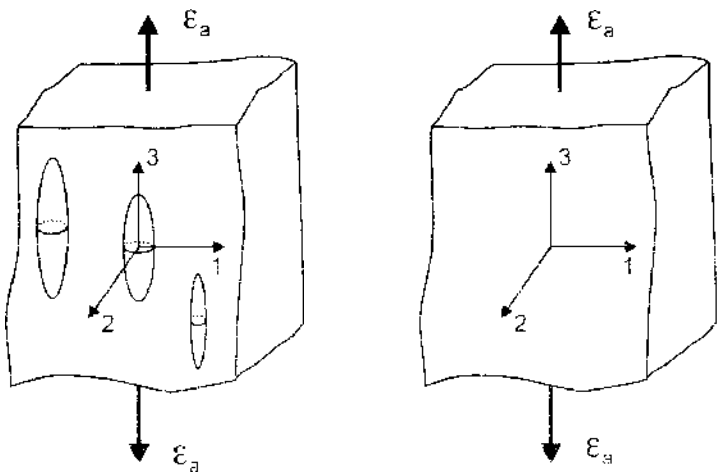


Figure A.1. Schematic of Mori-Tanaka method. (left) Multiphase composite material. (right) Comparison material.

Table A.1. Summary of related articles available in the literature. (key: E = experimental, M = modeling, R = review, NRP = nanotube-reinforced polymer, VGCF = vapor-grown carbon fiber).

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
Agag et al., Polymer **42**, 3399 (2001)	E	Nanoclay	Polyimide	Improvements in tensile modulus, higher T_g [221]
Ajayan et al., Science **265**, 1212 (1994)	E	MWNTs	Epoxy	Alignment of NTs via cutting [153]
Ajayan et al., Adv. Mater. **12**, 750 (2000)	E	SWNTs	Epoxy	Effectiveness of reinforcement governed by SWNT bundle behavior [171]
(*) Alexandre et al., Mater. Sci. Eng. R **28**, 1 (2000)	R	Nanoclay	Various	Review article [92]
Alexandre et al., Polymer **43**, 2123 (2002)	E	Nanoclay	PE	Impact of polymerization process on mechanical properties [223]
Allaoui et al., Compos. Sci. Technol. **62**, 1993 (2002)	E	MWNTs	Rubbery epoxy	Increase in mechanical and electrical properties [206]
Anastasiadis et al., Phys. Rev. Lett. **84**, 915 (2000)	E	Nanoclays	PMPS	Dielectric spectroscopy to probe role of confinement on intercalated polymer dynamics [245]
(*) Andrews et al., Macromol. Mater. Eng. **287**, 395 (2002)	E	MWNTs, CNFs	Various	Shear mixing of composite samples; electrical and mechanical properties [139]
Arrighi et al., Polymer **44**, 6259 (2003)	E	Nanosilica	SBR	Change in tan δ spectra due to interphase formation [121]
Avella et al., Nano Lett. **1**, 213 (2001)	E	Nano-$CaCO_3$	PMMA	Average abrasive weight loss halved with 2% nanoparticles [136]
Bai et al., Composites A **34**, 689 (2003)	E	MWNTs	Epoxy	Effect of NT length and aggregate size on mechanical and electrical properties [347]
(*) Barber et al., Appl. Phys. Lett. **82**, 4140 (2003); Barber et al., Compos. Sci. Technol. **64**, 2283 (2004)	E	MWNTs	Polyethylene-butene	Nanopullout experiments in AFM; change in interphase yield behavior [179, 180]
Barraza et al., Nano Lett. **2**, 797 (2002)	E	SWNTs	PS, Styrene-isoprene	Strong interactions between the NTs and the polymer chains [348]
Baschnagel et al., Adv. Polym. Sci. **152**, 41 (2000)	R	--	Model polymers	Bridging the gap between atomistic and coarse-grained polymer models [262]
Becker et al., Macromolecules **36**, 1616, (2003)	E	Nanoclay	Various epoxies	Decreased T_g and cross-link density in interfacial regions [239]
Benveniste, Mech. Mater. **6**, 147, (1987)	M	--	—	A new approach to the Mori-Tanaka theory for composite materials [310]
Bharadwaj et al., Polymer **41**, 7209 (2000)	M	POSS	Polynorborene	MD simulations, increase in T_g [349]
Bharadwaj et al., Polymer **43**, 3699 (2002)	E	Nanoclay	Polyester	Exfoliation leads to a *decrease* in cross-linking and mechanical properties [224]
Bhattacharayya et al., Polymer **44**, 2373 (2003)	E	SWNTs	PP	Nanocomposite exhibits faster crystallization rates due to SWNTs [192]
Bhimaraj et al., Wear 58, 1437 (2005)	E	Nanoalumina	PET	Wear resistance doubles at 2 wt% loading [134]
Biercuk et al., Appl. Phys. Lett. **80**, 2767 (2002)	E	SWNT, VGCF	Epoxy	Thermal and electrical properties show greater improvement with SWNTs [211]
Biswas et al., Adv. Polym. Sci. **155**, 167 (2001)	R	Nanoclay	Various	Review article [91]
Bohme et al., J. Chem. Phys. **116**, 9939 (2002)	M	—	Model linear polymer	Model of nanoscopic polymeric structures [276]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
Bower et al., Appl. Phys. Lett. **74**, 3317 (1999)	E	MWNTs	Thermoplastic polymer	Alignment by mechanical stretching above polymer T_g [157]
Brechet et al., Adv. Eng. Mater. **3**, 571 (2001)	R	Various	Various	Filler percolation and network in polymer nanocomposites [350]
Breuer et al., Polym. Compos. **25**, 630 (2004)	R	NT	Various	Review article [143]
Brown et al., Macromol. Mater. Eng. **36**, 1395 (2003)	M	Nanosilica	Model polymer	MD study of variation of structure and dynamics as a function of distance from interface [351]
Burnside et al., J. Polym. Sci. B Phys. **38**, 1595 (2000)	E	Nanoclay	Polysiloxane	Bound polymer chains showed an increase and broadening of T_g and loss of crystallization transition [242]
Cadek et al., Appl. Phys. Lett. **81**, 5123 (2002)	E	MWNTs	PVA, PVK	NTs nucleate crystallization of PVA, enhancement of matrix-nanotube stress transfer [185]
(*) Cadek et al., Nano Lett. **4**, 353 (2004)	E	Various NTs	PVA	Reinforcement scales with nanotube surface area [257]
Carrado, Appl. Clay Sci. **17**, 1 (2000).	R	Nanoclays	Various	Review article [93]
Chatterjee et al., J. Chem. Phys. **121**, 11420 (2004)	M	—	Model	Spatially adaptive lattice coarse-grained Monte Carlo simulations [285]
Chen et al., Comp. Mat. Sci. **29**, 1 (2004)	M	CNTs	Model polymer	Square representative volume elements for carbon nanotube-based composites [303]
Cho et al., Polymer **42**, 1083 (2001)	E	Nanoclay	Nylon 6	Direct melt compounding with conventional twin screw extruder [225]
Cho et al., Macromol. Mater. Eng. **290**, 179 (2005)	E	Expanded graphite	Polyimide	Dynamic mechanical and thermal properties of graphite nanoplatelet nanocomposites [109]
Christensen, J. Mech. Phys. Sol. **38**, 379 (1990)	R	—	—	Review of various micromechanics methods [315]
Chung, J. Mater. Sci. 37, 1475 (2002)	R	Graphite	-	Review of graphite [107]
Clancy et al., NASA/TM 2004-213030 (2004)	M		Polyimide	Coarse-grained and atomistic modeling of polyimides [284]
(*) Coleman et al., Appl. Phys. Lett. **82**, 1682 (2003)	E	SWNTs	PVA, PVP, PS	Nanocomposites formed by soaking NT buckysheets in polymer solution [197]
(*) Cooper et al., Appl. Phys. Lett. **81**, 3873 (2002)	E	SWNTs	Epoxy	Experimental nanopullout: SWNT-polymer adhesion exceed SWNT rope strength [352]
Cooper et al., Compos. Sci. Technol. **62**, 1105 (2002)	E	Various	PMMA	Tensile modulus and fracture toughness as function of nanoinclusion [161]
Cooper et al., Composites A **32**, 401 (2001)	E	SWNT, MWNT	Epoxy	Raman spectroscopy to study stress transfer to the NTs [181]
Delozier et al. Compos. Sci. Technol. **65**, 749 (2005)	E	SWNT	Polyimide	Preparation and characterization of space durable polymer nanocomposite films [201]
(*) Ding et al., Nano Lett. **3**, 1593 (2003)	E	MWNTs	PC	Balling-up of adhered polymer interphase on CNTs [251]
Doruker et al., Macromol. Theory Simul. **8**, 463 (1999)	M	—	Model polymer	Mapping/reverse mapping between coarse-grained and fully atomistic polymer model [353]
Dror et al., Langmuir **19**, 7012 (2003)	E	CNTs	Electrospun polymer nanofibers	Embedded CNTs in orientated polymer nanofibers by electrospinning [88]
Du et al., J. Polym. Sci., B **41**, 3333 (2003)	E	SWNTs	PMMA	Coagulation method provided better dispersion and enhanced property improvement [152]

Du et al., Macromolecules, **37**, 9048 (2004)	E	SWNTs	PMMA	Effect of nanotube networks in polymer nanocomposites [203]
Dufresne et al., J. Mater. Sci. **37**, 3915 (2002)	E	MWNTs	Poly(S-co-BuA)	Percolation around 1.5%; reinforcement up to 3 wt% without lowering failure strain [324]
(*) Dzenis, Science **304** 1917, (2004)	R		Various	Review of polymer nanofiber electrospinning [81]
(*) Eitan et al., Compos. Sci. Technol. (to be published, 2006)	E	MWNTs	PC	Interphase provide additional reinforcement to nanocomposite [116]
Ellison et al., Phys. Rev. Lett. **92**, 095702 (2004); Ellison et al., Nature Materials, **2**, 695, (2003)	E	—	Thin polymer films	Experimental measurement of change in polymer behavior near surfaces [10]
Enomoto et al., New Diamond Front. Carbon Technol. **14**, 11 (2004)	E	MWNTs, VGCFs	PS	Friction/wear properties of molded samples [355]
Fan et al., J. Mater. Res. **17**, 1622 (2002)	E	Nanoclay	Polyelectrolyte	Characterization of nanoclay-polymer LBL films [104]
Fertig et al., Compos. Sci. Technol. **64**, 2577 (2004)	M	Nanoclay	Model polymer	Multiscale FEM-micromechanics model of clay nanocomposite [305]
Fisher et al., Appl. Phys. Lett., **80**, 4647 (2002); Bradshaw et al., Compos. Sci. Technol. **63**, 1705, (2003); Fisher et al., Compos. Sci. Technol. **63**, 1689 (2003)	M	NTs	Model polymer	Hybrid FEM-micromechanics model of nanotube waviness on reinforcement [298–300]
Fisher et al., Adv. Compos. Lett. **13**, 105 (2004)	E	MWNTs	PC	Effective relaxation spectra due to reduced mobility interphase [117]
Fong et al. Polymer **43**, 775, (2002)	E	Nanoclays	Nylon 6	Electrospinning of nanoclay-nylon-6 nanocomposites [83]
Fornes et al., Polymer **44**, (2003)	M	Nanoclays	Model polymer	Mori-Takana and Halpin-Tsai model of impact of intercalation [297]
Foygel et al., Phys. Rev. B **71**, 104201 (2005)	M	CNTs	Model polymer	Monte Carlo simulations of electrical and thermal conductivity [218]
(*) Frankland et al., J. Phys. Chem. B **106**, 3046 (2002)	M	SWNTs	PE	MD analysis of influence of cross-links on matrix-NT shear strength [184]
Frankland et al., Compos. Sci. Technol. **63**, 1655 (2003)	M	SWNTs	PE	MD simulation to generate stress-strain curve of nanocomposite [277]
Frogley et al., Compos. Sci. Technol. **63**, 1647 (2003)	E	SWNTs, VGCFs	Silicone elastomer	Enhancement of the initial modulus, reduction of the ultimate properties [356]
Frohlich et al., Composites A **36**, 449 (2005)	E	Carbon black, silicates	Elastomer	Effect of filler-filler and filler-elastomer interaction on reinforcement [357]
Gao et al., Int. J. Sol. Struct. **42**, 1649 (2005)	M	CNTs	Model polymer	Shear-lag model for NRPs [290]
(*) Garg et al., Chem. Phys. Lett. **295**, 273 (1998)	M	CNTs	—	MD simulation of impact of functionalization on CNT mechanical properties [264]
Gauthier et al., Polymer **45**, 2761 (2004)	E	Nanosilica	SBR	Payne effect occurs at low silica content and dependent on surface treatment [127]
Geng et al., Adv. Mater. **14**, 1387 (2002)	E	SWNTs	PEO	Increase in mechanical properties, decrease in melting temperature [330]
Gersappe, Phys. Rev. Lett. **89**, 058301 (2002)	M	Nanosphere	Linear polymer	MD simulations showing nanofillers improve toughness [270]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
(*) Giannelis, Adv. Mater. **8**, 29 (1996); (*) Giannelis, Appl. Organomet. Chem. **12**, 675 (1998)	R	Nanoclays	Various	Review articles [97, 358]
(*) Giannelis et al., Adv. Polym. Sci. **138**, 107 (1999)	E, M	Nanoclays	Various	Nanocomposites as model systems for confined polymers [12]
Gilman, Appl. Clay Sci. **15**, 31 (1999)	R	Nanoclays	Various	Review article [359]
Gilman et al., Chem. Mater. **12**, 1866 (2000)	E	Nanoclays	PP, PS	Flammability improvements in clay nanocomposites [231]
(*) Glotzer et al., Annu. Rev. Mater. Res. **32**, 401 (2002)	M	—	Various	Review of computational simulations of polymers [283]
Goh et al., Chem. Phys. Lett. **373**, 277 (2003)	E	MWNT	Phenoxy	DMA mechanical behavior of *in situ* functionalized composites [360]
Gong et al., Chem. Mater. **12**, 1049 (2000)	E	MWNTs	Epoxy	Surfactant in processing resulted in better mechanical properties [361]
Gorga et al., J. Polym. Sci. B **42**, 2690 (2004)	E	MWNTs	PMMA	Orientation necessary to improve toughness of composite [190]
Govindjee et al., Solid State Commun. **110**, 227 (1999)	M	NTs		Discussion of continuum approximation for modeling NTs [61]
(*) Griebel et al., Comput. Methods Appl. Mech. Eng. **193**, 1773 (2004)	M	SWNTs	PE	MD simulation of elastic properties of NRP [362]
Gumen et al., Polymer **42**, 5717 (2001)	M	—	Epoxy	Computer simulation of T_g and mechanical properties [363]
Hackett et al., Chem. Mater. **12**, 2161 (2000)	M	Nanoclays	PEO	MD to study the atomic scale structure and dynamics of intercalated nanocomposites [281]
Haggenmueller et al., Chem. Phys. Lett. **330**, 219 (2000); Haggenmueller et al., J. Nanosci. Nanotechnol. **3**, 105 (2003)	E	SWNTs	PMMA, PS, PE	Melt-mixing/spinning to achieve alignment and exceptional properties [158, 167]
Halpin, J. Comput. Mat. **3**, 732 (1969); Halpin et al., Poly. Eng. Sci. **16**, 344 (1976)	M	—	Model polymer	Halpin-Tsai micromechanics methods [312, 313]
Hone et al., Appl. Phys. A **74**, 339 (2002)	E	SWNT, VGCF	Epoxy	High thermal conductivity polymer composites [212]
Hu et al., Compos. Sci. Technol. **63**, 1663 (2003); Hu et al., J. Mater. Chem. **14** (2004)	M	SWNTs	PS	MD simulation of chemical modification via polyatomic ion beam deposition [364]
Hu et al., Polym. Degrad. Stab. **84**, 545 (2004).	E	Nanosilica	PMMA	Increase in T_g, enhancement of thermal characteristics [118]
Hu et al., Proc. Royal Soc. London, Ser. A **461**, 1685 (2005)	M	SWNTs	PS	Three-phase RVE with interphase properties obtained from MD simulation [365]
Huang et al., Polym. Test. **23**, 9 (2004)	E	Nanosilica	LLDPE	Dynamic mechanical properties as a function of surface treatment [124]
Huhtala et al. Phys. Rev. B **70**, 045404 (2004)	M	MWNTs	—	MD simulation of effect of ion irradiation on load transfer within MWNT [366]
Hyun et al., Macromolecules **34**, 8084 (2001)	E	Nanoclays	PEO	Effect of surfactants on organoclay surfaces in nanocomposites [367]

Table A.1. Continued.

in het Panhuis et al., J. Phys. Chem. B **107**, 478 (2003)	M	SWNTs	PmPV	MD simulation of selective interactions between SWNT and a conjugated polymer [269]
Inceoglu et al., Polym. Eng. Sci. **43**, 661 (2003)	E	Nanoclays	Polyester	Impact of nanoclay chemical modification on mechanical properties [368]
Izumisawa et al., J. Chem. Phys. **117**, 3972 (2002)	M	—	Model polymer	Monte Carlo simulation of polymer interactions with surfaces [369]
Jia et al., Mater. Sci. Eng. A **271**, 395 (1999)	E	MWNTs	PMMA	CNTs may participate in polymerization: strong interface with matrix [170]
Jin et al., Appl. Phys. Lett. **73**, 1197 (1998)	E	MWNTs	PHAE	Alignment of NTs by mechanical stretching [156]
Jin et al., Chem. Phys. Lett. **318**, 505 (2000)	E	MWNTs	PEG, P2VP, P4VP, PVPh, PEO	Investigation of nonlinear optical properties [370]
Jin et al., Chem. Phys. Lett. **37**, 43 (2001)	E	MWNTs	PMMA	DMA analysis, shift/broadening of T_g peak [186]
Jin et al., Mater. Res. Bull. **37**, 271 (2002)	E	PVDF-coated MWNTs	PMMA	PVDF coating leads to property improvements [371]
Kao et al., Compos. Sci. Technol. **64**, 2291 (2004)	E	SWNTs	Epoxy	Raman spectroscopy study of nanocomposite deformation at 0.1 wt% loading [372]
Kawasumi, J. Polym. Sci. Pol. Chem. **42**, 819 (2004)	R	Nanoclays	Various	History of polymer nanocomposite development at Toyota [100]
Kilbride et al., J. Appl. Phys. **92**, 4024 (2002)	E	CNTs	Various	Experimental observation of scaling laws for *ac* and *dc* conductivity in NRP thin films [210]
Kim et al., J. Appl. Polym. Sci. **87**, 2106 (2003)	E	Nanoclays	PS	Rheological characterization of the nanocomposite [373]
Kimura et al., Adv. Mater. **14**, 1380 (2002)	E	CNTs	Various	Alignment of CNTs in NRP via magnetic field [165]
Kinloch et al., J. Adhes. **79**, 867 (2003)	E	Nanosilica	Epoxy adhesive	Significant increases in toughness of adhesive [129]
Ko et al., Adv. Mater. **15**, 1161 (2003)	E	SWNTs	PLA, PAN	Co-electrospinning of nanotube-polymer composites [86]
Koerner et al., Nature Mater. **3**, 115 (2004)	E	MWNTs, carbon black	Elastomer	Characterization of the stimuli-reactive response [374]
Kojima et al., J. Polym. Sci. B **32**, 625 (1994); Kojima et al., J. Polym. Sci. B **33**, 1039 (1995); Usuki et al., J. Mater. Res., **8**, 1179, 1993;	E	Nanoclays	Nylon 6	Original clay nanocomposite work at Toyota [90, 375]
Kong et al., Chem. Mater. **15**, 419 (2003)	E	Nanoclays	Epoxy	Real-time exfoliation behavior [376]
Koratkar et al., Adv. Mater. **14**, 997 (2002)	E	MWNTs	Cyanacrylate adhesive	Film provide structural damping via frictional/dissipative losses [377]
Kornmann et al., Polymer **42**, 4493 (2001)	E	Nanoclays	Epoxy	Impact of curing conditions on exfoliation [378]
(*) Krishnamoorti et al., Curr. Opin. Colloid Interface Sci. **6**, 464 (2001)	R	Nanoclays	Various	Viscoelastic measurements to probe strength of particle interactions [238]
Kumar et al., Polymer **43**, 1701 (2002)	E	VGCFs	PP	Conventional melt-spinning, increase in mechanical properties [379]
Lau et al., Composites B, **33**, 263 (2002)	R	CNTs	Various	Review article [142]
Lau, Chem. Phys. Lett. **370**, 399 (2003)	M	SWNTs, MWNTs	Model polymer	Theoretical study of stress transfer between NT and matrix [337]
Lau et al., Compos. Sci. Technol., **65**, 719 (2005)	E	SWNTs	Epoxy	Role of solvent in dispersion, cure of NRP [146]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
LeBaron et al., Appl. Clay Sci. **15**, 11 (1999)	R	Nanoclays	Various	Review article [94]
Lee et al., Macromolecules **36**, 804 (2003)	E	Nanoclays	Diblock copolymer	Linear viscoelastic properties of block copolymer nanocomposites [246]
(*) Leger et al., Adv. Polym. Sci. **138**, 185 (1999)	R	—	Various	Role of surface-anchored polymer layers in adhesion and friction [380]
Li et al., J. Nanosci. Nanotechnol. **3**, 423 (2003)	M	CNTs	Model polymer	Stress analysis of nanocomposite using molecular structural mechanics and FEM [289]
Li et al., Compos. Sci. Technol. **64**, 2089 (2004)	E	MWNTs, carbon black	PP	Antistatic enhancement greater for CNTs than for carbon black [381]
Li et al., Nanotechnology, **15**, 1416 (2004)	E	SWNTs	Epoxy	Nanoindentation, nanoscratch, nanoDMA characterization [195]
(*) Liang et al., Mater. Sci. Eng. A **365**, 228 (2004)	M	SWNT	Epoxy	MD simulation of interactions between SWNT and epoxy components [382]
(*) Liao et al., Appl. Phys. Lett. **79**, 4225 (2001)	M	CNTs	PS	Molecular mechanics simulations of interface; *IFSS* of 160 MPa [177]
Liu et al., Polymer **44**, 7529 (2003)	M	CNTs	PC	Rheological analysis of gelation, correspondence with percolation [383]
Liu et al., Compos. Sci. Technol. **63**, 331 (2003)	E	Nanoclay	Polyamide	Morphology, thermal, and mechanical behavior [219]
Liu et al., Polymer **43**, 1933 (2002)	E	Nanoclay	Nylon 6	Effect of confined polymer on phase transitions [243]
Liu et al., Appl. Phys. Lett. **84**, 4248 (2004)	E	CNTs, carbon black	Silicone elastomer	Study of thermal conductivity enhancement [213]
Liu et al., Polymer **42**, 10013 (2001)	E	Nanoclays	PP	Melt-grafting intercalation and tethering [384]
Liu et al., Mech. Mater. **35**, 69 (2003)	M	CNTs	Model polymer	Continuum model of composite using RVE [302]
Liu et al., Polymer **45**, 2051 (2004)	M	—	Various epoxies	Group interaction modeling (GIM) of glass transition [385]
Liu et al., Polymer **44**, 5159 (2003)	E	Nanosilica	Epoxy	Improved thermal stability, depression of T_g [132]
Liu et al., Appl. Phys. Lett. **86**, 123106 (2005)	E	CNTs	Various	Effects of chemical modifications on the thermal conductivity of NRPs [214]
Liu et al., Compos. Sci. Technol. **65**, 307 (2005)	M	Nanoclays	Epoxy	Fracture toughness and viscoelastic behavior of clay nanocomposites [222]
(*) Liu et al., Comput. Meth. Appl. Mech. Eng. **193**, 1529 (2004)	R	Various	Various	Introduction to computational nanomechanics and materials [338]
Long et al., Appl. Phys. Lett. **85**, 1796 (2004)	E	MWNTs	Polyaniline (PANI)	Suggest strong coupling between CNTs and the tightly coated polymer chains enhances the electronic properties of the composites [207]
Lordi et al., J. Mater. Res. **15**, 2770 (2000)	M	CNTs	Various	Molecular mechanics; helical polymer conformations [illegible] [173]
Lourie et al., Compos. Sci. Technol. **59**, 975 (1999)	E	MWNTs	Epoxy	Formation of fracture clusters in NRPs [386]
Lozano et al., J. Appl. Polym. Sci. **79**, 125 (2001)	E	VGCFs	PP	Strength increase limited by enhanced crystallization brought on by nanofiber [172]
(*) Luo et al., Compos. Sci. Technol. **63**, 1607 (2003)	M	Nanoclays	Epoxy	Three-phase micromechanics model accounting for intercalation level [296]

Table A.1. Continued.

Lusti et al., Modell. Simul. Mater. Sci. Eng. **12**, S107 (2004)	M	CNTs	Model polymer	Continuum FEM model of thermoelastic properties [304]
Mack et al., Adv. Mater. **17**, 77 (2005)	E	Graphite nanoplatelet	PAN	Reinforcement of electrospun nanofibers [85]
(*) Mamedov et al., Nature Mater. **1**, 190 (2002); Olek et al., Nano Lett. **4**, 1889 (2004)	E	SWNTs, MWNTs	Polyelectrolytes	Layer-by-layer (LBL) assembly of nanocomposites [387]
Manevitch et al., J. Phys. Chem. B **108**, 1428 (2004)	M	Nanoclays	—	MD simulation of elastic properties of a single lamella [388]
Manias et al., Macromolecules **33**, 7955 (2000)	E	Nanoclays	PS	Intercalation kinetics of polymers in 2 nm confinements [326]
McCarthy et al., J. Mater. Sci. Lett. **19**, 2239 (2000); McCarthy et al., Synth. Met. **121**, 1225 (2001); McCarthy et al., J. Phys. Chem. B **106**, 2210 (2002)	E	MWNTs, SWNTs	PmPV	Nucleation and crystalline growth of semiconjugated polymer from NT defects [389]
Meincke et al., Polymer **45**, 739 (2004)	E	MWNTs, carbon black	Polyamide-6 and ABS	PA6-NT nanocomposites and blends with ABS [187]
Messersmith et al., Chem. Mater. **6**, 1719 (1994)	E	Nanoclays	Epoxy	Broadened T_g at slightly higher temperature [240]
Messersmith et al., J. Polym. Sci. Pol. Chem. **33**, 1047 (1995)	E	Nanoclays	Poly(epsilon-caprolactone)	Order of magnitude reduction in permeability [7]
Minisini et al., Composites A **36**, 531 (2005); Minisini et al., Composites A **36**, 539 (2005)	M	Nanoclays	PP	MD/MM study of interactions of nanoclay with surfactants and PP [390]
Musto et al., Polymer **45**, 1697 (2004)	E	Nanosilica	Polyimide	Mechanical properties of micro and nanocomposites [391]
Nan et al., Appl. Phys. Lett. **85**, 3549 (2004)	M	CNTs	Model polymer	Thermal conductivity enhancement limited by the interface thermal resistance [215]
Natsuki et al., Carbon **42**, 39 (2004)	M	SWNT	—	Continuum-shell model of discrete molecular structures linked by C–C bonds [392]
Norris, Int. J. Solids Struct. **26**, 663 (1990)	M	Platelets	Model polymer	Micromechanics models of platelet-reinforced composites [393]
Odegard et al., Compos. Sci. Technol. **62**, 1869 (2002)	M	SWNTs, graphene sheet	--	Equivalent continuum model linking computational chemistry and solid mechanics [268]
(*) Odegard et al., Compos. Sci. Technol. **63**, 1671 (2003)	M	SWNTs	Polyimide	Equivalent continuum model for nanotube-polymer systems [286]
Odegard et al., Polymer **46**, 553 (2005)	M	Nanosilica	Polyimide	Equivalent continuum modeling; impact of nanoparticle functionalization [287]
Ogasawara et al., Composites A **35**, 67 (2004)	E	MWNTs	Polyimide	NTs act as cross-links, immobilizing PI chains at higher temperatures [188]
Ounaies et al., Compos. Sci. Technol. **63**, 1637 (2003)	E	SWNTs	Polyimide	Conductivity described by a percolation-like power law with a threshold of 0.05 vol.% [209]
Ozmusul et al., Polym. Compos., **23**, 110 (2002); Ozmusul et al., Polymer **43**, 4657 (2002); Picu et al., J. Chem. Phys. **118**, 11239 (2003)	M	Nanoinclusion	Model polymer	Nanocomposite graded interfaces due to preferential orientation of polymer [282, 394]
Pan et al., Appl. Phys. Lett. **74**, 3152, (1999)	E	MWNT	—	For CVD MWNTs, modulus and tensile strength were 450 and 1.72 GPa [37]
Panchagnula et al., Phys. Rev. Lett. **93**, 037801 (2004)	M	Spherical nanoparticle	Polyelectrolytes	MD simulation of multilayer assembly around a particle [395]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
(*) Pantano et al., J. Mech. Phys. Solids **52**, 789. (2004)	M	SWNTs, MWNTs	—	Continuum/FE model using shell and interaction elements for tubes and van der Waals forces [396]
Park et al., Chem. Phys. Lett. **364**, 303 (2002)	E	SWNTs	Polyimide	Conductivity enhancement 10 orders of magnitude at 0.1 vol%; shift in T_g [196]
Park et al., Nanotechnology **14**, L11 (2003)	E	SWNTs	Polyimide	Wetting of NT surfaces by polyimide [397]
Park et al., Polymer **44**, 2091 (2003)	E	Nanoclay	Epoxy, PMMA	Epoxy-aided dispersion of nanoclay in PMMA [229]
(*) Park et al., Comput. Meth. Appl. Mech. Eng. **193**, 1733 (2004)	R	Various	Model polymer	Introduction/tutorial to multiple scale modeling methods [339]
Park et al., Polymer **45**, 4507 (2004)	M	—	Copolymer films	Chemical interactions between polymer chains and substrate dictate whether increase or decrease in T_g [241]
Paul, Polymer **45**, 3901 (2004)	M	—	Various	MD simulations of the glass transition [398]
Peeterbroeck et al., Compos. Sci. Technol. **64**, 2317 (2004)	E	Nanoclays, MWNTs	EVA copolymer	Ternary nanocomposite with improved thermal and flame retardant properties [232]
Pipes et al., Compos. Sci. Technol. **62**, 419 (2002); Pipes et al., Compos. Sci. Technol., **63** 1571 (2003)	M	SWNT arrays	Model polymer	Helical twisted SWNT arrays for nanocomposite reinforcement [399]
Pipes et al., Compos. Sci. Technol. **63**, 1349 (2003)	M	SWNT, SWNT arrays	Model polymer	Self-consistent physical properties for nanocomposite applications [68]
(*) Potschke et al., Polymer **43**, 3247 (2002); Potschke et al., Compos. Interfaces **10**, 389 (2003); Potschke et al., Polymer **44**, 5023 (2003); Potschke et al., Eur. Polym. J. **40**, 137 (2004)	E	MWNT masterbatch	PC	Melt mixing of Hyperion masterbatch with PC [204, 400]
Potschke et al., Polymer Prepriats **44**, 760 (2003); Potschke et al., Polymer, **44**, 8061 (2003)	E	MWNT masterbatch	PC, PE	Blending of PC-MWNT nanocomposite with PE [205, 401]
Prasad et al., Geophys. Res. Lett., **29**, 1172 (2002)	E	Nanoclay	—	Modulus of clay using atomic force acoustic *microscopy* [402]
Prasse et al., Compos. Sci. Technol. **63**, 1835 (2003)	E	Pyrograf III	Epoxy	Anisotropy due to *ac* alignment of NTs [164]
Putz et al., J. Polym. Sci., B **42**, 2286 (2004)	E	SWNTs	PMMA	Suggest changes in mechanical properties of interphase [194]
Qian et al., Appl. Phys. Lett. **76**, 2868 (2000); Qian et al., J. Microsc. **204**, 39 (2001)	E	MWNTs	PS	*In situ* TEM study of deformation mechanisms in NRPs [36, 403]
(*) Qian et al., Appl. Mech. Rev. **55**, 495 (2002)	R	CNTs	...	Review article [30]
Qian et al., Comput. Meth. Appl. Mech. Eng. **193**, 1603 (2004)	M	CNTs	—	Multiscale projection methods for CNT systems [342]
Ramanathan et al., J. Polym. Sci., B **43**, 2269 (2005)	E	SWNTs	PMMA	Large change in viscoelastic behavior due to functionalization [182]
Ramanathan et al., (submitted to Polymer, 2006)	E	Nanographite	PMMA	Mechanical properties of expanded graphite—PMMA nanocomposites [237]
(*) Ramasubramaniam et al., Appl. Phys. Lett. **83**, 2928 (2003)	E	SWNTs	PS, PC	Noncovalent SWNT functionalization to increase dispersion [147]

Table A.1. Continued.

(*) Ray et al., Prog. Polym. Sci. 28, 1539 (2003)	R	Nanoclays	Various	Review of preparation and processing of nanoclay composites [99]
Ren et al., Carbon **41**, 2177 *(2003)*	E	SWNTs	Epoxy	Tension-tension fatigue behavior of NRPs [404]
(*) Ren et al., Macromolecules **33**, 3739 (2000)	E	Nanoclays	Block copolymer	Altered viscoelastic behavior at long times: pseudosolid-like behavior at large loadings [405]
Roberge et al., Polymer, **45**, 1401 (2004)	M	—	PET	MD simulation of uniaxial deformation of amorphous PET [406]
Rodlert et al., J. Rheol. **48**, 1049 (2004)	E	Nanoclays	Hyperbranched polyesters	Rheological properties as a function of level of intercalation [407]
Rong et al., Polymer **42** 3301 (2001)	E	Nanosilica	PP	Surface treatment leads to a double percolation of yielded zones [131]
Rossi et al., Colloids Surf. A **215**, 1 (2003)	E	Nanoclays	PEO, PEO-PPO-PEO triblock	Rheology of non-ionic polymers with nanoclays: adsorption depends on polymer structure [408]
Rouse et al., Nano Lett. **3**, 59 (2003)	E	SWNTs	Polyelectrolytes	Characterization of layer-by-layer (LBL) composite films [409]
Ruan et al., Polymer **44**, 5643 (2003)	E	MWNTs	UHMWPE	Enhanced toughness and ductility with MWNTs [191]
Saether et al., Compos. Sci. Technol., **63** 1543 (2003)	M	SWNT bundles	—	Transverse mechanical properties of SWNT bundles [67]
Safadi et al., J. Appl. Polym. Sci. **84**, 2660, (2002)	E	MWNTs	PS	MWNT alignment via spin-casting [159]
(*) Salaniwal et al., Phys. Rev. Lett. **89**, 258301 (2002)	M	Nanoplatelets	Model polymer	Monte Carlo simulations of rheology changes in polymer melts loaded with platelets [274]
Salvetat-Delmotte et al., Carbon **40**, 1729 (2002)	R	CNTs	—	Condensed review of mechanical properties [410]
Sammalkorpi et al., Phys. Rev. B **70**, 245416 (2004)	M	SWNTs	—	MD study of impact of vacancies on SWNT mechanical properties [411]
Sandler et al., Polymer **40**, 5967 (1999)	E	MWNTs	Epoxy	Enhancement of electrical properties [199]
Sandler et al., Composites A **33**, 1033 (2002); Sandler et al., J. Mater. Res. **38**, 2135 (2003)	E	VGCFs	PEEK	Interphase morphology could add additional reinforcement [256]
Sandler et al., Polymer **44**, 5893 (2003)	E	MWNTs	Epoxy	Electrical percolation as low as 0.001 wt% [208]
(*) Sandler et al., Polymer **45**, 2001 (2004)	E	Various MWNTs, CNFs	Polyamide-12	Comparative study of spun nanocomposite fibers [189]
Sato et al., J. Phys. Chem. B **105**, 7990 (2001)	M	Single clay layer	—	MD simulation of flexibility of single exfoliated layer [412]
Sawyer et al., Wear **254**, 573 (2003)	E	Nanoalumina	PTFE	Solid lubricant; at 20 wt% Al2O3, wear resistence increased 600 times [133]
(*) Schadler et al., Appl. Phys. Lett. **73**, 3842 (1998)	E	MWNTs	Epoxy	Load transfer in MWNT-polymer composite different in tension/compression [39]
Schmidt et al., Curr. Opin. Solid State Mater. Sci. **6**, 205 (2002)	R	Nanoclays	Various	Review article [95]
Shaffer et al., Adv. Mater., **11**, 937 (1999)	E	MWNTs	PVOH	Experimental evidence of interphase formation [169]
Shah et al., Polymer **45**, 2991 (2004)	E	Nanoclay	Nylon 6	Masterbatch processing of Nylon 6-organoclay nanocomposites [226]
Sharaf et al., Polymer **45**, 3943 (2004)	M	Spherical nanoparticles	Amorphous PE	Monte Carlo simulation of polymer chain deformations in filled systems [279]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
Shen et al., Phys. Rev. B **69**, 045414 (2004); Shen et al., Phys. Rev. B **71**, 035412 (2005)	M	SWNTs, MWNTs		Transversely isotropic elastic properties [65, 66]
(*) Shen et al., Phys. Rev. B **69** 045414 (2004)	M	Nanoclays	Various	Multiscale micromechanical modeling using effective clay particle [65]
Shen et al., Polymer **45**, 3341 (2004)	E	Nanoclay	Nylon 66	Nanoindentation study: creep behavior dominated by morphological changes in polymer [247]
Shen et al., J. Appl. Polym. Sci. **88**, 1864 (2003)	E	Expanded graphite	PMMA	Thermal conductivity enhancements using expanded graphite [234]
Shenogin et al., Appl. Phys. Lett. **85**, 2229 (2004)	M	SWNTs	Model polymer	MD model of impact of functionalization on thermal transport in NRPs [216]
Shenogin et al., J. Appl. Phys. **95**, 8136 (2004)	M	CNTs	Model polymer	MD simulation of role of thermal boundary resistance on heat flow in NRPs [217]
Shi et al., Appl. Phys. Lett. **81**, 5216 (2002); Shi et al., Appl. Phys. Lett. **83**, 5301 (2003)	E	Pyrograf III	Pyrrole, PS	Plasma coating of thin layer to enhance dispersion and load transfer [148, 149]
Shi et al., J. Eng. Mater. Technol. **126**, 250 (2004)	M	CNTs	Model polymer	Micromechanics model of waviness and agglomeration [413]
Shia et al., Polym. Compos. **19**, 608 (1998)	M	Nanoclays	Model polymer	Imperfect interface model; effective aspect ratio [414]
Shofner et al., Composites A **34**, 1207 (2003)	E	SWNTs, Pyrograf III	ABS	Extrusion freeform fabrication EFF [415]
Siegel et al., Scripta Mater. **44**, 2061 (2001)	E	Nanoalumina	PMMA	Large increase in ductility with addition of nanophase alumina [128]
Singh et al., Adv. Funct. Mater. **13**, 868 (2003)	E	SWNTs	PC	Entangled SWNT bundle network key for enhanced properties [193]
(*) Sinsawat et al., J. Polym. Sci. B **41**, 3272 (2003)	M	Model sheet	Model polymer	MD analysis of formation of intercalated/exfoliated nanocomposites [416]
(*) Smith et al., J. Chem. Phys. **117**, 9478 (2002)	M	Spherical nanoparticles	Model polymer	MD study of influence of interface on viscoelastic properties [15]
(*) Smith et al., Polymer **45**, 825 (2004)	E	SWNTs	Polyimide	Electrical and optical nanocomposite properties [200]
Smith et al., Compos. Sci. Technol. **63**, 1599 (2003)	M	Nanoparticle	Model polymer	MD study of nanoparticle interactions in a model polymer-nanoparticle composite [272]
Soldera et al., Polymer **45**, 1307 (2004); Soldera et al., Composites A **36**, 521 (2005)	M	—	PMMA	MD modeling of chain behavior for different PMMA tacticites [417]
(*) Sreekumar et al., Adv. Mater. **16**, 58 (2004)	E	SWNTs	PAN	Mechanical characterization of solution spun fibers (~60 μm diameter) [253]
Srivastava et al., J. Phys. Chem. B. **103**, 4330 (1999)	M	SWNT	—	Kinky chemistry: effects of conformational strain on the chemical reactivity [334]
(*) Srivastava et al. Appl. Mech. Rev. **56**, 215 (2003)	*R*	*CNTs*	*Various*	*Review of computational nanomechanics for NTs and* NRPs [29]

continued

Table A.1. Continued.

(*) Starr et al., Phys. Rev. E **64**, 021802 (2001); Starr et al., Macromolecules **35**, 4481 (2002)	M	Nanofiller	Model polymer	MD simulation of change in polymer dynamics close to surface causing change in T_g [13, 14]
Starr et al., J. Chem. Phys. **119**, 1777 (2003)	M	Equiaxed nanoparticles	Model polymer	MD simulation suggests the nanoparticle-matrix interaction strength controlled clustering, with higher interaction leading to dispersed states [273]
Stéphan et al., Synth. Met. **108**, 139 (2000)	E	SWNTs	PMMA	Intercalation of polymer within SWNT bundles in spin-coated samples [174]
(*) Sun et al., Acc. Chem. Res. **35**, 1096 (2002)	R	CNTs	–	Review of functionalization of CNTs [418]
Tang et al., Carbon **41**, 2779 (2003)	E	MWNTs	HDPE	Melt processing and mechanical properties [419]
(*) Tang et al., Nature Mater. **2**, 413 (2003)	E	Nanoclays	PEs	Nanostructured artificial nacre [105]
Tate et al., J. Phys. Chem. **115**, 9982 (2001)	E	—	Thin polymer films	Elevation of T_g of thin grafted polymer films [420]
Theodorou et al., Macromolecules **19**, 139 (1986)	M	—	Model polymer	Atomistic modeling of mechanical properties of polymer glasses [261]
Thompson et al., Macromolecules **35**, 1060 (2002)	M	Nanoparticles	Diblock copolymers	Model of the order in diblock copolymer-nanoparticle nanocomposites [294]
(*) Thostenson et al., Compos. Sci. Technol. **61**, 1899 (2001)	R	CNT	Various	Review of CNTs and their composites [141]
Thostenson et al., J. Appl. Phys., **91**, 6034 (2002)	E	CVD CNTs grown on CFs	Epoxy	Hybrid multiscale composites [421]
Thostenson et al., J. Phys. D **35**, 1.77 (2002)	E	MWNTs	PS	MWNT alignment via extrusion/drawing [160]
Thostenson et al., J. Phys. D **36**, 573 (2003)	M	MWNTs	PS	Micromechanics modeling: effective NT modulus based on outer-shell load transfer [38]
Thostenson et al., Compos. Sci. Technol. **65**, 491 (2005)	R	Various	Various	Nanocomposites in context [144]
Torres et al., Phys. Rev. Lett. **85**, 3221 (2000)	M	—	Model polymer	Change in T_g and diffusion coefficient due to substrate interactions [422]
Tsagaropoulos et al., Macromolecules **28**, 6067 (1995); Tsagaropoulos et al., Macromolecules **28**, 396 (1995)	E	Nanosilica	Various	Mechanical behavior characterized by two T_g's [123]
Tschop et al., Acta Polym. **49**, 61 (1998); Tschop et al., Acta Polym. **49**, 75 (1998)	M	—	PC	Coarse-grained and atomistic modeling of polymer melts [423]
Tu et al., Phys. Rev. B **65**, 233407 (2002)	M	SWNTs, MWNTs	—	Continuum limit from a local density approximation model [424]
Tucker et al., Compos. Sci. Technol. **59**, 655 (1999)	R	—	Model polymer	Review of micromechanical models for unidirectional short-fiber composites [316]
Uribe-Arocha et al., Polymer **44**, 2441 (2003)	E	Nanoclays	PA6	Differences in skin-core properties for injection molded nanocomposite samples [228]
Usuki et al., Nano Lett. **1**, 271 (2001)	E	Nanoclay	Nylon 6	Plasma treatment for 3D observation of nanocomposite cross section [425]
(*) Vaia et al., J. Polym. Sci. Part B **35**, 59 (1997)	E	Nanoclays	PEO	Relaxations of confined intercalated chains in polymer nanocomposites [244]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
Vaia et al., Polymer **42**, 1281 (2001)	R	Nanoclays	Various	Review article [426]
Valavala et al., Rev. Adv. Mater. Sci. **9**, 34 (2005)	R	Various	Various	Review of computational modeling approaches for polymer nanocomposites [259]
Valentini et al., J. Appl. Polym. Sci. **89**, 2657 (2003)	E	SWNTs	PP and ethylene-propylene-diene rubber (EPDM)	Polymer intercalation into SWNT bundles if low NT loading [198]
Valentini et al., Compos. Sci. Technol. **64**, 23 (2004)	E	SWNTs	Epoxy	Dielectric response; impact of SWNTs on curing [427]
van Dommelen et al., J. Mech. Phys. Solids **51**, 519 (2003); van Dommelen et al., Polymer **44**, 6089 (2003)	M	—	Semicrystalline polymers	Micromechanical modeling of semicrystalline polymers [428]
van Workum et al., Nano Lett. **3**, 1405 (2003)	M	—	Amorphous polymers	MC/MD simulation of mechanical properties of amorphous polymer nanostructures [429]
van Zanten et al., Phys. Rev. E **53**, 2053 (1996)	E	—	P2VP	T_g increases as film thickness decreases if strong polar attraction [430]
Varghese et al., Polymer **44**, 3977 (2004)	E	Nanoclays	Epoxidized natural rubber	Intercalation/exfoliation as related to mechanical properties [431]
Velasco-Santos et al., Chem. Mater. **15**, 4470 (2003)	E	MWNTs	PMMA	Study of functionalization on mechanical properties [183]
Velasco-Santos et al., J. Phys. D **36**, 1423 (2003)	E	MWNTs	Methyl-ethyl methacrylate co-polymer	Thermomechanical properties as a function of processing aid [432]
Vigolo et al., Science **290**, 1331 (2000)	E	SWNTs	PVA	Macroscopic fibers and ribbons of oriented CNTs [433]
(*) Wagner et al., Appl. Phys. Lett. **72**, 188 (1998)	E	MWNTs	Urethane acrylate	Stress-transfer efficiency order of magnitude larger than traditional fiber composites [176]
Wagner, Chem. Phys. Lett. **361**, 57 (2002)	M	SWNTs	Model polymer	Model of interfacial shear stress for NRPs [336]
Wang et al., Compos. Sci. Technol. **64**, 935 (2004); Wang et al., Compos. Sci. Technol. **64**, 925 (2004)	M	Nanoclays	Various	Micromechanical model for modulus, comparison with experimental data [434]
Wang et al., Compos. Sci. Technol. **65**, 793 (2005)	E	MWNTs	UHMWPE	CNTs result in a more regular alignment of the UHMWPE morphology [435]
(*) Wei et al., Nano Lett. **2** 647 (2002)	M	SWNTs	PE	MD simulation of T_g and thermal expansion/diffusion coefficients in NRPs [254]
Wei et al., Nano Lett. **4**, 1949 (2004)	M	SWNTs	PE	Orientation of local polymer aligned along the CNT [255]
Weng, Int. J. Eng. Sci. **28**, 1111 (1990)	M	—	—	Discussion of the Mori-Tanaka method [311]
Wetzel et al., Compos. Sci. Technol. **63**, 2055 (2003)	E	Ceramic nanoparticles	Epoxy	Tribological performance of nanocomposite [137]
(*) Wong et al., Polymer **44**, 7757 (2003)	E	MWNTs	PS, PF	Interfacial shear stress in NRPs [278]
Wood et al., Phys. Rev. B **62**, 7571 (2000); Wood et al., Composites A, **32**, 391 (2001); Zhao et al., Appl. Phys. Lett. **78**, 1748 (2001); Zhao et al., Composites A **34**, 1219 (2003)	E	SWNTs	Urethane acrylate	SWNTs as nanoscale strain sensors via micro-Raman spectroscopy [154, 178, 436]
Wu et al., Compos. Sci. Technol. **64**, 1345 (2004); Zhong et al., Compos. Sci. Technol. **64**, 1353 (2004).	M	Nanoparticle	Model polymer	Micromechanical bounds for particle composites with inhomogeneous interphase [437]

continued

Table A.1. Continued.

Xia et al., Phys. Rev. B **69**, 233408 (2004)	M	MWNTs	—	MD simulation of pullout/friction within MWNTs [438]
Xiao et al., Polymer **42**, 4813 (2001); Xiao et al., Polymer **43**, 2245 (2002)	E	Expanded graphite	PS	Molecular weight increases/distribution widens, two T_g transitions appear [248]
Xie et al., J. Phys. Chem. Solids **61**, 1153 (2000)	M	MWNTs	—	CVD modulus ~460 GPa [58]
Xin et al., Phys. Rev. B **62**, 13692 (2000)	M	SWNT	—	Continuum elasticity theory for SWNT modulus [63]
Xu et al., Mater. Phys. Mech., **4**, 148 (2001)	M	Graphene sheet		Comparison of MD, FEM, and closed-form elasticity solutions [439]
Xu et al., Nano Lett. **3**, 1135 (2003)	E	Bone-shaped CNTs	—	Bone-shaped CNTs for nanocomposites applications [20]
Yang et al., Polym. Degrad. Stab. **83**, 383 (2004)	E	SWNTs	ABS	SWNTs increase T_g but lower degradation temperature [250]
Yano et al., J. Polym. Sci. Pol. Chem. **31**, 2493 (19930	E	Nanoclays	Polyimide	Large decrease in permeability with addition of nanoclay [6]
Yao et al., Comput. Mater. Sci. **22**, 180 (2001)	M	SWNTs	—	MD simulation of SWNTs [266]
Yasmin et al., Scripta Mater. **49**, 81 (2003)	E	Nanoclay	Epoxy	Three-roll mill used to disperse/exfoliate nanoclays [220]
Yasmin et al., Polymer **45**, 8211 (2004)	E	Nanographite	Epoxy	Mechanical properties of nanographite-reinforced epoxy [233]
Yatsuyanagi et al., Polymer **42**, 9253 (2001)	E	Nanosilica	SBR	Formation of secondary silica structure in nanocomposite [120]
Ye et al., Appl. Phys. Lett. **85**, 1775 (2004)	E	SWNTs, MWNTs	PAN	Two-stage rupture of electrospun NFs under tension: polymer crazing and NT pullout [87]
Yoon et al., Polymer **44**, 5323 (2003); Yoon et al., Polymer **44**, 5341 (2003)	E	Nanoclays	PC	Melt-mixing processing-structure-properties study [227, 440]
Yu et al., J. Appl. Polym. Sci. **77**, 2404 (2000)	E	Nanocopper	Polyoxymethylene	Investigation of wear and friction behavior of micron- and nanoscale copper-reinforced polymer composites [135]
Yu et al., Phys. Rev. Lett., 84, 5552 (2000); Yu et al., Science **287**, 637 (2000); Yu et al., J. Phys. Chem. B **104**, 8764 (2000)	E	SWNT ropes, MWNTs	—	Direct tensile loading via nanomanipulator within an SEM [42, 56, 57]
Zanetti et al., Macromol. Mater. Eng. **279**, 1 (2000)	R	Nanoclays	Various	Review article [98]
Zeng et al., Composites B **35**, 245 (2004)	E	Pyrograf-III	PMMA	Melt-spinning of ~60 μm composite fibers with two different diameter VGCFs [441]
Zhang et al., Composites A **35** 1385 (2004)	E	Nano-TiO_2	Epoxy	Wear resistance enhancements of nano-TiO_2 due to a mechanism of nanoscale rolling [138]
Zhang et al., J. Appl. Mech. **69**, 454 (2002); Zhang et al., Int. J. Solids Struct. **39**, 3893 (2002); Zhang et al., J. Mech. Phys. Solids **52**, 977 (2004)	M	SWNTs		Nanoscale continuum theory incorporating interatomic potentials [442]
(*) Zhang et al., Polymer **45**, 3481 (2004).	E	Nano-TiO_2	PA6.6	Significant improvement of creep resistance [130]
Zhang et al., Nano Lett. **3**, 1285 (2003)	E	SWNTs	PVA	SWNTs treated with PVP and SDS to aid dispersion [150]
Zhang et al., J. Chem. Phys., **121**, 10814 (2004)	M	Spherical nanoparticles	Model polymer	Monte Carlo simulation of polymer-particle interaction [280]

continued

Table A.1. Continued.

Citation	Type (E/M/R)	Inclusion	Polymer	Notes
Zhao et al., Phys. Rev. B **65**, 144105 (2002)	M	SWNTs	—	MD and *ab initio* calculation of strain-induced defect formation in SWNTs [335]
Zheng et al., Polymer **73**, 6767 (2002); Zheng et al., Compos. Sci. Technol. **63**, 225 (2003)	E	Expanded graphite	PMMA	Electrical percolation at much lower loadings; shift in T_g for expanded graphite [235, 249]
Zheng et al., J. Appl. Polym. Sci. **91**, 2781 (2004)	E	Expanded graphite	HDPE	Electrical and mechanical properties of nanographite-HDPE nanocomposites [236]
Zhong et al., J. Rheol. **47**, 483 (2003)	E	Nanoclay	—	Rheological behavior of exfoliated dispersions of nanoclay in solvent [443]
(*) Zhou et al., Compos. Sci. Technol. **64** 2425 (2004)	M	SWNTs	Epoxy	Micromechanics model of stick-slip interface and impact on damping [444]
Zhu et al., Nano Lett. **3**, 1107 (2003)	E	SWNTs	Epoxy	Acid treatment and fluorination to improve dispersion [151]
Zhu et al., Compos. Sci. Technol. **63**, 1113 (2003)	E	Nanosilica	PVA	Impact of surface functionality on nonlinear viscoelastic behavior [125]
Zou et al., Carbon **42**, 271 (2004)	E	MWNTs	HDPE	Screw extrusion at high speed to promote shear thinning leading to better dispersion [445]

Given that the average stress in each phase is given as $\bar{\sigma}_r = C_r \bar{\varepsilon}_r$, using the equivalent inclusion method one can show that the average stress in the rth inclusion can be written in terms of the matrix stiffness,

$$\bar{\sigma}_r = C_r \bar{\varepsilon}_r = C_0(\bar{\varepsilon}_r - \varepsilon_r^*) \tag{A4}$$

where ε_r^* is the fictitious eigenstrain of the rth inclusion. For a single ellipsoidal inclusion in an infinite matrix, Eshelby showed that the eigenstrain and perturbed strain of the rth inclusion can be related via

$$\varepsilon_r^{\text{pl}} = S_r \varepsilon_r^* \tag{A5}$$

where S_r is the Eshelby tensor. General forms of the Eshelby tensor are provided in Ref. [446].

Solving for ε_r^* in Eq. (A4) and then using Eqs. (A3) and (A5), one can find the dilute strain-concentration factor of the rth phase, A_r^{dil}, which relates the average strain in the rth inclusion to the average strain in the matrix, such that

$$\bar{\varepsilon}_r = A_r^{\text{dil}} \bar{\varepsilon}_0 \tag{A6}$$

where

$$A_r^{\text{dil}} = [\mathbf{I} + S_r C_0^{-1}(C_r - C_0)]^{-1} \tag{A7}$$

and $\mathbf{I}$ is the fourth-order identity tensor. We further require that the volume-weighted average phase strains must equal the far-field applied strain, such that

$$f_0 \bar{\varepsilon}_0 + \sum_{r=1}^{N-1} f_r \bar{\varepsilon}_r = \varepsilon_a \tag{A8}$$

Given Eq. (A8) and using Eq. (A6), we can now define the strain-concentration factor A_0, which accounts for inclusion interaction by relating the average matrix strain in the composite to the uniform applied strain,

$$\bar{\varepsilon}_0 = A_0 \varepsilon_a \tag{A9}$$

where

$$A_0 = [f_0 I + \sum_{r=1}^{N-1} f_r A_r^{\text{dil}}]^{-1} \tag{A10}$$

Thus in the non-dilute composite, the average strain in the rth inclusion can be related to the applied strain such that

$$\bar{\varepsilon}_r = A_r^{\text{dil}} \varepsilon_a \tag{A11}$$

where from Eqs. (A6) and (A9) the strain-concentration factor A_r for the rth inclusion phase in the *non-dilute composite* is

$$A_r = A_r^{\text{dil}} A_0 \tag{A12}$$

To find the effective stiffness C for a unidirectionally aligned composite, we require that the average stress $\bar{\sigma}$ of the composite be equal to the sum of the weighted average stresses in each phase,

$$\bar{\sigma} = f_0 \bar{\sigma}_0 + \sum_{r=1}^{N-1} f_r \bar{\sigma}_r = C \varepsilon_a \tag{A13}$$

Through straightforward substitution and manipulation, the effective stiffness of the unidirectionally aligned composite is found to be

$$C = f_0 C_0 A_0 + \sum_{r=1}^{N-1} f_r C_r A_r \tag{A14}$$

Recalling Eqs. (A10) and (A12), we can also express Eq. (A14) in a slightly different form (see Ref. [311]),

$$C = \left(f_0 C_0 + \sum_{r=1}^{N-1} f_r C_r A_r^{\text{dil}}\right)\left(f_0 I + \sum_{r=1}^{N-1} f_r A_r^{\text{dil}}\right)^{-1} \tag{A15}$$

Weng further simplifies this expression using the relationship that $A_0^{\text{dil}} = I$. Such an expression is sensible given that by definition A_0 relates the average strain in the matrix to the uniform applied strain. In the dilute sense these strains will be equal [311].

Both (A14) and (A15) thus provide the effective stiffness of a multiphase composite with aligned inclusions.

When the inclusion phases are randomly orientated in the matrix, determination of the effective composite stiffness can be accomplished by taking the orientational averages of appropriate quantities [311]. In this case the strain consistency condition in Eq. (A8) can be written as

$$\left(f_0 I + \sum_{r=1}^{N-1} f_r \{A_r^{\text{dil}}\}\right)\bar{\varepsilon}_0 = \bar{\varepsilon}_a \tag{A16}$$

where brackets $\{\}$ represent the average of a quantity over all possible orientations. (Note that the derivation for unidirectional inclusions presented above is a subset of the more general derivation presented here). Because of this averaging process, the average strain in the matrix will be different from that in the unidirectional composite due to the random alignment of inclusions.

Similarly one can rewrite Eq. (A13) for the case of randomly orientated inclusions, with the understanding that the matrix stress and strain are orientation-independent due to the isotropy of the matrix (an appropriate assumption for polymer matrices) as

$$\bar{\sigma} = f_0 \bar{\sigma}_0 + \sum_{r=1}^{N-1} f_r \{\bar{\sigma}_r\} = C\bar{\varepsilon}_a \tag{A17}$$

This expression can be simplified using the relationships established earlier in this section, namely,

$$\begin{aligned} \bar{\sigma} &= f_0 C_0 \bar{\varepsilon}_0 + \sum_{r=1}^{N-1} f_r \{C_r \bar{\varepsilon}_r\} \\ &= f_0 C_0 \bar{\varepsilon}_0 + \sum_{r=1}^{N-1} f_r \{C_r A_r^{\text{dil}}\} \bar{\varepsilon}_0 \\ &= \left(f_0 C_0 + \sum_{r=1}^{N-1} f_r \{C_r A_r^{\text{dil}}\}\right) \bar{\varepsilon}_0 = C\bar{\varepsilon}_a \end{aligned} \tag{A18}$$

From Eqs. (A16) and (A18), the effective stiffness of a composite with randomly orientated inclusions can be written as

$$C = \left(f_0 C_0 + \sum_{r=1}^{N-1} f_r \{C_r A_r^{\text{dil}}\}\right)\left(f_0 \mathbf{I} + \sum_{r=1}^{N-1} f_r \{A_r^{\text{dil}}\}\right)^{-1} \tag{A19}$$

which is the direct analog of Eq. (A15) except that appropriate averaging is used here to account for the inclusion orientation. While the above formulations are concerned with the effective properties of multiphase composite materials, analytical expressions have been derived for the Mori-Tanaka solution for two-phase composites with 2D and 3D randomly orientated inclusions [446].

For completeness, a brief description of the orientational averaging of a fourth-order tensor is included here. A more complete derivation can be found elsewhere [140]. The transformation of a fourth-order tensor B_{ijkl} from local to global coordinates can be

written as

$$B_{ijkl}(\phi_1, \Phi) = a_{ir}a_{js}a_{kt}a_{lu}B'_{rstu} \tag{A20}$$

where the angular dependence of the a_{ij} terms representing the tranformation of a vector from the local to global coordinate system is implied and the standard convention that double indices implies summation is used. Thus the orientationally-averaged fourth-order tensor $\bar{B}_{ijkl}$ for a fourth-order tensor B_{ijkl} in 3D and 2D space can be written, respectively, as

$$\bar{B}_{ijkl} = \{B_{ijkl}\} = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{2\pi} B_{ijkl}(\phi_1, \Phi)\sin(\Phi)\, d\Phi\, d\phi_1 \tag{A21}$$

$$\bar{B}_{ijkl} = \{B_{ijkl}\} = \frac{1}{\pi}\int_0^{\pi} B_{ijkl}(\phi_1)\, d\phi_1 \tag{A22}$$

If standard contracted notation is used (see Ref. [271]), the resulting tensor component transformations can be expressed in matrix form. For a 3D random orientation of inclusions,

$$\begin{bmatrix} \{\bar{B}_{11}\} \\ \{\bar{B}_{22}\} \\ \{\bar{B}_{33}\} \\ \{\bar{B}_{12}\} \\ \{\bar{B}_{21}\} \\ \{\bar{B}_{13}\} \\ \{\bar{B}_{31}\} \\ \{\bar{B}_{23}\} \\ \{\bar{B}_{32}\} \\ \{\bar{B}_{44}\} \\ \{\bar{B}_{55}\} \\ \{\bar{B}_{66}\} \end{bmatrix} = \frac{1}{120}\begin{bmatrix} 9 & 45 & 24 & 5 & 5 & 6 & 6 & 10 & 10 & 40 & 24 & 20 \\ 9 & 45 & 24 & 5 & 5 & 6 & 6 & 10 & 10 & 40 & 24 & 20 \\ 64 & 0 & 24 & 0 & 0 & 16 & 16 & 0 & 0 & 0 & 64 & 0 \\ 3 & 15 & 8 & 15 & 15 & 2 & 2 & 30 & 30 & -40 & 8 & -20 \\ 3 & 15 & 8 & 15 & 15 & 2 & 2 & 30 & 30 & -40 & 8 & -20 \\ 8 & 0 & 8 & 0 & 40 & 12 & 32 & 20 & 0 & 0 & -32 & 0 \\ 8 & 0 & 8 & 40 & 0 & 32 & 12 & 0 & 20 & 0 & -32 & 0 \\ 8 & 0 & 8 & 0 & 40 & 12 & 32 & 20 & 0 & 0 & -32 & 0 \\ 8 & 0 & 8 & 40 & 0 & 32 & 12 & 0 & 20 & 0 & -32 & 0 \\ 8 & 0 & 8 & 0 & 0 & -8 & -8 & 0 & 0 & 20 & 28 & 40 \\ 8 & 0 & 8 & 0 & 0 & -8 & -8 & 0 & 0 & 20 & 28 & 40 \\ 3 & 15 & 8 & -5 & -5 & 2 & 2 & -10 & -10 & 40 & 8 & 20 \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{22} \\ B_{33} \\ B_{12} \\ B_{21} \\ B_{13} \\ B_{31} \\ B_{23} \\ B_{32} \\ B_{44} \\ B_{55} \\ B_{66} \end{bmatrix} \tag{A23}$$

Note that the 3D randomly oriented fourth-order tensor will be isotropic. For a 2D random orientation of inclusions, with the global 3-direction representing the out-of-plane direction, the contracted stiffness components can be written as

$$\begin{aligned} \{B_{11}\} &= \{B_{22}\} = \frac{1}{8}[3B_{22} + 3B_{33} + B_{23} + B_{32} + 4B_{44}] \\ \{B_{33}\} &= B_{11} \\ \{B_{12}\} &= \{B_{21}\} = \frac{1}{8}[B_{22} + B_{33} + 3B_{23} + 3B_{32} - 4B_{44}] \\ \{B_{13}\} &= \{B_{23}\} = \frac{1}{2}[B_{21} + B_{31}] \\ \{B_{31}\} &= \{B_{32}\} = \frac{1}{2}[B_{12} + B_{13}] \\ \{B_{44}\} &= \{B_{55}\} = \frac{1}{2}[B_{55} + B_{66}] \\ \{B_{66}\} &= \frac{1}{8}[B_{22} + B_{33} - B_{23} - B_{32} + 4B_{44}] \end{aligned} \tag{A24}$$

Other known orientational distributions of inclusions can be accounted for by using the appropriate orientational distribution function; see, for example, Ref. [447].

A.2. Summary of Related Literature

In Table A.1, we have assembled a large number of references related to the work and directions described within this chapter. We have attempted to classify each citation by its primary focus (experimental, modeling, or review) and the nanocomposite constitutive phases studied, as well as provide a brief summary or highlight of the work. While these brief summaries cannot possibly capture the essence of each reference, and such a compiliation cannot possibly include every reference applicable in this large and expanding field, we hope that such a collection will be useful to readers interested in pursuing additional information within the literature. Citations marked (*) represent in our opinion particularly useful references.

REFERENCES

1. Committee on Durability and Life Prediction for Polymer Matrix Composites in Extreme Environments, "Going to Extremes: Meeting the Emerging Demand for Durable Polymer Matrix Composites." National Research Council, Washington, D.C., 2005.
2. J. B. Donnet, R. C. Bansal, and M. J. Wang, (Eds.), "Carbon Black: Science and Technology," 2nd Edn. Marcel Dekker, New York, 1993.
3. Y. Kojima, A. Usuki, M. Kawasumi, A. Okada, T. Kurauchi, and O. Kamigaito, *J. Polym. Sci.* 31, 983 (1993).
4. GM press release, Warren, MI (2004).
5. S. Iijima, *Nature (London)* 354, 56 (1991).
6. K. Yano, A. Usuki, A. Okada, T. Kurauchi, and O. Kamigaito, *J. Polym. Sci.* 31, 2493 (1993).
7. P. B. Messersmith and E. P. Giannelis, *J. Polym. Sci.* 33, 1047 (1995).
8. J. L. Thomason, *Composites* 26, 487 (1995).
9. K. Mai, E. Mader, and M. Muhle, *Composites A* 29, 1111 (1998).
10. C. J. Ellison, R. T. Ruszkowski, N. J. Fredin, and J. M. Torkelson, *Phys. Rev. Lett.* 92, 095702 (2004); C. J. Ellison and J. M. Torkelson, *Nat. Mater.* 2, 695 (2003).
11. J. A. Forrest and R. A. L. Jones, in "Polymer Surfaces Interfaces and Thin Films" (A. Karim and S. Kumar, Eds.). World Scientific Publishing, Singapore, 2000.
12. E. P. Giannelis, R. Krishnamoorti, and E. Manias, *Adv. Polym. Sci.* 138, 107 (1999).
13. F. Starr, T. Schroder, and S. Glotzer, *Phys. Rev. E* 64, 021802 (2001).
14. F. W. Starr, T. B. Schroder, and S. C. Glotzer, *Macromolecules* 35, 4481 (2002).
15. G. D. Smith, D. Bedrov, L. Li, and O. Byutner, *J. Chem. Phys.* 117, 9478 (2002).
16. F. T. Fisher and L. C. Brinson, in "Society for Experimental Mechanics Conference Proceedings," Milwaukee, 2002.
17. K. Wise and J. Hinkley, "American Physical Society Spring Meeting," Seattle, 2001.
18. http://www.nanogram.com/technology/nm.htm.
19. L. S. Schadler, in "Nanocomposite Science and Technology" (P. M. Ajayan, L. S. Schadler, and P. V. Braun, Eds.). Wiley-VCH, Weinheim, Great Britain, 2003.
20. T. T. Xu, F. T. Fisher, L. C. Brinson, and R. S. Ruoff, *Nano Lett.* 3, 1135 (2003).
21. X. Q. Chen, S. L. Zhang, D. A. Dikin, W. Q. Ding, R. S. Ruoff, L. J. Pan, and Y. Nakayama, *Nano Lett.* 3, 1299 (2003).
22. B. G. Demczyk, J. Cumings, A. Zettl, and R. O. Ritchie, *Appl. Phys. Lett.* 78, 2772 (2001); E. Bengu and L. D. Marks, *Phys. Rev. Lett.* 86, 2385 (2001); D. Golberg, Y. Bando, L. Bourgeois, K. Kurashima, and T. Sato, *Appl. Phys. Lett.* 77, 1979 (2000); D. Srivastava, M. Menon, and K. Cho, *Phys. Rev. B* 63, 195413 (2001); K. Shelimov and M. Moskovits, *Chem. Mater.* 12, 250 (2000).
23. Y. Wang and J. J. Santiago-Aviles, *Nanotechnology* 15, 32 (2004).
24. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic, New York, 1996.
25. P. J. F. Harris, "Carbon Nanotubes and Related Structures: New Materials for the 21st Century." Cambridge University Press, Cambridge, 1999.
26. P. G. Collins and P. Avouris, *Sci. Am.* 62 (2000).
27. L. Forro and C. Schonenberger, *Topics Appl. Phys.* 80, 329 (2001); R. Saito, M. S. Dresselhaus, and G. Dresselhaus, "Physical Properties of Carbon Nanotubes," Imperial College Press, London, 1998.
28. B. I. Yakobson and P. Avouris, *Topics Appl. Phys.* 80, 287 (2001).
29. D. Srivastava, C. Wei, and K. Cho, *Appl. Mech. Rev.* 56, 215 (2003).
30. D. Qian, G. J. Wagner, W. K. Liu, M.-F. Yu, and R. S. Ruoff, *Appl. Mech. Rev.* 55, 495 (2002).
31. J. Bernholc, D. Brenner, M. B. Nardelli, V. Meunier, and C. Roland, *Annu. Rev. Mater. Res.* 32, 347 (2002).
32. P. M. Ajayan and O. Z. Zhou, *Topics Appl. Phys.* 80, 391 (2001).
33. M. Dresselhaus, G. Dresselhaus, P. Eklund, and R. Saito, *Phys. World* 11, 33 (1998).
34. L. C. Venema, J. W. G. Wildöer, C. Dekker, G. A. Rinzler, and R. E. Smalley, *Appl. Phys. A* 66, 153 (1998).
35. R. C. Haddon, J. Sippel, A. G. Rinzler, and F. Papadimitrakopoulos, *MRS Bull.* 29, 252 (2004); M. Zheng, A. Jagota, E. D. Semke, B. A. Diner, R. S. McLean, S. R. Lustig, R. E. Richardson, and N. G. Tassi, *Nat. Mater.* 2, 338 (2003).
36. D. Qian, E. C. Dickey, R. Andrews, and T. Rantell, *Appl. Phys. Lett.* 76, 2868 (2000).
37. Z. W. Pan, S. S. Xie, L. Lu, B. H. Chang, L. F. Sun, W. Y. Zhou, G. Wang, and D. L. Zhang, *Appl. Phys. Lett.* 74, 3152 (1999).

38. E. T. Thostenson and T. W. Chou, *J. Phys. D* 36, 573 (2003).
39. L. S. Schadler, S. C. Giannaris, and P. M. Ajayan, *Appl. Phys. Lett.* 73, 3842 (1998).
40. A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tománek, J. E. Fischer, and R. E. Smalley, *Science* 273, 483 (1996).
41. M. J. López, A. Rubio, J. A. Alonso, L.-C. Qin, and S. Iijima, *Phys. Rev. Lett.* 86, 3056 (2001).
42. M.-F. Yu, B. I. Yakobson, and R. S. Ruoff, *J. Phys. Chem. B* 104, 8764 (2000).
43. T. W. Ebbesen and P. M. Ajayan, *Nature* 358, 220 (1992).
44. C. Journet, W. K. Maser, P. Bernier, A. Loiseau, M. L. de la Chapelle, S. Lefrants, P. Deniard, R. Lee, and J. E. Fischer, *Nature* 388, 756 (1997).
45. M. J. Bronikowski, P. A. Willis, D. T. Colbert, K. A. Smith, and R. E. Smalley, *J. Vac. Sci. Tech. A* 19, 1800 (2001).
46. A. M. Cassell, J. A. Raymakers, J. Kong, and H. Dai, *J. Phys. Chem. B* 103, 6484 (1999).
47. G. Che, B. B. Lakshmi, C. R. Martin, and E. R. Fisher, *Chem. Mater.* 10, 260 (1998).
48. M. Endo, *Chemtech.* 18, 568 (1988).
49. J.-P. Salvetat, A. J. Kulik, J.-M. Bonard, G. D. A. Briggs, T. Stöckli, K. Métérier, S. Bonnamy, F. Béguin, N. A. Burnham, and L. Forró, *Adv. Mater.* 11, 161 (1999).
50. G. Gao, T. Çagin, and W. A. Goddard III, *Nanotechnology* 9, 184 (1998); E. Hernández, C. Goze, P. Bernier, and A. Rubio, *Phys. Rev. Lett.* 80, 4502 (1998).
51. J. Che, T. Çagin, and W. Goddard III, *Theor. Chem. Acc.* 102 (1999).
52. S. Sánchez-Portal, E. Artacho, J. M. Soler, A. Rubio, and P. Ordejón, *Phys. Rev. B* 59, 12678 (1999).
53. T. Belytschko, S. P. Xiao, G. C. Schatz, and R. Ruoff, *Phys. Rev. B* 65, 235430 (2002).
54. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996); A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14103 (1998); J.-P. Salvetat, J.-M. Bonard, N. H. Thomson, A. J. Kulik, L. Forró, W. Benoit, and L. Zuppiroli, *Appl. Phys. A* 69, 255 (1999); F. Li, M. Cheng, S. Bai, G. Su, and M. S. Dresselhaus, *Appl. Phys. Lett.* 77, 3161 (2000).
55. M. R. Falvo, G. J. Clary, R. M. Taylor II, V. Chi, F. P. Brooks, Jr., S. Washburn, and R. Superfine, *Nature* 389, 582 (1997); J.-P. Salvetat, G. Briggs, J.-M. Bonard, R. Basca, A. Kulik, T. Stöckli, N. Burnham, and L. Forró, *Phys. Rev. Lett.* 82, 944 (1999); M.-F. Yu, T. Kowalewski, and R. S. Ruoff, *Phys. Rev. Lett.* 85, 1456 (2000).
56. M.-F. Yu, B. S. Files, S. Arepalli, and R. S. Ruoff, *Phys. Rev. Lett.* 84, 5552 (2000).
57. M.-F. Yu, O. Lourie, M. Dyer, K. Moloni, T. F. Kelly, and R. S. Ruoff, *Science* 287, 637 (2000).
58. S. Xie, W. Li, Z. Pan, B. Chang, and L. Sun, *J. Phys. Chem. Solids* 61, 1153 (2000).
59. E. T. Fisher, D. A. Dikin, X. Chen, and R. S. Ruoff, in "Applied Physics of Carbon Nanotubes: Fundamentals of Theory, Optics and Transport Devices" (S. V. Rotkin and S. Subramoney, Eds.). Springer Series in Nanoscience and Technology, 2005.
60. E. W. Wong, P. E. Sheehan, and C. M. Lieber, *Science* 277, 1971 (1997).
61. S. Govindjee and J. Sackman, *Solid State Commun.* 110, 227 (1999).
62. C. Ru, *Phys. Rev. B* 62, 10405 (2000); C. Q. Ru, *J. Appl. Phys.* 89, 3426 (2001); C. Q. Ru, *Phys. Rev. B* 62, 9973 (2000); C. Q. Ru, *Phys. Rev. B* 62, 16962 (2000).
63. Z. Xin, Z. Jianjun, and O.-Y. Zhong-can, *Phys. Rev. B* 62, 13692 (2000).
64. B. I. Yakobson, M. P. Campbell, C. J. Brabec, and J. Bernholc, *Comput. Mater. Sci.* 8, 341 (1997).
65. L. Shen and J. Li, *Phys. Rev. B* 69, 045414 (2004).
66. L. Shen and J. Li, *Phys. Rev. B* 71, 035412 (2005).
67. E. Saether, S. J. V. Frankland, and R. B. Pipes, *Compos. Sci. Technol.* 63, 1543 (2003).
68. R. B. Pipes, S. J. V. Frankland, P. Hubert, and E. Saether, *Compos. Sci. Technol.* 63, 1349 (2003).
69. L. J. Matienzo, D. W. Wang, and F. Emmi, in "Characterization of Polymers" (N. J. Chou, S. P. Kowalczyk, R. Saraf, and H.-M. Tong, Eds.). Butterworth-Heinnmann, Boston, 1994.
70. C. R. Martin, *Science* 266, 1961 (1994); D. Routkevitch, A. A. Tager, J. Haruyama, D. Almawlawi, M. Moskovits, and J. M. Xu, *IEEE Trans. Electron Devices* 43, 1646 (1996).
71. N. Wang, Z. K. Tang, G. D. Li, and J. S. Chen, *Nature (London)* 408, 50 (2000).
72. J. Li, C. Papadopoulos, J. M. Xu, and M. Moskovits, *Appl. Phys. Lett.* 75, 367 (1999); Y. C. Sui, B. Z. Cui, R. Guardian, D. R. Acosta, L. Martinez, and R. Perez, *Carbon* 40, 1011 (2002).
73. T. Kyotani, L. F. Tsai, and A. Tomita, *Chem. Mater.* 8, 2109 (1996).
74. H. Masuda and K. Fukuda, *Science* 268, 1466 (1995).
75. J. S. Suh and J. S. Lee, *Appl. Phys. Lett.* 75, 2047 (1999); J. S. Lee, G. H. Gu, H. Kim, K. S. Jeong, J. Bae, and J. S. Suh, *Chem. Mater.* 13, 2387 (2001); Z. H. Yuan, H. Huang, L. Liu, and S. S. Fan, *Chem. Phys. Lett.* 345, 39 (2001).
76. G. Che, B. B. Lakshmi, C. R. Martin, E. R. Fisher, and R. S. Ruoff, *Chem. Mater.* 10, 260 (1998).
77. I. J. Beyerlein, Y. T. Zhu, and S. Mahesh, *Compos. Sci. Technol.* 61, 1341 (2001); Y. T. Zhu, J. A. Valdez, I. J. Beyerlein, S. J. Zhou, C. Liu, M. G. Stout, D. P. Butt, and T. C. Lowe, *Acta Mater.* 47, 1767 (1999); R. C. Wetherhold and F. K. Lee, *Compos. Sci. Technol.* 61, 517 (2001); X. Zhao, B. Zhou, and X. Tian, *Biominetics* 3, 1 (1995).
78. Y. T. Zhu and I. J. Beyerlein, *Mater. Sci. Eng. A* 326, 208 (2002).
79. V. K. Varadan and J. N. Xie, *Smart Mater. Struct.* 11, 728 (2002); C. Kuzuya, W. In-Hwang, S. Hirako, Y. Hishikawa, and S. Motojima, *Chem. Vapor Depos.* 8, 57 (2002); M. Zhang, Y. Nakayama, and L. J. Pan, *Jpn. J. Appl. Phys. Pt. 2* 39, L1242 (2000).
80. A. Volodin, D. Buntinx, M. Ahlskog, A. Fonseca, J. B. Nagy, and C. Van Haesendonch, *Nano Lett.* 4, 1775 (2004).

81. Y. Dzenis, *Science* 304, 1917 (2004).
82. D. Li and Y. Xia, *Adv. Mater.* 16, 1151 (2004).
83. H. Fong, W. K. Liu, C. S. Wang, and R. A. Vaia, *Polymer* 43, 775 (2002).
84. J. Mack, L. Viculis, A. Ali, R. Luoh, G. Yang, R. Kaner, T. Hahn, and F. Ko, in "American Society for Composites Conference Proceedings." Lafayette, Indiana, 2002.
85. J. J. Mack, L. M. Viculis, A. Ali, R. Luoh, G. Yang, T. Hanhn, F. K. Ko, and R. B. Kaner, *Adv. Mater.* 17, 77 (2005).
86. F. Ko, Y. Gogotsi, A. Ali, N. Naguib, H. Ye, G. Yang, C. Y. Li, and P. Willis, *Adv. Mater.* 15, 1161 (2003).
87. H. Ye, H. Lan, N. Titchenal, Y. Gogotsi, and F. Ko, *Appl. Phys. Lett.* 85, 1775 (2004).
88. Y. Dror, W. Salalha, R. L. Khalfin, Y. Cohen, E. Zussman, and A. Yarin, *Langmuir* 19, 7012 (2003).
89. D. Reneker and I. Chun, *Nanotechnology* 7, 216 (1996); H. Fong and D. H. Reneker, in "Structure Formation in Polymeric Fibers" (S. D. R and M. V. Sessman, Eds.). Hanser Publishers, Munich, 2001; Z. M. Huang, Y. Z. Zhang, M. Kotaki, and S. Ramakrishna, *Compos. Sci. Technol.* 63, 2226 (2003).
90. Y. Kojima, A. Usuki, M. Kawasumi, A. Okada, T. Kurauchi, O. Kamigaito, and K. Kaji, *J. Poly. Sci. B* 32, 625 (1994); Y. Kojima, A. Usuki, M. Kawasumi, A. Okada, T. Kurauchi, O. Kamigaito, and K. Kaji, *J. Poly. Sci. B* 33, 1039 (1995).
91. M. Biswas and S. S. Ray, *Adv. Polym. Sci.* 155, 167 (2001).
92. M. Alexandre and P. Dubois, *Mater. Sci. Eng. Rep.* 28, 1 (2000).
93. K. A. Carrado, *Appl. Clay Sci.* 17, 1 (2000).
94. P. C. LeBaron, Z. Wang, and T. J. Pinnavaia, *Appl. Clay Sci.* 15, 11 (1999).
95. D. Schmidt, D. Shah, and E. P. Giannelis, *Curr. Opin. Solid State Mater. Sci.* 6, 205 (2002).
96. R. A. Vaia and E. P. Giannelis, *MRS Bull.* 26, 394 (2001); T. J. Pinnavaia and G. Beall, (Eds.), "Polymer-clay Nanocomposites." Wiley, Chichester, 2000.
97. E. P. Giannelis, *Adv. Mater.* 8, 29 (1996).
98. M. Zanetti, S. Lomakin, and G. Camino, *Macromol. Mater. Eng.* 279, 1 (2000).
99. S. S. Ray and M. Okamoto, *Prog. Polym. Sci.* 28, 1539 (2003).
100. M. Kawasumi, *J. Polym. Sci. A* 42, 819 (2004).
101. L. E. McNeil and M. Grimsditch, *J. Phys.* 5, 1681 (1993).
102. R. D. Piner, T. T. Xu, F. T. Fisher, Y. Qiao, and R. S. Ruoff, *Langmuir* 19, 7995 (2003).
103. Y. L. Zhou, Z. Li, N. F. Hu, Y. H. Zeng, and J. F. Rusling, *Langmuir* 18, 8573 (2002).
104. X. Fan, M. Park, C. Xia, and R. Advincula, *J. Mater. Res.* 17, 1622 (2002).
105. Z. Tang, N. A. Kotov, S. Magonov, and B. Ozturk, *Nat. Mater.* 2, 413 (2003).
106. B. T. Kelly, "The Physics of Graphite." Applied Science, London, 1981.
107. D. D. L. Chung, *J. Mater. Sci.* 37, 1475 (2002).
108. A. Yasmin, J. J. Luo, and I. M. Daniel, in "Proceedings of the 11th Annual International Conference on Composites/Nano Engineering (ICCE-11)." Hilton Head Island, South Carolina, 2004.
109. D. Cho, S. Lee, G. Yang, H. Fukushima, and L. T. Drzal, *Macromol. Mater. Eng.* 290, 179 (2005).
110. J. J. Aklonis and W. J. MacKnight, "Introduction to Polymer Viscoelasticity," 2nd Edn. Wiley, New York, 1983.
111. J. D. Ferry, "Viscoelastic Properties of Polymers," 3rd Edn. Wiley, New York, 1980.
112. N. W. Tschogel, "The Phenomenological Theory of Linear Viscoelastic Behavior." Springer, Berlin, 1989.
113. H. F. Brinson and L. C. Brinson, "Polymer Engineering Science and Viscoelasticity," in preparation, 2006.
114. R. J. Young and P. A. Lovell, "Introduction to Polymers," 2nd Edn. Chapman & Hall, London, 1991.
115. G. B. McKenna, Y. Leterrier, and C. R. Schultheisz, *Polym. Eng. Sci.* 35, 403 (1995).
116. A. Eitan, F. T. Fisher, R. Andrews, L. C. Brinson, and L. S. Schadler, *Compos. Sci. Technol.* (to be published, 2006).
117. F. T. Fisher, A. Eitan, R. Andrews, L. S. Schadler, and L. C. Brinson, *Adv. Compos. Lett.* 13, 105 (2004).
118. Y. H. Hu, C. Y. Chen, and C. C. Wang, *Polym. Degrad. Stabil.* 84, 545 (2004).
119. C. Becker, P. Mueller, and H. K. Schmidt, *Proc. SPIE Int. Soc. Opt. Eng.* 3469, 88 (1998).
120. E. Yatsuyanagi, N. Suzuki, M. Ito, and H. Kaidou, *Polymer* 42, 9253 (2001).
121. V. Arrighi, I. J. McEwen, H. Qian, and M. B. Serrano Prieto, *Polymer* 44, 6259 (2003).
122. B. J. Ash, L. S. Schadler, and R. W. Siegel, *Mater. Lett.* 55, 83 (2002).
123. G. Tsagaropoulos and A. Eisenberg, *Macromolecules* 28, 396 (1995); G. Tsagaropoulos and A. Eisenberg, *Macromolecules* 28, 6067 (1995).
124. Y. Huang, S. Jiang, L. Wu, and Y. Hua, *Polym. Test.* 23, 9 (2004).
125. A. J. Zhu and S. S. Sternstein, *Compos. Sci. Technol.* 63, 1113 (2003).
126. S. S. Sternstein and A. J. Zhu, *Macromolecules* 35, 7262 (2002).
127. C. Gauthier, E. Reynaud, R. Vassoille, and L. Ladouce-Stelandre, *Polymer* 45, 2761 (2004).
128. R. Siegel, S. Chang, B. Ash, J. Stone, P. Ajayan, R. Doremus, and L. Schadler, *Scripta Mater.* 44, 2061 (2001).
129. A. J. Kinloch, J. H. Lee, A. C. Taylor, S. Sprenger, C. Eger, and D. Egan, *J. Adhes.* 79, 867 (2003).
130. Z. Zhang, J. L. Yang, and K. Friedrich, *Polymer* 45, 3481 (2004).
131. M. Z. Rong, M. Q. Zhang, Y. X. Zheng, H. M. Zeng, and K. Friedrich, *Polymer* 42, 3301 (2001).
132. Y. L. Liu, C. Y. Hsu, W. L. Wei, and R. J. Jeng, *Polymer* 44, 5159 (2003).
133. W. G. Sawyer, K. D. Freudenberg, P. Bhimaraj, and L. S. Schadler, *Wear* 254, 573 (2003).
134. P. Bhimaraj, D. L. Burris, J. Action, W. G. Sawyer, C. L. Toney, R. W. Siegel, and L. S. Schadler, *Wear* 258, 1437 (2005).
135. L. Yu, S. Yang, H. Wang, and Q. Xue, *J. Appl. Polym. Sci.* 77, 2404 (2000).
136. M. Avella, M. E. Errico, and E. Martuscelli, *Nano Lett.* 1, 213 (2001).

137. B. Wetzel, F. Haupert, and M. Q. Zhang, *Compos. Sci. Technol.* 63, 2055 (2003).
138. Z. Zhang, C. Breidt, L. Chang, F. Haupert, and K. Friedrich, *Composites A* 35, 1385 (2004).
139. R. Andrews, D. Jacques, M. Minot, and T. Rantell, *Macromol. Mater. Eng.* 287, 395 (2002).
140. F. T. Fisher, Ph.D. dissertation, Northwestern University, 2002.
141. E. T. Thostenson, Z. F. Ren, and T. W. Chou, *Compos. Sci. Technol.* 61, 1899 (2001).
142. K. T. Lau and D. Hui, *Composites B* 33, 263 (2002).
143. O. Breuer and U. Sundararaj, *Polym. Compos.* 25, 630 (2004).
144. E. T. Thostenson, C. Y. Li, and T. W. Chou, *Compos. Sci. Technol.* 65, 491 (2005).
145. X. Gong, J. Liu, S. Baskaran, R. D. Voise, and J. S. Young, *Chem. Mater.* 12, 1049 (2000).
146. K. T. Lau, M. Lu, C. K. Lam, H. Y. Cheung, F. L. Sheng, and H. L. Li, *Compos. Sci. Technol.* 65, 719 (2005).
147. R. Ramasubramaniam, J. Chen, and H. Y. Liu, *Appl. Phys. Lett.* 83, 2928 (2003).
148. D. L. Shi, J. Lian, P. He, L. M. Wang, W. J. van Ooij, M. J. Schulz, Y. Liu, and D. B. Mast, *Appl. Phys. Lett.* 81, 5216 (2002).
149. D. L. Shi, J. Lian, P. He, L. M. Wang, F. Xiao, L. Yang, M. J. Schulz, and D. B. Mast, *Appl. Phys. Lett.* 83, 5301 (2003).
150. X. Zhang, T. Liu, T. V. Sreekumar, S. Kumar, V. C. Moore, R. H. Hauge, and R. E. Smalley, *Nano Lett.* 3, 1285 (2003).
151. J. Zhu, J. D. Kim, H. Q. Peng, J. L. Margrave, V. N. Khabashesku, and E. V. Barrera, *Nano Lett.* 3, 1107 (2003).
152. F. M. Du, J. E. Fischer, and K. I. Winey, *J. Poly. Sci. B* 41, 3333 (2003).
153. P. Ajayan, O. Stephan, C. Colliex, and D. Trauth, *Science* 265, 1212 (1994).
154. J. R. Wood, Q. Zhao, and H. D. Wagner, *Composites A* 32, 391 (2001).
155. Y. H. Zhao and G. J. Weng, *Acta Mechan.* 116, 29 (1996).
156. L. Jin, C. Bower, and O. Zhou, *Appl. Phys. Lett.* 73, 1197 (1998).
157. C. Bower, R. Rosen, J. Lin, J. Han, and O. Zhou, *Appl. Phys. Lett.* 74, 3317 (1999).
158. R. Haggenmueller, H. H. Gommans, A. G. Rinzler, J. E. Fischer, and K. I. Winey, *Chem. Phys. Lett.* 330, 219 (2000).
159. B. Safadi, R. Andrews, and E. A. Grulke, *J. Appl. Polym. Sci.* 84, 2660 (2002).
160. E. T. Thostenson and T. W. Chou, *J. Phys. D: Appl. Phys.* 35, L77 (2002).
161. C. A. Cooper, D. Ravich, D. Lips, J. Mayer, and H. D. Wagner, *Compos. Sci. Technol.* 62, 1105 (2002).
162. X. Q. Chen, T. Saito, H. Yamada, and K. Matsushige, *Appl. Phys. Lett.* 78, 3714 (2001).
163. B. W. Smith, Z. Benes, D. E. Luzzi, J. E. Fischer, D. A. Walters, M. J. Casavant, J. Schmidt, and R. E. Smalley, *Appl. Phys. Lett.* 77, 663 (2000).
164. T. Prasse, J. Y. Cavaille, and W. Bauhofer, *Compos. Sci. Technol.* 63, 1835 (2003).
165. T. Kimura, H. Ago, M. Tobita, S. Ohshima, M. Kyotani, and M. Yumura, *Adv. Mater.* 14, 1380 (2002).
166. B. Wang, Z. Liang, R. Shankar, and C. Zhang, in "Society for Advancement of Material and Process Engineering (SAMPE) Conference Proceedings." Long Beach, California, 2004.
167. R. Haggenmueller, W. Zhou, J. E. Fischer, and K. I. Winey, *J. Nanosci. Nanotechnol.* 3, 105 (2003).
168. S. Badaire, V. Pichot, C. Zakri, P. Poulin, P. Launois, J. Vavro, C. Guthy, M. Chen, and J. E. Fischer, *J. Appl. Phys.* 96, 7509 (2004).
169. M. S. P. Shaffer and A. H. Windle, *Adv. Mater.* 11, 937 (1999).
170. Z. Jia, Z. Wang, C. Xu, J. Liang, B. Wei, D. Wu, and S. Zhu, *Mater. Sci. Eng. A* 271, 395 (1999).
171. P. M. Ajayan, L. S. Schadler, C. Giannaris, and A. Rubio, *Adv. Mater.* 12, 750 (2000).
172. K. Lozano and E. Barrera, *J. Appl. Polym. Sci.* 79, 125 (2001).
173. V. Lordi and N. Yao, *J. Mater. Res.* 15, 2770 (2000).
174. C. Stéphan, T. P. Nguyen, M. L. de la Chapelle, S. Lefrant, C. Journet, and P. Bernier, *Synth. Met.* 108, 139 (2000).
175. O. Lourie and H. D. Wagner, *Appl. Phys. Lett.* 73, 3527 (1998).
176. H. D. Wagner, O. Lourie, Y. Feldman, and R. Tenne, *Appl. Phys. Lett.* 72, 188 (1998).
177. K. Liao and S. Li, *Appl. Phys. Lett.* 79, 4225 (2001).
178. Q. Zhao, J. R. Wood, and H. D. Wagner, *Appl. Phys. Lett.* 78, 1748 (2001).
179. A. H. Barber, S. R. Cohen, S. Kenig, and H. D. Wagner, *Compos. Sci. Technol.* 64, 2283 (2004).
180. A. H. Barber, S. R. Cohen, and H. D. Wagner, *Appl. Phys. Lett.* 82, 4140 (2003).
181. C. A. Cooper, R. J. Young, and M. Halsall, *Composites A* 32, 401 (2001).
182. T. Ramanathan, H. Liu, and L. C. Brinson, *J. Polym. Sci. B* 43, 2269 (2005).
183. C. Velasco-Santos, A. L. Martinez-Hernandez, F. T. Fisher, R. Ruoff, and V. M. Castano, *Chem. Mater.* 15, 4470 (2003).
184. S. J. V. Frankland, A. Caglar, D. W. Brenner, and M. Griebel, *J. Phys. Chem. B* 106, 3046 (2002).
185. M. Cadek, J. N. Coleman, V. Barron, K. Hedicke, and W. J. Blau, *Appl. Phys. Lett.* 81, 5123 (2002).
186. Z. Jin, K. P. Pramoda, G. Xu, and S. H. Goh, *Chem. Phys. Lett.* 337, 43 (2001).
187. O. Meincke, D. Kaempfer, H. Weickmann, C. Friedrich, M. Vathauer, and H. Warth, *Polymer* 45, 739 (2004).
188. T. Ogasawara, Y. Ishida, T. Ishikawa, and R. Yokota, *Composites Part A* 35, 67 (2004).
189. J. K. W. Sandler, S. Pegel, M. Cadek, F. Gojny, M. van Es, J. Lohmar, W. J. Blau, K. Schulte, A. H. Windle, and M. S. P. Shaffer, *Polymer* 45, 2001 (2004).
190. R. E. Gorga and R. E. Cohen, *J. Poly. Sci.* 42, 2690 (2004).
191. S. L. Ruan, P. Gao, X. G. Yang, and T. X. Yu, *Polymer* 44, 5643 (2003).

192. A. R. Bhattachararyya, T. V. Sreekumar, T. Liu, S. Kumar, L. M. Ericson, H. Hauge, and R. E. Smalley, *Polymer* 44, 2373 (2003).
193. S. Singh, Y. Q. Pei, R. Miller, and P. R. Sundararajan, *Adv. Funct. Mater.* 13, 868 (2003).
194. K. W. Putz, C. A. Mitchell, R. Krishnamoorti, and P. F. Green, *J. Poly. Sci. B* 42, 2286 (2004).
195. X. Li, H. Gao, W. A. Scrivens, D. Fei, X. Xu, M. A. Sutton, A. P. Reynolds, and M. L. Myrick, *Nanotechnology* 15, 1416 (2004).
196. C. Park, Z. Ounaies, K. A. Watson, R. E. Crooks, J. Smith, S. E. Lowther, J. W. Connell, E. J. Siochi, J. S. Harrison, and T. L. S. Clair, *Chem. Phys. Lett.* 364, 303 (2002).
197. J. N. Coleman, W. J. Blau, A. B. Dalton, E. Munoz, S. Collins, B. G. Kim, J. Razal, M. Selvidge, G. Vieiro, and R. H. Baughman, *Appl. Phys. Lett.* 82, 1682 (2003).
198. L. Valentini, J. Biagiotti, J. M. Kenny, and M. A. L. Manchado, *J. Appl. Polym. Sci.* 89, 2657 (2003).
199. J. Sandler, M. S. P. Shaffer, T. Prasse, W. Bauhofer, K. Schulte, and A. H. Windle, *Polymer* 40, 5967 (1999).
200. J. G. Smith, J. W. Connell, D. M. Delozier, P. T. Lillehei, K. A. Watson, Y. Lin, B. Zhou, and Y. P. Sun, *Polymer* 45, 825 (2004).
201. D. M. Delozier, K. A. Watson, J. G. Smith, and J. W. Connell, *Compos. Sci. Technol.* 65, 749 (2005).
202. S. A. Curran, P. M. Ajayan, W. J. Blau, D. L. Carroll, J. N. Coleman, A. B. Dalton, A. P. Davey, A. Drury, B. McCarthy, S. Maier, and A. Strevens, *Adv. Mater.* 10, 1091 (1998).
203. F. Du, R. C. Scogna, W. Zhou, S. Brand, J. E. Fischer, and K. I. Winey, *Macromolecules* 37, 9048 (2004).
204. P. Potschke, S. M. Dudkin, and I. Alig, *Polymer* 44, 5023 (2003).
205. P. Potschke, A. R. Bhattacharyya, and A. Janke, *Polymer* 44, 8061 (2003).
206. A. Allaoui, S. Bai, H. M. Cheng, and J. B. Bai, *Compos. Sci. Technol.* 62, 1993 (2002).
207. Y. Long, Z. Chen, X. Zhang, J. Zhang, and Z. H. Liu, *Appl. Phys. Lett.* 85, 1796 (2004).
208. J. K. W. Sandler, J. E. Kirk, I. A. Kinloch, M. S. P. Shaffer, and A. H. Windle, *Polymer* 44, 5893 (2003).
209. Z. Ounaies, C. Park, K. E. Wise, E. J. Siochi, and J. S. Harrison, *Compos. Sci. Technol.* 63, 1637 (2003).
210. B. E. Kilbride, J. N. Coleman, J. Fraysse, O. Fournet, M. Cadek, A. Drury, S. Hutzler, S. Roth, and W. J. Blau, *J. Appl. Phys.* 92, 4024 (2002).
211. M. J. Biercuk, M. C. Llaguno, M. Radosavljevic, J. K. Hyun, A. T. Johnson, and J. E. Fischer, *Appl. Phys. Lett.* 80, 2767 (2002).
212. J. Hone, M. C. Llaguno, M. J. Biercuk, A. T. Johnson, B. Batlogg, Z. Benes, and J. E. Fischer, *Appl. Phys. A* 74, 339 (2002).
213. C. H. Liu, H. Huang, Y. Wu, and S. S. Fan, *Appl. Phys. Lett.* 84, 4248 (2004).
214. C. H. Liu and S. S. Fan, *Appl. Phys. Lett.* 86, 123106 (2005).
215. C. W. Nan, G. Liu, Y. Lin, and M. Li, *Appl. Phys. Lett.* 85, 3549 (2004).
216. S. Shenogin, A. Bodapati, L. Xue, R. Ozisik, and P. Keblinski, *Appl. Phys. Lett.* 85, 2229 (2004).
217. S. Shenogin, L. Xue, R. Ozisik, P. Keblinski, and D. G. Cahill, *J. Appl. Phys.* 95, 8136 (2004).
218. M. Foygel, R. D. Morris, D. Anez, S. French, and V. L. Sobolev, *Phys. Rev. B* 71, 104201 (2005).
219. T. X. Liu, Z. H. Liu, K. X. Ma, L. Shen, K. Y. Zeng, and C. B. He, *Compos. Sci. Technol.* 63, 331 (2003).
220. A. Yasmin, J. L. Abot, and I. M. Daniel, *Scripta Mater.* 49, 81 (2003).
221. T. Agag, T. Koga, and T. Takeichi, *Polymer* 42, 3399 (2001).
222. W. Liu, S. V. Hoa, and M. Pugh, *Compos. Sci. Technol.* 65, 307 (2005).
223. M. Alexandre, P. Dubois, T. Sun, J. M. Garces, and R. Jerome, *Polymer* 43, 2123 (2002).
224. R. K. Bharadwaj, A. R. Mehrabi, C. Hamilton, C. RTrujillo, M. Murga, R. Fan, A. Chavira, and A. K. Thompson, *Polymer* 43, 3699 (2002).
225. J. W. Cho and D. R. Paul, *Polymer* 42, 1083 (2001).
226. R. K. Shah and D. R. Paul, *Polymer* 45, 2991 (2004).
227. P. J. Yoon, D. L. Hunter, and D. R. Paul, *Polymer* 44, 5323 (2003).
228. P. Uribe-Arocha, C. Mehler, J. E. Puskas, and V. Altstadt, *Polymer* 44, 2441 (2003).
229. J. H. Park and S. C. Jana, *Polymer* 44, 2091 (2003).
230. J. Nielsen, *J. Macromol. Sci. Chem.* 1, 929 (1967).
231. J. W. Gilman, C. L. Jackson, A. B. Morgan, J. R. Harris, E. Manian, E. P. Giannelis, M. Wuthenow, D. Hilton, and S. H. Phillips, *Chem. Mater.* 12, 1866 (2000).
232. S. Peeterbroeck, M. Alexandre, J. B. Nagy, C. Pirlot, A. Fonseca, N. Moreau, G. Philippin, J. Delhalle, Z. Mekhalif, R. Sporken, G. Beyer, and P. Dubois, *Compos. Sci. Technol.* 64, 2317 (2004).
233. A. Yasmin and I. M. Daniel, *Polymer* 45, 8211 (2004).
234. J. W. Shen, X. M. Chen, and W. Y. Huang, *J. Appl. Polym. Sci.* 88, 1864 (2003).
235. W. Zheng and S. C. Wong, *Compos. Sci. Technol.* 63, 225 (2003).
236. W. Zheng, X. H. Lu, and S. C. Wong, *J. Appl. Polym. Sci.* 91, 2781 (2004).
237. T. Ramanathan, S. Stankovich, D. Dikin, H. Liu, S. T. Nguyen, R. S. Ruoff, and L. C. Brinson, 2006, (submitted to Polymer).
238. R. Krishnamoorti and K. Yurekli, *Curr. Opin. Coll. Interface Sci.* 6, 464 (2001).
239. O. Becker, Y. B. Cheng, R. J. Varley, and G. P. Simon, *Macromolecules* 36, 1616 (2003).
240. P. B. Messersmith and E. P. Giannelis, *Chem. Mater.* 6, 1719 (1994).
241. C. H. Park, J. H. Kim, M. Ree, B. H. Sohn, J. C. Jung, and W. C. Zin, *Polymer* 45, 4507 (2004).
242. S. D. Burnside and E. P. Giannelis, *J. Poly. Sci. B* 38, 1595 (2000).
243. X. Liu and Q. Wu, *Polymer* 43, 1933 (2002).
244. R. A. Vaia, B. B. Sauer, O. K. Tse, and E. P. Giannelis, *J. Poly. Sci. B* 35, 59 (1997).
245. S. H. Anastasiadis, K. Karatasos, G. Vlachos, E. Manias, and E. P. Giannelis, *Phys. Rev. Lett.* 84, 915 (2000).

246. K. M. Lee and C. D. Han, *Macromolecules* 36, 804 (2003).
247. L. Shen, I. Y. Phang, L. Chen, T. Liu, and K. Zeng, *Polymer* 45, 3341 (2004).
248. F. Xiao, M. Xiao, and K. Gong, *Polymer* 42, 4813 (2001); M. Xiao, L. Sun, J. Liu, Y. Li, and K. Gong, *Polymer* 43, 2245 (2002).
249. W. Zheng, S. C. Wong, and H. J. Sue, *Polymer* 43, 6767 (2002).
250. S. Yang, J. R. Castilleja, E. V. Barrera, and K. Lozano, *Polym. Degrad. Stab.* 83, 383 (2004).
251. W. Ding, A. Eitan, F. T. Fisher, X. Chen, D. A. Dikin, R. Andrews, L. C. Brinson, L. S. Schadler, and R. S. Ruoff, *Nano Lett.* 3, 1593 (2003).
252. F. T. Fisher and L. C. Brinson, *Compos. Sci. Technol.* 61, 731 (2001).
253. T. V. Sreekumar, T. Liu, B. G. Min, H. Guo, S. Kumar, R. H. Hauge, and R. E. Smalley, *Adv. Mater.* 16, 58 (2004).
254. C. Wei, D. Srivastava, and K. Cho, *Nano Lett.* 2, 647 (2002).
255. C. Wei and D. Srivastava, *Nano Lett.* 4, 1949 (2004).
256. J. Sandler, P. Werner, M. S. P. Shaffer, D. Demchuk, V. Altstadt, and A. H. Windle, *Composites Part A* 33, 1033 (2002); J. Sandler, A. H. Windle, P. Werner, V. Altstadt, M. V. Es, and M. S. P. Shaffer, *J. Mater. Res.* 38, 2135 (2003).
257. M. Cadek, J. N. Coleman, K. P. Ryan, V. Nicolosi, G. Bister, A. Fonseca, J. B. Nagy, K. Szostak, F. Beguin, and W. J. Blau, *Nano Lett.* 4, 353 (2004).
258. R. D. Bradshaw and L. C. Brinson, *Mech. of Time-Depend. Mater.* 1, 85 (1997).
259. P. K. Valavala and G. M. Odegard, *Rev. Adv. Mater. Sci.* 9, 34 (2005).
260. D. Raabe, "Computational Materials Science: The Simulation of Materials Microstructures and Properties," Wiley-VCH, New York, 1998; M. E. J. Newman and G. T. Barkema, "Monte Carlo Methods in Statistical Physics," Oxford University Press, New York, 1999.
261. D. Theodorou and U. Suter, *Macromolecules* 19, 139 (1986).
262. J. Baschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Muller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, and V. Tries, *Adv. Polym. Sci.* 152, 41 (2000).
263. D. Frenkel and B. Smit, "Understanding Molecular Simulation: From Algorithms to Applications," Academic Press, San Diego, 1996.
264. A. Garg and S. B. Sinnott, *Chem. Phys. Lett.* 295, 273 (1998).
265. J. Bernholc, C. Brabec, M. Nardelli, A. Maiti, C. Roland, and B. I. Yakobson, *Appl. Phys. A* 67, 39 (1998).
266. Z. H. Yao, C. C. Zhu, M. Cheng, and J. H. Liu, *Comp. Mater. Sci.* 22, 180 (2001).
267. S. J. V. Frankland, A. Caglar, D. W. Brenner, and M. Griebel, *J. Phys. Chem. B* 106, 3046 (2002); V. Cannillo, C. Leonelli, and A. R. Boccaccini, *Mater. Sci. Eng. A* 323, 246 (2002).
268. G. M. Odegard, T. S. Gates, L. M. Nicholson, and K. E. Wise, *Compos. Sci. Technol.* 62, 1869 (2002).
269. M. in het Panhuis, A. Maiti, A. B. Dalton, A. van den Noort, J. N. Coleman, B. McCarthy, and W. J. Blau, *J. Phys. Chem. B* 107, 478 (2003).
270. D. Gersappe, *Phys. Rev. Lett.* 89, 058301 (2002).
271. I. Daniel and O. Ishai, "Engineering Mechanics of Composite Materials," Oxford University Press, New York, 1994.
272. J. S. Smith, D. Bedrov, and G. D. Smith, *Compos. Sci. Technol.* 63, 1599 (2003).
273. F. W. Starr, J. F. Douglas, and S. C. Glotzer, *J. Chem. Phys.* 119, 1777 (2003).
274. S. Salaniwal, S. K. Kumar, and J. F. Douglas, *Phys. Rev. Lett.* 89, 258301 (2002).
275. S. Kumar, (unpublished work, 2005).
276. T. R. Bohme and J. J. de Pablo, *J. Chem. Phys.* 116, 9939 (2002).
277. S. J. V. Frankland, V. M. Harik, G. M. Odegard, D. W. Brenner, and T. S. Gates, *Compos. Sci. Technol.* 63, 1655 (2003).
278. M. Wong, M. Paramsothy, X. J. Xu, Y. Ren, S. Li, and K. Liao, *Polymer* 44, 7757 (2003).
279. M. A. Sharaf and J. E. Mark, *Polymer* 45, 3943 (2004).
280. Q. Zhang and L. A. Archer, *J. Chem. Phys.* 121, 10814 (2004).
281. E. Hackett, E. Manias, and E. Giannelis, *Chem. Mater.* 12, 2161 (2000).
282. M. S. Ozmusul and R. C. Picu, *Polymer* 43, 4657 (2002); R. C. Picu and M. S. Ozmusul, *J. Chem. Phys.* 118, 11239 (2003).
283. S. C. Glotzer and W. Paul, *Annu. Rev. Mater. Res.* 32, 401 (2002).
284. T. C. Clancy and J. Hinkley, NASA/TM-2004-213030 (2004).
285. A. Chatterjee and D. G. Vlachosa, *J. Chem. Phys.* 121, 11420 (2004).
286. G. M. Odegard, T. S. Gates, K. E. Wise, C. Park, and E. J. Siochi, *Compos. Sci. Technol.* 63, 1671 (2003).
287. G. M. Odegard, T. C. Clancy, and T. S. Gates, *Polymer* 46, 553 (2005).
288. G. M. Odegard, in "Handbook of Theoretical and Computational Nanotechnology" (M. Reith and W. Schommers, Eds.). American Scientific Publishers, Stevenson Ranch, CA, 2006.
289. C. Y. Li and T. W. Chou, *J. Nanosci. Nanotechnol.* 3, 423 (2003).
290. X. L. Gao and K. Li, *Int. J. Sol. Struct.* 42, 1649 (2005).
291. C. Li and T. W. Chou, in "44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference," Norfolk, Virginia, 2003.
292. S. P. Timoshenko and J. N. Goodier, "Theory of Elasticity," 3rd Edn. McGraw-Hill, New York, 1970.
293. M. L. Dunn and H. Ledbetter, *J. Appl. Mech.* 62, 1023 (1995).
294. R. B. Thompson, V. V. Ginzburg, M. W. Matsen, and A. C. Balazs, *Macromolecules* 35, 1060 (2002).
295. X. Xu, M. M. Thwe, C. Shearwood, and K. Liao, *Appl. Phys. Lett.* 81, 2833 (2002).

296. J. J. Luo and I. M. Daniel, *Compos. Sci. Technol.* 63, 1607 (2003).
297. T. D. Fornes and D. R. Paul, *Polymer* 44, (2003).
298. F. T. Fisher, R. D. Bradshaw, and L. C. Brinson, *Appl. Phys. Lett.* 80, 4647 (2002).
299. F. T. Fisher, R. D. Bradshaw, and L. C. Brinson, *Compos. Sci. Technol.* 63, 1689 (2003).
300. R. D. Bradshaw, F. T. Fisher, and L. C. Brinson, *Compos. Sci. Technol.* 63, 1705 (2003).
301. N. Sheng, M. C. Boyce, D. M. Parks, G. C. Rutledge, J. I. Abes, and R. E. Cohen, *Polymer* 45, 487 (2004).
302. Y. J. Liu and X. L. Chen, *Mech. Mater.* 35, 69 (2003).
303. X. L. Chen and Y. J. Liu, *Comput. Mater. Sci.* 29, 1 (2004).
304. H. R. Lusti and A. A. Gusev, *Mod. Simul. Mater. Sci. Eng.* 12, S107 (2004).
305. R. S. Fertig and M. R. Garnich, *Compos. Sci. Technol.* 64, 2577 (2004).
306. P. Poncharal, Z. L. Wang, D. Ugarte, and W. A. de Heer, *Science* 283, 1513 (1999).
307. J. D. Eshelby, *Proceed. Royal Soc. London A* 241, 376 (1957).
308. B. Budiansky, *J. Mech. Phys. Solids* 13, 223 (1965); R. Hill, *J. Mech. Phys. Solids* 13, 213 (1965); R. M. Christensen and K. H. Lo, *J. Mech. Phys. Solids* 34, 639 (1986); R. M. Christensen and K. H. Lo, *J. Mech. Phys. Solids* 27, 315 (1979).
309. T. Mori and K. Tanaka, *Acta Metall.* 21, 571 (1973).
310. Y. Benveniste, *Mech. Mater.* 6, 147 (1987).
311. G. J. Weng, *Int. J. Eng. Sci.* 28, 1111 (1990).
312. J. C. Halpin, *J. Compos. Mater.* 3, 732 (1969).
313. J. C. Halpin and J. L. Kardos, *Polym. Eng. Sci.* 16, 344 (1976).
314. J. Aboudi, "Mechanics of Composite Materials- A Unified Micromechanics Approach," Elsevier, Amsterdam, 1991.
315. R. M. Christensen, *J. Mech. Phys. Solids* 38, 379 (1990).
316. C. L. Tucker and E. Liang, *Compos. Sci. Technol.* 59, 655 (1999).
317. G. Dvorak, *J. Mech. Phys. Solids* 47, 899 (1999); R. M. Christensen, H. Schantz, and J. Shapiro, *J. Mech. Phys. Solids* 40, 69 (1992).
318. T. Mura, "Micromechanics of Defects in Solids," 2nd Edn. Kluwer, Dordrecht, 1982.
319. G. J. Weng, *Int. J. Eng. Sci* 22, 845 (1984).
320. Z. Hashin, *J. Appl. Mech.* 32, 630 (1965); Z. Hashin, *Int. J. Sol. Struct.* 6, 797 (1970).
321. W. Lin, Ph.D. dissertation, Northwestern University, 1996.
322. L. C. Brinson and W. G. Knauss, *J. Appl. Mech.* 59, 730 (1992).
323. F. T. Fisher and L. C. Brinson, (manuscript in preparation, 2006).
324. E. Manias, A. Subbotin, G. Hadziioannou, and G. ten Brinke, *Molecul. Phys.* 85, 1017 (1995).
325. I. Bitsanis and C. Pan, *J. Chem. Phys.* 99, 5520 (1993).
326. E. Manias, H. Chen, R. Krishnamoorti, J. Genzer, E. J. Kramer, and E. P. Giannelis, *Macromolecules* 33, 7955 (2000).
327. L. Zhu and K. A. Narh, *J. Polym. Sci. B* 42, 2391 (2004).
328. P. K. Mallick, "Fiber-reinforced Composites," Marcel Dekker, New York, 1993.
329. H. Krenchel, "Fiber Reinforcement," Akademisk, Copenhagen, 1964.
330. H. Geng, R. Rosen, B. Zheng, H. Shimoda, L. Fleming, J. Liu, and O. Zhou, *Adv. Mater.* 14, 1387 (2002).
331. H. Liu, T. Ramanathan, and L. C. Brinson, *J. Appl. Mech.* (to be published, 2006).
332. H. M. Shodja and A. S. Sarvestani, *J. Appl. Mech.* 68, 3 (2001); Y. Benveniste, G. J. Dvorak, and T. Chen, *Mech. Mater.* 7, 305 (1989); M. Hori and S. Nemat-Nasser, *Mech. Mater.* 14, 189 (1993); H. A. Luo and G. J. Weng, *Mech. Mater.* 8, 347 (1987); H. A. Luo and G. J. Weng, *Mech. Mater.* 8, 77 (1989).
333. M. B. Nardelli, J.-L. Fattebert, D. Orlikowski, C. Roland, Q. Zhao, and J. Bernholc, *Carbon* 38, 1703 (2000); B. C. Pan, W. S. Yang, and J. Yang, *Phys. Rev. B* 62, 12652 (2000).
334. D. Srivastava, D. W. Brenner, J. D. Schall, K. D. Ausman, M.-F. Yu, and R. S. Ruoff, *J. Phys. Chem. B* 103, 4330 (1999).
335. Q. Z. Zhao, M. B. Nardelli, and J. Bernholc, *Phys. Rev. B* 65, 144105 (2002).
336. H. D. Wagner, *Chem. Phys. Lett.* 361, 57 (2002).
337. K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003).
338. W. K. Liu, E. G. Karpov, S. Zhang, and H. S. Park, *Comput. Meth. Appl. Mech. Eng.* 193, 1529 (2004).
339. H. S. Park and W. K. Liu, *Comput. Meth. Appl. Mech. Eng.* 193, 1733 (2004).
340. G. Friesecke and R. D. James, *J. Mech. Phys. Solids* 48, 1519 (2000).
341. M. Arroyo and T. Belytschko, *J. Mech. Phys. Solids* (2002); D. Qian, W. K. Liu, S. Subramoney, and R. S. Ruoff, *J. Nanosci. Nanotechnol.* 3, 185 (2003).
342. D. Qian, G. J. Wagner, and W. K. Liu, *Comput. Meth. Appl. Mech. Eng.* 193, 1603 (2004).
343. D. Walters, L. Ericson, M. Casavant, J. Liu, D. Colbert, K. Smith, and R. Smalley, *Appl. Phys. Lett.* 74, 3803 (1999); W. D. Shen, B. Jiang, B. S. Han, and S. S. Xie, *Phys. Rev. Lett.* 84, 3634 (2000); I. Chasiotis and W. G. Knauss, *Exper. Mech.* 42, 51 (2002); S. Sundararajan and B. Bhushan, *Sens. Actuat. A* 101, 338 (202); M. Guthold, M. Falvo, W. G. Matthews, S. Paulson, J. Mullin, S. Lord, D. Erie, S. Washburn, R. Superfine, F. P. Brooks, and R. M. Taylor, *J. Molecul. Graph. Model.* 17, 187 (1999).
344. S. Lu, D. A. Dikin, S. Zhang, F. T. Fisher, J. Lee, and R. S. Ruoff, *Rev. Sci. Inst.* 75, 2151 (2004); D. M. Burns and V. M. Bright, *SPIE Int. Soc. Opt. Eng.* (1997); J. H. Comtois and V. M. Bright, *Sens. Actuat. A* 58, 19 (1997); T. Moulton and G. K. Ananthasuresh, *Sens. Actuat. A* 90, 38 (2001).
345. M.-F. Yu, M. J. Dyer, G. D. Skidmore, H. W. Rohrs, X. Lu, K. D. Ausman, J. R. von Ehr, and R. S. Ruoff, *Nanotechnology* 10, 244 (1999); K. Fukushima, S. Kawai, D. Sayu, and H. Kawakatsu, *Rev. Sci. Inst.* 73, 2647

(2002); P. A. Williams, S. J. Papadakis, M. R. Falvo, A. M. Patel, M. Sinclair, A. Seeger, A. Helser, R. M. Taylor, S. Washburn, and R. Superfine, *Appl. Phys. Lett.* 80, 2574 (2002); H. T. Miyazaki, Y. Tomizawa, K. Koyano, T. Sato, and N. Shinya, *Rev. Sci. Inst.* 71, 3123 (2000).
346. H. W. P. Koops, R. Weiel, D. Kern, and T. H. Baum, *J. Vac. Sci. Tech. B* 6, 477 (1988).
347. J. B. Bai and A. Allaoui, *Composites Part A* 34, 689 (2003).
348. H. J. Barraza, F. Pompeo, E. A. O'Rear, and D. E. Resasco, *Nano Lett.* 2, 797 (2002).
349. R. Bharadwaj, R. Berry, and B. Farmer, *Polymer* 41, 7209 (2000).
350. Y. Brechet, J. Cavaille, E. Chabert, L. Chazeau, R. Dendieval, L. Flandin, and C. Gauthier, *Adv. Eng. Mater.* 3, 571 (2001).
351. D. Brown, P. Mele, S. Marceau, and N. D. Alberola, *Macromolecules* 36, 1395 (2003).
352. C. A. Cooper, S. R. Cohen, A. H. Barber, and H. D. Wagner, *Appl. Phys. Lett.* 81, 3873 (2002).
353. P. Doruker and W. L. Mattice, *Macromol. Theor. Sim.* 8, 463 (1999).
354. A. Dufresne, M. Paillet, J. L. Putaux, R. Canet, F. Carmona, P. Delhaes, and S. Cui, *J. Mater. Sci.* 37, 3915 (2002).
355. K. Enomoto, T. Yasuhara, S. Kitakata, H. Murakami, and N. Ohtake, *New Diam. Front. Carbon Technol.* 14, 11 (2004).
356. M. D. Frogley, D. Ravich, and H. D. Wagner, *Compos. Sci. Technol.* 63, 1647 (2003).
357. J. Frohlich, W. Niedermeier, and H.-D. Luginsland, *Composites A* 36, 449 (2005).
358. E. P. Giannelis, *Appl. Organometal. Chem.* 12, 675 (1998).
359. J. W. Gilman, *Appl. Clay Sci.* 15, 31 (1999).
360. H. W. Goh, S. H. Goh, G. Q. Xu, K. P. Pramoda, and W. D. Zhang, *Chem. Phys. Lett.* 373, 277 (2003).
361. X. Y. Gong, J. Liu, S. Baskaran, R. D. Voise, and J. S. Young, *Chem. Mater.* 12, 1049 (2000).
362. M. Griebel and J. Hamaekers, *Comput. Meth. Appl. Mech. Eng.* 193, 1773 (2004).
363. V. Gumen, F. Jones, and D. Attwood, *Polymer* 42, 5717 (2001).
364. Y. Hu, I. Jang, and S. B. Sinnott, *Compos. Sci. Technol.* 63, 1663 (2003); Y. Hu and S. B. Sinnott, *J. Mater. Chem.* 14 (2004).
365. N. Hu, H. Fukunaga, C. Lu, M. Makeyama, and B. Yan, *Proc. R. Soc. London Ser. A* 461, 1685 (2005).
366. M. Huhtala, A. V. Krasheninnikov, J. Aittoneimi, S. J. Stuart, K. Nordlund, and K. Kaski, *Phys. Rev. B* 70, 045404 (2004).
367. Y. H. Hyun, S. T. Lim, H. J. Choi, and M. S. Jhon, *Macromolecules* 34, 8084 (2001).
368. A. B. Inceoglu and U. Yilmazer, *Polym. Eng. Sci.* 43, 661 (2003).
369. S. Izumisawa and M. S. Jhon, *J. Chem. Phys.* 117, 3972 (2002).
370. Z. Jin, X. Sun, G. Xu, S. Goh, and W. Ji, *Chem. Phys. Lett.* 318, 505 (2000).
371. Z. X. Jin, K. P. Pramoda, S. H. Goh, and G. Q. Xu, *Mater. Res. Bull.* 37, 271 (2002).
372. C. C. Kao and R. J. Young, *Compos. Sci. Technol.* 64, 2291 (2004).
373. T. H. Kim, S. T. Lim, C. H. Lee, H. J. Choi, and M. S. Jhon, *J. Appl. Polym. Sci.* 87, 2106 (2003).
374. H. Koerner, G. Price, N. A. Pearce, M. Alexander, and R. A. Vaia, *Nat. Mater.* 3, 115 (2004).
375. A. Usuki, Y. Kojima, M. Kawasumi, A. Okada, Y. Fukushima, T. Kurauchi, and O. Kamigaito, *J. Mater. Res.* 8, 1179 (1993).
376. D. Kong and C. E. Park, *Chem. Mater.* 15, 419 (2003).
377. N. Koratkar, B. Q. Wei, and P. M. Ajayan, *Adv. Mater.* 14, 997 (2002).
378. X. Kornmann, H. Lindberg, and L. A. Berglund, *Polymer* 42, 4493 (2001).
379. S. Kumar, H. Doshi, M. Srinivasarao, J. O. Park, and D. A. Schiraldi, *Polymer* 43, 1701 (2002).
380. L. Leger, E. Raphael, and H. Hervet, *Adv. Polym. Sci.* 138, 185 (1999).
381. C. Li, T. Liang, W. Lu, C. Tang, X. Hu, M. Cao, and J. Liang, *Compos. Sci. Technol.* 64, 2089 (2004).
382. Z. Liang, J. Gou, C. Zhang, B. Wang, and L. Kramer, *Mater. Sci. Eng. A* 365, 228 (2004).
383. C. Y. Liu, J. Zhang, J. S. He, and G. H. Hu, *Polymer* 44, 7529 (2003).
384. X. H. Liu and Q. J. Wu, *Polymer* 42, 10013 (2001).
385. H. Liu, A. Uhlherr, and M. K. Bannister, *Polymer* 45, 2051 (2004).
386. O. Lourie and H. Wagner, *Compos. Sci. Technol.* 59, 975 (1999).
387. A. A. Mamedov, N. A. Kotov, M. Prato, D. M. Guldi, J. P. Wicksted, and A. Hirsch, *Nat. Mater.* 1, 190 (2002); M. Olek, J. Ostrander, S. Jurga, H. Mohwald, N. A. Kotov, K. Kempa, and M. Giersig, *Nano Lett.* 4, 1889 (2004).
388. O. L. Manevitch and G. C. Rutledge, *J. Phys. Chem. B* 108, 1428 (2004).
389. B. McCarthy, J. N. Coleman, S. A. Curran, A. B. Dalton, A. P. Davey, Z. Konya, A. Fonseca, J. B. Nagy, and W. J. Blau, *J. Mater. Sci Lett.* 19, 2239 (2000); B. McCarthy, J. N. Coleman, R. Czerw, A. B. Dalton, D. L. Carroll, and W. J. Blau, *Syn. Metals* 121, 1225 (2001); B. McCarthy, J. N. Coleman, R. Czerw, A. B. Dalton, M. I. H. Panhuis, A. Maiti, A. Drury, P. Bernier, J. B. Nagy, B. Lahr, H. J. Byrne, D. L. Carroll, and W. J. Blau, *J. Phys. Chem. B* 106, 2210 (2002).
390. B. Minisini and F. Tsobnang, *Composites A* 36, 531 (2005); B. Minisini and F. Tsobnang, *Composites A* 36, 539 (2005).
391. P. Musto, G. Ragosta, G. Scarinzi, and L. Mascia, *Polymer* 45, 1697 (2004).
392. T. Natsuki, K. Tantrakarn, and M. Endo, *Carbon* 42, 39 (2004).
393. A. N. Norris, *Int. J. Solids Struct.* 26, 663 (1990).
394. M. S. Ozmusul and R. C. Picu, *Polym. Compos.* 23, 110 (2002).
395. V. Panchagnula, J. Jeon, and A. V. Dobrynin, *Phys. Rev. Lett.* 93, 037801 (2004).
396. A. Pantano, D. M. Parks, and M. C. Boyce, *J. Mech. Phys. Solids* 52, 789 (2004).

397. C. Park, R. E. Crooks, E. J. Siochi, J. S. Harrison, N. Evans, and E. Kenik, *Nanotechnology* 14, L11 (2003).
398. W. Paul, *Polymer* 45, 3901 (2004).
399. R. B. Pipes and P. Hubert, *Compos. Sci. Technol.* 62, 419 (2002); R. B. Pipes and P. Hubert, *Compos. Sci. Technol.* 63, 1571 (2003).
400. P. Potschke, T. D. Fornes, and D. R. Paul, *Polymer* 43, 3247 (2002); P. Potschke, A. R. Bhattacharyya, and A. Janke, *Europ. Poly. J.* 40, 137 (2004); P. Potschke, A. R. Bhattacharyya, A. Janke, and H. Goering, *Compos. Interfaces* 10, 389 (2003).
401. P. Potschke and A. R. Bhattacharyya, *Polym. Preprints* 44, 760 (2003).
402. M. Prasad, M. Kopycinska, U. Rabe, and W. Arnold, *Geophys. Res. Lett.* 29, 1172 (2002).
403. D. Qian and E. C. Dickey, *J. Microscopy* 204, 39 (2001).
404. Y. Ren, F. Y. Li, H. M. Cheng, and K. Liao, *Carbon* 41, 2177 (2003).
405. J. Ren, A. S. Silva, and R. Krishnamoorti, *Macromolecules* 33, 3739 (2000).
406. M. Roberge, R. E. Prud'homme, and J. Brisson, *Polymer* 45, 1401 (2004).
407. M. Rodlert, C. J. G. Plummer, Y. Leterrier, and J. E. Manson, *J. Rheol.* 48, 1049 (2004).
408. S. Rossi, P. F. Luckham, and T. F. Tadros, *Colloids Surf. A* 215, 1 (2003).
409. J. H. Rouse and P. T. Lillehei, *Nano Lett.* 3, 59 (2003).
410. J. P. Salvetat-Delmotte and A. Rubio, *Carbon* 40, 1729 (2002).
411. M. Sammalkorpi, A. V. Krasheninnikov, A. Kuronen, K. Nordlund, and K. Kaski, *Phys. Rev. B* 70, 245416 (2004).
412. H. Sato, A. Yamagishi, and K. Kawamura, *J. Phys. Chem. B* 105, 7990 (2001).
413. D. L. Shi, X. Q. Feng, Y. Y. Huang, K. Hwang, and H. Gao, *J. Eng. Mater. Technol.* 126, 250 (2004).
414. D. Shia, C. Y. Hui, S. D. Burnside, and E. P. Giannelis, *Polym. Compos.* 19, 608 (1998).
415. M. L. Shofner, F. J. Rodriguez-Macias, R. Vaidyanathan, and E. V. Barrera, *Composites A* 34, 1207 (200-).
416. A. Sinsawat, K. L. Anderson, R. A. Vaia, and B. L. Farmer, *J. Poly. Sci. B* 41, 3272 (2003).
417. A. Soldera and Y. Grohens, *Polymer* 45, 1307 (2004); A. Soldera and N. Metatla, *Composites A* 36, 521 (2005).
418. Y. P. Sun, K. F. Fu, Y. Lin, and W. J. Huang, *Acc. Chem. Res.* 35, 1096 (2002).
419. W. Z. Tang, M. H. Santare, and S. G. Advani, *Carbon* 41, 2779 (2003).
420. R. S. Tate, D. S. Fryer, S. Pasqualini, M. F. Montague, J. J. De Pablo, and P. F. Nealey, *J. Phys. Chem.* 115, 9982 (2001).
421. E. T. Thostenson, W. Z. Li, D. Z. Wang, Z. F. Ren, and T. W. Chou, *J. Appl. Phys.* 91, 6034 (2002).
422. J. A. Torres, P. F. Nealey, and J. J. de Pablo, *Phys. Rev. Lett.* 85, 3221 (2000).
423. W. Tschop, K. Kremer, J. Batoulis, T. Burger, and O. Hahn, *Acta Polym.* 49, 61 (1998); W. Tschop, K. Kremer, O. Hahn, J. Batoulis, and T. Burger, *Acta Polym.* 49, 75 (1998).
424. Z. C. Tu and Z. C. Ou-Yang, *Phys. Rev. B* 65, 233407 (2002).
425. A. Usuki, N. Hasegawa, H. Kadoura, and T. Okamoto, *Nano Lett.* 1, 271 (2001).
426. R. A. Vaia and E. P. Giannelis, *Polymer* 42, 1281 (2001).
427. L. Valentini, D. Puglia, F. Frulloni, I. Armentano, J. M. Kenny, and S. Santucci, *Compos. Sci. Technol.* 61, 23 (2004).
428. J. A. W. van Dommelen, D. M. Parks, M. C. Boyce, W. A. M. Brekelmans, and F. P. T. Baaijens, *J. Mech. Phys. Solids* 51, 519 (2003); J. A. W. van Dommelen, D. M. Parks, M. C. Boyce, W. A. M. Brekelmans, and F.P. T. Baaijens, *Polymer* 44, 6089 (2003).
429. K. van Workum and J. J. de Pablo, *Nano Lett.* 3, 1405 (2003).
430. J. H. van Zan[illegible], W. E. Wallace, and W. L. Wu, *Phys. Rev. E* 53, 2053 (1996).
431. S. Varghese, J. Karger-Kocsis, and K. G. Gatos, *Polymer* 44, 3977 (2004).
432. C. Velasco-Santos, A. L. Martinez-Hernandez, F. Fisher, R. Ruoff, and V. M. Castano, *J. Phys. D: Appl. Phys.* 36, 1423 (2003).
433. B. Vigolo, A. Penicaud, C. Coulon, C. Sauder, R. Pailler, C. Journet, P. Bernier, and P. Poulin, *Science* 290, 1331 (2000).
434. J. Wang and R. Pyrz, *Compos. Sci. Technol.* 64, 935 (2004); J. Wang and R. Pyrz, *Compos. Sci. Techno* 64, 925 (2004).
435. Y. Wang, R. Cheng, L. Liang, and Y. Wang, *Compos. Sci. Technol.* 65, 793 (2005).
436. J. R. Wood, Q. Zhao, M. D. Frogley, E. R. Meurs, A. D. Prins, T. Peijs, D. J. Dunstan, and H. D. Wagner, *Phys. Rev. B* 62, 7571 (2000); Q. Zhao and H. D. Wagner, *Composites A* 34, 1219 (2003).
437. Y. M. Wu, Z. P. Huang, Y. Zhong, and J. Wang, *Compos. Sci. Technol.* 64, 1345 (2004); Y. Zhong, J. Yang, Y. M. Wu, and Z. P. Huang, *Compos. Sci. Technol.* 64, 1353 (2004).
438. Z. Xia and W. A. Curtin, *Phys. Rev. B* 69, 233408 (2004).
439. X. Xu and K. Liao, *Mater. Phys. Mech.* 4, 148 (2001).
440. P. J. Yoon, D. L. Huter, and D. R. Paul, *Polymer* 44, 5341 (2003).
441. J. J. Zeng, B. Saltysiak, W. S. Johnson, D. A. Schiraldi, and S. Kumar, *Composites B* 35, 245 (2004).
442. P. Zhang, Y. Huang, P. H. Geubelle, P. A. Klein, and K. C. Hwang, *Int. J. Solids Struct.* 39, 3893 (2002); P. Zhang, Y. Huang, H. Gao, and K. C. Hwang, *J. Appl. Mech.* 69, 454 (2002); P. Zhang, H. Jiang, Y. Huang, P. H. Geubelle, and K. C. Hwang, *J. Mech. Phys. Solids* 52, 977 (2004).
443. Y. Zhong and S. Q. Wang, *J. Rheol.* 47, 483 (2003).
444. X. Zhou, E. Shin, K. W. Wang, and C. E. Bakis, *Compos. Sci. Technol.* 64, 2425 (2004).
445. Y. Zou, Y. Feng, L. Wang, and X. Liu, *Carbon* 42, 271 (2004).
446. G. P. Tandon and G. J. Weng, *Compos. Sci. Technol.* 27, 111 (1986).
447. J. Schjodt-Thomsen and R. Pyrz, *Mech. Mater.* 33, 531 (2001).

CHAPTER 7

Modeling and Simulation of Carbon Nanotube/Polymer Composites

Jihua Gou
Department of Mechanical Engineering, University of South Alabama, Mobile, Alabama, USA

Kin-tak Lau
Department of Mechanical Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

CONTENTS

1. Introduction 362
2. Interfacial Bonding of Carbon Nanotube/Polymer Composites 363
   2.1. Molecular Mechanics Simulation 364
   2.2. Molecular Dynamics Simulation 368
   2.3. Analytical Modeling with Classical Continuum Mechanics 369
3. Mechanical Behavior of Carbon Nanotube/Polymer Composites 377
   3.1. Molecular Mechanics and Molecular Dynamics Simulations 378
   3.2. Continuum Modeling 380
   3.3. Multiscale Modeling 383
4. Experimental Investigation of Carbon Nanotube/Polymer Composites 388
   4.1. Interfacial Bonding of Carbon Nanotube/Polymer Composites 388
   4.2. Mechanical Properties of Carbon Nanotube/Polymer Composites 390
5. Conclusion 391
   References 391

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (361–393)

1. INTRODUCTION

In 1985, a team led by Dr. Richard Smalley at Rice University discovered a new form of carbon, buckyministerfullerene C_{60}, a geometric cage-like structure of carbon atoms that is composed of hexagonal and pentagonal faces [1]. Smalley's work resulted in the discovery of carbon nanotubes by Iijima in 1991 [2] and his receiving the 1997 Nobel Prize in chemistry. Since the discovery of carbon nanotubes, significant research is underway to study the unique structural, mechanical, electrical, thermal, and chemical properties of carbon nanotubes and explore their potential applications.

The fullerene structure of carbon nanotubes, derived from the honeycomb lattice representing a single atomic layer of crystalline graphite and held together by strong and high conductive sp^2 bonds, is vastly different from conventional graphite or carbon structures. Theoretical modeling [3–15] and experimental measurements [16–23] have showed that carbon nanotubes possess extraordinary mechanical properties. Specifically, the strength of carbon nanotubes is predicted to be 30 to 120 times stronger than steel, yet only 1/6 the weight of steel. Further, they also can sustain large elastic deformation without breaking. For example, their fracture strains are estimated to be 10–30%, a factor of 10–100 times better than those of carbon fibers, such as IM7, a fiber commonly used in military and aerospace applications. These evaluations of carbon nanotubes indicate that they possess amazing mechanical properties greater than those of graphite fiber. Besides their extraordinary mechanical properties, carbon nanotubes offer either metallic or semiconducting characteristics based on the chiral structure of fullerene. They also possess superior thermal and electrical properties: thermal stability up to 2800°C in a vacuum and 750°C in air, thermal conductivity about twice as high as diamond, with electric current transfer capacity 1000 times greater than copper wires.

Composite materials containing at least one phase with constituents of less than 100 nm in size can be termed nanocomposite materials. The discovery of carbon nanotubes opens the door to enhance the properties of polymer composites by adding them to the matrix materials for structural and multifunctional applications. Carbon nanotubes have large interfacial area per volume and possess extraordinary mechanical, thermal as well as electrical properties. Therefore, it is anticipated that the resulting nanocomposite materials will have enhanced strength, modulus, fracture toughness, electrical conductivity, dimensional stability, and resistance to thermal degradation. In the past few years, significant progress has been achieved to develop carbon nanotube-based nanocomposites [24–29]. Despite these successes, several critical issues as follows must be solved before the full potential of carbon nanotubes is realized in nanocomposite materials applications:

- Proper selection of the polymer matrix;
- High purity of carbon nanotubes used in the composites;
- Uniform dispersion of carbon nanotubes within the polymer matrix during and after manufacturing process;
- Controllable alignment of carbon nanotubes in the composites;
- Good interfacial bonding between the nanotubes and polymer matrix in the composites.

First of all, the polymer matrix should be properly chosen before processing of carbon nanotube/polymer composites. Both thermosetting and thermoplastic polymers have been used as matrix materials in carbon nanotube/polymer composites. Research reports are conflicting regarding the interfacial strength and mechanical properties of carbon nanotube/polymer composites. Depending on the polymer matrix and processing conditions, large variations in their measured properties are found. Therefore, the properties of carbon nanotube/polymer composites are highly polymer specific. Regarding the issues of purity, dispersion, and alignment of carbon nanotubes, these could be achieved by precisely controlling the manufacturing processes of carbon nanotubes and nanotube/polymer composites. Due to the van der Waals interaction, the nanotubes tend to aggregate to form bundles or ropes and further agglomerate when dispersed in the polymer matrix. The high surface area of carbon nanotubes also results in a high viscosity of the nanotube/polymer mixture particularly when fabricating composites with high loading level of nanotubes, which makes the dispersion of

carbon nanotubes extremely difficult. The dispersion of carbon nanotubes within the polymer matrix can be controlled through *in-situ* polymerization of nanotube/polymer monomer mixture [30–31], solution casting of suspension of carbon nanotubes in the dissolved polymers [32–34], melt mixing of carbon nanotubes/nanofibers with polymers [35–37], and the use of surfactants [38]. Carbon nanotubes can be produced with reasonable quality by several techniques. However, the mechanical and novel electronic properties of carbon nanotubes are not only sensitive to their diameter and chirality but are also highly anisotropic. Therefore, it is crucial to obtain the controllable alignment of carbon nanotubes in the final composites. Regarding the alignment of carbon nanotubes, several techniques such as slice cutting [39], mechanical stretching [25], melt spinning [26], and alignment in the magnetic field [40] have appeared in the past few years.

Another challenging issue is the interfacial bonding between the nanotubes and polymer matrix in the composites. A small number of research papers have now appeared in the literature concerning the assessment of the extent and efficiency of load transfer at the nanotube/polymer interface. An efficient load transfer from the polymer matrix to the nanotubes is required to take advantage of the very high Young's modulus and strength of carbon nanotubes in the composites. Unlike conventional fiber-reinforced polymer composites, large interfacial areas are available for load transfer in carbon nanotube/polymer composites due to high aspect ratio of the nanotubes. However, the strength of the interface between the nanotubes and polymer matrix is not well understood, and there is no well-developed experimental method for its direct measurement. Therefore, both the mechanisms and magnitudes of load transfer between the nanotubes and polymer matrix remain unclear at the current stage. Moreover, theoretically and computationally predicting the mechanical behavior of carbon nanotube/polymer composites will be of crucial importance before they are used in real structural applications. Technically, analyzing the interfacial bonding and predicting the mechanical behavior of carbon nanotube/polymer composites can be classified as multiscale and multiphysics problems due to the small dimension of carbon nanotubes in the nanocomposites. Therefore, a combination of analytical modeling, computational method, and experimental characterization should be used to solve this problem for the future development of nanocomposite materials.

The rapid pace of research development and industrial applications of nanocomposite materials has made it necessary to summarize what we currently know and what we don't know about this interesting nanostructure. Some technical reviews on carbon nanotube and nanotube-based composites have been conducted in the past few years, with varying purpose and level of detail [41–44]. The purpose of this review is to compare and contrast the details of several computational models developed for the analysis of the interfacial bonding and the prediction of the mechanical behavior so as to provide a better understanding of the processing–morphology–property inter-relationships of carbon nanotube/polymer composites. The modeling and simulation of carbon nanotube/polymer composites can provide initial guidelines for the development of nanocomposites to help reduce the scope, cost, and time for the real experiments. However, the proper selection of mathematical models and computational methods is of importance, and this crucial issue will be addressed in this review. This review is organized as follows. In the next section, several models for the interfacial bonding of carbon nanotube/polymer composites are described in detail. In Section 3, the major methods to model the mechanical behavior of carbon nanotube/polymer composites are reviewed, and the limitations of each method are discussed. Section 4 addresses the recent experimental characterization of interfacial bonding and mechanical properties of carbon nanotube/polymer composites. Finally, some concluding remarks are offered in Section 5.

2. INTERFACIAL BONDING OF CARBON NANOTUBE/POLYMER COMPOSITES

It is widely accepted that the efficiency of load transfer is controlled by the interfacial characteristics of the fiber and matrix in fiber-reinforced polymer composites. The major mechanisms of load transfer include mechanical interlocking, chemical bonding, and nonbonded

interactions between the fiber and polymer matrix. These mechanisms may be applicable for carbon nanotube/polymer composites. In recent years, a few researchers have studied the interfacial bonding characteristics of carbon nanotube/polymer composites by using computational and theoretical modeling techniques. These modeling techniques can generally be classified into two categories: One is the atomistic modeling where the atom is used as the building block of materials. In the carbon nanotube/polymer composites, the dimension of the nanotube is about the same size as a polymer chain. Therefore, the discrete nature of the atomic interactions between the nanotube and surrounding polymer matrix should be taken into account. In the other words, the nature of the load transfer should be understood using physics-based analysis where molecular mechanics and molecular dynamics simulations will be the most important tools. The other approach is the analytical modeling based on classical continuum mechanics. In this section, we will review and evaluate the computational and analytical models that predict the interfacial strength of carbon nanotube/polymer composites.

2.1. Molecular Mechanics Simulation

For conventional fiber-reinforced polymer composites, the single fiber pullout test has been widely used to characterize the fiber–matrix interfacial properties under tensile loading condition. However, the pullout test of the nanotube is very challenging due to the technical difficulties involved in the manipulation of carbon nanotubes, leading to the use of theoretical and computational approach to characterize the interface of carbon nanotube/polymer composites. Because the diameter of the nanotubes is at the nanometer scale, the interaction at the nanotube/polymer interface is highly dependent on the local molecular structure and bonding. Molecular mechanics and molecular dynamics simulations have been used to investigate molecular interactions at the nanotube/polymer interface. The molecular mechanics study of interfacial binding of carbon nanotube/polymer composites was first conducted by Lordi and Yao [45]. They used force-field–based molecular mechanics to calculate the binding energies and sliding frictional stresses between the nanotubes and different polymer matrices and found that the binding energies and frictional forces play only a minor role in determining the strength of the interface, but that the helical polymer conformations are essential. They suggested that the strength of the nanotube/polymer interface may result from the molecular-level entanglement of the nanotube and polymer matrix.

The conformation of polymer matrix in the carbon nanotube/polymer composites has a significant influence on the interactions between the nanotube and polymer matrix. Gou et al. [46–48] examined the molecular interactions between the nanotube and thermosetting matrix, epoxy resin, during composite processing. In their subsequent work, they studied the interfacial bonding between the nanotube and epoxy resin after curing reaction [49–52]. The pullout simulations of a single-walled nanotube from the cured epoxy resin were conducted through molecular mechanics simulations. In these simulations, a periodic model of the nanotube/epoxy resin composite system needs to be constructed. To construct such a simulation cell, the following relationship for the nanotube, epoxy resin, and their composite must be satisfied.

$$\frac{n_{\mathrm{NT}}}{N_A} \cdot m_{\mathrm{NT}} + \frac{n_{\mathrm{polymer}}}{N_A} \cdot m_{\mathrm{polymer}} = a \cdot b \cdot c \cdot \rho_{\mathrm{composite}} \tag{1}$$

where n_{NT} and n_{polymer} are the number of the nanotube fragments and cured epoxy resin molecules, respectively. m_{NT} and m_{polymer} are the molecular weights of the nanotube fragment and cured epoxy resin molecule, respectively. N_A is Avogadro's number (6.023×10^{23} formula units/mol), a, b, and c are parameters of the simulation cell, and $\rho_{\mathrm{composite}}$ is the density of the resulting nanotube composite. The two terms on the left side of the equation are the masses of the polymer matrix and the nanotube in the composite, which are calculated by the molecular weight and Avogadro's number. The term on the right side of Eq. (1) is the mass of the single-walled carbon nanotube (SWNT)/polymer composite system, which is calculated by volume and density of the simulation cell. The density of the composite could

be conserved during molecular dynamics simulations. The initial density can be determined by the rule of mixture if the volume fraction of the nanotubes is known.

In their study, the composite was composed of a fragment of (10, 10) single-walled nanotube totally embedded into the amorphous polymer matrix of the cured epoxy resin. The (10, 10) nanotube has a diameter of 13.56 Å. A model of the composite with 21,288 atoms is shown in Fig. 1, which consisted of a supercell in the range of 50 Å × 50 Å × 100 Å. The configuration was initiated by randomly generating the epoxy resins surrounding the nanotube using an initial density of 1.2 g/cm^3. The matrix polymer was then equilibrated for approximately 20 ps with a time step of 0.2 fs while holding the nanotube rigid. The system was further equilibrated for 60 ps at a time step of 2 fs with a nonrigid nanotube to create a zero initial stress state. The energy of the nanotube/epoxy resin composite system was minimized during the calculations to achieve the strongest bonding between the nanotube and epoxy resin.

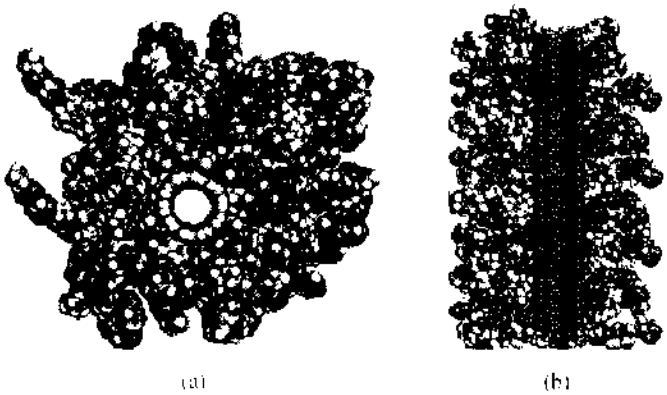
In the nanotube/polymer composites, the bonding strength between the nanotube and epoxy resin can be evaluated by interfacial bonding energy. The nature of interfacial bonding energy comes from the electrostatic and van der Waals forces in the molecular system. Generally, the interaction energy is estimated from the energy difference, ΔE, between the total energy of the composite and the sum of the energies of individual molecules as follows:

$$\Delta E = E_{\text{total}} - (E_{\text{nanotube}} + E_{\text{polymers}}) \tag{2}$$

where E_{total} is the total potential energy of the composite system, E_{nanotube} is the potential energy of the nanotube without the polymer, and E_{polymer} is the potential energy of polymer without the nanotube. In the other words, the interaction energy can be calculated as the difference between the minimum energy and the energy at an infinite separation of the nanotube and polymer matrix. The interaction energy of the nanotube/epoxy resin composite system was calculated using molecular mechanics after energy minimization was applied, as shown in Table 1. In this case, the interaction energy came from both van der Waals interaction and electrostatic interaction between the nanotube and epoxy resin. The negative values of interaction energies were assumed to be attractive forces between the nanotube and epoxy resin. The electrostatic forces resulted from coulombic attraction between the positive hydrogen on both ends of the nanotube and epoxy resin. However, carbon atoms of the nanotube were neutrally charged. Therefore, the electrostatic forces were much smaller than the van der Waals forces. The total interaction energy ΔE is −900 kcal/mol, resulting in attractive forces between the nanotube and epoxy resin. The total interaction energy, ΔE, is twice the interfacial bonding energy γ scaled by the contact area A [45]:

$$\gamma = \frac{\Delta E}{2A} \tag{3}$$

For the nanotube/epoxy composite system, the interfacial bonding energy was 0.1 kcal/mol·Å^2, calculated from Eq. (3).



(a) (b)

Figure 1. Molecular model of single-walled nanotube/epoxy resin composite system: (a) top view and (b) sectional view. Reprinted with permission from [49], J. Gou et al., *Comput. Mater. Sci.* 31, 225 (2004). © 2004, Elsevier.

Table 1. Interaction energies of single-walled nanotube/epoxy resin composite system (kcal/mol).

	Electrostatic		van der Waals		Nonbonded interaction energy	
	Resin	SWNT	Resin	SWNT	Resin	SWNT
Resin	7851.58	−4.202	7308.72	−445.554	15160.3	−449.756
SWNT	−4.202	72.350	−445.554	1861.82	−449.756	1934.17

The mechanical properties of nanotube composites are known to strongly depend on the magnitude of interfacial shear stress from the polymer matrix to the nanotubes. It is also clear that the level of physical–chemical interactions established at their interface plays an important role on the magnitude of load transfer. To characterize the interfacial shear strength of the composites, the pullout simulations of a nanotube were performed. The nanotube was pulled out of the epoxy resin along the tube axis direction z. The snapshots of the pullout simulations are shown in Fig. 2. The potential energy, interaction energy, and interfacial bonding energy of the composite system were plotted against the displacement of the nanotube from the epoxy resin, as shown in Fig. 3.

The potential energy of the nanotube/epoxy resin composite system was increased as the nanotube was pulled out of the epoxy resin, as shown in Fig. 3a. In fact, the increase of the potential energy resulted from the decrease of the interaction energy and the increase of the energies of both the nanotube and epoxy resin. The interaction energy increased due to the decrease of contact area during the pullout. In the pullout simulation, the nanotube and epoxy resin were not held fixed. The potential energy of the nanotube and epoxy resin increased due to the changes of their configurations during the pullout. The deformation of the nanotube and epoxy resin during the pullout has influence on the pullout energy. During the pullout, the interaction energy changed with the displacement linearly, as shown in Fig. 3b. This is due to the stable interfacial binding interaction between the nanotube and epoxy resin. The interfacial binding energy kept constant with a value of 0.1 kcal/mol · A^2 during the pullout, as shown in Fig. 3c. After the nanotube was completely pulled out of the epoxy resin, the potential energy of the system was leveled off because there was no interaction between the nanotube and epoxy resin. The interaction energy then kept zero, and there was no change in the potential energies of the nanotube and epoxy resin.

The pullout energy, E_{pullout}, is defined as the energy difference between the fully embedded nanotube and the complete pullout configuration. The pullout energy was divided into three terms, which included the energy change in the nanotube, polymer, and their interaction as follows:

$$E_{\text{Pullout}} = E_2 - E_1 = (E_{\text{NT2}} - E_{\text{NT1}}) + (E_{resin2} - E_{resin1}) + (\Delta E_2 - \Delta E_1) \tag{4}$$

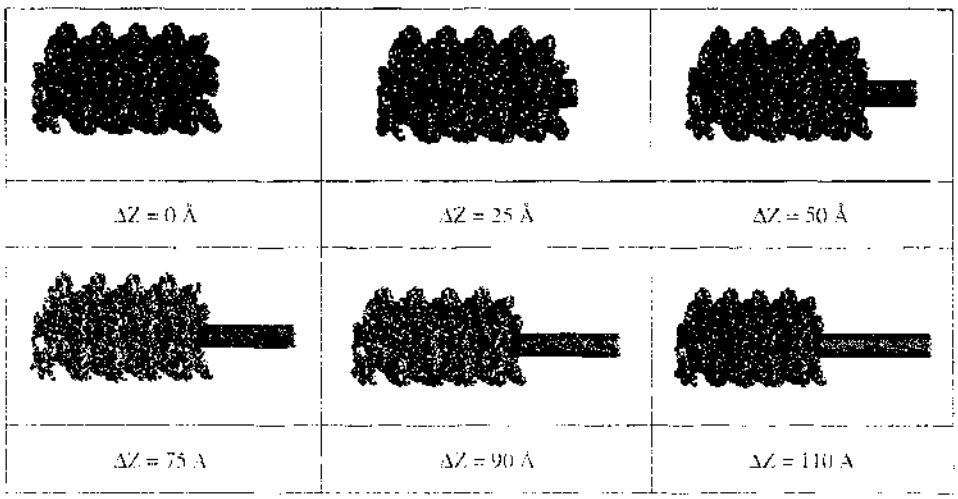


Figure 2. Snapshots from the pullout simulations of the nanotube. Reprinted with permission from [49], J. Gou et al., *Comput. Mater. Sci.* 31, 225 (2004), © 2004, Elsevier.

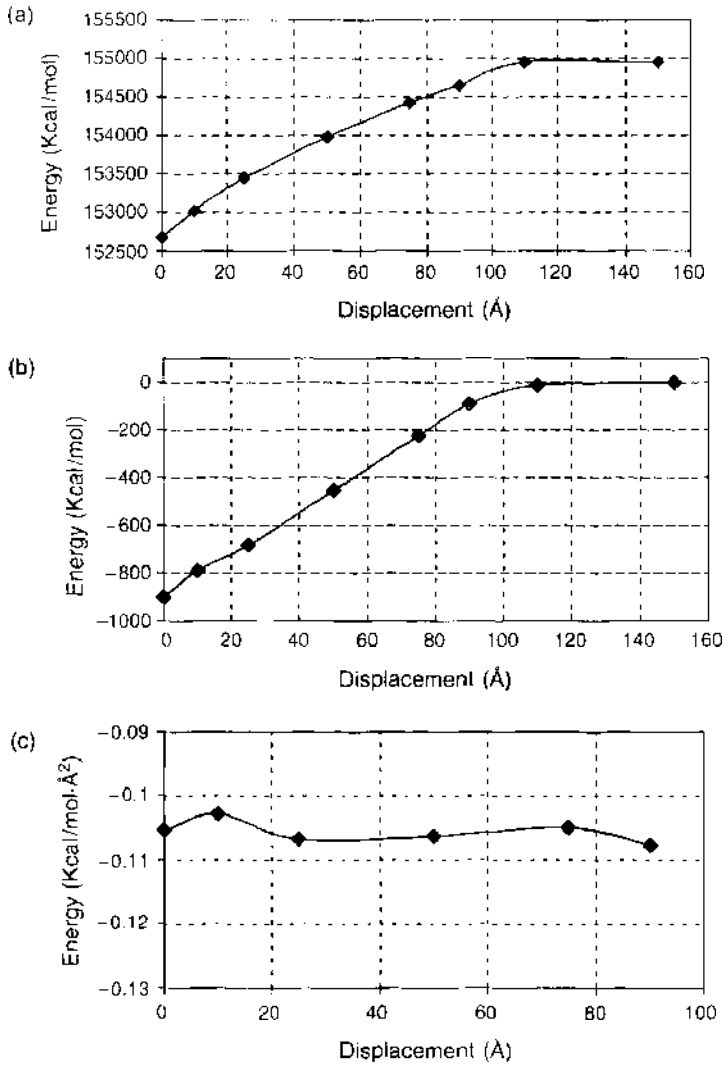


Figure 3. Energy plots during the pullout of the nanotube: (a) potential energy, (b) interaction energy, and (c) interfacial bonding energy. Reprinted with permission from [49], J. Gou et al., *Comput. Mater. Sci.* 31, 225 (2004). © 2004, Elsevier.

where E_{NT}, E_{resin} is the potential energy of the nanotube and epoxy resin, respectively, and ΔE is the interaction energy between the nanotube and epoxy resin. The pullout energy can be related to the interfacial shear stress, τ_i, by the following relation

$$E_{\text{pullout}} = \int_0^L 2\pi r(L-x)\tau_i\,dx = \pi r\,\tau_i L^2 \tag{5}$$

$$\tau_i = \frac{E_{\text{pullout}}}{\pi r L^2} \tag{6}$$

where r and L are the radius and length of the nanotube, respectively, and x is the displacement of the nanotube. At the initial stage of the pullout, the potential energy of the composite was 152,665 kcal/mol. After the pullout, the potential energy increased to 154,951 kcal/mol. From the pullout simulations, the interfacial shear strength between the nanotube and epoxy resin was about 75 MPa.

Liao and Li [53] studied the mechanical interlocking due to the mismatch in the coefficient of thermal expansion (CTE) between the nanotube and polystyrene (PS) in the composites. The thermal residual radial stress was estimated to be about −40 MPa/K from the three-phase concentric cylinder model of elasticity for a unidirectional composite [54]. In addition, they performed the pullout simulations for both single-walled and double-walled nanotubes from the PS matrix based on molecular mechanics. In the absence of atomic bonding between

the nanotube and PS matrix, the physical attraction came from the nonbonded interactions, which consisted of electrostatic and van der Waals interaction, and deformation induced by these forces. The pullout energies for single-walled and double-walled nanotubes are 158.7 and 154.4 kcal/mol, respectively. Based on Eq. (6), the interfacial shear stress between the nanotube and the PS matrix, τ_i, estimated from the molecular simulation, was calculated to be 160 MPa, which is significantly higher than most carbon fiber-reinforced polymer composites.

2.2. Molecular Dynamics Simulation

A second approach to modeling the interfacial bonding between the nanotube and polymer matrix is based on molecular dynamics simulations. Molecular dynamics simulations can generate information at the molecular level, including atomic positions and velocities. They can be performed to analyze the interfacial sliding between the nanotube and polymer matrix in the composites during the pullout process. At the molecular level, the interfacial sliding is related to the fundamental origins of the sliding friction. The physical principles of sliding friction can be found in [55].

Frankland et al. [56–57] studied the interfacial sliding during the entrie pullout process of the nonfunctionalized and functionalized nanotube/polyethylene (PE) composite systems through molecular dynamics simulations. In the functionalized nanotube composite system, a total of six cross-links containing two methylene units each were created between the nanotube and PE matrix. In their study, a unidirectional force was applied to each atom of the nanotube along the nanotube axis. The applied force was increased incrementally over time. After the molecular dynamics simulations, the velocity and displacement of the nanotube were recorded to characterize the interfacial interactions during the pullout.

The conversion of molecular dynamics simulation results to the interfacial shear stress requires appropriate frictional models. Frankland et al. further developed an interfacial friction model with a relation between the applied force and the velocity of the nanotube. A detailed derivation of their interfacial friction model can be found in [56], from which we summarized the main points as follows.

They adopted the Newton's constitutive law to calculate the shear stress τ [58]:

$$\tau = \tau_0 + \mu_m \dot{\gamma} \tag{7}$$

where τ_0 is the yield stress and μ_m and $\dot{\gamma}$ are the viscosities of a fluid and a viscous matrix material, respectively. Based on Newton's constitutive law, the nanoscale friction model was developed:

$$\langle \tau_{rz} \rangle_{\text{pull}} = \tau_0 + \mu_{\text{eff}} \frac{\partial \langle V_z \rangle}{\partial r} = \tau_0 + \mu_{\text{eff}} \frac{\langle V_z \rangle}{\bar{h}_{\text{vdW}}} \tag{8}$$

where $\langle \tau_{rz} \rangle_{\text{pull}}$ is the total shear stress, μ_{eff} is the effective viscosity, V_z is the average nanotube velocity in the z-direction, and $\bar{h}_{\text{vdW}}$ is an average interfacial separation.

From the molecular dynamics simulation data, the average nanotube velocity $\langle V_z \rangle$ is linearly related to the average applied force, $\langle f \rangle$.

$$\langle f \rangle = \chi_{\text{eff}} \langle V_z \rangle \tag{9}$$

where χ_{eff} is the viscosity coefficient. The applied force, $\langle f \rangle$, is related to the shear stress, $\langle \tau_{rz} \rangle$, by the force balance:

$$\langle f \rangle = \langle \tau_{rz} \rangle A_{ss} \tag{10}$$

where A_{ss} is the interfacial area during the steady sliding, which is given by:

$$A_{ss} = 2\pi (R_{NT} + \bar{h}_{\text{vdW}}/2) L_{ss} \tag{11}$$

where R_{NT} is the radius of nanotube and L_{NT} is the length of nanotube. Therefore, the effective viscosity can be related to the viscosity coefficient as follows:

$$\mu_{eff} = \frac{\chi_{eff}\bar{h}_{vdW}}{2\pi(R_{NT} + \bar{h}_{vdW}/2)L_{NT}} \tag{12}$$

Then, the interfacial friction model for the entire pullout process is

$$\langle f \rangle_{pull} = f_0 + \chi_{eff}\langle V_z \rangle \tag{13}$$

where f_0 is the critical pullout force, which can be obtained directly from the molecular dynamics simulation.

Based on the molecular dynamics simulation data, the critical pullout force, f_0, and the force-velocity dependence, χ_{eff}, can be found. Because the average interfacial separation, $\bar{h}_{vdW}$, length of nanotube, L_{NT}, and the radius of nanotube, R_{NT}, are known, the effective viscosity, μ_{eff}, can be calculated by using Eq. (12). For the nonfunctionalized nanotube/crystalline PE composite system, the value of τ_c was calculated to be 2.8 MPa, which provided an estimate of the lower limit of shear strength because no electrostatic interactions were included to supplement the relatively weak Lennard-Jones forces in these simulations. For the functionalized nanotube/crystalline PE composite system, two sets of τ_c values were reported. The lower τ_c value of 6.8 MPa corresponded to the applied force at which a nanotube alone began to shear with respect to the PE matrix. The higher value of τ_c of 110 MPa was determined from the initial force at which the nanotube pulled the PE chains with it through the PE matrix. These results suggest that even a relatively low density of cross-link could have a large influence on the properties of the nanotube/polyethylene interface.

2.3. Analytical Modeling with Classical Continuum Mechanics

Much literature has addressed that classical elastic theories could be used to interpret the mechanical behavior of nanotube-based structures. Krishnan et al. [59] have first estimated the Young's modulus of nanotubes using a well-known vibration theory. A cantilevered nanotube was treated as an elastic rod subjected to a traverse vibration. Its vibrational behavior is governed by a fourth-order differential equation in space and is analogous in nature:

$$\frac{\partial^2 w(x)}{\partial t^2} + \frac{EI}{\rho A}\frac{\partial^4 w(x)}{\partial x^4} = 0 \tag{14}$$

and the solution to Eq. (14) is

$$w(x) = \cos(n^2 t)[B\cos(nx) + C\sin(nx) + D\cosh(nx) + E\sinh(nx)] \tag{15}$$

where E, I, ρ, and A denote the Young's modulus, the second moment of inertia, the density and cross-sectional area of the nanotube, respectively. B, C, D, and E are arbitrary constants that are dependent on the boundary conditions of the nanotube. By appropriately considering those boundary conditions of the clamped-free configuration of the nanotubes with the experimentally measured vibration amplitude, Young's modulus can be calculated by using the regression method. It was found that Young's modulus of the nanotubes with the diameter range of 1.0 to 1.5 nm was as high as 1.25 TPa. Ru [60] has also mentioned that the elastic rod or beam models can be adequately used to model the overall mechanical performances of the nanotubes in both static and dynamic loading conditions. Due to the aspect ratio of the nanotubes usually being large, a high possibility of failure mode of the nanotubes is on the buckling-driven debonding in composite structures. Euler buckling formula in Eq. (16) has been used in Ru's paper, to study the critical load of a nanotube, by treating it as a hinged elastic column, particularly for using these nanostructural elements as atomic force microscopy (AFM) tips.

$$P_{CR} = \frac{\pi^2 EI}{L_{CR}} \tag{16}$$

The buckling of a nanotube not only happens when they are subjected to a pure compressive load at its free end. Wang et al. [61] have analytically and numerically found that buckling under bending of nanotubes also occurs on their compressive side. The linear elastic properties of the nanotubes were used in the simulation, and the structural performance of the nanotubes after being bent compared well with experimental observation. Because much literature has addressed that the nanotubes possess elastic mechanical properties under small strain conditions [62–63], it is reasonable to use the linear elastic mechanical properties of the nanotubes at their low straining condition to study the global mechanical properties of their related composite structures. In the following, the stress transfer behavior of different types of carbon nanotube based on an analytical study through the use of elastic shell and local density approximation method is reviewed.

In the early stage of carbon nanotube–related research, some diverse results on the interfacial-bonding characteristic between the nanotube and polymer-based matrix have been found. Xu et al. [64] have addressed that high interfacial shear stress between a multiwalled nanotube and an epoxy matrix was observed from a fractured sample. Wagner [65] first used the Kelly-Tyson model, which has been widely used to study the matrix–fiber stress transfer mechanism in micrometer-size fiber composites, to study the interfacial shear strength between the nanotube and polymer matrix. Because it was found that the binding force between the inner layers is very low, and the sliding failure always occurs, only a single-walled system is of interest in his work. In his study, it was assumed that an externally applied stress to a nanotube/polymer composite is fully transferred to the nanotubes via a nanotube–matrix interfacial shear mechanism at the molecular level. A single nanotube cylindrical model was used to study the stress transfer properties of the composite shown in Fig. 4.

To consider the force balance in the composite system, the following equation is formed [66]:

$$\tau_{NT} d_i d_t = (\sigma_{NT} + d\sigma_{NT})\left(\frac{d_o^2 - d_i^2}{4}\right) - \sigma_{NT}\left(\frac{d_o^2 - d_i^2}{4}\right) \tag{17}$$

where τ_{NT} is the interfacial shear strength between the nanotube and matrix, σ_{NT} is the tensile strength of the nanotube, and d_t, d_o, and d_i are the length, outer diameter, and inner diameter of the nanotube, respectively. After integrating Eq. (17) and considering the critical length of a typical short-fiber system in composite structures, the interfacial shear strength can be written as:

$$\tau_{NT} = \sigma_{NT}\left[\frac{1}{2}\left(\frac{l_c}{d_o}\right)^{-1}\left(1 - \frac{d_i^2}{d_o^2}\right)\right] \tag{18}$$

where l_c/d_o is the critical aspect ratio of the nanotube and d_i/d_o is the diameter ratio. As an external applied stress of 50 GPa was used, the interfacial strength was calculated for critical length values of 100, 200, and 500 nm. It is concluded that the interfacial shear stress is affected by several factors, including the critical length and the outer diameter of the

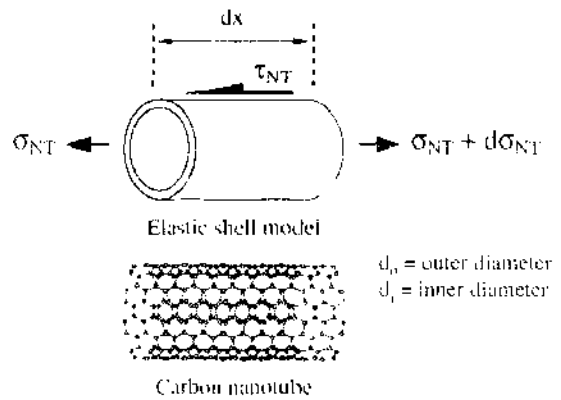


Figure 4. Schematic of elastic shell model that is used to simulate a single-walled nanotube structure.

nanotubes. The increase of the diameter of the nanotube results in increasing the interfacial shear strength at the bond interface.

Latterly, Lau [66] has conducted an analytical study on the interfacial bonding properties of nanotube/polymer composites using the well-developed local density approximation model [67], classical elastic shell theory [68], and conventional fiber-pullout model. Several parameters such as the wall thickness, Young's modulus, volume fraction, and chiral vectors of the nanotubes were considered in his calculations. In the past, most analytical studies used a constant Young's modulus E_{NT} to study the mechanical properties of nanotubes and nanotube/polymer composites. However, many experimental works have proved that the modulus of multiwalled nanotube was less than that of a single-walled structure. Tu and Ou-Yang [68] have used the local density approximation model cooperated with the elastic shell theory to estimate the Young's modulus of multiwalled carbon nanotube (MWNT) structures. They found that the Young's modulus of the nanotubes decreases with increasing the number of wall layers N. The Young's modulus E_{SWNT}, Poisson ratio ν_{NT}, and effective wall thickness h of the SWNT used in their study were 4.7 TPa, 0.34 and 0.75 Å, respectively. The radius of the first layer (an inner layer) of nanotubes could be determined by using the rolling graphene model:

$$\rho_0 = \frac{\sqrt{m^2 + n^2 + mn}}{\pi}\sqrt{3}a_0 \tag{19}$$

where ρ_0, m, n, and a_0 are the nonrelaxed radius and indices of the SWNTs and C–C bond distance (1.42 Å) in Fig. 5, respectively.

The Young's modulus E_{NT} of MWNTs estimated by using the local density approximation model and classical elastic shell theory can be expressed as [67]:

$$E_{NT} = \frac{N}{N - 1 + R}E_{SWNT}R \tag{20}$$

where R denotes the thickness to spacing ratio of the nanotubes (h/d). It is obvious that for $N = 1$, the $E_{NT} = E_{SWNT}$, which corresponds to the result of the SWNT. The parameter d is the spacing between each graphene layer (≈3.4 Å). The outer radius of the MWNTs could be determined by

$$R_{NT} = \rho_0 + d(N - 1) \tag{21}$$

In Fig. 6, the Young's modulus of the nanotubes with different number of wall layers is plotted. The Young's modulus of the nanotube decreases as the number of wall layers increases. When the number of layers exceeds 20, Young's modulus of the nanotubes becomes almost constant.

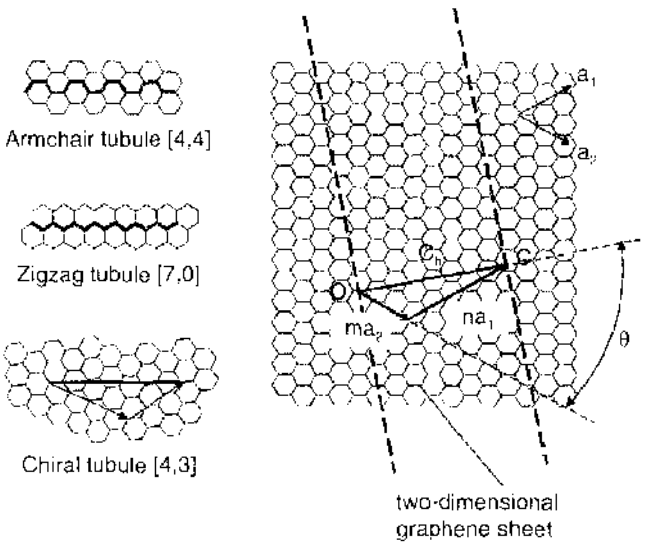


Figure 5. Chiral vector OC or $C_h = na_1 + ma_2$ is defined on the hexagonal lattice of carbon atoms (a graphene sheet) by unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and the chiral angle θ with respect to the zigzag axis.

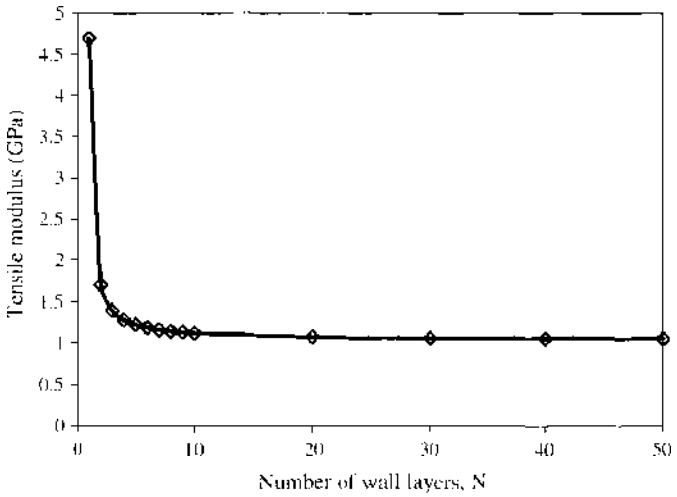


Figure 6. Tensile modulus of the nanotubes with different number of wall layers.

Lau [66] developed a simple pullout model for a MWNT/polymer composite. In Fig. 7a, the construction of the composite system is shown. The MWNT could be imagined as a group of coaxial circular shells packed together with a uniform interval spacing d and the effective wall thickness h. The effective cross sectional area of the nanotubes can be calculated by

$$A_{\text{eff}} = 2\pi h\left\{N\rho_0 + \sum_{C=1}^{N} d(C-1)\right\} \tag{22}$$

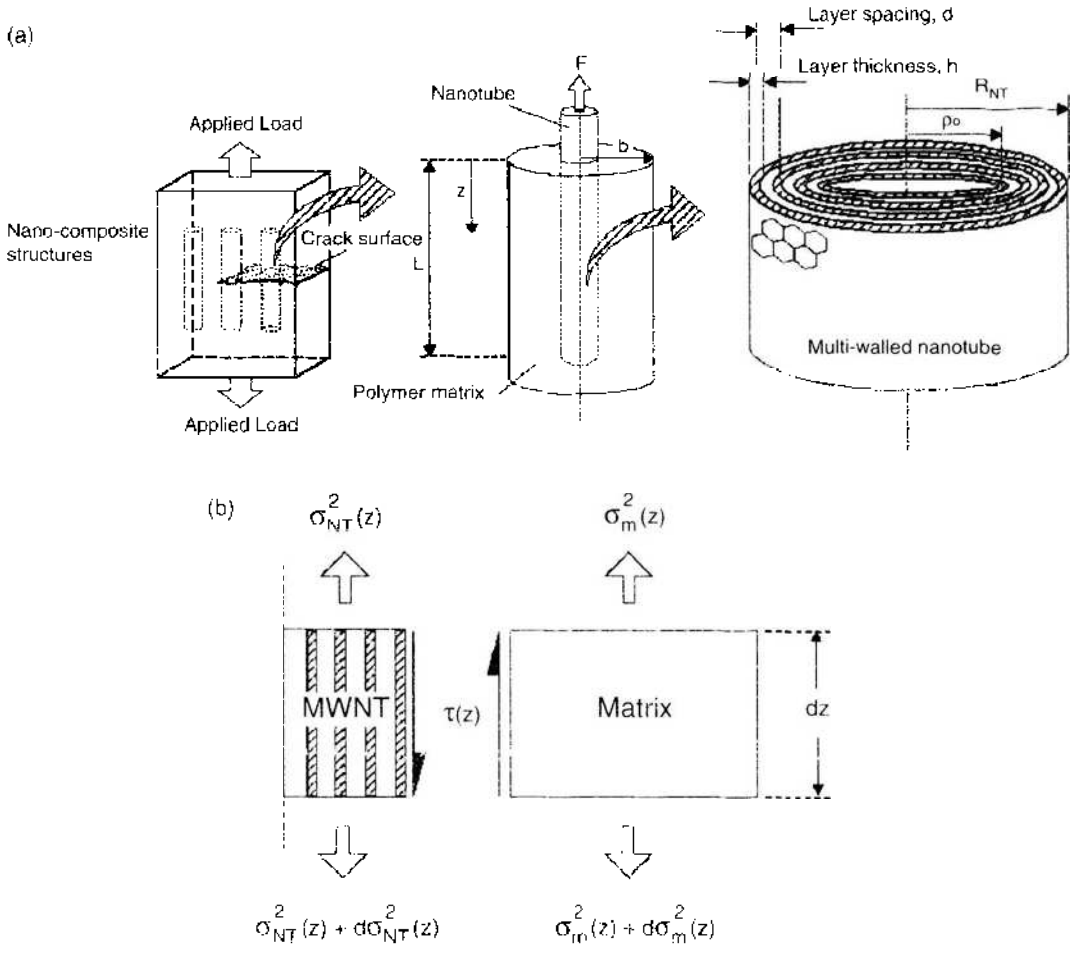


Figure 7. (a) A schematic diagram of the nanotube/polymer composites and (b) a simple fiber-pullout model. Reprinted with permission from [66], K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003), © 2003, Elsevier.

Because the role of the nanotubes in composite structures is to take up all stresses that are transferred from the matrix through the interfacial shear stress, their bonding properties between the nanotubes and matrix should be studied in detail. A well-known classical fiber-pullout model was used to study the interfacial bonding problem of nanotube/polymer composites. Three-dimensional, axisymmetric cylinder model was proposed to study the stress transfer behavior between the nanotubes and polymer matrix. The mechanical properties and geometrical factors that were used in Lau's study were followed according to the data given in [68] and Eqs. (19) and (20). The interfacial shear stress occurs when a nanotube is subjected to a pullout force, which is generated by a crack opening, propagating perpendicularly to the longitudinal axis of the nanotubes.

The nanotube with an outer radius of R_{NT} is located at the center of a coaxial cylinder matrix with an outer radius of b. The total embedding length of the nanotube is L. Assuming that the nanotube is undertaking the pullout force, F due to the existence of crack opening and the bonding force on the other end cap of the nanotube is very small, thus the axial stress applied on the other end is equal to zero. The boundary conditions at two ends of this model are:

$$\sigma_{NT}^{z}(0) = \sigma_{\text{pullout}} \tag{23a}$$

$$\sigma_{NT}^{z}(L) = 0 \tag{23b}$$

$$\sigma_{m}^{z}(0) = 0 \tag{24a}$$

$$\sigma_{m}^{z}(L) = \gamma\sigma_{\text{pullout}} \tag{24b}$$

where σ represents the stress, the superscripts denote the coordinates of the three principal axes (r, θ, z), and subscripts NT and m refer to the materials of the nanotube and matrix, respectively, as indicated in Fig. 7b. And γ is the area ratio of the nanotube and matrix, that is,

$$\gamma = \frac{R_{NT}^{2}}{b^{2} - R_{NT}^{2}} \tag{25}$$

The pullout stress, σ_{pullout}, can be determined by F/A_{eff}. The equilibrium between the axial stress of the nanotube and interfacial shear stress $\tau(z)$ in the nanotube's longitudinal direction can be expressed as [69–70]:

$$\frac{d\sigma_{NT}^{z}(z)}{dz} = -\frac{2}{R_{NT}}\tau(z) \tag{26}$$

$$\frac{d\sigma_{m}^{z}(z)}{dz} = \frac{2\gamma}{a}\tau(z) \tag{27}$$

By considering the boundary conditions given by Zhou [71], Lau et al. [69], Gao et al. [72], and the general equilibrium equations for a 3D axisymmetric problem [73] and the continuity of axial and radial deformations at any bonded interfaces, the axial stress in the nanotubes can be written in the following differential equation form:

$$\frac{d\sigma_{NT}^{z}(z)}{dz} - A_{1}\sigma_{NT}^{z}(z) - A_{2}\sigma_{\text{pullout}} = 0 \tag{28}$$

By using the boundary conditions at the ends of two nanotubes as indicated in Eqs. (23a), (23b) and (24a), (24b), and considering the relationship in Eq. (26), the equations of the nanotube's axial and the interfacial shear stresses can be obtained as:

$$\sigma_{NT}^{z}(z) = \omega_{1}\sinh(\lambda z) + \omega_{2}\cosh(\lambda z) - \frac{A_{2}}{A_{1}}\sigma_{\text{pullout}} \tag{29}$$

and

$$\tau(z) = \frac{-R_{NT}\lambda}{2}[\omega_{1}\cosh(\lambda z) + \omega_{2}\sinh(\lambda z)] \tag{30}$$

where the parameters of A_1, A_2, λ, ω_1, ω_2 are the functions of the mechanical properties and geometrical factors of the nanotube and matrix and are given as follows:

$$A_1 = \frac{\alpha(1-2k\upsilon_{NT}) + \gamma(1-2k\upsilon_m)}{U_2 - 2kU_1} \tag{31}$$

$$A_2 = \frac{-\gamma(1-2k\upsilon_m)}{U_2 - 2kU_1} \tag{32}$$

$$U_1 = \frac{\gamma}{8}\left\{2\eta_1 b^2 \ln\left(\frac{b}{R_{NT}}\right)\left[1+\gamma\left(\frac{b^2}{R_{NT}^2}+1\right)\right] - 2\eta_2(b^2+R_{NT}^2) + 4b^2 - 2\eta_1(b^2-R_{NT}^2)\right\} \tag{33}$$

$$U_2 = \frac{\upsilon_m\gamma}{4}\left[2\eta_1 b^2 \ln\left(\frac{b}{R_{NT}}\right)(1+\gamma) - \eta_2(b^2+R_{NT}^2) + 2b^2 - \eta_1(b^2-R_{NT}^2)\right] \tag{34}$$

$$k = \frac{\alpha\upsilon_{NT} + \gamma\upsilon_m}{\alpha(1-\upsilon_{NT}) + 1 + 2\gamma + \upsilon_m} \tag{35}$$

$$\alpha = \frac{E_{NT}}{E_m} \tag{36}$$

$$\eta_1 = \frac{2(1+\upsilon_m)}{\upsilon_m} \tag{37}$$

$$\eta_2 = \frac{1+2\upsilon_m}{\upsilon_m} \tag{38}$$

$$\lambda = \sqrt{A_1} \tag{39}$$

$$\omega_1 = \frac{\frac{A_2}{A_1} - \left(1+\frac{A_2}{A_1}\right)\cosh(\lambda L)}{\sinh(\lambda L)}\sigma_{\text{pullout}} \tag{40}$$

$$\omega_2 = \left(1+\frac{A_2}{A_1}\right)\sigma_{\text{pullout}} \tag{41}$$

For a simple fiber-pullout model, the maximum interfacial shear stress is located at a pull-out end region where $z = 0$ [69–72]. The magnitude of the shear stress is highly affected by the inherent properties and geometrical factors of the nanotubes and matrices. In Lau's study, the total embedding length of the nanotubes and the outer diameter of the cylinder were 10 μm and 4 μm, respectively. The Young's modulus and Poisson's ratio of the epoxy were 3.3 GPa and 0.48, respectively [74]. A pullout force of 1 nN is applied to the end of the nanotubes. In Fig. 8, it is clearly seen that the maximum shear stress of an SWNT/epoxy composite is comparatively higher than that of other composites made by MWNTs subjected to the same pullout force. The decreases of the maximum shear stress of MWNTs composites may be due to the decrease of the nanotube's modulus as the modulus ratio of the nanotube to the matrix (E_{NT}/E_m). Another reason may be due to the reduction of an axial stress in the nanotubes because of the increase of the effective cross-sectional area of the MWNTs. However, this equation fails to correct for the shear stress dropping to zero when $z = L$. In Fig. 9, plots of the nanotubes' modulus and volume fraction against the number of nanotube wall layers are shown. It is obvious that the decrease of the interfacial shear stress is not linearly proportional to the volume fraction of the nanotubes as the surface contact area. The figure shows that the increase in the amount of nanotube could eventually reduce the load taken by the nanotubes. Therefore, the interfacial bonding strength between the nanotube and matrix is reduced.

In Fig. 10, a plot of the interfacial shear stress of SWNTs with different chiralities is shown. It shows that the maximum shear stress of a zigzag nanotube is comparatively higher than that of chiral (5, 3) and armchair (5, 5) nanotubes. Because of the small cross-sectional area of the zigzag nanotubes as determined by Eq. (19), the total surface contact area at the bond interface of the zigzag nanotube is therefore comparatively smaller than that of the others, and hence, a higher interface shear stress is generated. Although it is shown

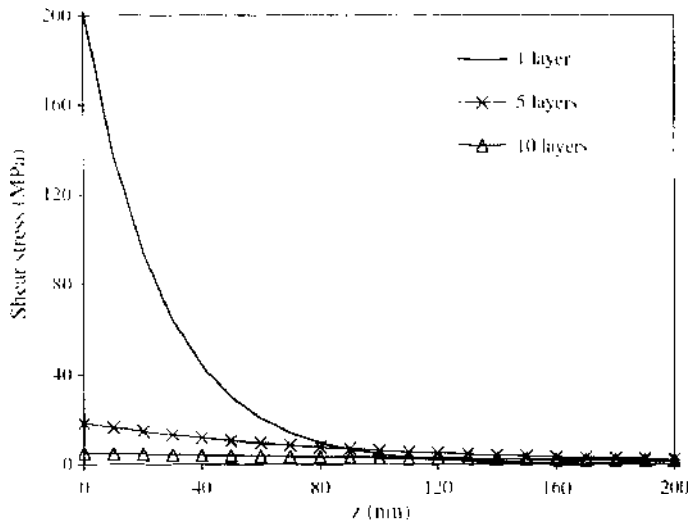


Figure 8. Shear stress distribution of the nanotube/polymer composites with different number of nanotube wall layers. Reprinted with permission from [66], K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003). © 2003, Elsevier.

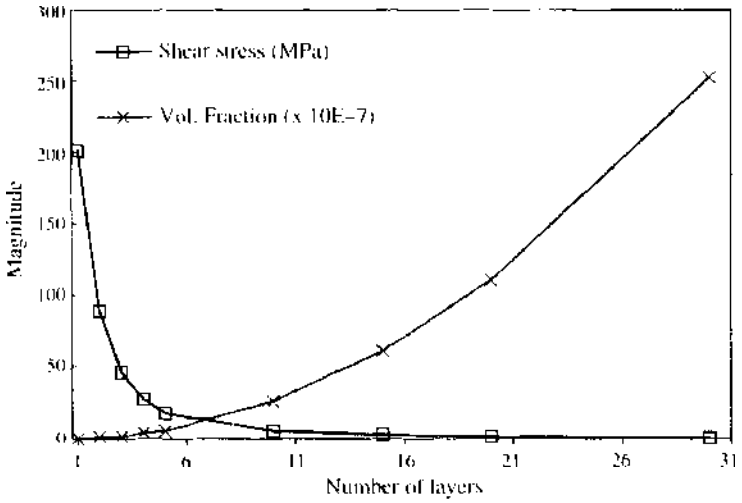


Figure 9. Relationship between the maximum shear stress, nanotube volume fraction, and number of nanotube wall layers. Reprinted with permission from [66], K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003). © 2003, Elsevier.

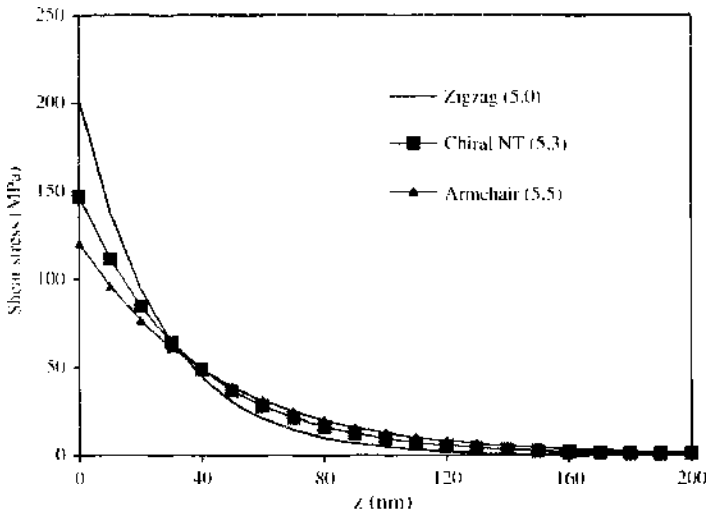


Figure 10. Shear stress distribution of the nanotube/polymer composites with embedding different chiral nanotubes. Reprinted with permission from [66], K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003). © 2003, Elsevier.

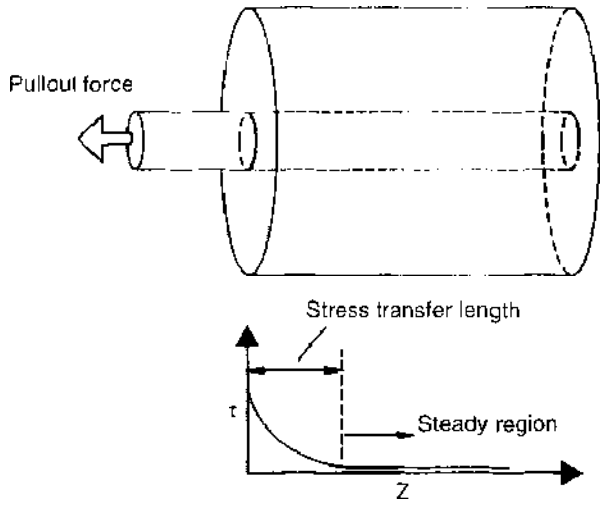


Figure 11. Schematic illustration of the stress transfer mechanism of the nanotube/polymer composite. Reprinted with permission from [66], K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003). © 2003, Elsevier.

that the maximum shear stresses of the chiral and armchair nanotubes are lower compared with the zigzag nanotubes, they required a long stress transfer length to allow all stresses to be fully transferred from the matrix to the nanotubes, as shown in Fig. 11. Therefore, according to the trend of the curves shown in the figure, the maximum shear stress at the interface of the nanotube is not only affected by the Young's modulus but also by the cross-sectional area of the nanotubes. In Fig. 12, the relationship between the diameter and Young's modulus of the single-walled nanotube and the interfacial shear stress between the nanotube and matrix is plotted. In this figure, it is clearly demonstrated that the interfacial shear stress is highly affected by the diameter and modulus of the nanotubes.

It is well-known that the maximum shear stress should be located at the starting point of the bond where $z = 0$, therefore the allowable pullout force (P_{cr}) and the tensile stress ($\sigma_{f,c}$) in the nanotube/polymer composite could be estimated by rearranging Eq. (30) once the allowable interfacial shear stress (τ_{max}) is known by using the following equations:

$$P_{cr} = \frac{2\tau_{max}\sinh(\lambda L)}{R_{NT}\lambda A_{eff}\left[\frac{A_2}{A_1} - \left(1 + \frac{A_2}{A_1}\right)\cosh(\lambda L)\right]} \tag{42}$$

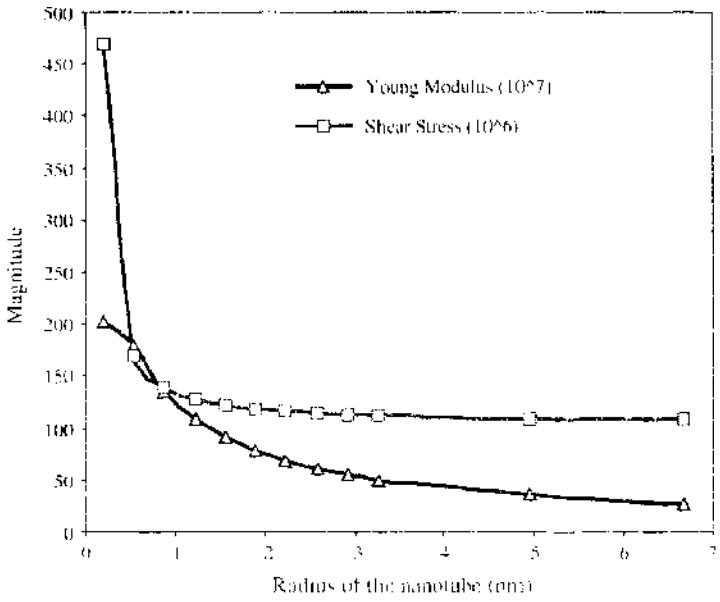


Figure 12. Relationship between the Young's modulus and radius of the nanotubes and the maximum shear stress of the nanotube polymer composites.

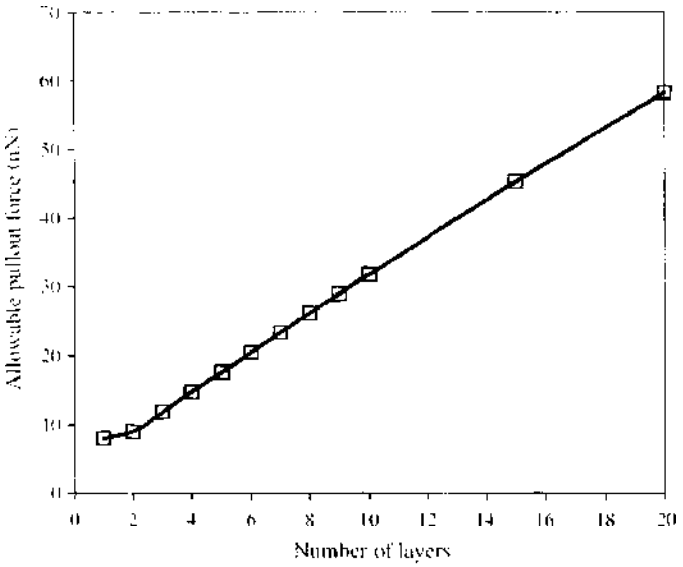


Figure 13. Maximum allowable pullout force of the nanotubes. Reprinted with permission from [66]. K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003). © 2003, Elsevier.

and

$$\sigma_{cr}^{*} = \frac{P_{cr}}{A_{\text{eff}}} v_{\text{NT}} + \sigma_{m}^{*}(1 - v_{\text{NT}}) \tag{43}$$

where v_{NT} and σ_{m}^{*} denote the volume fraction of the nanotubes, and the stress in the matrix at the nanotube stress is equal to P_{cr}/A_{eff}. Liao and Li [53] have recently determined by using molecular simulation that the maximum interfacial shear stress where a nanotube was pulled out of a nanotube/polymer system is about 160 MPa. By substituting this value into Eq. (43), it is possible to determine the maximum nanotube pullout force. In Fig. 13, it shows that the allowable pullout force of the nanotubes increases with increasing the number of the wall layers, thus decreasing the interfacial shear stress between the MWNTs and matrix. The increase in the number of layers and embedding length of nanotubes would result in decreasing the induced shear stress at the interface. Both factors are significant in the design of carbon nanotube/polymer composites.

3. MECHANICAL BEHAVIOR OF CARBON NANOTUBE/POLYMER COMPOSITES

Compared to carbon fiber-reinforced polymer composites, carbon nanotube/polymer composites hold the potential to provide significant improvements in strength, stiffness, and toughness. In addition, the remarkable property-to-weight ratio of carbon nanotube/polymer composites makes them ideal candidates for structural applications. To use successfully carbon nanotube/polymer composites in engineering applications, it is significant to predict the structural responses of this new class of nanocomposites under different loading conditions. The modeling and simulation of nanocomposite materials can play very important roles in this development and has been a matter of significant discussion in the literature. The focus of this section is on the computational methods to predict the overall mechanical responses of carbon nanotube/polymer composites.

There are two kinds of techniques in materials modeling: discrete modeling and continuum modeling. In discrete modeling, the material is an assemblage of atoms, which are considered as individual particles. It is obvious that discrete modeling (i.e. molecular dynamics simulations) can be applied in the study of mechanical behavior of nanocomposites. In continuum modeling, the region of materials is considered as a continuum. In the past few years, the mechanical behavior of carbon nanotubes and carbon nanotube/polymer composites has been studied with comprehensive computational approaches from discrete modeling to continuum modeling. In this section, we will review the major modeling and simulation methods.

including molecular mechanics and molecular dynamics simulations, continuum modeling, and multiscale modeling.

3.1. Molecular Mechanics and Molecular Dynamics Simulations

In the early stage of carbon nanotube–related research, molecular mechanics and molecular dynamics simulations have been used to estimate the mechanical properties of individual nanotubes or bundles [75–81]. Generally, the atomic interactions in carbon nanotubes are determined by the force potentials from the bonded and nonbonded interactions. The nonbonded interactions are either due to the van der Waals force or electrostatic interactions. The van der Waals force is most often modeled using the Lennard-Jones potential function [82] originally derived for inert gases. The general form of this potential is

$$\Phi(r) = \frac{\lambda_n}{r^n} - \frac{\lambda_m}{r^m} \tag{44}$$

For van der Waals forces arising from dipole–dipole interactions, the attractive part corresponds to $m = 6$. The most common form of this potential is the so-called (6–12) form:

$$\Phi(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{45}$$

The minimum of $\Phi(r)$ is determined by calculating the first-order derivative of $\Phi(r)$ *versus* r and equating it to zero. The van der Waals force between two carbon atoms can be estimated from:

$$F_{\text{vdW}} = -\frac{d\phi}{dr} = \frac{24\varepsilon}{r}\left[2\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{46}$$

The two parameters, σ and ε, can be estimated from experimental data such as the equilibrium bond length (lattice parameters at equilibrium), equilibrium bond energy (cohesive energy), and bulk modulus at equilibrium. The bonding energy (E_{bond}) is the sum of four different interactions among atoms, namely bond stretching (U_r), angle variation (U_θ), inversion (U_ω), and torsion (U_τ) [83].

$$E_{\text{Bond}} = U_r + U_\theta + U_\omega + U_\tau \tag{47}$$

A schematic illustration of each energy term and corresponding bond structure for a graphene cell is shown in Fig. 14. The most commonly used functional forms are:

$$U_r = \frac{1}{2}\sum_i K_i (dR_i)^2 \tag{48}$$

$$U_\theta = \frac{1}{2}\sum_j C_j (d\theta_j)^2 \tag{49}$$

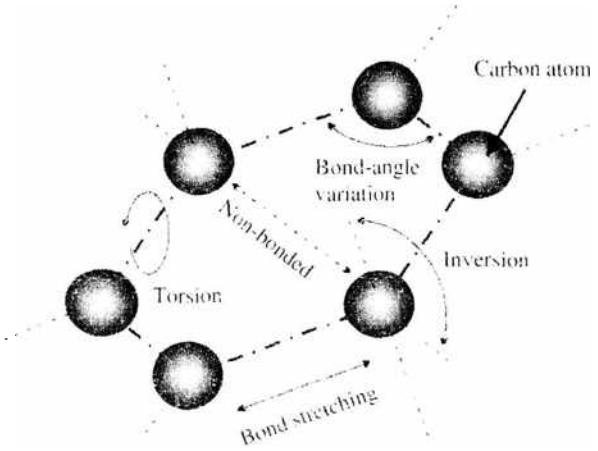


Figure 14. Bond structures and corresponding energy terms of a graphene cell.

$$U_{\omega} = \frac{1}{2}\sum_{k} B_k (d\omega_k)^2 \tag{50}$$

$$U_{\tau} = \frac{1}{2}\sum_{i} A_i [1 + \cos(n_i \tau_i - \phi_i)] \tag{51}$$

where dR_i is the elongation of the bond identified by the label i, K_i is the force constant associated with the stretching of the "i" bond, and $d\theta_j$ and $d\omega_k$ are the variance of bond angle j and inversion angle k, respectively. C_j and B_k are force constants associated with angle variance and inversion, respectively. A_i is the "barrier" height to rotation of the bond i; n_i is multiplicity, which gives the number of minimums as the bond is rotated through 2π [83].

To determine the tensile modulus of a single-walled nanotube subjected to uniaxial loadings, it is useful to observe that at small strains the torsion, the inversion, the van der Waals, and the electrostatic interactions energy terms are small compared with the bond stretching and the angle variation terms. Thus, the total energy of the single-walled nanotube can be reduced to:

$$E_{\text{Total}} = \frac{1}{2}\sum_{i} K_i (dR_i)^2 + \frac{1}{2}\sum_{j} C_j (d\theta_j)^2 \tag{52}$$

The force constants K_i and C_i can be obtained from quantum mechanics (*ab initio*). The average macroscopic elastic modulus and Poisson's ratio were estimated to be about 1.347 TPa and 0.261, respectively [83]. Such calculations may be performed either using the force or the energy approach, by measuring the mechanical forces developed between carbon atoms in nanotubes with different chiral arrangements.

Based on the calculations of those energy terms, molecular dynamics simulations have been used to predict the mechanical behavior of carbon nanotube/polymer composites. One example of recent work is the study of the stress–strain behavior of carbon nanotube/polyethylene composites under isostrain loading condition conducted by Frankland et al. [84]. In their study, the overall mechanical responses of two unidirectional nanotube/polyethylene composite systems, long-nanotube and short-nanotube composites, have been studied when subjected to longitudinal and transverse loading conditions. The application of strain was accomplished by uniformly expanding the dimensions of the length of the simulation cell and the coordinates of the atoms in the direction of the deformation. The molecular dynamics simulation or potential energy minimization was then carried out to equilibrate the composite system and to measure the corresponding stress. The method to determine the stress corresponding to the applied strain is summarized as follows.

For the linear-elastic material, the stress to the specific strain value is defined as the change in the internal energy with respect to the strain per unit volume [85]:

$$\sigma_{ij} = \frac{1}{V}\left(\frac{\partial E}{\partial \varepsilon_{ij}}\right)_S \tag{53}$$

where V is the volume of the solid, E is the total internal energy, ε_{ij} is the strain tensor, and the subscript S denotes constant entropy. At the atomic level, the total internal energy can be expressed as the summation of the energies of the individual atoms, E^{α}:

$$E^{\alpha} = T^{\alpha} + U^{\alpha} = \frac{1}{2} M^{\alpha} (v^{\alpha})^2 + \Phi^{\alpha}(r) \tag{54}$$

where for each atom α, T^{α} is the kinetic energy, U^{α} is the potential energy, M^{α} is the mass, v^{α} is the magnitude of its velocity, and $\Phi^{\alpha}(r)$ is the potential energy at the atom location r. Therefore, the stress components were calculated for each strain increment by using

$$\sigma_{ij} = -\frac{1}{V}\sum_{\alpha}\left(M^{\alpha} v_j^{\alpha} v_i^{\alpha} + \sum_{\beta} F_i^{\alpha\beta} r_j^{\alpha\beta}\right) \tag{55}$$

where V is the volume of the MD model and $V = \sum_{\alpha} V^{\alpha}$. Based on the molecular dynamics simulations results, the stress–strain curves of both long-nanotube and short-nanotube

composites were generated. It was found that the long-nanotube composite shows an increase in the stiffness relative to the polymer and behaves anisotropically under the different loading conditions. However, the short-nanotube composite shows no enhancement relative to the polymer, most probably because of its low aspect ratio. Griebel et al. [86] also examined the elastic moduli of long- and short-nanotube composites through molecular dynamics simulations. They used Parrinello-Rahman approach to apply external stress to the periodic model of the nanotube/PE composite. They compared the elastic moduli of carbon nanotube/polymer composites with the rule-of-mixture predictions and found the rule-of-mixture worked well for the long-nanotube composites but fails for the short-nanotube composites.

3.2. Continuum Modeling

Continuum modeling based on continuum mechanics has been applied for investigating the overall mechanical deformation of carbon nanotubes with the aspect ratio greater than 10^3. Sohelberg et al. [87] have addressed that the vibration analysis can be used to study the structural rigidity based on continuum methods of carbon nanotubes. Drexler [88] has pointed out that vibrations in nanostructures will give rise to the issue of positional uncertainty, and that high structural rigidity will minimize this uncertainty. This is actually the major argument for the use of diamondoid covalent solids in nanomachine construction. The discussion has illustrated that vibrational motions in the nanostructures can, in addition to predicting structural integrity, impair performance and introduce positional uncertainty. In fact, the operational performance of nanostructural systems is highly dependent on the ability to anticipate reliably the nanostructure vibrational properties. Sohelberg et al. [87] have first proposed the use of continuum methods to study the vibrational performance, such as natural frequency of carbon nanotubes. The nanotubes, as long narrow molecules composed of repeated units, can be simulated using classical trajectory methods. Sumpter and Noid [89] have studied the positional instability in graphitic nanotubes. In their work, they focused on three modes of vibration—longitudinal stretching, circumferential breathing, and transverse flexion—and considered the continuum description of five basic types of vibration in an unconstrained [illegible]be (free vibrated) as shown in Fig. 15.

In their study, as Timoshensko beam, a long circular shaft with a constant cross-sectional area, aligned along the shaft's longitudinal axis, z, was considered. A well-known wave equation was used to simulate the vibration behavior of the shaft:

$$\frac{\partial^2 u}{\partial z^2} = \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} \tag{56}$$

where t and u are the time and the longitudinal displacement of a point on the cross section at z, respectively. The constant a is given by:

$$a = \left(\frac{E}{\rho}\right)^{1/2} \tag{57}$$

where E and ρ are the Young's modulus and mass density per unit volume, respectively, of the shaft. Subjected to the boundary condition that both ends of the shaft are free to move,

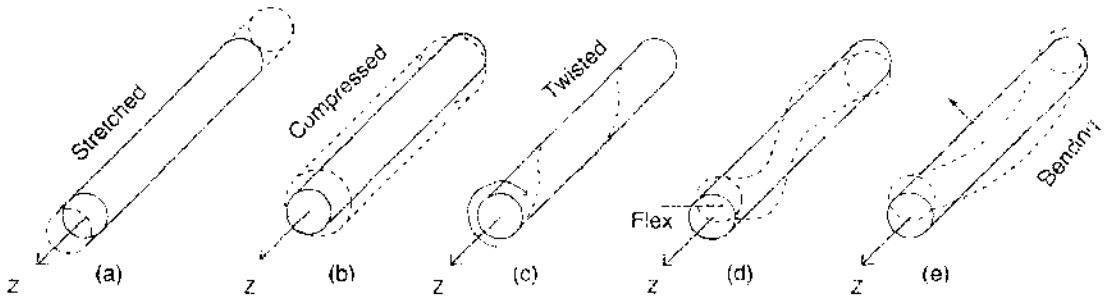


Figure 15. Some vibrational modes of a cylinder: (a) longitudinal stretching, (b) circumferential breathing, (c) torsional twist, (d) transverse flex with an odd number of nodes, and (e) transverse flex with an even number of nodes.

the solutions demand an integer number of half-wavelengths over the length L, which results in a discrete spectrum of allowed vibrational frequencies,

$$\omega_{\mathrm{l}}^{i} = \frac{i\pi}{L}\left(\frac{E}{\rho}\right)^{1/2} \tag{58}$$

By applying appropriate boundary conditions of different modes of vibration addressed in [89], the natural frequencies of the shaft at different vibration modes can be determined by the following equations:

Longitudinal stretching frequency:

$$\omega_{\mathrm{L}}^{0} = \frac{57}{L} \tag{59}$$

Circumferential breathing frequency:

$$\omega_{\mathrm{C}}^{0} = \frac{46}{2\pi r} \tag{60}$$

Torsional twisting frequency:

$$\omega_{\mathrm{T}}^{i} = \frac{i\pi}{L}\left(\frac{G}{\rho}\right)^{1/2} \tag{61}$$

where $i = 1, 2, 3\ldots$, r is the radius of the nanotube. From Eqs. (59) to (61), it is found that the fundamental frequencies of longitudinal stretching and torsional twisting modes are independent of the tube's diameter, while the circumferential breathing mode is independent of the tube length. However, we can see that resonance between longitudinal stretching [Eq. (59)] and torsional twisting [Eq. (61)] will occur when the harmonic indices are in a critical ratio related to the elastic and torsional moduli:

$$\frac{\omega_{\mathrm{L}}^{i}}{\omega_{\mathrm{T}}^{i}} = \left(\frac{G}{E}\right)^{1/2} \tag{62}$$

The transverse (also called lateral) flexions shown in Fig. 15d and Fig. 15e can be separated with an odd number of vibrational nodes (Fig. 15d) from those with an even number (Fig. 15e) because of their different symmetry properties. The transverse flexing vibrations are governed by the following differential equation:

$$\frac{\partial^2 v}{\partial z^4} = \frac{-1}{\alpha^2}\frac{\partial^2 v}{\partial t^2} \tag{63}$$

where v is the lateral displacement at position z and

$$\alpha = \sqrt{\frac{EI}{\rho A}} \tag{64}$$

where I and A are the second moment of inertia and cross-sectional area, respectively, of the nanotube. The boundary conditions for free ends give rise to a transcendental equation for the allowed vibrational frequencies ω_{F}^{i} owing to the fourth-order differential equation,

$$\cos\left[L\left(\frac{\omega_{\mathrm{F}}^{i}}{\alpha}\right)^{1/2}\right]\cosh\left[L\left(\frac{\omega_{\mathrm{F}}^{i}}{\alpha}\right)^{1/2}\right] = 1 \tag{65}$$

where $i = 1, 2, 3\ldots$ ($i = 0$ corresponds to tube translation). The above equation can easily be solved by numerical methods. The numerical methods, such as the finite element and boundary element methods, can be applied readily for solving the 3D multidomain elasticity problems.

In the carbon nanotube/polymer composites, the nanotubes are in different sizes and configurations, and their orientations in the polymer matrix could be aligned or random. Their structural characteristics make the prediction of the mechanical responses of carbon

nanotube/polymer composites extremely complicated. Recently, Chen and Liu [90–92] have applied the concept of representative volume elements (RVEs) to extract the mechanical properties of carbon nanotube/polymer composites based on 3D elasticity theory and finite element simulations. In their RVE approach, a single nanotube with surrounding polymer matrix was modeled, with properly applied boundary and interface conditions to account for the effects of the surrounding polymer matrix. According to the shape of the cross section, three types of representative volume elements are constructed: cylindrical, square, and hexagonal RVEs shown in Fig. 16. Specifically, the cylindrical RVE is used to model carbon nanotubes with different diameters. The square RVE is applied when carbon nanotubes are arranged evenly in a square pattern, while the hexagonal RVE is adopted when carbon nanotubes are in a hexagonal pattern. Compared to the square RVE, the cylindrical RVE tends to overestimate the effective Young's moduli of carbon nanotube/polymer composites. This can be explained by the fact that a cylindrical RVE overestimates the volume fraction of carbon nanotube due to the negligence of the small amount of polymer matrix. In this section, the formulas to extract the effective material constants of carbon nanotube/polymer composites for the square RVE are summarized as follows.

For the square RVE, four effective material constants are to be determined including Young's moduli E_x and E_z and Poisson's ratios ν_{xy} and ν_{zx}. To determine these four unknown material constants, four equations are required. Liu and Chen [90] devised the two loading cases, axial stretching and lateral loading, to provide the four such equations based on the elasticity theory. Generally, the strain–stress relation for a transversely isotropic material [93] is given as follows:

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{Bmatrix} = \begin{bmatrix} \frac{1}{E_x} & -\frac{\nu_{xy}}{E_x} & -\frac{\nu_{zx}}{E_z} \\ -\frac{\nu_{xy}}{E_x} & \frac{1}{E_x} & -\frac{\nu_{zx}}{E_z} \\ -\frac{\nu_{zx}}{E_z} & -\frac{\nu_{zx}}{E_z} & \frac{1}{E_z} \end{bmatrix} \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{Bmatrix} \tag{66}$$

For the square RVE under an axial stretch ΔL, by integrating and averaging the third equation in Eq. (66) on the plane $z = L/2$, the Young's modulus, E_z, can be obtained as:

$$E_z = \frac{\sigma_{\text{ave}}}{\varepsilon_z} = \frac{L}{\Delta L}\sigma_{\text{ave}} \tag{67}$$

where σ_{ave} is the averaged value of the stress σ_z and is given by:

$$\sigma_{\text{ave}} = \frac{1}{A}\int_A \sigma_z(x, y, L/2)\, dx\, dy \tag{68}$$

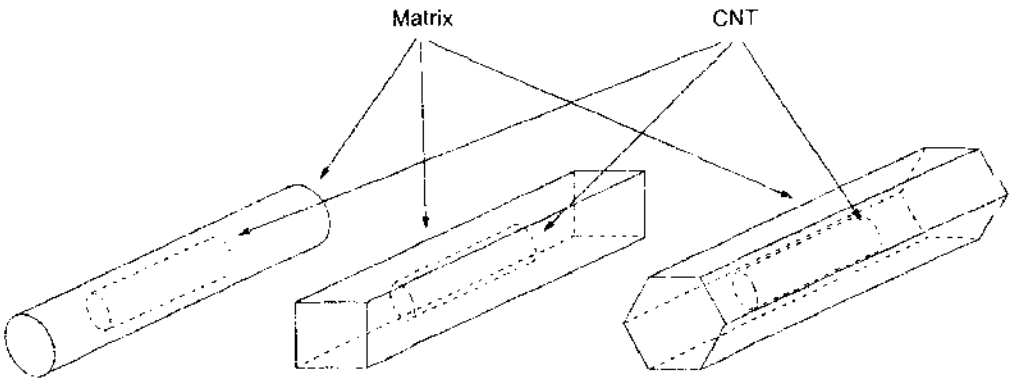


Figure 16. Schematic of three types of representative volume elements (RVEs): (a) cylindrical RVE, (b) square RVE, and (c) hexagonal RVE. Reprinted with permission from [90], X. L. Chen and Y. J. Liu, *Comput. Mater. Sci.* 29, 1 (2004), © 2004, Elsevier.

By applying the boundary conditions along $x = \pm a$ on the first equation in Eq. (66), the transverse strain, ε_x, can be written in the following form:

$$\varepsilon_x = -\frac{\nu_{zx}}{E_z}\sigma_z = -\nu_{xz}\varepsilon_x = -\nu_{zx}\frac{\Delta L}{L} = \frac{\Delta a}{a} \tag{69}$$

Then, the Poisson's ratio, ν_{zx}, can be calculated as follows:

$$\nu_{zx} = -\left(\frac{\Delta a}{a}\right)\Big/\left(\frac{\Delta L}{L}\right) \tag{70}$$

For square RVE under a lateral uniform load p, the plane strain condition is applied due to the constraint in the z-direction [$\varepsilon_z = 0$ and $\sigma_z = \nu_{zx}(\sigma_x + \sigma_y)$]. The stress–strain relation in (66) is reduced to

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \end{Bmatrix} = \begin{bmatrix} \frac{1}{E_x} - \frac{\nu_{zx}^2}{E_z} & -\frac{\nu_{xy}}{E_x} - \frac{\nu_{zx}^2}{E_z} \\ -\frac{\nu_{xy}}{E_x} - \frac{\nu_{zx}^2}{E_z} & \frac{1}{E_x} - \frac{\nu_{zx}^2}{E_z} \end{bmatrix} \begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} \tag{71}$$

Applying the boundary conditions along $x = \pm a$ and $y = \pm a$ on Eq. (71), the strains in x- and y-direction are calculated by

$$\varepsilon_x = -\left(\frac{\nu_{xy}}{E_x} + \frac{\nu_{zx}^2}{E_z}\right)p = \frac{\Delta x}{a} \tag{72}$$

$$\varepsilon_y = \left(\frac{1}{E_x} - \frac{\nu_{zx}^2}{E_z}\right)p = \frac{\Delta y}{a} \tag{73}$$

where Δx and Δy are the changes of dimensions in the x- and y-direction, respectively. By solving Eqs. (72) and (73), the effective Young's modulus and Poisson's ratio in the transverse direction can determined by the following equations:

$$E_x = E_y = \frac{1}{\frac{\Delta x}{pa} + \frac{\nu_{zx}^2}{E_z}} \tag{74}$$

$$\nu_{xy} = -\left(\frac{\Delta x}{pa} + \frac{\nu_{zx}^2}{E_z}\right)\Big/\left(\frac{\Delta y}{pa} + \frac{\nu_{zx}^2}{E_z}\right) \tag{75}$$

Chen and Liu [90] studied two numerical examples for carbon nanotube/polymer composites, one on a RVE with long nanotube and the other on a RVE with a short nanotube by using finite element method (FEM) simulations. The deformations and stresses were computed from the FEM simulations, and the effective Young's moduli and Poisson's ratios for carbon nanotube/polymer composites were then extracted using Eqs. (67), (70), (74), and (75).

3.3. Multiscale Modeling

Molecular dynamics simulations of carbon nanotube/polymer composites are currently limited to very small length and time scales and cannot deal with the larger length scales. However, carbon nanotube/polymer composites for real applications must expand from nano to micro and eventually to macro length scales. The continuum modeling has been employed for quite some time in the study of individual carbon nanotubes and nanotube/polymer composites. However, the validity of the continuum approach to modeling of carbon nanotubes and nanotube/polymer composites is still not well established, and the practice will continue to be questioned for some time to come. Therefore, the most promising method for simulations of carbon nanotubes and nanotube/polymer composites should be multiscale modeling techniques, where both molecular dynamics simulation and continuum modeling are integrated in a computing environment. Specifically, the multiscale modeling techniques

connect the atomistic models to continuum models by carrying the crucial information with the intrinsic nanoscale features from the molecular simulations to continuum simulations. For example, these modeling techniques are able to incorporate the interfacial characteristics of the nanotube and the surrounding polymer matrix at the molecular level into finite element modeling to predict the mechanical behavior of carbon nanotube/polymer composites. In this section, we will review two typical multiscale modeling techniques in the computational study of carbon nanotube/polymer composites, including molecular structural mechanics and equivalent-continuum mechanics.

3.3.1. Molecular Structural Mechanics

Li and Chou [94–95] first developed molecular structural mechanics to predict mechanical behavior of carbon nanotubes and nanotube/polymer composites. The concepts of molecular structural mechanics originated from the observation of geometric similarities between the nanoscopic fullerenes and macroscopic frame structures. Therefore, the fundamental to this modeling technique is the notion that a carbon nanotube can be regarded as a frame-like structure and the primary bonds between two nearest-neighboring atoms can be considered as load-bearing beam members, whereas an individual atom acts as the joints of the related load-bearing beam members, as shown in Fig. 17. Establishment of a linkage between structural mechanics and molecular mechanics enables the sectional property parameters of these beam members to be obtained. The major steps of the molecular structural mechanics approach are summarized as follows.

First, take atoms as joints and bonds as beams in the equivalent frame-like structure. The total potential energy of the molecular system is a sum of energies due to bonded and nonbonded interactions:

$$U = \sum U_r + \sum U_\theta + \sum U_\phi + \sum U_\omega + \sum U_{\text{vdW}} \tag{76}$$

where U_r is for bond stretch interaction, U_θ for bond angle bending, U_ϕ for dihedral angle torsion, U_ω for out-of-plane torsion, and U_{vdW} for nonbonded van der Waals interaction

$$U_r = \frac{1}{2}k_r(\Delta r)^2 \tag{77}$$

$$U_\theta = \frac{1}{2}k_\theta(\Delta\theta)^2 \tag{78}$$

$$U_\tau = U_\phi + U_\omega = \frac{1}{2}k_\tau(\Delta\phi)^2 \tag{79}$$

where k_r, k_θ, and k_τ are the bond stretching force constant, bond angle bending force constant, and torsional resistance, respectively. Δr, $\Delta\theta$, and $\Delta\phi$ represent the bond stretching increment, the bond angle change, and the angle change of bond twisting, respectively.

Second, determine beam sectional parameters from molecular force fields constants. Because the equivalent beam is assumed to be of round section, it can be assumed that $I_x = I_y = I$ and only three stiffness parameters, the tensile resistance EA, flexural rigidity EI, and torsional stiffness GJ, need to be determined for deformation analysis. From the

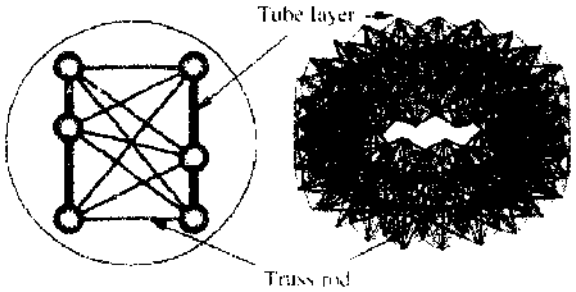


Figure 17. Truss model of double-walled carbon nanotube. Reprinted with permission from [96], C. Y. Li and T. W. Chou, *Composites Sci. Technol.* 63, 1517 (2003), © 2003, Elsevier.

viewpoint of structural mechanics, the deformation of a space frame results in the change of strain energy. The strain energy for a beam element is given by

$$U = \sum U_A + \sum U_M + \sum U_T \tag{80}$$

where U_A, U_M, and U_T are strain energies for axial tension, bending, and torsion, respectively.

$$U_A = \frac{1}{2}\frac{EA}{L}(\Delta L)^2 \tag{81}$$

$$U_M = \frac{1}{2}\frac{EI}{L}(2\alpha)^2 \tag{82}$$

$$U_T = \frac{1}{2}\frac{GJ}{L}(\Delta\beta)^2 \tag{83}$$

where L is the length of the equivalent beam, ΔL is the axial stretching deformation, α denotes the rotational angle at the ends of the beam, and $\Delta\beta$ is the relative rotation between the ends of the beam. By considering the energy equivalence between Eq. (76) and Eq. (80), a direct relationship between the structural mechanics parameters and molecular mechanics force constants can be established:

$$\frac{EA}{L} = k_r, \quad \frac{EI}{L} = k_\theta, \quad \frac{GJ}{L} = k_\tau \tag{84}$$

Third, establish elemental stiffness matrices and elemental load vectors and then assemble the global stiffness matrix and the global vector. Finally, solve the displacements of atoms and then compute the elastic properties of carbon nanotubes.

For the simulations of van der Waals interactions, the truss rod model is used [96]. The van der Waals force acting along the connecting line between two interacting atoms is simulated by a truss rod that connects the two interacting atoms with rotatable end joints. The truss rod, thus, transmits only tensile or compressive forces. Because the van der Waals force between two atoms is highly nonlinear, the truss rod connecting the atoms is a nonlinear element with its load–displacement relationship characterized by the van der Waals force, as shown in Fig. 18. The generalized displacement control method was used for simulating the nonlinearity of the truss rod. At the nanotube/polymer interface, the activation of a truss rod is determined by the distance between an atom in the nanotube and a node in the polymer. If the distance between an atom and a node in the polymer is less than 2.5σ ($\sigma = 0.34$ nm), a truss rod is activated. The center of the atoms of the nanotube is located in the midsection of the tube thickness, which is assumed to be 0.34 nm.

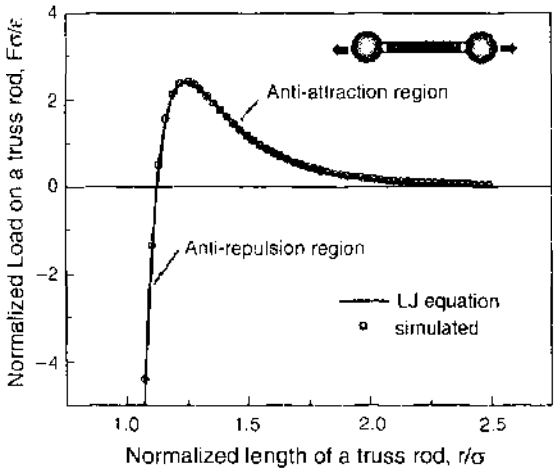


Figure 18. Load-displacement curve of the nonlinear truss rod. Reprinted with permission from [96], C. Y. Li and T. W. Chou, *Composites Sci. Technol.* 63, 1517 (2003) © 2003, Elsevier.

By combining the newly developed molecular structural mechanics approach and continuum finite element method, Li and Chou [97] further examined the effect of interfacial load transfer on the stress distribution in carbon nanotube/polymer composites. The nanotube was modeled through molecular structural mechanics approach at the atomistic scale. The polymer matrix was modeled through finite element method. In their study, the nanotube/polymer interface was assumed to be bonded either perfectly or by van der Waals. They examined the interfacial shear stress and axial stress distributions in the polymer matrix, axial stress profile in the nanotube, and the effect of nanotube aspect ratio on load transfer and concluded that the effective Young's modulus of the nanotube/polymer composite calculated from the simulation was in good agreement with that predicted from the rule-of-mixture. They also found the stress concentrations in the nanotube and the polymer matrix for various loading conditions, which may cause the failure of the nanotube and polymer matrix. The longer nanotube had higher load-carrying capability due to the larger contact region, although the maximum shear stress for various nanotube aspect ratios was close. It can be seen that the perfect interface had better load-carrying capability than the van der Waals interface due to the higher interfacial shear stress and axial stress in the nanotube.

3.3.2. Equivalent-Continuum Modeling

In their recent work, Odegard et al. [98] developed an equivalent-continuum method to determine the effective geometry and effective bending rigidity of a graphene structure. Based on this modeling technique, the nanotube/polymer interface was modeled as an effective continuum fiber to predict the bulk mechanical behavior of carbon nanotube/polymer composites. This method consists of three major steps shown in Fig. 19. Specifically, the equivalent-continuum modeling technique can be described as follows.

First, a suitable representative volume element of the nanotube/polymer composite system is constructed based on the molecular structure of nanotube and the surrounding polymer chains, which is obtained from molecular dynamics simulations. Therefore, the nanotube/polymer composite system is regarded as an assemblage of many atoms. According to molecular mechanics, the total potential energy, E^m, for the nanotube/polymer composite system is described by the sum of the individual energy contributions as follows:

$$E^m = \sum E^r + \sum E^\theta + \sum E^\tau + \sum E^\omega + \sum E^{nb} \tag{85}$$

where E^r, E^θ, E^τ, and E^ω are the energies associated with bond stretching, angle variation, torsion, and inversion, respectively, and E^{nb} is the energy of the nonbonded interactions, which includes van der Waals and electrostatic interactions. For example, the specific energy terms for bond stretching, bond-angle variance, and van der Waals interactions are

$$E^m = \sum_{\text{type bond}} \sum K^r (r - R)^2 + \sum_{\text{type bond}} \sum K^\theta (\theta - \Theta)^2 + \sum_{\text{type bond}} \sum \left(\frac{K^\alpha}{r^{12}} - \frac{K^\beta}{r^6} \right) \tag{86}$$

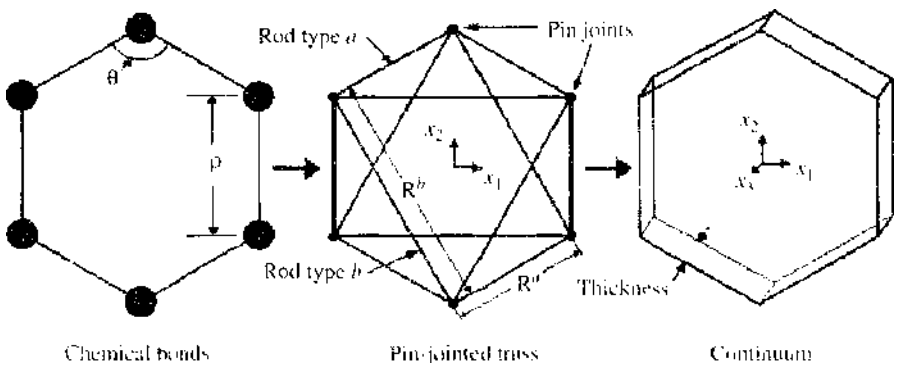


Figure 19. Representative volume elements for molecular, truss, and continuum models. Reprinted with permission from [98], G. M. Odegard et al., *Composites Sci. Technol.* 62, 1869 (2002). © 2002, Elsevier.

where P and Θ refer to the undeformed bond length and bond angle, respectively, and p and θ refer to the deformed quantities. The symbols K^r and K^θ represent the molecular mechanics force constants associated with the bond stretching and bond angle interactions, respectively, and both K^α and K^β represent the molecular mechanics force constants associated with the van der Waals interaction.

Second, an equivalent truss model of the RVE is developed to link the molecular model and equivalent-continuum model. In this modeling technique, each atom in the molecular model is represented by a pin-joint, and each truss element represents atomic bonded and nonbonded interactions. The values of the modulus of the truss elements are determined using the molecular mechanics force constants. Therefore, the total potential energy of the molecular model and the strain energy of the equivalent-truss model are equal under the same loading conditions. The total mechanical strain energy of the equivalent-truss model may take the form

$$E^t = \sum_{\text{type}} \sum_{\text{rod}} \frac{AY}{2R}(r - R)^2 \tag{87}$$

where A, Y, R, and r are the cross-sectional area, Young's modulus, and undeformed and deformed lengths of the truss elements, respectively. In order to represent the mechanical behavior with the truss model, Eq. (87) must be equated with Eq. (86) in a physically meaningful manner. For small deformations, the Young's moduli of the rods representing primary bond and the bond-angle variance interactions can then be determined as a function of the force constants:

$$Y^a = \frac{2K^r R^a}{A^a} \tag{88}$$

$$Y^b = \frac{3K^\theta}{2R^b A^b} \tag{89}$$

where the superscripts a and b are associated with the primary bonding and bond-angle variance interactions, respectively. However, the Young's modulus of the truss element that represents van der Waals force is more difficult to determine. Finally, an equivalent-continuum model of the RVE is constructed by equating the total strain energies under identical loading conditions. For example, the equivalent-truss model is replaced with an equivalent-continuous plate with a finite thickness, which is also called effective thickness.

Odegard et al. [99–100] established constitutive equations for both functionalized and nonfunctionalized nanotube composites by using equivalent-continuum modeling technique. In their study, the equivalent-continuum model is a continuum solid cylinder shown in Fig. 20. The effective mechanical properties or geometries of the cylinder are determined from equating strain energies. The continuum solid cylinder is further used as an effective fiber in the subsequent micromechanical analysis to determine the bulk properties of nanotube/polymer composites. The elastic, bulk composite behavior is described by:

$$\{\sigma\} = [C]\{\varepsilon\} \tag{90}$$

where $\{\sigma\}$ and $\{\varepsilon\}$ are column vectors that contain the components of stress and strain tensors for the nanotube/polymer composite system, and $[C]$ is the stiffness matrix. The components of the stiffness matrix are dependent on the properties, concentrations, orientations, and interactions of the constituents and are quantitatively determined by using micromechanical analysis.

In their study, the moduli of both nonfunctionalized and functionalized nanotube/polyethylene composite systems were examined for various nanotube lengths, volume fractions, and orientations. For a fixed nanotube volume fraction of 1% and various nanotube lengths, the Young's modulus of the random composite, the shear modulus of the random composite, and the longitudinal Young's modulus of the aligned composite have been shown to decrease up to 10% when the nanotube is functionalized. For a fixed nanotube length of 400 nm and various nanotube volume fractions, the longitudinal Young's modulus of the

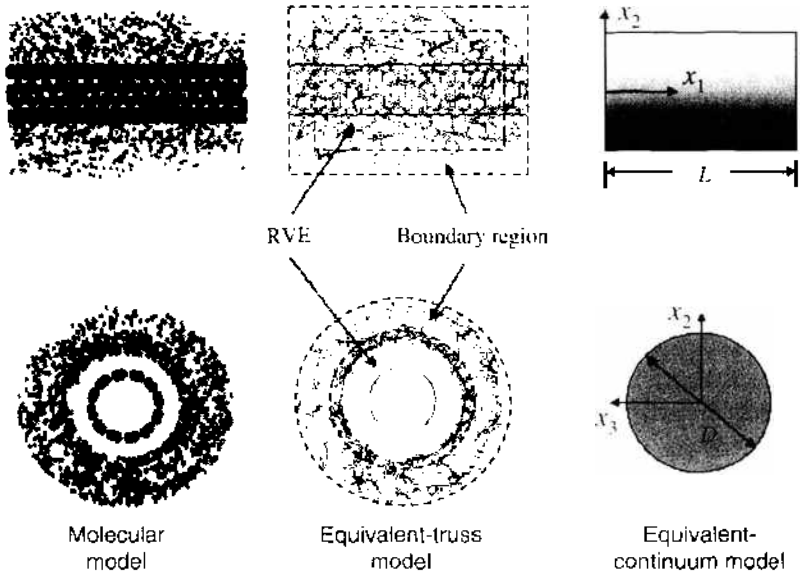


Figure 20. Equivalent-continuum modeling of the nanotube/polymer composite system [100].

aligned composite, the Young's modulus of the random composite, and the shear modulus of the random composite have also been shown to decrease up to 12% when the nanotube is functionalized. These results indicated that the functionalization has degraded the elastic stiffness of carbon nanotube/polyethylene composite.

4. EXPERIMENTAL INVESTIGATION OF CARBON NANOTUBE/POLYMER COMPOSITES

In the past few years, extensive experimental research has been conducted in processing and characterization of carbon nanotube/polymer composites. In the following section, we will review recent experimental work related to the characterization of interfacial bonding and mechanical behavior of carbon nanotube/polymer composites, which will provide direct or indirect evidences for the theoretical and computational results.

4.1. Interfacial Bonding of Carbon Nanotube/Polymer Composites

Because carbon nanotubes possess superior tensile modulus and strength, the possibility on the successful use of these nanostructural elements is highly dependent on the efficiency of the load transfer between the nanotubes and polymer matrix and the load transfer within the nanotube system itself. A few researchers have experimentally studied the efficiency of the interfacial stress transfer of carbon nanotube/polymer composites. Depending on the polymer matrix and processing conditions, the interfacial bonding properties of carbon nanotube/polymer composites seem to have large variations.

A few previous works have concluded that the nanotubes have a poor chemical interaction to polymeric materials. The addition of catalysts to enhance their interfacial bonding strength may eventually destroy the perfect surface atomic structure, to which is attributed the high strength of carbon nanotubes. In the study of SWNT/epoxy composites using Raman spectroscopy, Ajayan et al. [101] suggested that the nearly constant value of the Raman peak in tension is related to tube sliding within the nanotube ropes and indicated poor interfacial load transfer between the nanotubes. Schadler et al. [102] studied the load transfer of MWNT/epoxy composites in tension and compression. They found that the compressive modulus was higher than the tensile modulus of the composites, and the Raman peak shifted only in compression. These findings suggested that during load transfer, only the outer layers were stressed in tension because of inner-wall sliding effects, whereas all layers responded in compression. Lau and Shi [103] also investigated the fracture behavior of multiwalled carbon nanotube/polymer composites. They found that multiwalled carbon nanotubes were

pulled out after a flexural property test on multiwalled nanotube/epoxy composites. They mentioned that the use of multi-walled nanotubes may not be appropriate for advanced composite structures, as only the outermost layer of the nanotubes takes all applied loads. If only an outermost layer of carbon nanotubes takes all applied loads, the stress and strain in individual layers of the nanotubes are different. Therefore, the Young's modulus estimated by the theoretical and molecular simulations, which assumed that the force is applied uniformly to the ends of all tubes without considering the end caps' effects, cannot be used to predict the global tensile modulus of carbon nanotube/polymer composites. Moreover, multiwalled nanotubes with end caps are entirely embedded into the matrix of the composites. The bonding strength between the matrix and the end caps of the nanotubes is still unknown.

However, several experimental measurements have been conducted and obtained a relatively higher interfacial shear strength. Wagner et al. [104] examined the fragmentation of MWNTs in the polymer films and concluded that the interfacial bonding stress between the nanotubes and polymer matrix could be as high as 500 MPa. Lourie et al. [105] studied the expansion of elliptical hole spanned by the rope of SWNTs, which were well anchored at both ends. The fracture of the nanotubes occurred in tension within the hole rather than in shear within the gripping polymer region at the ends of the bundles. Cooper et al. [106] used Raman spectroscopy to detect stress transfer and concluded that the effective modulus of SWNTs dispersed in a composite could be over 1 TPa and that of MWNTs was about 0.3. In their recent work, Cooper et al. [107] directly measured the interfacial strength by drawing out individual SWNT ropes and MWNTs bridging across holes in an epoxy matrix using a scanning probe microscope tip. Based on these experiments, the interfacial shear strength between the MWNTs and the epoxy matrix were calculated to be in the range of 35–376 MPa, whereas most of SWNT ropes were fractured instead of being pulled out of the epoxy matrix. Barber et al. [108] also measured the adhesive interactions between the MWNT and the polyethylene–butene matrix by performing reproducible nano-pullout experiments using atomic force microscopy. Their experimental data resulted in a relatively high interfacial separation stress of 47 MPa.

Currently, much work is underway toward improving the interfacial bonding between the nanotube and polymer matrix in the composites. Thostenson et al. [109] have grown multiwalled nanotubes directly on the surface of carbon fibers by using chemical vapor deposition (CVD) method. This work aimed to improve the bonding strength of carbon fibers to polymer matrix by the generation of mechanical interlocks with providing a strongest interface. According to the measured fragment lengths and captured birefringence patterns, they found that a high bonding strength was achieved at the matrix/matrix interface for a fiber with nanotube grown. Mamedov et al. [110] have latterly developed a deposition technique called "layer-by-layer (LBL) assembly" to minimize the phase segregation that normally exists by using the conventional blending process for fabricating carbon nanotube/polymer thin films. This technique is based on the alternating absorption of monolayers of individual components attracted by electrostatic and van der Waals interactions. By using the LBL technique, the high structural homogeneity and interconnectivity of structural components of the LBL composite films could be achieved, and the amount of structural defects could be greatly reduced. The mechanical strength of LBL composite films is comparatively better than that of any conventional plastic materials. Recently, Lu et al. [111] have successfully produced multiwalled coiled carbon nanotubes (MCCNTs) using catalytic chemical vapor deposition (CCVD) on finely divided Cobalt nanoparticles supported on silica gel under reduced pressure and at lower gas flow rates. In their study, the irregular coils with various shapes of nanotubes were found. The coiled shape of carbon nanotubes can generate the mechanical interlocking between the nanotubes and polymer matrix at the nanometer scale when the MCCNT/polymer composites are subjected to different mechanical loadings. The MCCNTs with the specified coil diameter and coil pitch can be produced by controlling their growth conditions. The good interfacial bonding and mechanical properties of the MCCNT/polymer composites are expected by controlling the structural parameters of MCCNTs, such as the coil diameter and coil pitch.

4.2. Mechanical Properties of Carbon Nanotube/Polymer Composites

Many recent experimental studies have proved that adding a small amount of carbon nanotubes into polymers could enhance the mechanical properties of carbon nanotube/polymer composites. Qian et al. [24, 112] have reported that the addition of 1 wt% MWNTs into polystyrene resulted in 36–42% increase in tensile modulus and 25% increase in tensile strength. They also observed via transmission electron microscopy (TEM) micrographs that the nanotubes were able to bridge the crack surface of the composite once a crack was initiated. The crack was nucleated at the low nanotube density area and propagated along the weak nanotube/polymer interfaces or regions with the relatively low nanotube density regions. The pullout of the nanotubes was observed when the crack opening displacement reached ≈800 nm. Therefore, the function of carbon nanotubes was to bridge up the crack in the nanocomposites. Andrews et al. [113] have used the bundles of nanotubes and petroleum-derived pitch matrix to form composites. They found that the overall tensile strength increased with increasing the fraction of nanotubes in the composites.

Recent papers have reported that the Raman spectroscopy technique is an effective tool to nondestructively measure the strain of carbon nanotubes and nanotube/polymer composites by measuring a peak signal shift of the Raman spectra to identify the strain conditions [114]. Schadler et al. [27] and Wood et al. [115] studied the use of Raman spectroscopy technique to measure the strain of carbon nanotube/epoxy composites. They found that the compressive modulus of the nanotube was higher than the tensile modulus. Cooper et al. [116] have also demonstrated the Raman spectroscopy of a nanotube composite beam subjected to four-point bending. They found a Raman peak shift of the nanotube composites as the surface strain increased.

Lau et al. [117] have found that the hardness of carbon nanotube/polymer composite increased with increasing the nanotube's weight fraction. They also found that the hardness was dropped at the low nanotube's weight fraction samples because of the weak bonding interface between the nanotube and polymer matrix. Increasing the nanotube's weight fraction would result in forming a mesh-like networking structure by high aspect ratio nanotubes, which could enhance the hardness of the composites. Lau et al. [118–119] have studied the flexural strength of carbon nanotube/epoxy beams under different ambient temperature environments. They found that the flexural properties of the nanotube beams decreases compared with a beam without the nanotubes. The cause of the strength reduction was due to structural nonhomogeneity and/or the existence of a weak-bonding interface between the nanotubes and the surrounding matrix. Although the mesh-like structure was formed inside the beams, it did not improve the flexural strength because this is mainly determined by the nanotube/matrix bonding. The fracture surfaces of carbon nanotube/epoxy composites after flexural strength tests have shown different failure mechanisms for composites pretreated at different temperatures. It was found that the nanotubes within the composites after being treated at warm and cryogenic temperatures were aligned perpendicular and parallel to the fracture surfaces, respectively.

The potential applications of using carbon nanotubes as nanoreinforcements and nanoconductors in polymer or metallic-based composite structures are significant. Besides the structural applications, the electrical and electronic applications of using carbon nanotube/polymer composites are exclusive. Much work in this area has recently been conducted by many research groups around the world. The major focus has paid much attention on the determination of the resistance, which in turn could be converted to strain or other chemical quantities, of different types of carbon nanotube/polymer thin films [120]. The films could be used as sensors for any tiny instruments and coatings for electrostatic discharge protection for high-speed vehicle applications. The investigation on the durability and reliability of carbon nanotube/polymer composites subjected to different mechanical and thermal loading cycles is a key issue. In addition, the incorporation of carbon nanotubes into polymeric materials for wear-sensitive components is possible to decrease the generation of the wear debris. Attempts are being made to improve the wear resistance of polymers by introducing carbon nanotubes into the ultrahigh molecular weight polyethylene (UHMWPE), as UHMWPE is commonly used as the acetabular cup component of total hip prosthesis. Wear debris from

UHMWPE produced during articulation is known to cause serious problems such as implant loosening. The wear rate significantly decreases with increasing the content of carbon nanotubes. The wear rate as measured on simulated wear test in bovine serum approaches an undetectable level when 0.5 wt% carbon nanotubes are added. Therefore, carbon nanotube is potentially an attractive additive for artificial implant material and high-quality components that need excellent wear resistance.

5. CONCLUSION

The discovery of carbon nanotubes has initiated a number of scientific investigations to explore their unique properties and potential applications. The composite community considers carbon nanotubes as ideal reinforcements for structural and multifunctional composite applications. However, some crucial issues must be solved before the full potential of carbon nanotubes is realized in the nanocomposite materials. Fundamental understanding and highly accurate predictive methods for the interfacial bonding and mechanical behavior of carbon nanotube/polymer composites are crucial to realize successfully the extraordinary properties of this new class of nanocomposite materials. This review has compared several different computational methods to study the interfacial bonding and mechanical behavior of carbon nanotube/polymer composites. The comparison sheds light on the main differences and similarities between the methods and introduces the essential features of new methods.

Molecular mechanics and molecular dynamics simulations promotes science-based understanding of the properties of complex materials and phenomena. In the development of carbon nanotube/polymer composites, they offer insight into the local interactions among individual atoms based on the discrete models of the nanotubes and polymer matrix. They have provided enough detailed information for understanding the load transfer and mechanical behavior of carbon nanotube/polymer composites. However, these simulations are currently limited to very small length and time scales and therefore are not suitable for large-scale analysis in real-life applications. Continuum modeling based on continuum mechanics has been employed to study the mechanical responses of carbon nanotube/polymer composites under various loading conditions. Although the initial simulation results seem to be successful, the validity of the continuum modeling of carbon nanotube/polymer composites is still an open question. Therefore, multiscale modeling holds great promise for the continued advancement of modeling technique for carbon nanotube/polymer composites. This modeling technique is capable of taking into account the discrete nature of nanocomposite materials while being efficient to model the nanocomposite materials at larger length scales by integrating molecular simulation and continuum modeling. Regarding the modeling of carbon nanotube/polymer composites, it is required to incorporate more structural characteristics of carbon nanotubes into the computational models, such as structural defects on the nanotube surface, ropes or bundles of nanotubes, and waviness of the nanotubes in the nanocomposites.

REFERENCES

1. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
2. S. Iijima, *Nature* 354, 56 (1991).
3. G. Overney, W. Zhong, and D. Tomanek, *Zeitschrift Fur Physik A-Atoms Molecules and Clusters* 27, 93 (1993).
4. C. Q. Ru, Effective bending stiffness of carbon nanotubes, *Phys. Rev. B* 62, 9973 (2000).
5. L. Vaccarini, C. Goze, L. Henrard, E. Hernandez, P. Bernier, and A. Rubio, *Carbon* 38, 1681 (2000).
6. J. P. Lu, *Phys. Rev. Lett.* 79, 1297 (1997).
7. E. Hernandez, C. Goze, P. Bernier, and A. Rubio, *Phys. Rev. Lett.* 80, 4502 (1998).
8. C. Q. Ru, *Phys. Rev. B* 62, 10405 (2000).
9. V. N. Popov, V. E. V. Doren, and M. Balkanski, *Solid State Commun.* 114, 395 (2000).
10. S. Govindjee and J. L. Sackman, *Solid State Commun.* 110, 227 (1999).
11. C. Q. Ru, *J. Appl. Phys.* 87, 7227 (2000).
12. C. Q. Ru, *J. Appl. Phys.* 80, 3426 (2001).
13. A. N. Kolmogorov and V. H. Crespi, *Phys. Rev. Lett.* 85, 4727 (2000).

14. E. Saether, S. J. V. Frankland, and R. B. Pipes, *Composites Sci. Technol.* 63, 1543 (2003).
15. E. Saether, *Composites Sci. Technol.* 63, 1551 (2003).
16. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 680 (1996).
17. E. W. Wong, P. E. Sheehan, and C. M. Lieber, *Science* 277, 1971 (1997).
18. A. Krishnan, *Phys. Rev. B* 58, 14013 (1998).
19. J. P. Salvetat, A. J. Kulik, J. Bonard, G. A. Briggs, T. Stöckli, K. Méténier, S. Bonnamy, F. Béguin, N. A. Burnham, and L. Forró, *Adv. Mater.* 11, 161 (1999).
20. J. P. Salvetat, G. A. D. Briggs, J. M. Bonard, R. R. Bacsa, A. J. Kulik, T. Stöckli, N. A. Burnham, and L. Forró, *Phys. Rev. Lett.* 82, 944 (1999).
21. D. A. Walters, L. M. Ericson, M. J. Casavant, J. Liu, D. T. Colbert, K. A. Smith, and R. E. Smalley, *Appl. Phys. Lett.* 74, 3803 (1999).
22. M. F. Yu, O. Lourie, M. J. Dyer, K. Moloni, T. F. Kelly, and R. S. Ruoff, *Science* 287, 637 (2000).
23. M. F. Yu, B. S. Files, S. Arepalli, and R. S. Ruoff, *Phys. Rev. Lett.* 84, 5552 (2000).
24. D. Qian, E. C. Dickey, R. Andrews, and T. Rantell, *Appl. Phys. Lett.* 76, 2868 (2000).
25. L. Jin, C. Bower, and O. Zhou, *Appl. Phys. Lett.* 73, 1197 (1998).
26. R. Haggenmueller, H. H. Gommans, A. G. Rinzler, J. E. Fischer, and K. I. Winey, *Chem. Phys. Lett.* 330, 219 (2000).
27. L. S. Schadler, S. C. Giannaris, and P. M. Ajayan, *Appl. Phys. Lett.* 73, 3842 (1998).
28. S. Kumar, T. D. Dang, F. E. Arnold, A. R. Bhattacharyya, B. G. Min, X. F. Zhang, R. A. Vaia, C. Park, W. W. Adams, R. H. Hauge, R. E. Smalley, S. Ramesh, and P. A. Willis, *Macromolecules* 35, 9039 (2002).
29. Z. H. Fan, K. T. Hsiao, and S. G. Advani, *Carbon* 42, 871 (2004).
30. Z. J. Jia, Z. Y. Wang, C. L. Xu, J. Liang, B. Q. Wei, D. H. Wu, and S. W. Zhu, *Mater. Sci. Eng. A* 271, 395 (1999).
31. C. Park, Z. Ounaies, K. A. Watson, R. E. Crooks, J. Smith, S. E. Lowther, J. W. Connell, E. J. Siochi, J. S. Harrison, and T. L. S. Clair, *Chem. Phys. Lett.* 364, 303 (2002).
32. C. Stéphan, T. P. Nguyen, M. Lamy de la Chapelle, S. Lefrant, C. Journet, and P. Bernier, *Synth. Metals* 108, 139 (2000).
33. M. S. Shaffer and A. H. Windle, *Adv. Mater.* 11, 937 (1999).
34. J. Sandler, M. S. P. Shaffer, T. Prasse, W. Bauhofer, K. Schulte, and A. H. Windle, *Polymer* 40, 5967 (1999).
35. Z. X. Jin, K. P. Pramoda, C. Q. Xu, and S. H. Goh, *Chem. Phys. Lett.* 337, 43 (2001).
36. K. Lozano and E. V. Barrera, *J. Appl. Polym. Sci.* 79, 125 (2001).
37. K. Lozano, J. B. Rios, and E. V. Barrera, *J. Appl. Polym. Sci.* 80, 1162 (2001).
38. X. Y. Gong, J. Liu, S. Baskaran, R. D. Voise, and J. Young, *Chem. Phys. Lett.* 12, 8049 (2000).
39. P. M. Ajayan, O. Stephan, C. Colliex, and D. Trauth, *Science* 265, 1212 (1994).
40. D. A. Walters, M. J. Casavant, X. C. Qin, C. B. Huffman, P. J. Boul, L. M. Ericson, E. H. Haroz, M. J. O'Connell, K. Smith, D. T. Colbert, and R. E. Smalley, *Chem. Phys. Lett.* 338, 14 (2001).
41. K. T. Lau and D. Hui, *Composite Part B: Engineering* 33, 263 (2003).
42. E. T. Thostenson, Z. F. Ren, and T. W. Chou, *Composites Sci. Technol.* 61, 1899 (2001).
43. V. Antonucci, K. T. Hsiao, and S. G. Advani, in "Advanced Polymeric Materials: Structure Property Relationships" (G. O. Shonaike and S. G. Advani, Eds.), p. 397. CRC Press, Boca Raton, 2003.
44. D. Qian, G. J. Wagner, W. K. Liu, M. F. Yu, and R. S. Ruoff, *Appl. Mechanics Rev.* 55, 495 (2002).
45. V. Lordi and N. Yao, *J. Mater. Res.* 15, 2770 (2000).
46. Z. Y. Liang, T. Wu, C. Zhang, B. Wang, and L. Kramer, *Mater. Sci. Eng. A* 365, 228 (2004).
47. J. Gou, "Single-Walled Carbon Nanotube Bucky Paper/Epoxy Composites: Molecular Dynamics Simulation and Process Development," Ph.D. dissertation, Florida State University, Tallahassee, FL, 2002.
48. J. Gou, K. Anumakonda, Z. Y. Liang, C. Zhang, and B. Wang, in "Proceedings of the 22th Southeastern Conference on Theoretical and Applied Mechanics (SECTAM XXII)" (H. Mahfuz and M. V. Hosur, Eds.), Tuskegee University, Tuskegee, AL, 2004.
49. J. Gou, B. Minaie, B. Wang, Z. Y. Liang, and C. Zhang, *Comput. Mater. Sci.* 31, 225 (2004).
50. J. Gou, S. L. Jiang, B. Minaie, Z. Y. Liang, C. Zhang, and B. Wang, in "Proceedings of 2003 ASME International Mechanical Engineering Congress and Exposition (IMECE'03)," CD-ROM, Washington, DC, 2003.
51. J. Gou, B. Minaie, Z. Y. Liang, C. Zhang, and B. Wang, in "Proceedings of the Tenth International Conference on Composites/Nano Engineering" (D. Hui, Ed.), p. 207. New Orleans, LA, 2003.
52. J. Gou, in "Proceedings of the 11th Annual International Conference on Composites/Nano Engineering" (D. Hui, Ed.), CD-ROM, Hilton Head Island, SC, 2004.
53. K. Liao and S. Li, *Appl. Phys. Lett.* 79, 4225 (2001).
54. K. Lafdi and Y. M. Tan, *Composites Part B: Engineering* 32, 365 (2001).
55. B. N. J. Persson, "Sliding Friction: Physical Principles and Applications," Springer, Berlin, 1998.
56. S. J. V. Frankland and V. M. Harik, *Surf. Sci.* 525, L 103 (2003).
57. S. J. V. Frankland, A. Caglar, D. W. Brenner, and M. Griebel, *J. Phys. Chem. B* 106, 3046 (2002).
58. G. K. Batchelor, "Introduction to Fluid Dynamics," Cambridge University Press, Cambridge, 1980.
59. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14012 (1998).
60. C. Q. Ru, Elastic models for carbon nanotubes, in "Encyclopedia of Nanoscience and Nanotechnology" (H. S. Nalwa, Ed.), Vol. X, pp. 1–14. American Scientific Publishers, 2003.
61. X. Y. Wang and X. Wang, *Composites Part B: Engineering* 35, 79 (2004).
62. J. P. Lu, *J. Phys. Chem. Solids* 58, 1649 (1997).
63. K. T. Lau, C. Gu, G. H. Gao, H. Y. Ling, and S. R. Reid, *Carbon* 42, 423 (2004).

64. X. J. Xu, M. M. Thwe, C. Shearwood, and K. Liao, *Appl. Phys. Lett.* 81, 2833 (2002).
65. H. D. Wagner, *Chem. Phys. Lett.* 361, 57 (2002).
66. K. T. Lau, *Chem. Phys. Lett.* 370, 399 (2003).
67. T. Lenosky, X. Gonze, M. Teter M, and V. Elser, *Nature* 355, 333 (1992).
68. Z. C. Tu and Z. C. Ou-Yang, *Phys. Rev. B* 65, 233407 (2002).
69. K. T. Lau, W. K. Chan, S. Q. Shi, and L. M. Zhou, *Mater. Design* 23, 265 (2002).
70. K. T. Lau, C. K. Poon, L. M. Zhou, and L. C. Yam, *Mater. Sci. Forum* 394, 527 (2002).
71. L. M. Zhou, Y. W. Mai, and L. Ye, *Composites Engineering* 5, 10 (1995).
72. Y. C. Gao, Y. W. Mai, and B. Cotterell, *Zeitschrift für angewandte Mathematik und Physik* 39, 550, (1988).
73. S. P. Timoshenko and J. N. Goodier, "Theory of Elasticity," McGraw-Hill, New York, 1970.
74. P. K. Mallick, "Composites Engineering Handbook," Marcel Dekker, New York, 1997.
75. N. Yao and V. Lordi, *J. Appl. Phys.* 84, 1939 (1998).
76. S. Iijima, C. Brabec, A. Maiti, and J. Bernholc, *J. Chem. Phys.* 104, 2089 (1996).
77. S. B. Sinnott, O. A. Shenderova, C. T. White, and D. W. Brenner, *Carbon* 36, 1 (1998).
78. B. I. Yakobson, C. J. Brabec, and J. Bernholc, *Phys. Rev. Lett.* 76, 2511 (1996).
79. A. Garg and S. B. Sinnott, *Chem. Phys. Lett.* 295, 273 (1998).
80. Y. Jin and F. G. Yuan, *Composites Sci. Technol.* 63, 1507 (2003).
81. D. Qian, W. K. Liu, and R. S. Ruoff, *Composites Sci. Technol.* 63, 1561 (2003).
82. J. E. Lennard-Jones, *Proc. Royal Soc.* A106, 441 (1924).
83. T. C. Chang and H. J. Gao, *J. Mechanics Phys. Solids*, 51, 1059 (2003).
84. S. J. V. Frankland, V. M. Harik, G. M. Odegard, D. W. Brenner, and T. S. Gates, *Composites Sci. Technol.* 63, 1655 (2003).
85. C. Y. Fung, "Foundations of Solid Mechanics," Prentice-Hall, Englewood Cliffs, NJ, 1965.
86. M. Griebel and J. Hamaekers, *Comp. Methods Appl. Mechanics Eng.* 193, 1773 (2004).
87. K. Sohlberg, B. G. Sumpter, R. E. Tuzun, and D. W. Noid, *Nanotechnology* 9, 30 (1999).
88. K. E. Drexler, "Nanosystems: Molecular Machinery, Manufacturing and Computation," Wiley, New York, 1992.
89. B. G. Sumpter and D. W. Noid, *J. Chem. Phys.* 102, 6619 (1995).
90. X. L. Chen and Y. J. Liu, *Comput. Mater. Sci.* 29, 1 (2004).
91. Y. J. Liu and X. L. Chen, *Mechanics Mater.* 35, 69 (2003).
92. Y. J. Liu and X. L. Chen, *Electronic Boundary Elements* 1, 316 (2003).
93. M. W. Hyer, "Stress Analysis of Fiber-Reinforced Composite Materials," McGraw-Hill, Boston, 1998.
94. C. Y. Li and T. W. Chou, *Int. J. Solids Structures* 40, 2487 (2003).
95. C. Y. Li and T. W. Chou, *Appl. Phys. Lett.* 84, 121 (2004).
96. C. Y. Li and T. W. Chou, *Composites Sci. Technol.* 63, 1517 (2003).
97. C. Y. Li and T. W. Chou, *J. Nanosci. Nanotechnol.* 3, 1 (2003).
98. G. M. Odegard, T. S. Gates, L. M. Nicholson, and K. E. Wise, *Composites Sci. Technol.* 62, 1869 (2002).
99. G. M. Odegard, T. S. Gates, K. E. Wise, C. Park, and E. J. Siochi, *Composites Sci. Technol.* 63, 1671 (2003).
100. G. M. Odegard, S. J. V. Frankland, and T. S. Gates, in "Proceedings of the 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials," AIAA 2603-1701, American Institute of Aeronautics and Astronautics, Reston, VA, 2003.
101. P. M. Ajayan, O. Stephan, C. Colliex, and D. Trauth, *Science* 265, 1212 (1994).
102. J. Sandler, M. S. P. Shaffer, T. Prasse, W. Bauhofer, K. Schulte, and A. H. Windle, *Polymer* 40, 5967 (1999).
103. K. T. Lau and S. Q. Shi, *Carbon* 40, 2965 (2002).
104. H. D. Wagner, O. Lourie, Y. Feldman, and R. Tenne, *Appl. Phys. Lett.* 72, 188 (1998).
105. O. Lourie and H. D. Wagner, *Appl. Phys. Lett.* 73, 3527 (1998).
106. C. A. Cooper and R. J. Young, *J. Raman Spectrosc.* 30, 929 (1999).
107. C. A. Cooper, S. R. Cohen, A. H. Barber, and H. D. Wagner, *Appl. Phys. Lett.* 81, 3873 (2002).
108. A. H. Barber, S. R. Cohen, and H. D. Wagner, *Appl. Phys. Lett.* 82, 4140 (2003).
109. E. T. Thostenson, W. Z. Li, D. Z. Wang, Z. F. Ren, and T. W. Chou, *J. Appl. Phys.* 91, 6034 (2002).
110. A. A. Mamedov, N. A. Kotov, M. Prato, D. M. Guldi, J. P. Wicksted, and A. Hirsch, *Nature* 1, 190 (2002).
111. M. Lu, W. M. Liu, X. Y. Guo, and H. L. Li, *Carbon* 42, 805 (2004).
112. D. Qian and E. C. Dickey, *J. Microsc.* 204, 39 (2001).
113. R. Andrews, D. Jacques, A. M. Rao, T. Rantell, F. Derbyshire, Y. Chen, J. Chen, and R. C. Haddon, *Appl. Phys. Lett.* 75, 1329 (1999).
114. M. S. Dresselhaus, A. Jorio, A. G. Souza Filho, G. Dresselhaus, and R. Saito, *Phys. Rev. B* 323, 15 (2002).
115. J. R. Wood, Q. Zhao, and H. D. Wagner, *Composites Part A: Applied Science and Manufacturing* 32, 391 (2001)
116. C. A. Cooper and R. J. Young, *J. Raman Spectrosc.* 30, 929 (1999).
117. K. T. Lau, S. Q. Shi, and L. M. Zhou, *J. Composite Mater.* 37, 365 (2003).
118. K. T. Lau and D. Hui, *Carbon* 40, 1605 (2002).
119. K. T. Lau and S. Q. Shi, *Carbon* 40, 2961 (2002).
120. D. S. Lim, J. W. An, and H. J. Lee, *Wear* 252, 512 (2002).

CHAPTER 8

Nano-Characterization of Materials: Silicon, Copper, Carbon Nanotubes, and Diamond Thin Films

Liangchi Zhang

School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, New South Wales, Australia

CONTENTS

1. Introduction 396
2. Molecular Dynamics Modeling of Materials 396
   2.1. Initial Model 396
   2.2. Simulation Temperature 397
   2.3. Size of Control Volume 398
   2.4. Integration Time Steps 398
   2.5. Temperature Conversion 399
   2.6. Stress Analysis 399
   2.7. Potential Function 401
3. Nanoindentation on Monocrystalline Silicon 402
   3.1. Introduction 402
   3.2. Modeling 403
   3.3. Phase Transformation 405
   3.4. Deformation Characteristics 409
   3.5. Effect of Oxygen Penetration 413
   3.6. Effect of Water 415
   3.7. Cyclic Indentations 417
4. Nanotribology 420
   4.1. Introduction 420
   4.2. Moving Control Volume 421
   4.3. Diamond–Copper Sliding Systems 422
   4.4. Scale Effect of Contact Size on Friction Transition 429

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (395–456)

4.5. Diamond–Silicon Sliding Systems 433
4.6. Multiasperity Sliding 437
5. Characterization of Carbon Nanotubes 440
5.1. Introduction 440
5.2. Modeling 441
5.3. Deformation Characteristics 441
5.4. Potential 441
5.5. Number of Thermostat Atoms 443
5.6. Influence of Thermostat Schemes 444
5.7. Integral Time Step, Displacement Step, and Relaxation Step 445
5.8. Summary 446
6. Deformation of Copper Nanowhiskers 446
6.1. Introduction 446
6.2. Modeling 447
6.3. Effects of Atomic Orientation and Specimen Shape . . . 447
6.4. Elastic Modulus and Ultimate Tensile Strength 448
6.5. Shear Banding and Necking 448
7. Buckling of Diamond Thin Films 449
7.1. Background 449
7.2. Theory and Simulation Method 450
7.3. Uniaxial Stress 452
7.4. Biaxial Stresses 452
8. Concluding Remarks 453
References 454

1. INTRODUCTION

The development of microelectromechanical systems (MEMS) and nanoelectromechanical systems (NEMS) requires a deep understanding of the deformation mechanisms of materials at the atomistic and nanometric levels. Although quantum mechanics is ideal for very small models on the atomic scale and micro/continuum mechanics is powerful for analyzing the objects of micro and macroscopic dimensions, molecular dynamics simulation provides a useful means of detailed characterization of materials on the nanometer scale. The advantage of using molecular dynamics lies in its capacity to handle relatively large molecular systems, which are hard for quantum mechanics to tackle, and its reliability in exploring atomistic deformation mechanisms such as phase transformations and dislocation emitions, to which micromechanics and continuum mechanics are not applicable.

Fundamentally, a molecular dynamics analysis involves calculating the phase-space trajectories of each atom based on its interaction with other atoms in accordance with Newtonian dynamics. The phase-space trajectory describes the motion of an atom by describing its position in the Cartesian coordinate and its momentum. Although the principle of molecular dynamics is simple, a reliable application of the technique requires sophisticated consideration on the details of every step in the modeling and simulation.

This chapter will elaborate on some fundamental concepts and features of molecular dynamics modeling for characterizing the nanodeformation mechanisms of advanced materials, with the phase transformations in monocrystalline silicon and the mechanical deformation of carbon nanotubes and diamond-like thin films as the examples.

2. MOLECULAR DYNAMICS MODELING OF MATERIALS

2.1. Initial Model

The very first step in the nanodeformation characterization of a material subjected to a nanoprocessing operation using the molecular dynamics analysis is to generate an initial

molecular model of the material consistent with the operation. Here the nanoprocessing operation can be any process corresponding to external loading, either simple or complex, such as tension, bending, indentation, tribological sliding, and mechanical cutting. One way to generate the initial model is to locate atoms on a perfect lattice structure of a material that represents the real atomic structure of the material [1]. An alternative is to take the initial positions of atoms from the end of an earlier simulation. However this may not be viable, as one may often change the size and shape of a model. For example, atoms in a model to simulate copper will be positioned in accordance to a face-centred cubic (FCC) lattice structure with its lattice constant [2], whereas a model to simulate diamond will be positioned in accordance to a diamond cubic structure [3].

The model having atoms positioned in such a way implies that additional potential energy between atoms has been artificially applied to the system, because in a real material, atoms are actually vibrating around their equilibrium positions. Before simulation can go on, therefore, it is necessary to "relax" this constructed initial model from its artificially assigned conditions to its natural equilibrium status, consistent with the environmental temperature. This involves the execution of the constant-temperature molecular dynamics program for a specified number of time steps with a chosen potential function. During the relaxation process, the velocities of Newtonian and thermostat atoms that are initially assigned on the basis of a normal distribution will gradually reach equilibrium at the specified environmental temperature of the simulation by a velocity scaling. The number of time steps required to run for such a relaxation is determined by the time needed for the model to arrive at the specified equilibrium temperature. Depending on the specific materials and their atomic lattice structures, essential relaxation time steps vary. The details will be discussed in the specific examples later in the chapter.

2.2. Simulation Temperature

The portion of the material simulated by the molecular dynamics model, often called the control volume, is only a small part of the material in the neighborhood of the deformation zone of interest. This portion in reality is part of the rest of the material or the surrounding environment. As such, any heat generated during nanoprocessing within the control volume will be conducted away. If this heat conduction process cannot be simulated in a molecular dynamics analysis, the simulation results can be incorrect. To achieve a reasonable heat conduction out from the control volume, special layers of atoms, called thermostat atoms, must normally be arranged to surround the model, as illustrated in Fig. 1, the control volume for a nanoindentation setup. Then, based on the temperature conversion rule, to be discussed later in this chapter, temperature regulation is made so that the temperature caused by the kinetic energy of these atoms will always be consistent with the environmental temperature.

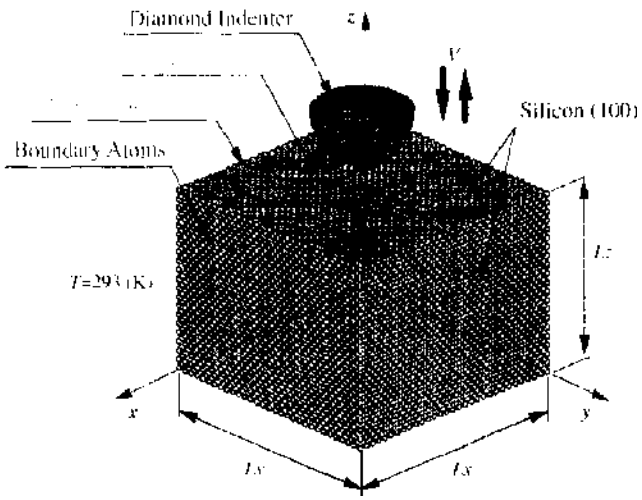


Figure 1. Molecular dynamics model of silicon specimen with a hemispherical indenter. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999), © 1999, Japan Society of Mechanical Engineers.

When the dimension of the control volume is sufficiently large, the arrangement of the thermostat atoms provides an efficient and reliable way for natural heat conduction in the solid.

The scaling of velocities [2] can be performed by

$$V_{new} = V_{old}\sqrt{\frac{\text{Kinetic energy corresponding to the environmental temperature}}{\text{Kinetic energy of a thermostat atom before scaling}}} \quad (1)$$

where V_{new} is the scaled velocity of a thermostat atom and V_{old} is its original velocity.

2.3. Size of Control Volume

There are three major issues that must be considered in the dimension selection of a control volume of interest. To reduce computational cost, the volume should be as small as possible. However, a too-small volume will bring about significant boundary effects that make the results unreliable. These include the boundary temperature effect and the boundary displacement effect. To eliminate these effects, the dimension of the control volume must be sufficiently large that the temperature at the boundary of the control volume is close to the environmental temperature. Then the application of thermostat atoms can make the natural heat conduction happen in simulation. Similarly, the volume should also be sufficiently large that the motion of the boundary atoms does not affect the atoms in the deformation zone of interest. Normally, an error analysis is necessary to generate a suitable dimension. Two examples of the selection process of the control volumes for nanoindentation [3], which has fixed boundary atoms, and those for nano-tribological sliding and nanomachining (nanocutting, nanopolishing, and nanogrinding), which use moving boundary atoms [2, 4, 5], will be discuss in detail later.

2.4. Integration Time Steps

The prediction of the phase-space trajectories of atoms in molecular dynamics simulation is based on Newton's second law of motion. To solve the differential equation, the finite difference method is often necessary because collisions between atoms are not instantaneous; rather, they are strong repulsive and attractive interactions that occur over a finite duration. However, the use of the finite difference method implies that the size of the integration time step, Δt, must be determined very carefully, because both the global truncation error and global round-off error depend on Δt. Moreover, these two types of errors are affected differently by changes in Δt, as shown in Fig. 2 [5]. The global truncation error decreases with decreasing Δt, whereas the other depends on the number of calculations. This means that the smaller the size of the time steps, the more calculations to be done and the greater the global round-off error.

A way to reduce the round-off error is to have an efficient code and use high-precision arithmetic. To reduce the truncation error, however, it is necessary to reduce the size of Δt. It must also be noted that a smaller time step is usually associated with a greater computational cost.

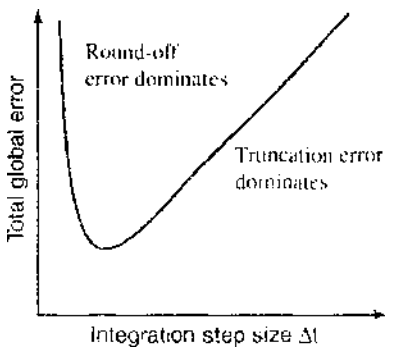


Figure 2. Different types of errors against the size of time steps.

In addition to the above problems, it is essential to note that the finite difference algorithm commonly used in molecular dynamics, the Verlet's method [1–5], is conditionally stable. That is to say, it will become unstable when Δt reaches a critical value, though the threshold can be determined through a series of experiments. In general, a suitable time step is recommended to be from 1/28 to 1/32 of the vibration period of an atom [3, 5]. Hence, the optimum time step is dependent on both the specific material and the potential function used. For instance, with the Tersoff potential [6, 7], an individual atom of silicon or diamond can be forced to move in a direction to determine the corresponding stiffness k, as shown in Fig. 3, so that the period of vibration of the atom in the direction, T, can be determined by $T = 2\pi(m/k)^{1/2}$ [3, 5], where m is the mass of the atom.

2.5. Temperature Conversion

Another important factor in a successful molecular dynamics analysis is the reliable conversion between the kinetic energy and temperature of an atom. An inappropriate conversion will result in an error in the velocities of the atoms and hence render the simulation incorrect. It is always essential that for a given material, a temperature conversion model is identified carefully before carrying out a molecular dynamics simulation. In studying the deformation mechanisms of monocrystalline silicon under the nanoindentation and nanoscratching with diamond tools [3, 4, 8], for instance, there are three models available [9] for the conversion. They are Dulong–Petit's model, which takes into account the independent lattice vibration, Einstein's model, which is based on the consideration of the single characteristic frequency, and Debye's model, which involves a range of frequencies. A comparison with the experimental measurement, as shown in Fig. 4, shows that in the temperature regime encountered in the nanoindentation and nanoscratching, the Debye's model is the best for silicon and the Einstein's model is the most suitable for diamond. Clearly, if these models are incorrectly used, the result of molecular dynamics simulation cannot be correct.

2.6. Stress Analysis

The analysis of stress in a material is an important part of our understanding of characteristics such as deformation or transformation of the material under various machining processes. Stress analysis also allows useful criteria to be set for the prediction of phase transformation or plastic deformation of the material under processing.

However, on a fine scale, materials cannot be treated as a continuum, and the conventional definition of stress is no longer valid [3, 8, 10]. Here, let us take the deformation

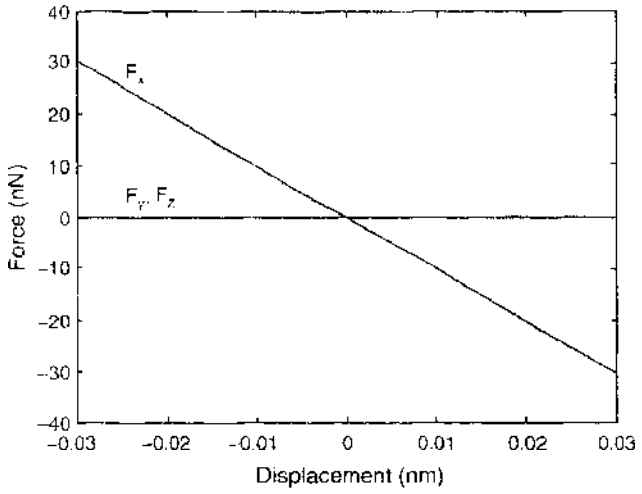


Figure 3. Force-displacement relationship (Tersoff potential) when moving a silicon atom in the [100] direction. The material stiffness k can be obtained from the gradient of the plot. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

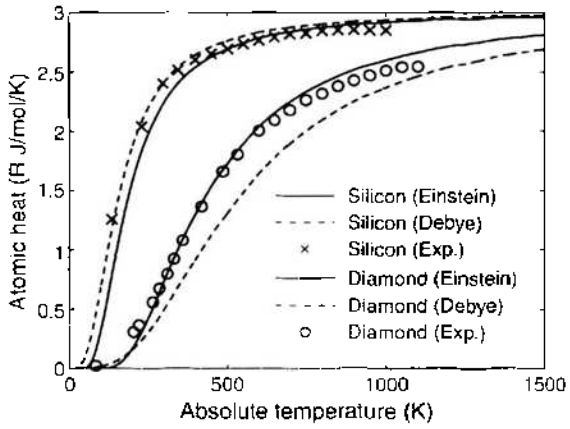


Figure 4. Comparison of temperature conversion models. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

of monocrystalline silicon as an example to introduce the concept of stress analysis on the atomic scale.

Figure 5 shows a conceptual atomistic model for calculating stresses. Assume that a solid is divided into an upper part Ω_1 and a lower part Ω_2 by a plane A. Consider a small element Γ in Ω_1 with a base area S in plane A. The stress on S is defined by

$$\sigma = \frac{F}{S} \tag{2}$$

where $\mathbf{F}$ is the resultant force on S induced by the interaction between the atoms in Ω_2, and those in Γ and should be calculated by

$$F = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} f_{ij} - \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} f_{0ij} \tag{3}$$

in which N_A is the number of atoms in Γ, N_B is the number of atoms in Ω_2, f_{ij} is the interatomic force vector during indentation between atom i in Γ and atom j in Ω_2, and f_{0ij} is the interatomic force vector before indentation between atom i in Γ and atom j in Ω_2. The contribution of interatomic forces before indentation, f_{0ij}, must be eliminated, as the stress analysis considers only the effect of deformation resulting from external forces. This consideration ensures consistency with the definition of stress in continuum mechanics.

It is important to note that in continuum mechanics, a stress vector is defined at a mathematical point, but in the definition here on the atomic scale, area S should always be finite.

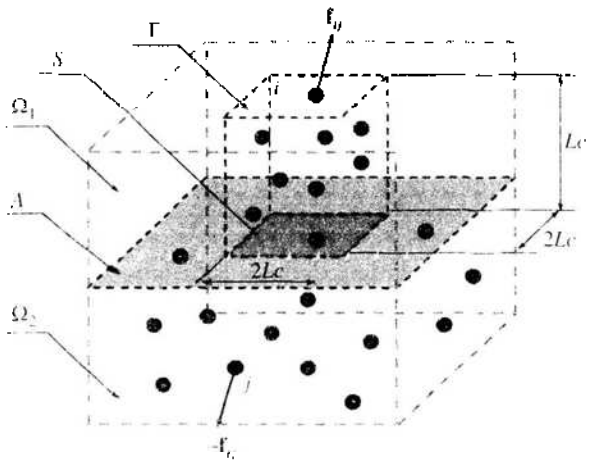


Figure 5. Concept of stress vector on an atomic scale. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

Moreover, to obtain a representative stress vector on S, the height of the element must be selected in a way that the interaction force between atom j in Ω_2 and atom i in Ω_1, but beyond the top surface of Γ, is negligible. In the stress analysis of silicon specimen subjected to nanoindentation, for instance, it is found that the optimum size for element Γ is $2L_c \times 2L_c \times 2L_c$, as shown in Fig. 5, where $L_c = 0.543$ nm is the side length of a unit cell of silicon.

There are two questions that arose from the above stress definition on a molecular scale. The first pertains to the momentum flux of atoms within the material. That is, whether it is necessary to consider the contribution of this element in the calculation of stresses. In fact, each element of the stress tensor will compose of a kinetic part and a potential part, that is,

$$J_{\alpha\beta} = m \sum_{i}^{N} v_{i\alpha} v_{i\beta} + \frac{1}{2} \sum_{i \neq j}^{N} r_{ij\beta} F_{ij\alpha} \tag{4}$$

where m is the atomic mass, $v_{i\alpha}$ is the α-component of the velocity of atom i, $r_{ij\beta}$ is the β-component of the vector $\mathbf{r}_{ij}$ separating atoms i and j, and $F_{ij\alpha}$ is the α-component of the force exerted on atom i by atom j. However, simulation results have shown that for a solid, the contribution resulting from the momentum flux of atoms within the body is very small compared with the effects of interatomic forces between the atoms. Hence, if the simulation is done for a solid, this term can be neglected.

The second question is that when a material is under an external load applied by another solid, for example, in the case of nanoindentation involving both the workpiece material (monocrystalline silicon) and the tool (diamond indenter), is necessary to include the interaction forces between the workpiece and tool atoms in the stress analysis of the workpiece? It has been shown that the effect of this interaction is small when the work–tool interactive potential is low and, hence, is negligible if the region of interest for the stress analysis is not in close proximity with the tool. Stress calculation of the workpiece in such cases can therefore be achieved by considering the atoms of the workpiece alone. However, if the work-tool interactive potential is high, the influence of the tool atoms must be considered.

2.7. Potential Function

Given the adequacy of the simulation techniques and the appropriate selection of the control volume and time step size, the reliability of the results of a simulation depends on the quality of the atomic interaction potential used. If the potential function does not model the behavior of the atoms correctly, the results produced will be unable to simulate the actual deformation correctly.

It would be desirable to take the interactions directly from first-principle calculations. However, this would take up far too much computational time. Thus, in order to obtain useful atomic interactions, empirical potentials have been developed.

In general, the potential energy of a system of N atoms is a function of the atomic coordinates. Thus, the potential energy may be expressed in a series of n-body interatomic potentials

$$\phi_{\text{total}} = \phi_o + \Sigma\phi_1(i) + \Sigma\phi_2(ij) + \Sigma\phi_3(ijk) + \cdots + \phi_N(ijkl, \ldots) \tag{5}$$

where ϕ_n is the n-body interaction potential that is a function of the positions of n atoms $ijk \ldots$. The sums in Eq. (5) are over all combinations (excluding redundant contributions) of n atoms in the system. The potential ϕ_n, and the number of terms, which are retained in a practical application, depend on the nature of the system under investigation.

In s–p-bonded metals, that is, metals with free electrons outside the ion cores, the structure-dependent part of the system energy can be accurately expressed in terms of the second-order (pair potential) terms of Eq. (5). Hence pair potentials (such as Lennard–Jones, Morse, etc.) have a firm theoretical basis in simple metals [11]. Examples of such metals are copper, sodium, magnesium and aluminium. Zhang and coworkers have employed the Morse pair potential in the modeling of copper atoms and find the results of the simulation satisfactory [2, 12]. A detailed explanation and expression for the Morse potential used in the modeling of copper can be found in Ref. [2].

When using the pair potential function, the omission of angle dependent forces and many-body effects results in a severe restriction on its application to more covalent systems such as silicon, in which the directionality of the localization of charge in the bonding region is important. In such cases, it will require an empirical potential incorporating two- and three-body interactions. The parameters used in the potential function can be determined by least-square fits to a database of calculated energies of real and hypothetical atomic structures. However, because of the exclusion of some atomic structures, it is often difficult to evaluate the generality of such potential. In addition, this method of simulation requires much computational time.

To address these shortcomings of the above-mentioned potential function, the Tersoff potential was developed. This replaces the traditional two- and three-body expansion of the interaction energy with a simple pair-like potential, where the bond order of the atoms is affected by its local environment [6, 7]. In this way, structural chemistry is included in a classical empirical potential. It has been found that energies and geometries for silicon are very well described using the Tersoff potential. A detailed explanation and expression for the Tersoff potential can be found in the following sections.

3. NANOINDENTATION ON MONOCRYSTALLINE SILICON

3.1. Introduction

Indentation mechanics has long been an established method to understand the constitutive behavior of metals in various engineering fields [13]. In recent years, it has also been regarded as an appropriate method to characterize hard and brittle materials such as silicon, ceramics, and glass, which are hard to investigate using a conventional tensile test because of their brittle nature.

The response of silicon to indentation has received extensive attention during the past decade. According to these studies, the initial plastic deformation is accompanied by a densifying semiconductor-to-ductile metal phase transformation. On unloading, a tetragonal body centred to amorphous semiconductor phase transformation is observed [14]. The conclusions are supported by several studies carried out using electrical resistance [15–19], x-ray diffraction [20–24], Raman spectroscopy, and optical properties [25–27]. These experimental studies revealed that silicon transformed from its diamond cubic structure to a metallic-body centred tetragonal structure, known also as β-tin. In the experiments, the metallic nature of the β-tin phase is inferred by the fall in the resistivity of silicon by a few orders of magnitude during indentation. Diffraction patterns obtained during indentation [28] also lend credence to the transformation of silicon from its diamond cubic to the body-centred tetragonal phase. Recent development of unique *in situ* stage for transmission electron microscopy (TEM) has also made it possible to image nanoindentation in real time. Although these experiments have furnished us with a wealth of information, combining the information into a convincing picture of the microscopic mechanisms of the phase transition is impeded by possible ambiguities in interpreting these experimental findings.

There are some fundamental problems in characterizing the properties of a silicon monocrystal by nanoindentation. First, an experimental result may not correctly reveal the true behavior of the material, because of factors like surface roughness, surface contamination, surface microstructural change, subsurface microcracks, and profile accuracy of the indenter, which may affect the response of silicon under indentation on the nanometer scale. Zhang, Tanaka, and Zarudi [28–30] pointed out that a specimen of a silicon monocrystal prepared by grinding and polishing normally contains an amorphous surface layer of SiO_x about 10–80 nm thick, followed by a region with dislocations, as shown in Fig. 6. Hence, experiments using such specimens cannot provide exact information about silicon monocrystal under nanoindentation. Second, it is unclear how amorphous phase transformation and other microstructural changes in silicon monocrystal influence its indentation properties. There have been many reports on the nanoindentation of both ceramics and silicon (e.g., Ref. [31]). The load-displacement curves showed common features of "pop-in" and pop-out." However, the mechanisms of deformation of these materials under nanoindentation have not been fully understood, and experimental findings have often led

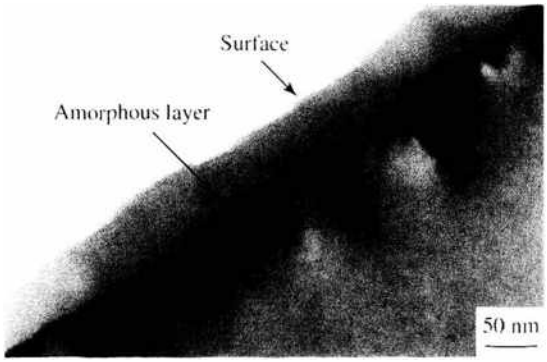


Figure 6. The subsurface microstructure of silicon monocrystals after a two-body contact sliding. Note the top amorphous layer and the dislocations in the crystal zone. (sliding velocity = 23.95 m/s, tip radius= 1 μm, indentation depth = 15.2 nm, sliding in [100] direction). Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

to controversial and conflicting explanations of microcracking, plastic deformation, or phase transformation. Nevertheless, it was found that the stresses in a silicon specimen during unloading do not suggest microcracking. By comparing them with the critical hydrostatic stress for phase transformation based on experiments by Hu and Spain [23], Weppelmannm, Field, and Swain [31] concluded that the phenomenon of "pop-out" was caused by amorphous phase transformation.

All the above studies did not consider the effect of surface roughness on nanoindentation. On the nanometer scale, surface asperities play a central role. For example, surface asperities always deform first before a full contact takes place between indenter and specimen. In the transient process, localized microstructural change may occur and alter the overall response of indentation. However, if the indenter and specimen surfaces are sufficiently smooth, the indentation will be influenced by surface energy, as pointed out by Johnson, Kendall, and Roberts [32]. When the surface roughness is comparable to or larger than the indentation scale, the surface energy effect becomes immeasurable. In addition, probably because of the capacity of the load measurement devices in ordinary nanoindentation tests, unloading indentation in all the above studies was considered to be complete when the load reduced to zero. This ignored the adhesion effect resulting from surface energy on the possible microstructural change that would in turn alter the mechanical properties of the specimen subjected to repeated loading. A thorough theoretical investigation into the mechanics and physics in silicon on the nanometer scale is therefore necessary.

In this section, we will follow the steps outlined in Section 2 in establishing the model and carrying out a reliable simulation of nanoindentation of monocrystalline silicon. The findings from the simulation will also be discussed.

3.2. Modeling

Figure 1 shows the model of a specimen of silicon monocrystal and diamond indenter used in the simulation. The hemispherical diamond indenter has a radius of 2.14 nm, made up of 1818 atoms. The dimension of the control volume of the silicon specimen has to be made sufficiently large to eliminate boundary effects. Taking this into consideration, an optimum control volume is chosen on the basis of an iterative process of increasing the control volume size until further increases do not affect the displacement and velocities of the atoms resulting from the indentation process. An optimum size of 6.5 × 10.3 × 10.3 nm, made up of 36,341 atoms, is obtained for this simulation. For example, Figure 7 shows the displacement field of the atoms, using such a control volume, at the maximum indentation for the optimum control volume size, where the direction of an arrow indicates the direction of an atom displacement, and the length of the arrow is the magnitude of displacement. It can be seen that the atoms affected by the indenter are primarily those near the indenter-silicon interface. Atoms away from that region do not displace from their equilibrium positions.

To restrict the rigid body motion of the specimen, layers of boundary atoms that are fixed to space are used to contain the Newtonian atoms, with the exception of the top (100)

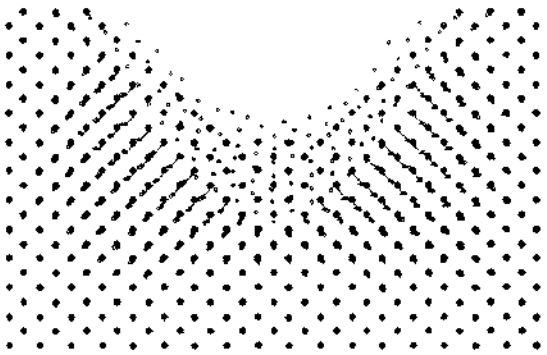


Figure 7. Displacement field of the atoms at the maximum indentation. The arrows represent the displacement of the atoms from their initial position (before indentation) to their new positions at the maximum indentation. The filled circles represent the equilibrium positions of the atoms, whereas the smaller, hollow circles represent the displaced positions of the atoms at the maximum indentation. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

surface, which is exposed to the indenter. Thermostat atoms are also used to ensure reasonable outward heat conduction away from the control volume.

For covalent systems such as silicon, the directionality of bonding is important. Tersoff [6, 7] proposed a simple pairlike potential in which the bond order of the atoms is affected by its local environment. This replaces the two- and three-body potential conventionally employed when directionality of bonding is a concern. On the basis of empirical data, Tersoff also verified that the Tersoff potential is capable of predicting stable phases of diamond cubic silicon and body-centred tetragonal β-tin. Hence Tersoff potential is used in this simulation to dictate the interaction between the silicon atoms. When that assuming j and k are the neigboring atoms of atom i, the atomic bond lengths of atoms i–j and i–k are r_{ij} and r_{ik}, and the angle between bonds i–j and i–k is θ_{ijk}, then the total Tersoff energy E_T can be expressed as

$$E_T = \sum_i E_i = \frac{1}{2} \sum_{i \neq j} W_{ij} \tag{6}$$

where W_{ij} is the bond energy, so that the summation in the equation is about all the atomic bonds in the control volume, W_{ij} is a function of repulsive pair potential f_R and attractive pair potential f_A and has the form of

$$W_{ij} = f_C(r_{ij})[f_R(r_{ij}) + b_{ij} f_A(r_{ij})] \tag{7}$$

where

$$\begin{aligned}
&f_R(r_{ij}) = A_{ij} \exp(-\lambda_{ij} r_{ij}), \qquad f_A(r_{ij}) = -B_{ij} \exp(-\mu_{ij} r_{ij}) \\
&f_C(r_{ij}) = \begin{cases} 1 & r_{ij} \leq R_{ij} \\ \dfrac{1}{2} + \dfrac{1}{2} \cos[\pi (r_{ij} - R_{ij})/(S_{ij} - R_{ij})], & R \leq r_{ij} \leq S_{ij} \\ 0 & r_{ij} \geq R_{ij} \end{cases} \\
&b_{ij} = \chi_{ij} (1 + \beta_i^{ni} \zeta_i^{ni})^{-1/2ni}, \quad \zeta_{ij} = \textstyle\sum_{k \neq i,j} f_C(r_{ik}) g(\theta_{ijk}) \\
&g(\theta_{ijk}) = 1 + \frac{c_i^2}{d_i^2} - c_i^2/[d_i^2 + (h_i - \cos\theta_{ijk})^2] \\
&\lambda_{ij} = (\lambda_i + \lambda_j)/2, \quad \mu_{ij} = (\mu_i + \mu_j)/2, \quad A_{ij} = (A_i A_j)^{1/2} \\
&B_{ij} = (B_i B_j)^{1/2}, \quad R_{ij} = (R_i R_j)^{1/2}, \quad S_{ij} = (S_i S_j)^{1/2}
\end{aligned} \tag{8}$$

Other parameters such as A, B, R, S, λ, χ, and μ, as listed in Table 1, are Tersoff potential parameters, depending on individual materials. With Eqs. (6) and (7), the interaction forces between silicon atoms can be obtained by calculating the gradient of E_T.

Table 1. Parameters in Tersoff potential for carbon and silicon.

	Carbon	Silicon
A (eV)	1.3936×10^3	1.8308×10^3
B (eV)	3.4670×10^2	4.7118×10^2
$\lambda\mu$	34.879	24.799
μ	22.119	17.322
β	1.5724×10^{-7}	1.1000×10^{-6}
n	7.2751×10^{-1}	7.8734×10^{-1}
c	3.8049×10^4	1.0039×10^5
d	4.384×10^0	1.6217×10^1
h	-5.7058×10^{-1}	-5.9825×10^{-1}
R (nm)	0.18	0.27
S (nm)	0.21	0.30
$\chi_{C-C} = 1.0$	$\chi_{Si-Si} = 1.0$	$\chi_{C-Si} = 0.9776$

The interaction between the silicon and diamond indenter atoms is modeled by the modified Morse potential [2] given by

$$\phi(r_{ij}) = \lambda_1 D[e^{-2\lambda_2\alpha(r_{ij}-r_0)} - 2e^{\lambda_2\alpha(r_0-r_0)}] \tag{9}$$

The parameters such as D, α, and r_0 are shown in Table 2. Interaction force is calculated by the gradient of ϕ.

The choice of these potentials is supported by previous simulations and tests, which showed good agreement between simulation results and experimental data [2–4, 8, 12].

To simulate the nanodeformation under room-temperature conditions, the silicon atoms were initially arranged in a perfect diamond cubic structure, with the lattice parameters equal to its equilibrium value at an ambient temperature of 23°C. The ambient temperature is maintained by the use of the thermostat atoms that surround the control volume, as shown in Fig. 1. During the simulation process, the temperature of the thermostat atoms is kept at 23°C by scaling their velocities at every time step, using Eq. (1).

On the basis of the method of time-step selection described in Section 2.4, it was found that a time step of 1.0 fs for diamond and 2.5 fs for silicon would provide sufficiently accurate integrations.

3.3. Phase Transformation

Snapshots of the location of the atoms of a silicon specimen at different stages during the indentation are shown in Fig. 8. The size of the spheres that depict silicon atoms has been deliberately reduced so that we can clearly see any changes to the crystalline order of diamond cubic silicon. At the maximum indentation (Fig. 8a), it can be observed that the order of atoms beneath the indenter differs considerably from its original diamond cubic pattern. However, these transformed atoms still maintain a long-ranged crystalline order. This indicates that a displacive phase transformation of one crystalline form of silicon to another has occurred.

An investigation into the coordination number of the atoms reveals such a transformation [33]. There is a significant increase in the number of six-coordinated atoms. Figure 9

Table 2. Parameters in the standard Morse potential.

Parameter	C–Si
D (eV)	0.435
α(nm^{-1})	46.487
r_0 (nm)	0.19475
λ_1	1
λ_2	1

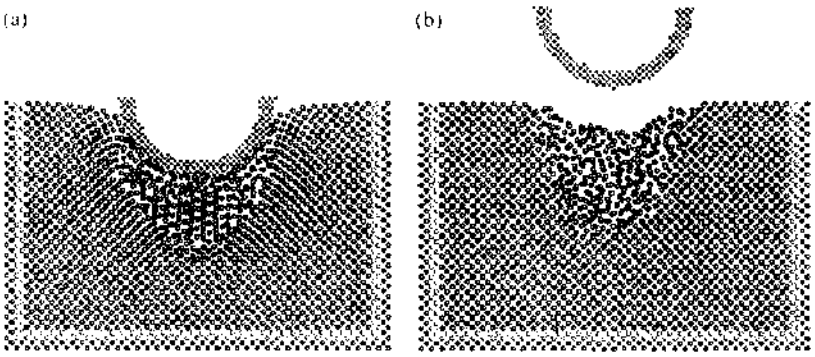


Figure 8. Phase transformations during indentation. (a) Positions of atoms in silicon specimen at the maximum indentation. Atoms beneath the indenter (circled by the dotted line) have a crystalline order different from that of original diamond cubic structure. (b) Residual amorphous silicon after indentation. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

shows the variation in the coordination numbers of the silicon atoms during the indentation process. This is consistent with the theoretical coordination number of an atom in the β-tin phase. It can also be seen that the atoms that are six coordinated are formed in the region just beneath the indenter during maximum indentation. As a further investigation into the nature of the phase transformation, the bond length distributions of the atoms within the transformation zone are calculated. The results of these calculations show that during the indentation, the average atomic distance between the atoms that have undergone transformation changed from 2.35 Å (diamond cubic structure) to 2.43 and 2.58 Å (β-tin) (Fig. 10). From the simulation, it is found that the change is caused by the flattening of the tetrahedron structure in diamond cubic silicon. The displacive transformation at progressive time steps is demonstrated in Fig. 11. By determining the spatial coordinates of the atoms, it is found that these four atoms of the flattened tetrahedron and another two atoms at a slightly further distance of 2.58 Å form the six nearest neighbors of the six-coordinated atoms. Figure 12(a) shows one of these atoms with its six nearest neighbors. At the maximum indentation, about 730 atoms transform from the four-coordinated diamond cubic phase to the six-coordinated β-tin phase. A portion of the transformed six-coordinated atoms beneath the indenter obtained from the simulation is shown in Fig. 13. The atoms form a repetitive crystal structure with lattice parameters $a = 4.684$ Å and $c = 2.585$ Å. These parameters of the new phase formed are in complete agreement to Donohue's description [34] of high-pressure

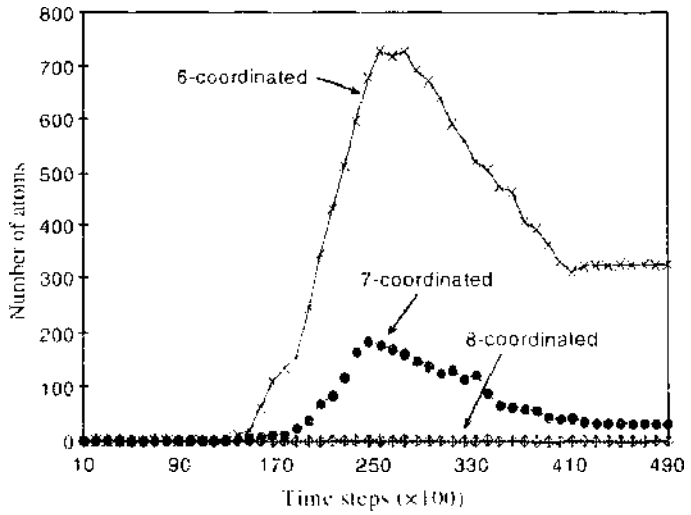


Figure 9. Graph of number of atoms with specified nearest number of neighbors against time. Note that here is a significant increase in the number of atoms having six nearest neighbors during indentation. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

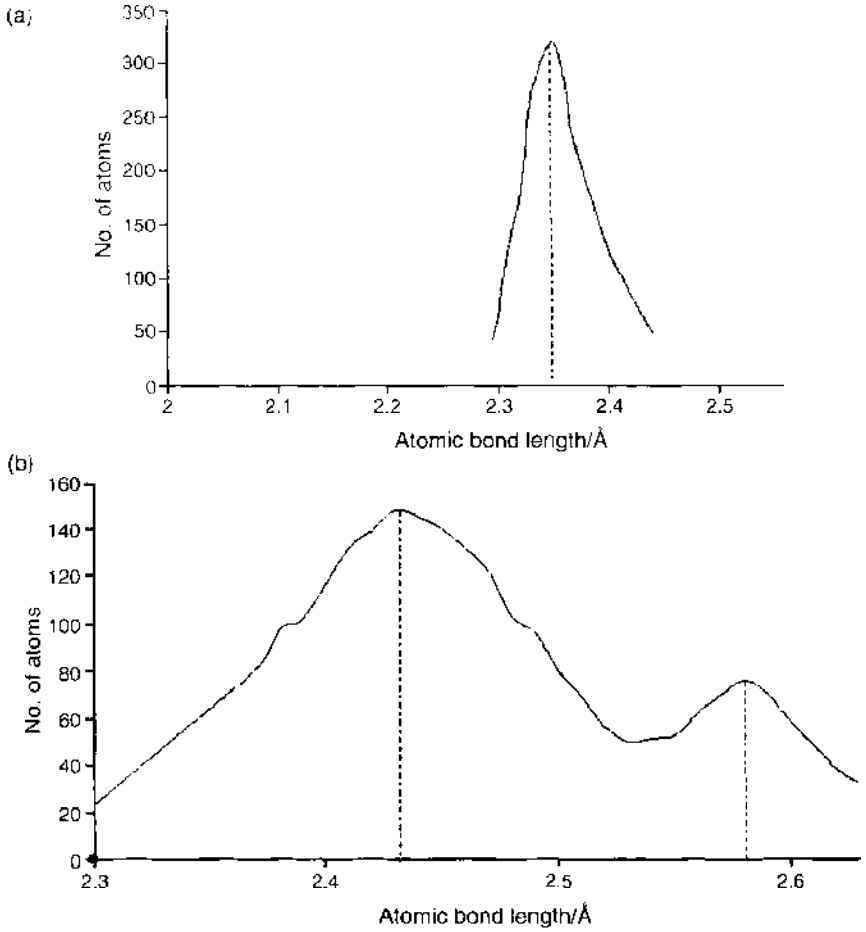


Figure 10. Length of bond. (a) At the start of indentation; (b) at the maximum indentation. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

β-tin. The change of structure from diamond cubic to β-tin is accomplished by displacing atoms along the c-axis with an increase in bond length and a decrease in volume. Hence, this simulation indicates that β-tin forms beneath the indenter in the compressive stress region through displacive phase transformation during indentation.

To ensure that the β-tin phase is not an intermediate phase obtained as a result of the short simulation time, the indenter's position is kept unchanged for at least 250,000 time

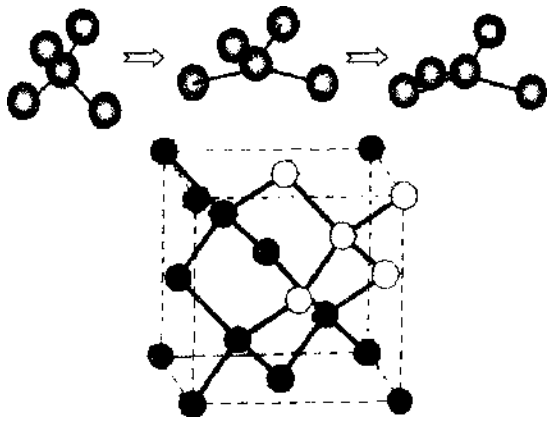


Figure 11. Flattening of the tetrahedron in the diamond cubic structure. The atoms with a lighter color form the tetrahedron in the diamond cubic structure. The three diagrams of the top row show, respectively, the shape changes of the tetrahedron before indentation, during indentation, and at the maximum indentation.

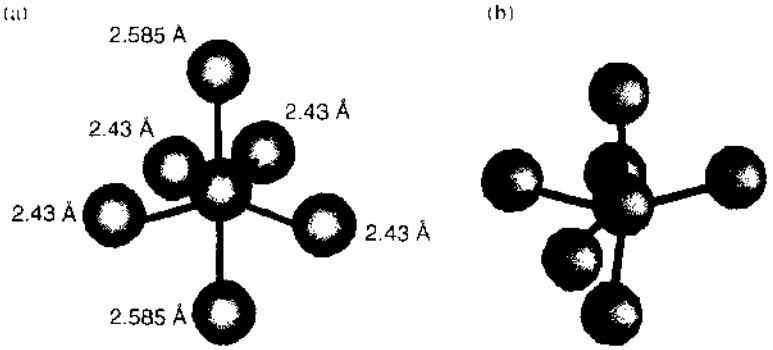


Figure 12. Atoms in β-tin silicon structures. (a) An atom of β-tin silicon with its six nearest neighbors. Four atoms are at a distance of 2.43 Å, and another two are at a slightly further distance of 2.585 Å. (b) An atom of the distorted β-silicon obtained during the second indentation, with its six nearest neighbors. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

steps at a stage with β-tin formed (e.g., at the maximum indentation). By determining the spatial coordinates of these atoms, it is found that the β-tin structure in the transformed zone remains unchanged in that period of time. The coordination number of the atoms is once again recorded during the simulation. Figure 14 shows the variation in the coordination number of the atoms with the number of time steps. Clearly, the number of atoms that are six coordinated (β-tin atoms) remains constant for the entire period during which the indenter's position is held unchanged. It also indicates that the new phase formed is not an intermediate unstable phase. However, when the indenter is unloaded, the β-tin transforms to amorphous silicon. (Details will be discussed later.) This means that the β-tin phase obtained is stable as long as the required stress field is maintained.

Experimental studies [35–39] have shown that phase transformation in silicon from its normal, diamond cubic structure to the denser, β-tin structure will take place under pure hydrostatic pressure in the range of 11–12 GPa. However, it was also found that under more complex conditions, such as in indentation, the transformation pressure may be reduced to as low as 8 GPa [40]. In this simulation, the maximum hydrostatic pressure attained is 12 GPa, which is consistent with these experimental findings.

Referring once to Fig. 8b, it is observed that the crystalline order of the atoms is lost on the unloading of the indenter, showing that a body centred tetragonal to amorphous phase transformation has occurred. This is consistent with Clarke et al.'s observation [37–39] that proposed a possible explanation for the formation of the amorphous silicon after indentation in this study: At the relatively rapid unloading rate employed (40 m/s) and the nonhydrostatic constraint imposed on the transformed region, the high-pressure, crystalline form cannot transform back fast enough, and without complications, the amorphous phase forms metastably.

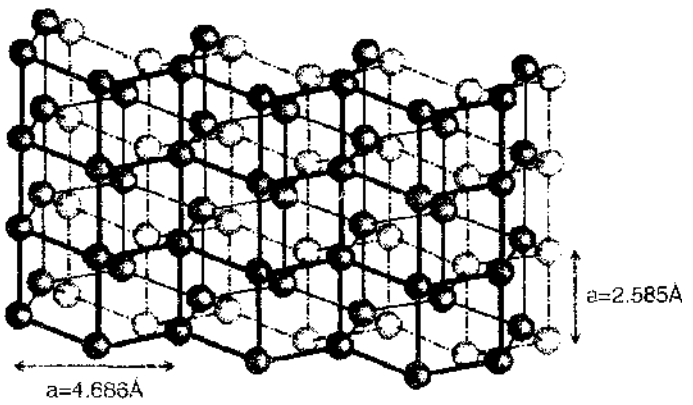


Figure 13. Crystal structure of β-tin silicon at the maximum indentation. Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

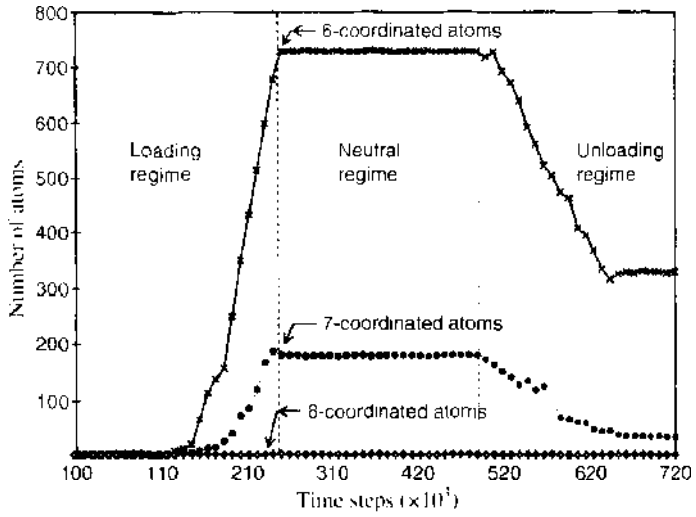


Figure 14. Number of atoms with specified nearest number of neighbors against time. In this simulation, the indenter is deliberately held at the maximum indentation for 250,000 time steps, denoted as the neutral regime, to show the stability of the transformed β-tin silicon phase when the Tersoff potential is used Reprinted with permission from [33], W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000). © 2000, Institute of Physics Publishing.

It is interesting to note that within this amorphous region, most of the atoms are four coordinated with the exception of some atoms that are six coordinated. This indicates that the amorphous phase consists of four-coordinated atoms, but without any long-range order. The six-coordinated atoms are observed to be the crystallite remnants of the β-tin phase, maintaining the tetragonal body–centered crystal structure, interspersed within the amorphous region. Experimental findings also lend credence to this claim [38–39].

In his experiments with silicon specimens subjected to hydrostatic pressure in a diamond-anvil pressure cell, Minomura [41] found that the β-tin to amorphous phase transformation is in fact reversible. To examine whether this is also the case in high-speed nanoindentation, Cheong and Zhang [42] carried out a repeated indentation simulation. It was found that the β-tin phase could indeed have recovered on the second indentation loading. This conclusion was drawn from the fact that there is once again an increase in the number of six-coordinated atoms during the second indentation. However, it was found that the phase transformation during the second indentation is heterogeneous, with mixing of tetragonal body-centered and amorphous phases. In addition to that, the β-tin structure obtained from the indentation of amorphous silicon (the second indentation) is distorted compared to that obtained from the indentation of diamond cubic silicon during the first indentation. Figure 12b shows the six nearest neighbors of the distorted β-tin structure. A comparison between Figs. 12a and 12b shows the extent of the distortion. This is also in agreement with experimental claims.

In the simulation, there is an absence of the body-centered cubic structure that is expected on unloading of the β-tin, as observed by Hu and coworkers in their hydrostatic loading tests [23]. This can be a result of the fact that the hydrostatic pressure under the indenter during loading is not high enough to affect such a transformation on unloading. In this simulation, the maximum hydrostatic pressure under the indenter is only 12 GPa. Minomura [41] stated that for silicon specimens subjected to pressure above 15 GPa, transformation into the body-centered cubic structure occurs on unloading. For a pressure (11–15 GPa) lower than that, the β-tin will reversibly transform into the amorphous phase, as is the case for the present simulation. In addition, the rapid rate of unloading and low load favors phase transformation of β-tin silicon to amorphous silicon without other crystalline phases, as stated by Ge, Domnich, and Gogotsi [43].

3.4. Deformation Characteristics

Zhang and Tanaka [3] investigated carefully the deformation characteristics of monocrystalline silicon subjected to nanoindentation, using the molecular dynamics analysis [3]. A complete load-displacement curve of indentation, as shown in Fig. 15(a), consists of the loading

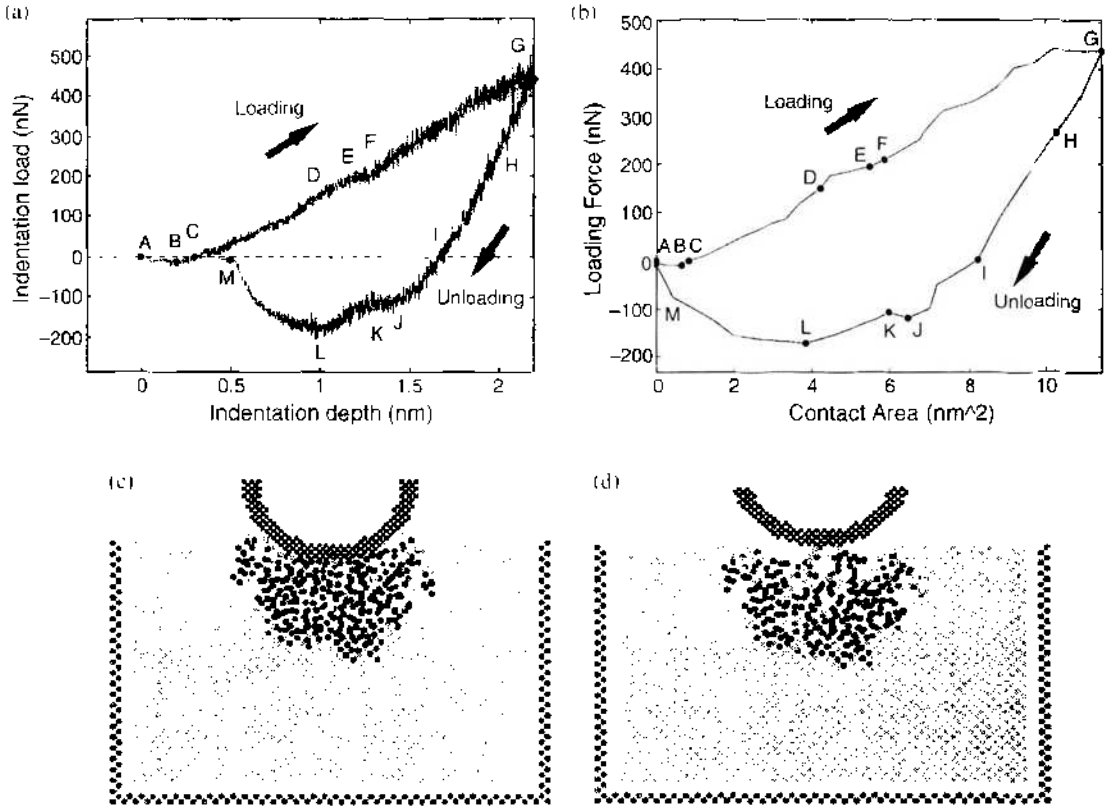


Figure 15. Overall response of silicon nanoindentation in a complete loading–unloading cycle. The indentation conditions are $d^{min} = 2.1947$ nm and $R = 2.14$ nm. The diamond indenter in this case contains 1818 atoms. In (c) and (d), compare the difference of adhesion between the diamond and silicon surfaces. (a) Load-displacement curve ($V = 40$ m/s); (b) load-contact area curve ($V = 40$ m/s); (c) contact during unloading at $P = -128$ nN ($V = 200$ m/s, $d = 0.6947$ nm); (d) contact during unloading at $P = -16.4$ nN ($V = 200$ m/s, $d = 0.1947$ nm). Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

path, ABCDEFG, and the unloading path, GHIJKLM. The significant deviation between loading and unloading paths indicates that a remarkable inelastic deformation occurs during indentation. Compared with any experimental measurements of nanoindentation (e.g., those presented in Refs. [21, 38]), the present curve shows two distinguished features demonstrated by part ABC in the loading path and part IJKLM in the unloading path. From state A to C, the indentation load is negative (tensile), which means that when the diamond indenter approaches the silicon specimen, the surface atoms in silicon and diamond first experience attractive forces before compression. Thus, even under a large tensile indentation load, as shown in Fig. 15b, the contact area between the indenter and specimen is finite. However, at the end of loading (state G in Fig. 15a), the contact area becomes very large. Thus, at the state with zero indentation load during unloading (state I in Fig. 15a), the contact area is still large because of the attraction between silicon and diamond atoms. The attraction keeps increasing until state *L*, although the contact area is continuously shrinking (Fig. 15b). Such a process of contact area shrinkage can be more clearly seen in Fig. 15c and 15d through a cross-sectional view of the atomic lattice deformation.

Under static and purely elastic conditions, Johnson, Kendall, and Roberts [32] developed the well-known JKR theory to modify the Hertzian prediction of contact area by considering the effect of surface adhesion energy. For the present indentation configuration, the JKR theory gives rise to

$$\left[\frac{P}{P_c} - \left(\frac{a}{a_c}\right)^3\right]^2 = 4\left(\frac{a}{a_c}\right)^3 \tag{10}$$

and indicates that contact breaks when $P = -5P_c/9$ at $a = a_c/3^{2/3}$, where P_c is the maximum tensile indentation load and a_c is the corresponding contact area. Figure 16 compares the

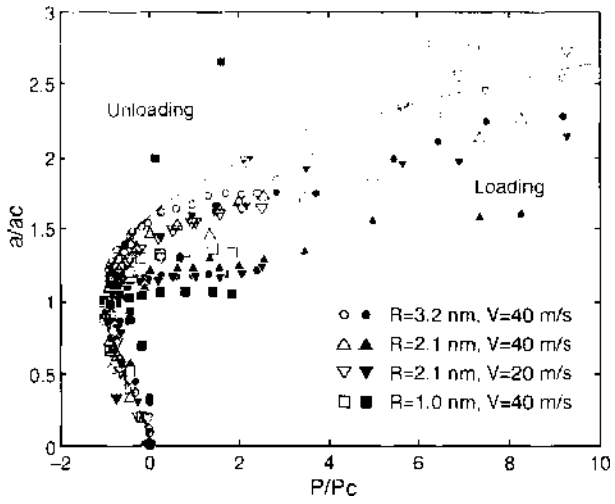


Figure 16. Comparison with the JKR theory. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999), © 1999, Japan Society of Mechanical Engineers.

prediction from Eq. (10) and the result from molecular dynamics simulation, where P_c and a_c are taken as the indentation loads and contact areas at states B and L for loading and unloading, respectively. It is interesting to note that in the vicinity of $P = -P_c$, the predictions of the JKR theory are in good agreement with the present results, although the theory is only valid for an elastic and static indentation, whereas significant inelastic deformation occurred in these cases and indentation speed varied largely from 20 to 200 m/s.

Under purely elastic indentation, the contact will become unstable when P reaches $-5P_c/9$ during unloading, as pointed out by the JKR theory. In an inelastic indentation, however, the occurrence of such contact instability depends on the strength of atomic interaction of the material subjected to indentation and that between the material and the indenter. In all the indentation tests in this study, the instability of contact during unloading never happens, as shown in Fig. 16. When indenting some other materials, as reported by Durig and Stalder [44], for instance, instability still occurs, although significant plastic deformation appears.

After the initial attraction caused by surface adhesion (A–C in Fig. 15a), the indentation is elastic as reflected in the load displacement relationship from C to E on Fig. 15a. When indentation proceeds to a certain level, say state E for the case in Fig. 15a, phase transformation from the diamond cubic to the β-tin silicon occurs. From E to F, the number of transformed atoms increases by 218%, and the volume of transformed zone increases by 202% (see Fig. 17a and Fig. 17b). The fast phase change from E to F causes a rapid temperature rise (Fig. 17c). The β-tin silicon is much softer than the crystal silicon, introducing a local softening phenomenon from E to F. On the load-displacement curve of indentation (Fig. 15a), the softening brings about a clear slope change from CDE to EF. As a result, the transformed phase is condensed quickly so that the density of the phase increases considerably from D to F (see Fig. 17d). After state F, the variation of all the above (transformed phase growth, temperature rise, contact area development, and density increment of the transformed phase) becomes steady and linear. Thus from F to G, both the elastic deformation in the crystal zone (bending of atomic lattice) and the further development of transformed phase have linear contributions to the overall deformation and give rise to a linear load-displacement relationship from F to G, as shown in Fig. 15a. In this stage, the average relative density of the transformed phase increases from 1.55 to 1.70 (Fig. 17d).

During unloading, the deformation behavior from G to I is mainly the result of elastic recovery of the crystal lattice. This phenomenon is more evident at the very beginning of unloading from G to H. In this period, the total volume of the transformed zone decreases linearly by 18.2% (Fig. 17b), and the total number of transformed atoms reduces by 14.8% (Fig. 17a), but the average density of transformed phase increases by 1.2% (Fig. 17d). This means that the deformation is dominated by elastic recovery, so that the load-displacement curve behaves also linearly. It is interesting to note that the contact area (Fig. 15b) and

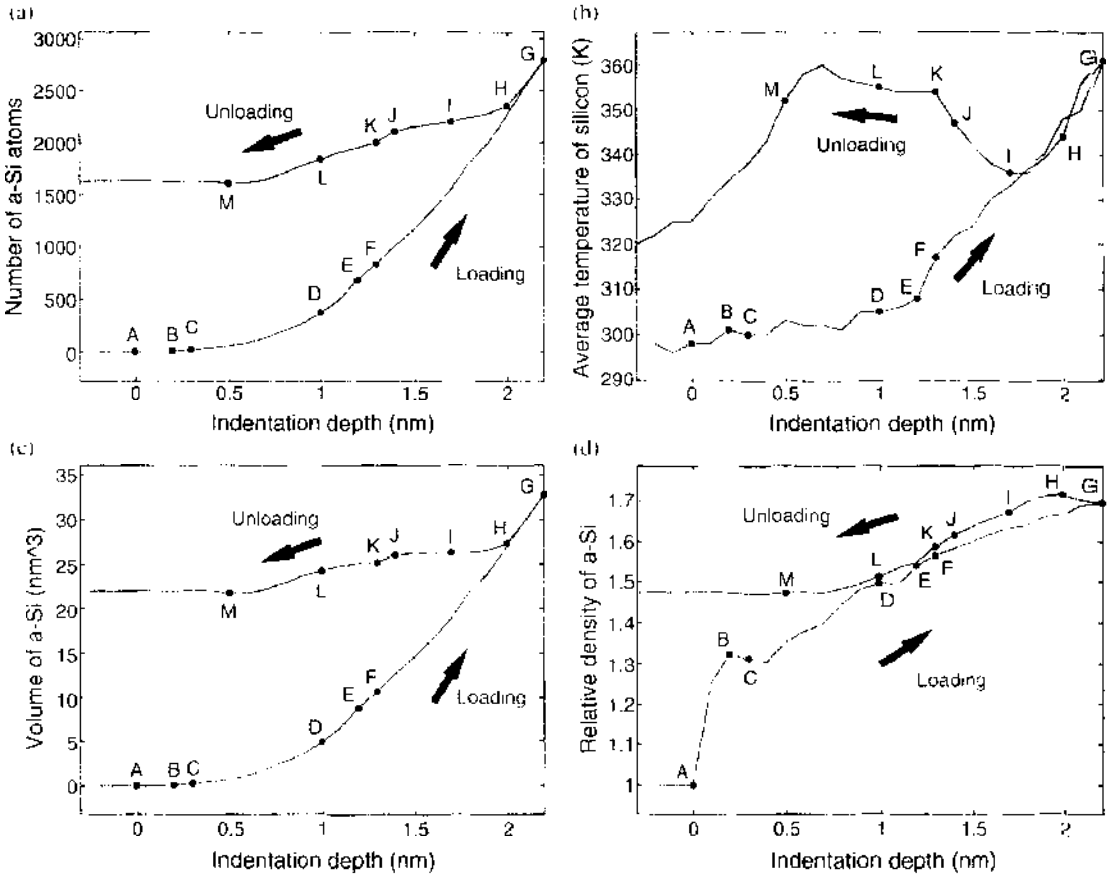


Figure 17. Variation of key quantities during indentation. Indentation conditions are $d^{\max} = 2.1947$ nm, $R = 2.14$ nm, and $V = 40$ m/s. (a) Number of amorphous silicon; (b) volume of amorphous zone; (c) average temperature rise in the specimen; (d) ratio of the average density of amorphous silicon to the density of crystal silicon. The notation a-Si in the figure stands for amorphous silicon. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

temperature of the specimen (Fig. 17b) also decreases linearly. After state I, the effect of elastic recovery is no longer dominant. The continuous recrystallization and density change with almost a constant amorphous volume make the load-displacement curve vary in a complex manner from I to M. Because of this, temperature rises again. From J to K, recrystallization becomes dominant, leading to a fast decrease of density, a drop of amorphous volume, and a weaker interaction with the diamond indenter. This makes the load-displacement curve bend from J to K. The recrystallization process finishes shortly after state L, when the number of amorphous atoms becomes constant, as shown in Fig. 17a. It is also clear that state L is critical, beyond which the indenter–specimen contact becomes purely adhesive because both the volume and density of amorphous zone become constant. After L, contact area shrinks quickly and finally vanishes at M with an residual indentation depth. An irreversible indentation mark thus remains as a result of the local density change of silicon.

A silicon monocrystal is anisotropic, as its atomic structure implies. The anisotropy can be understood more directly by monitoring the development of the amorphous zone. Figure 18 shows the profile change of the amorphous zone in xoy-plane (see Fig. 1) when observing in the positive z-direction. Clearly, amorphous phase grows with different rates in different directions, and the difference becomes greater when indentation proceeds. However, during unloading, recrystallization also occurs with different rates in different directions. As shown in Fig. 18, in the direction that amorphous phase grows faster during loading, recrystallization also goes faster during unloading. As a result, the amorphous zone during and after unloading becomes quite axisymmetric. Nevertheless the deformation anisotropy is not remarkable in the whole loading and unloading process. Indeed, the anisotropy ratio of a silicon monocrystal under pure elastic deformation is 0.64, which deviates from an ideally isotropic material by only 36% [10, 45]. All the above has been detailed in Ref. [3].

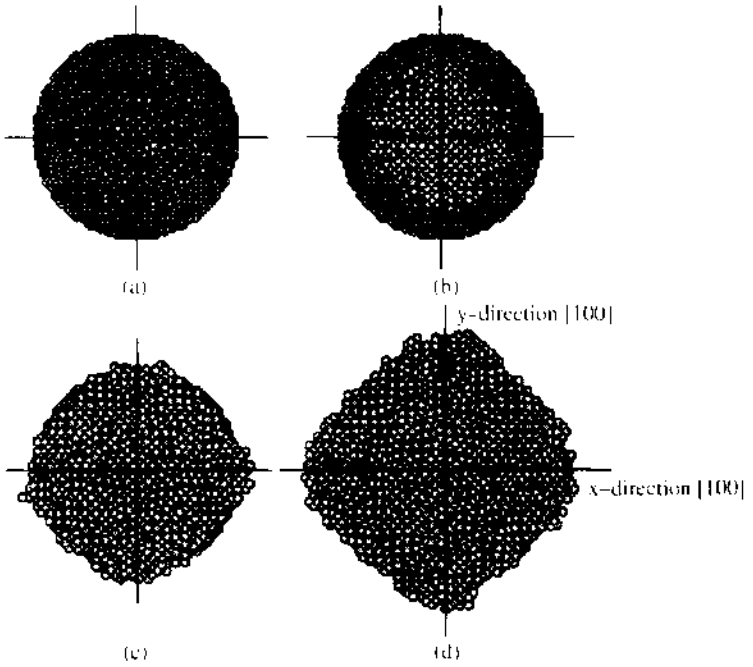


Figure 18. Bottom view of the development of amorphous zone during loading when indentation depth increases. Reprinted with permission from [3], L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999). © 1999, Japan Society of Mechanical Engineers.

3.5. Effect of Oxygen Penetration

This section discusses the oxidation of silicon during an indentation in a nonvacuum environment.

The oxidation of silicon without mechanical loading has been studied to a certain extent [46–51]. For example, Kato et al. [46] reported that the back bond oxidation was energetically favored. However, they found that this oxidation requires an activation barrier of 0.8–2.4 eV. Bu and Rabalais [47] used the time-of-flight scattering and recoiling spectrometry technique to study the chemisorptions of O_2 on a Si (100) surface. They found that the exposure to O_2 resulted in dissociative chemisorptions, and at higher coverage the O atoms were adsorbed at the bridge positions between first and second silicon layers. Pasquarello et al. [48] studied the oxidation of the Si (001)–SiO_2 interface and observed that three layers of crystalline silicon were oxidized. They also observed that the oxygen diffusion involved an intermediate configuration in which the oxygen atom is three-fold coordinated by silicon. With external mechanical stresses, the effect of oxygen penetration becomes more significant. In an experimental study on contact sliding, Zhang and Zarudi [30] reported that the oxygen penetrates into the amorphous layer, changes the atomic bonding of silicon, alters the threshold of phase transformation, and accelerates wear. However, the mechanism of this chemical process is not yet known.

To capture the effect of position of the oxygen molecule, Mylvaganam and Zhang [52] investigated the following special indentation cases.

First, the O_2 molecule was laterally (i.e., parallel to the surface) right below the indenter and 2 Å above the surface. In this case, the oxygen molecule dissociated into atoms, and both atoms went into the substrate and formed three or four coordinations with Si atoms. At the early stages of indentation, these oxygen atoms were within the first two or three layers of silicon. As the loading proceeded, these oxygen atoms went in farther, but they were close to the indenter tip, and there were no significant changes in the silicon–oxygen coordination. Toward the maximum loading, and at unloading, the addresses of the nearest neighbors showed that the oxygen atoms were coordinated to different silicon atoms. However, the oxygen atoms were three- or four-coordinated with silicon atoms throughout loading and unloading. At the end of the indentation process, the oxygen atoms were found about 8 Å below the surface in the amorphous phase.

The difference in the total energy during the latter stages of loading and on unloading of the indentation process with and without oxygen would give the energy released in the formation of new Si–O bonds. This analysis showed that altogether, 10 new Si–O bonds should have formed. This is higher than the number of Si–O bonds that are worked out by looking at the nearest neighbors. However, it was found that there were some Si atoms that would have weak interactions with the penetrated oxygen atoms, which would account for the excess energy.

Second, the O_2molecule was 1 Å away from the center. During the loading process in this case, the oxygen molecule dissociated into atoms, but only one oxygen atom went into the substrate, coordinated with silicon atoms, and formed three/four coordinations with silicon atoms as in Case (i).

Third, the O_2 molecule was 5 Å above the surface. In this case, at the initial stages the O_2 molecule rotated around and then dissociated into atoms, but only one oxygen atom went into the substrate and formed three or four coordination with silicon atoms.

These results indicate that the position and orientation of the oxygen molecule play a key role in the penetration of the oxygen atoms into the substrate. The structural changes during the loading and unloading of the indentation process when the O_2 is placed right below the indenter are shown in Fig. 19. It is clear that there is a significant difference in the number of four- five-, and six-coordinated silicon atoms. At the maximum indentation, there is a remarkable increase in the number of five-coordinated silicon atoms, which could be an intermediate in the formation of six-coordinated β-tin phase, when the indentation was carried out in the presence of O_2. This is a result of the attraction of silicon atoms toward the penetrated oxygen atoms and is confirmed by Fig. 20, which shows the number of silicon atoms that are within 10 Å of the O atom/dummy O atom. Throughout the indentation process (except at the very beginning), there are more Si atoms around the O atom compared to the number of Si atoms in the same region when the indentation is carried out without O_2 (i.e., in vacuum).

On unloading, as discussed in the prevous section [33], the β-tin silicon transforms into the amorphous phase, and as a result, the number of four-coordinated silicon atoms may increase, which reflected in Fig. 19a. The fact that after unloading there are more Si atoms around oxygen clearly indicates that the oxygen penetration causes damage to the silicon substrate.

In summary, mechanical process can cause silicon monocrystal to get oxidized, and the oxidation can damage its subsurface structure. This explains qualitatively why oxidation and its associated subsurface damage take place in nanotribological sliding and precision surfacing, such as grinding, polishing, and lapping. In these processes, the many moving surface

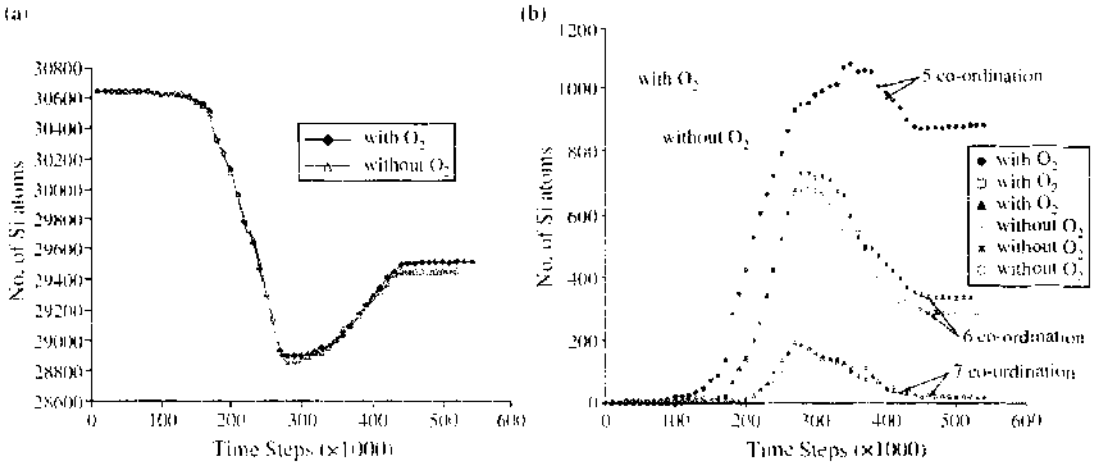


Figure 19. Structural changes of the substrate. (a) Comparison of the number of four-coordinated Si atoms; (b) comparison of the number of five-, six-, and seven-coordinated Si atoms during the indentation with and without O_2. Reprinted with permission from [52], K. Mylvaganam and L. C. Zhang, *Nanotechnology* 13, 623 (2002). © 2002, Institute of Physics Publishing.

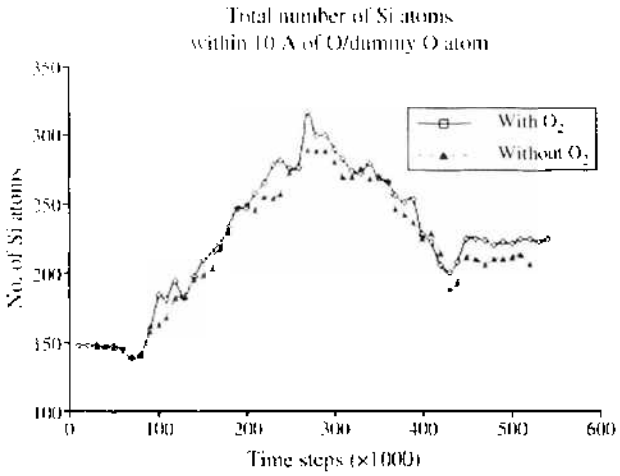


Figure 20. Comparison of the total number of Si atoms around one of the penetrated O atom/dummy O atom during the indentation process with and without O_2. Reprinted with permission from [52], K. Mylvaganam and L. C. Zhang, *Nanotechnology* 13, 623 (2002). © 2002, Institute of Physics Publishing.

asperities or abrasive grits, which can be viewed as the moving indenters, would increase greatly the chance of oxidation.

The above discussion concerns the oxidation of an initially clean surface. An analysis with a preoxidized layer on a pure silicon sample can be carried out without difficulty.

3.6. Effect of Water

Polishing and grinding of monocrystalline silicon are usually done in open environment with water-based coolant or abrasive slurry [30, 53–55]. The existance of water will introduce changes of mechanical properties of silicon. Figures 21 and 22 show the experimental results of monocrystalline silicon subjected to cycle indentations in air and in water under the same loading and unloading conditions [56]. It was found that in the first cycle, the load-displacement curve of indentation in water is almost identical with that in air. However, a different response appears on the second cycle (Fig. 21). The gradient of the loading curve in water becomes higher, indicating that the mechanical properties of silicon have been altered because of the involvement of water in the first indentation cycle, and the discrepancy between the load-displacement curves becomes greater with further indentation cycles. The difference is also reflected by the dimension of the residual indentation marks and by the amount of material extruded during the cyclic loading, as shown in Fig. 22, where silicon in water exhibits less plastic deformation. The detailed mechansm is unclear so far.

We can use molecular dynamics simulation to achieve an understanding. Figure 23 shows a cross section of the initial model of nanoindentation under water, consisting of 36,285

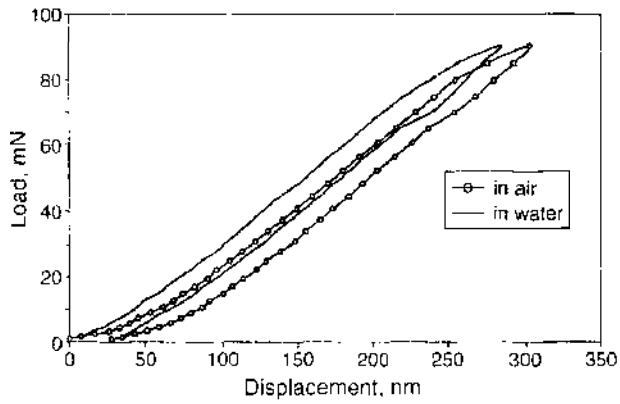


Figure 21. The load-displacement curves of the second indentation cycle when tested in air and water. The maximum indentation load is 90 mN, and the radius of the diamond indenter is 5 μm.

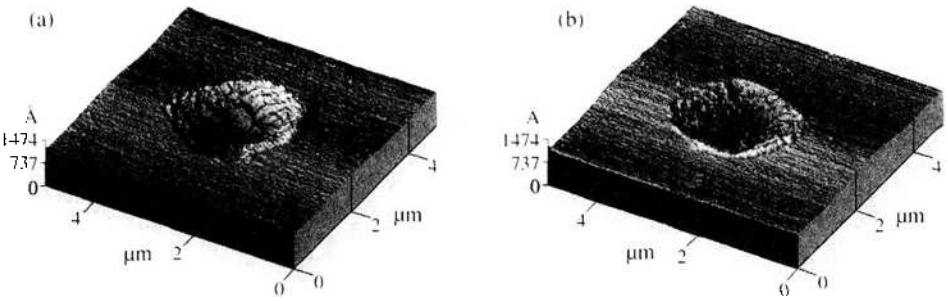


Figure 22. Difference in residual deformation in silicon when indented in air and in water, respectively. (a) Residual indentation mark after the fifth indentation cycle in air; (b) residual indentation mark after the fifth indentation cycle in water. Reprinted with permission from [56], I. Zarudi et al., *Key Eng. Mater.* 233–236, 609 (2003). © 2003, Trans Tech Publications.

silicon atoms, 3,906 carbon atoms (diamond indenter), and 3,468 water molecules. Similarly, the Tersoff potential is used to determine the interactions between silicon atoms, and the Morse potential is applied to describe those between the diamond indenter and the silicon workpiece. To introduce the effect of water, a rigid water model [57–59] is employed to determine the interactions between water molecular sites (i.e., two positively charged hydrogen atoms, one noncharged oxygen atom, and one negatively charged site) located in a planar configuration. However, the Si–H_2O and C–H_2O interactions are described by the long-range Lennard–Jones potential

$$v(r) = 4\varepsilon\left[\left(\frac{\delta}{r}\right)^{12} - \left(\frac{\delta}{r}\right)^{6}\right] \tag{11}$$

where ε and δ are the Lennard–Jones parameters that can be determined by the Lorentz–Berthelot mixing rules [60]. Table 3 lists their values for C–O and Si–O interactions based on the study by Yang and Kim [61].

It was found that as the indenter was pushed into silicon, water molecules were trapped at the indenter–silicon interface, thereby forming a cavity in the silicon surface (Fig. 24). The damaged zone grows in size and forms a rougher residual indentation mark. The tapped water changed the local contact stress distribution and gave rise to a smaller phase

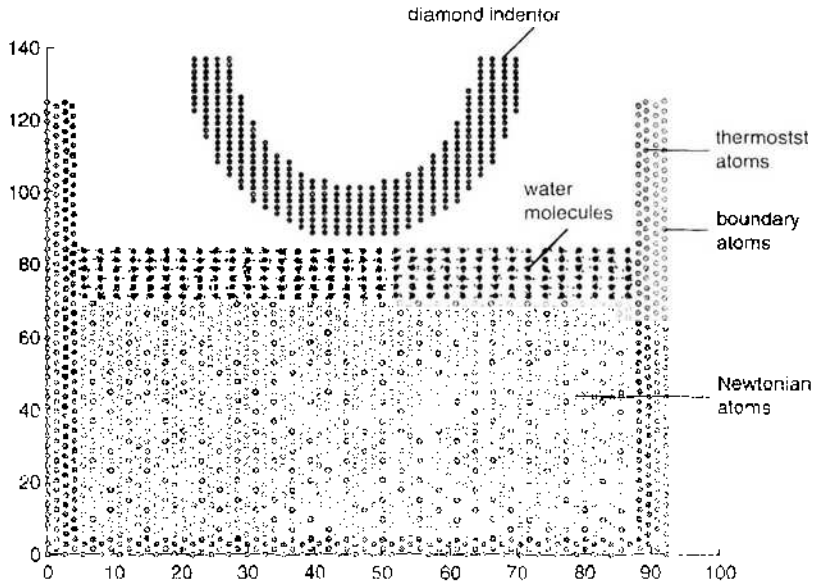


Figure 23. Initial two-dimensional cross-sectional view of the molecular dynamics model for nanoindentation in water. The water molecules are originally placed in a lattice configuration and allowed the system to relax for 5000 time steps before the diamond indenter is pushed downward.

Table 3. Lennard–Jones Potential Parameters.

Inter-molecular pair	ε(kJ mol^{-1})	δ(Å)
C–O	0.4785	3.275
Si (4-coordination)–O	0.9286	3.388
Si(2-coordination)–O	53.6616	1.3807

transformation zone when compared with the indentation without water. In addition, water at the interface reduced the indenter–silicon adhesion considerably. All these changes contributed to the change of the overall response of silicon to indentation.

3.7. Cyclic Indentations

The previous sections have shown the phase transformation details in silicon during a single nanoindentation cycle and provided some understanding of the overall response of silicon in a complete loading–unloading process. However, the results obtained are not sufficient to uncover the deformation mechanisms of many processes. For example, in grinding and polishing wafers of monocrystalline silicon, the material is actually subjected to repeated tool-workpiece interactions. When the first interaction has created a damaged zone, the material may deform differently under subsequent operations. The purpose of this section is to understand how monocrystalline silicon will behave under repeated indentations using molecular dynamic analysis [42]. The model used is the same as that for the previous single nanoindentation.

It shows that the residual amorphous phase after the first indentation seems to remain amorphous throughout the second and third indentations. There is no significant change in size of the amorphous zone. The residual indentation depths also appear to be consistent after the first, second, and third indentations. Figure 25 shows the volume of the transformed zone during indentation. There is only a small increase in the volume of amorphous silicon at the second and third indentations. However, by considering the number of the nearest-neighbor atoms, it can be observed that there is an increase of atoms with six nearest neighbors during the loading phase of each indentation. This indicates the recovery of the β-tin phase mentioned earlier.

Figure 26 shows the load-displacement curves for all three indentation cycles. The indentation depths are taken from the same reference, which is the initial untouched surface of the monocrystalline silicon, and that is why the load-displacement curves of the second and third indentations start at the depth of about 0.6 nm, which is the depth of the residual indentation mark after the first indentation. It is clear that the load-displacement curve of the first indentation is quite different from those of the second and third. First, there is a

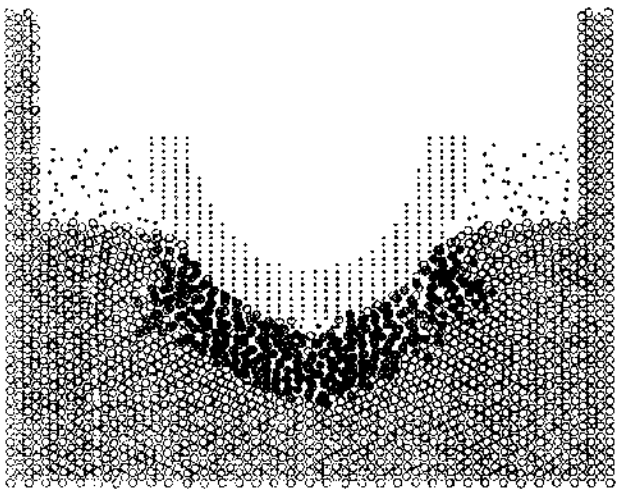


Figure 24. Cross-sectional view of simulation of nanoindentation with water at the indentation depth of 2.58 nm.

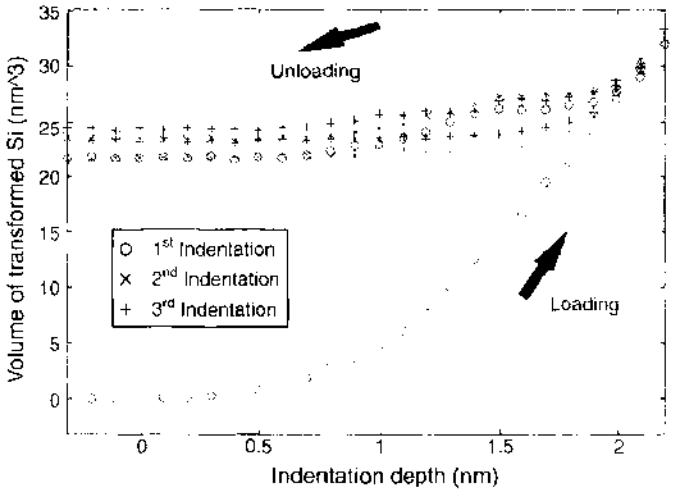


Figure 25. Volume of transformed silicon versus indentation depth during repeated indentations. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000). © 2000, Kluwer Academic Publishers.

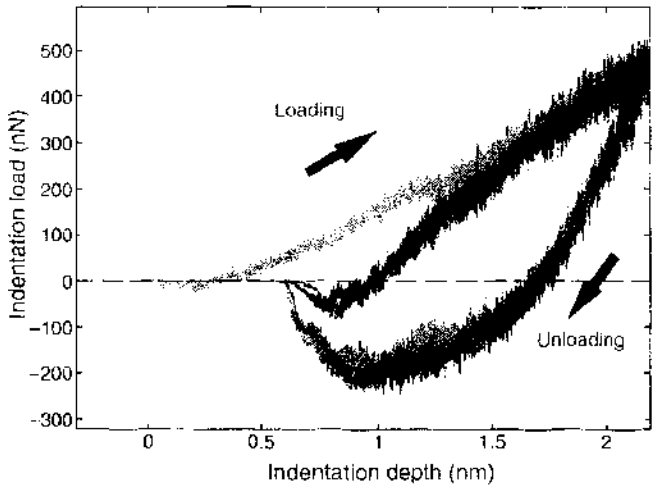


Figure 26. Load-displacement curves of repeated indentations. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000). © 2000, Kluwer Academic Publishers.

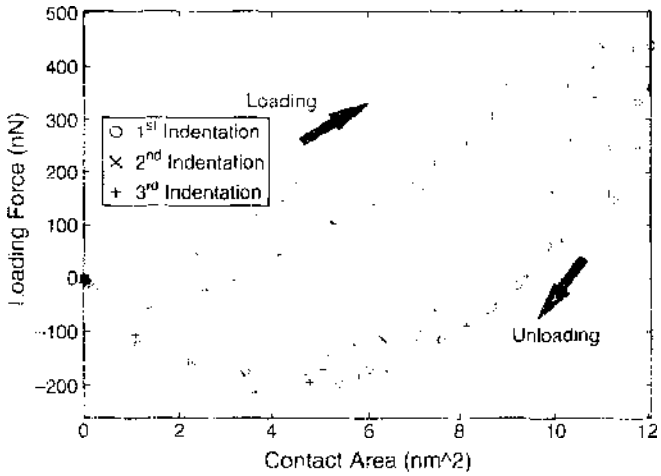


Figure 27. Indentation force versus contact area between indenter and silicon specimen. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000). © 2000, Kluwer Academic Publishers.

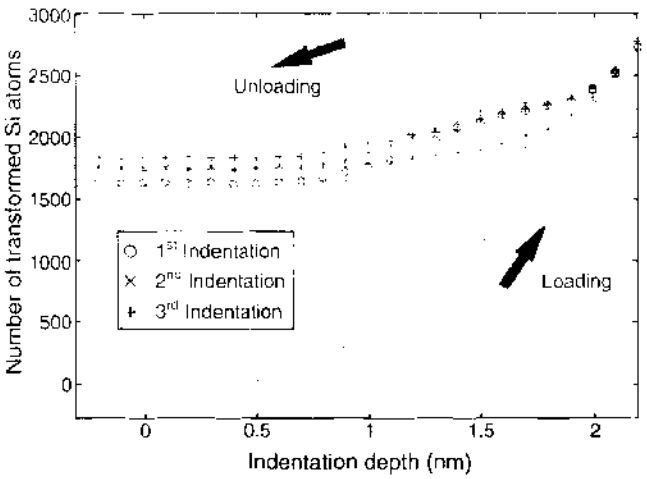


Figure 28. Number of silicon atoms that have transformed from diamond cubic to other structures when indentation depth varies. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000), © 2000, Kluwer Academic Publishers.

marked increase in the resistance to the indentation during the loading phase of the second and third indentation, demonstrated by the gradient increase of the load-displacement curve. This is partly a result of the residual stresses within the residual amorphous zone after the first indentation. Second, although all three indentations start with an initially attractive stage as the indenter approaches the workpiece, the attractive forces during the second and third indentations are considerably greater than that of the first, indicating that the amorphous silicon atoms experience greater attraction to the diamond indenter compared to diamond-cubic silicon atoms. The contact area between the workpiece and the indenter is also greater in the second and third indentations, as shown in Fig. 27.

Figure 28 shows the number of silicon atoms that have transformed from their original diamond cubic structure to other forms during the indentations. Compared to the first indentation, where the number of amorphous silicon atoms increases from 0 to 1624, the increase in the number of amorphous silicon atoms in the second and third indentations is very small, merely an increase of 8.6% (from 1622 to 1761 atoms) and 4.9% (from 1748 to 1834 atoms), respectively. In fact, Fig. 28 shows that the loading and unloading paths of the second and third indentations almost follow the unloading path of the first, indicating that once the atoms have already transformed to its amorphous state, further indentation does little to change it. As a result, the amorphous zone does not change much.

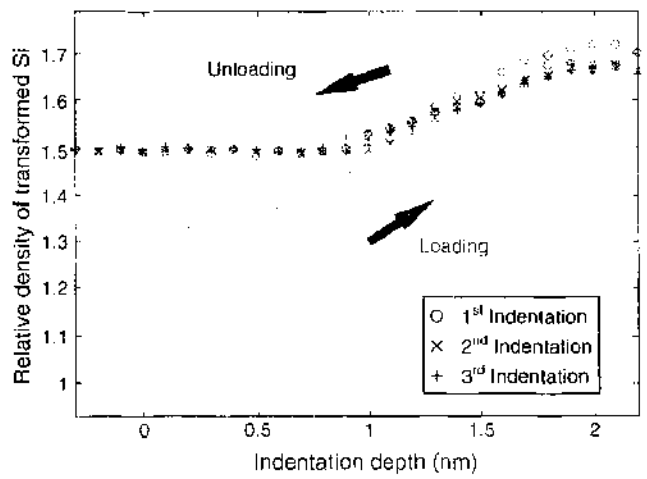


Figure 29. Relative density of transformed silicon versus indentation depth. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000), © 2000, Kluwer Academic Publishers.

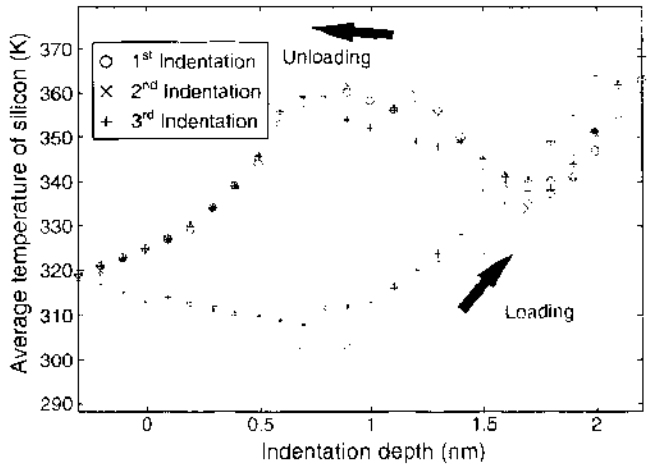


Figure 30. The variation of average temperature of silicon specimen with the indentation depth. Reprinted with permission from [42], W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000). © 2000, Kluwer Academic Publishers.

Because the increase in the volume of the transformed silicon and the increase in the number of transformed silicon atoms are proportional, the relative density of the amorphous silicon remains almost constant. Figure 29 demonstrates that the relative density of the transformed silicon during the second and third indentations varies in a manner very similar to the unloading stage of the first indentation. This again shows that the second and third indentations do not do much to change the existing residual amorphous region.

Figure 30 presents the average temperature variation of the silicon workpiece during indentations. In the second and third cycles, once the indenter comes into contact with the workpiece at the depth of 0.6 nm (the residual indentation mark caused by the first indentation), there is an increase in the temperature of the workpiece, corresponding possibly to the microstructural changes in silicon from amorphous to metallic phase. On unloading, the temperature drops initially, followed by another increase. This could be caused by other microstructural changes reversing from the metallic phase back to the amorphous form. Thus, the similarities in temperature changes described above indicate that the microstructural phase transformations that occurred during the first indentation may have also occurred during the second and third indentations.

4. NANOTRIBOLOGY

4.1. Introduction

A full understanding of friction and wear on the atomic scale is of fundamental importance to the development of nanotechnology. Immediate examples in which atomic friction and wear play a central role are the optimal design, fabrication, and operation of devices with atomic resolution, such as micromachines and high-density magnetic recording systems. Over the last 16 y or so, many studies have been carried out to explore the mechanisms of nanofriction and nanowear, both theoretically and experimentaly. For instance, an early experimental observation of atomic-scale friction using an atomic force microscope was reported by a research group at IBM [62]. The friction coefficient between a tungsten tip of radius 300 nm and a basal plane of a graphite grain was found to be 0.012 at a normal load of 10 μN. Another early interesting study [63] was about the sliding of a tungsten tip of radius 10 μm on a carbon-sputtered surface. A frictional force of about 1 μN at zero normal force was measured, indicating an infinite friction coefficient. In the meantime, theoretical studies using either molecular dynamics simulation and the first-principle calculations [64–68] were also carried out and showed that friction coefficient could vary significantly, from 0.01 to 0.07, under different sliding conditions. Belak and Stowers [69] studied the indentation and scraping of a copper monocrystal by a rigid diamond tip using the molecular dynamics simulation. While focusing mainly on some issues of indentation with an observation of

dislocation initiation, they also simulated a cutting under a specific condition and observed that the friction coefficient was about 1.0. However, the mechanisms of friction and wear and the effect of deformation, sliding conditions, and dislocation motions were not studied.

A more systematic research was conducted later by Homola [70], who proposed the concept of interfacial sliding to describe the sliding of two perfect, molecularly smooth, and undamaged mica surfaces. Mechanisms of wear were also addressed. The investigation demonstrated that the Bowden–Tabor formula [71], which states that the frictional force is proportional to the real molecular contact area, could well describe the frictional behavior during atomic sliding. In fact, the importance of the atomic contact area to atomic friction is not difficult to understand if the JKR theory [13, 32] is recalled. This theory, although considering the effect of surface energy in its analysis, has implicitly indicated that the real contact area must be of great concern to sliding loads on the atomic scale.

If looking into the details of contact sliding, we can have two primary situations. When two surfaces are in sliding without foreign particles, they are in two-body contact sliding, as shown in Fig. 31a. In this case, the interactions among surface asperities play a central role in the process of wear and friction. However, if some particles appear between the surfaces, which could be the debris from worn surfaces or foreign particles resulting from contamination, a three-body contact sliding occurs, as shown in Fig. 31b. Under such circumstances, the kinetics and properties of the particles contribute to the tribology of the surfaces. The above sliding processes are common in nanotribological systems.

4.2. Moving Control Volume

To simulate a sliding or cutting process, the distance of relative motion of an asperity to a workpiece is significant, and in grinding and polishing, the situation is similar. An apparent way of simulating the steady-state behaviour of the workpiece material is to take a large portion of the work material as the control volume for simulation. However, this is infeasible because the cost and capacity of computation limit the number of atoms in a model. The moving control volume technique [2, 4, 5, 72] is then necessary and advantageous, in which the dimension of the control volume is similar to that of a nonmoving one but with special atom removal and addition techniques. Shown in Fig. 32 is an example of using the moving control technique for simulating nanocutting, in which zone 1, the zone affected by cutting deformation, is equivalent to the normal control volume size discussed previously. Zones 2–5 are those unaffected by the cutting deformation. When the cutting proceeds, atoms in zone 5 can be removed because the removal will not affect the behavior of the atoms in zone 1. However, zone 4 becomes a transition zone because the boundary between zones 4 and 5, which is originally inside the workpiece, becomes a free surface. The dimensions of both $\delta 4$ and $\delta 5$ must be determined by an error analysis such that the removal of zone 5 and the creation of the new free surface do not influence the behavior of atoms in Zone 1. Similarly, because atoms are added in zone 2 after the removal of atoms from zone 5, zone 3 becomes a transition zone. Because of the same reason, the dimensions of both $\delta 2$ and $\delta 3$ must also be determined by an error analysis. Zone 6 consists of the boundary and thermostat atoms.

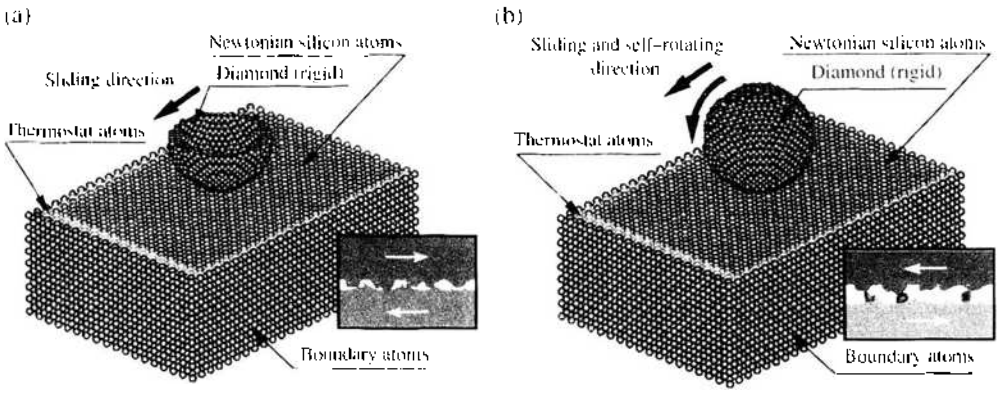


Figure 31. Molecular dynamics modeling of the sliding processes. (a) Two-body sliding; (b) three-body sliding.

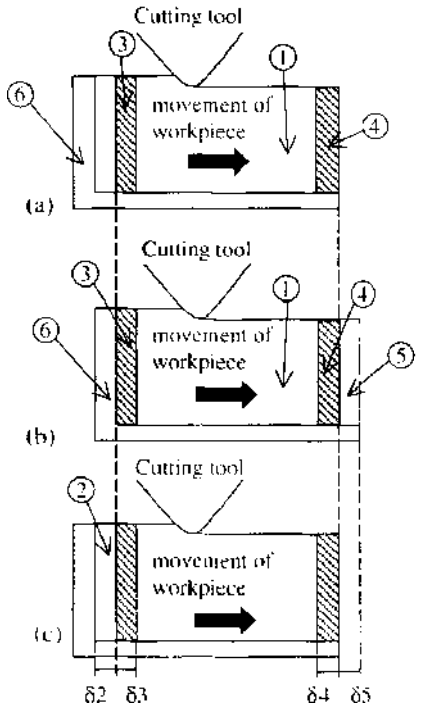


Figure 32. The concept of moving control volume of a workpiece subject to nanocutting. (a) Initial control volume; (b) control volume before the operation of removal and addition; (c) control volume after the operation of removal and addition. Reprinted with permission from [72], W. C. D. Cheong, L. Zhang, and H. Tanaka, *Key Eng. Mater.* 196, 31 (2001). © 2001, Trans Tech Publications.

4.3. Diamond–Copper Sliding Systems

4.3.1. Methods of Modeling and Analysis

For simplicity, let us consider an atomically smooth diamond asperity sliding on an atomically smooth surface of a copper monocrystal in its (111) plane. The variables of interest are the sliding speed V, indentation depth d, degree of surface lubrication or contamination, and tip radius of asperity R, that is, the radius of the envelope of centers of the surface atoms. The environmental temperature of the sliding system is 293 K, and the asperity rake angle is -60 degrees. In addition, we assume that d keeps constant in a sliding process, which implies that the sliding system has an infinite loop stiffness.

The interactions between copper atoms can be described by the modified Morse potential given in Eq. (9), where r_{ij} is the interatomic separation between atoms i and j, and r_0 is the equilibrium separation at which the potential minimizes. D and α are the material constants listed in Table 4. The physical meaning of D is the cohesive energy between the two atoms. Teems λ_1 and λ_2 are nondimensional parameters indicating the cohesive strength change between the atom pair i–j. For copper–copper atom pairs, $\lambda_1 = \lambda_2 = 1$, and hence Eq. (9) becomes identical to the standard Morse potential. For copper–diamond atom pairs, however, λ_1 takes a value in interval (0, 1], whereas $\lambda_2 \geq 1$. This is because an application of lubrication or contamination on the diamond-copper interface weakens the cohesive strength between copper and diamond atoms. Hence, the effect of surface lubrication or contamination on the friction and wear can be qualitatively investigated by

Table 4. Parameters in the standard Morse potential.

Parameter	Cu–Cu	Cu–C
D (eV)	0.342	0.087
α(nm^{-1})	13.59	51.40
r_0	0.287	0.205

varying λ_1 in (0, 1] and ignoring the detailed effects of lubricant or contaminant atoms. However, when λ_1 varies in (0, 1], the potential curve between a pair of copper-diamond atoms will distort toward the negative direction of r. Bearing in mind that the atomic distance between any copper and diamond atoms increases, rather than decreases, when a lubricant or contaminant atom appears to separate them from direct contact, the potential curve should distort toward the positive direction of r. Hence, λ_2 in Eq. (9) must be larger than 1. In this analysis, we determine λ_2, when λ_1 changes, by keeping the spring constant of a copper–diamond atomic pair unchanged. This means that when λ_1 is given, λ_2 is determined by $(\lambda_1 D)(\lambda_2 \alpha)^2 = D\alpha^2$; hence $\lambda_2 = \lambda_1^{-1/2}$.

With the above potential function available, the forces on atom i resulting from the interaction of all the other atoms can be calculated by

$$F_i = -\sum_{i=1, j\neq i}^{N} \Delta_i \phi(r_{ij}) \tag{12}$$

where N is the total number of atoms in the model, including thermostat, boundary, and diamond atoms. (Here, $12{,}000 \leq N \leq 15{,}000$ is used in conjunction with the technique of moving control volume.) As a consequence, the motion of all the Newtonian atoms in the control volume, including their instant position and velocity vectors, can be obtained by following the standard procedures of molecular dynamics analysis, described previously.

In principle, an asperity is three dimensional, and thus a three-dimensional molecular dynamics analysis would be more appropriate. However, we found, based on a careful comparison [73], that a two-dimensional model can lead to accurate enough results in terms of the variations of temperature and sliding forces and easier characterization of deformation. We will therefore focus on the two-dimensional, plane-strain analysis in this section.

When an instant configuration of the copper atomic lattice during sliding is obtained by the molecular dynamics analysis, the distribution of dislocations in the deformed lattice can be determined by the standard dislocation analysis [45]. Figure 33 shows an example of identifying edge dislocations when two extra half planes of atoms (dark atoms) appear in the deformed atomic lattice. The Burgers vector can be obtained by application of the Burgers circuit. Because we are now carrying out a two-dimensional molecular dynamics simulation, we can only study edge dislocations [73]. For convenience, however, we call them dislocations from here on.

4.3.2. Mechanisms of Wear

The deformation of the copper specimen has four distinct regimes under sliding. They are the no-wear regime, adhering regime, ploughing regime, and cutting regime, as shown in Fig. 34. In the figure, the transition of deformation regimes is characterized by the nondimensional indentation depth δ, which is defined as d/R and can be viewed, when contact sliding takes place, as a measure of the strain imposed by the diamond asperity.

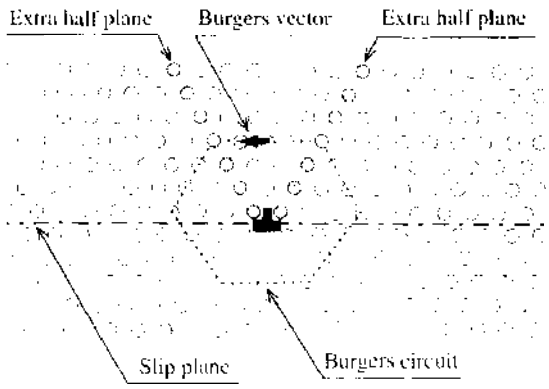


Figure 33. Determination of edge dislocations. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997). © 1997, Elsevier Ltd.

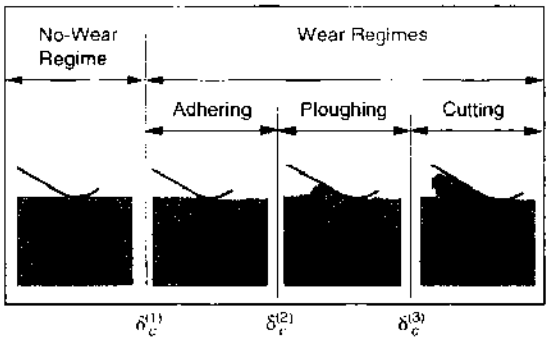


Figure 34. Transition diagram of deformation regimes. (The diamond slides from the right to the left.) Reprinted with permission from [2]. L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997). © 1997, Elsevier Ltd.

In the no-wear regime, the atomic lattice of copper is deformed purely elastically. After the diamond asperity slides over, the deformed lattice recovers completely. In this case, sliding does not introduce any wear or initiate any dislocation.

When δ increases and reaches its first critical value, $\delta_c^{(1)}$, adhering occurs. The atomic bonds of some surface copper atoms are broken by the diamond sliding. These copper atoms then adhere to the asperity surface and move together with it. However, they may form new bonds with other surface atoms of copper and return to the atomic lattice. The above process repeats again and again during sliding, causes a structure change of the copper lattice near the surface, and creates surface roughness of the order of one to three atomic dimensions. In the meantime, some dislocations are also activated in the subsurface see (Fig. 35a).

If δ increases further, to its second critical value, $\delta_c^{(2)}$, the above adhering deformation will be replaced by ploughing (Figs. 34 and 35b). An apparent feature of deformation in the stage is that a triangular atom-cluster always exists in front of the leading edge of the diamond asperity and appears as a triangular wave being pushed forward. In this regime, the deformation zone in the subsurface becomes very large, and a great number of dislocations are activated. Moreover, the motion of dislocations and their interactions in the subsurface become extremely complex. Grain boundaries can also be generated by ploughing, as shown, for instance, by the continuous orange curve in Fig. 35b.

When δ reaches its third critical value, $\delta_c^{(3)}$, a new deformation state, cutting, appears, characterized by chip formation. Compared with the ploughing regime, the dimension of the deformation zone during cutting is smaller. Dislocations distribute much more closely to the sliding interface (Fig. 35c).

The above figure of deformation regimes and their transition represents the most general case. Under some specific sliding conditions, not all the regimes would appear except the no-wear regime.

For example, if the tip radius of the diamond asperity keeps unchanged but the sliding speed changes, then at lower sliding speeds all the four regimes described above appear. At higher speeds, however, ploughing regime vanishes (see Fig. 36a). In contrast at a given sliding speed, if the tip radius of the asperity is very small, say 1 nm, only no-wear and cutting regimes emerge, as shown in Fig. 36b. However, with relatively larger tip radii, adhering appears as a transition from no-wear to cutting.

Another important factor that alters the deformation transition is the effect of surface lubrication or contamination. If the sliding interface is chemically clean, $\lambda_1 = \lambda_2 = 1$ in Eq. (9). In this case, as shown in Fig. 36c, ploughing does not happen at a given sliding speed and tip radius. If the surface is lubricated, $\lambda_1 \leq 1$ with $\lambda_2 \geq 1$, and all the four regimes occur.

It is obvious from Fig. 36 that the no-wear regime exists in a wide range of indentation depths. In addition, a smaller radius, a lower sliding speed, or a better surface lubrication (i.e., smaller λ_1) enlarges the no-wear regime. This highly indicates that a no-wear design of sliding systems may be possible in practice. Moreover, it is important to note that the size of the no-wear regime is a strong function of sliding speed and surface lubrication. Therefore, sliding speed and lubrication should be taken into specific account in an attempt to design no-wear sliding systems.

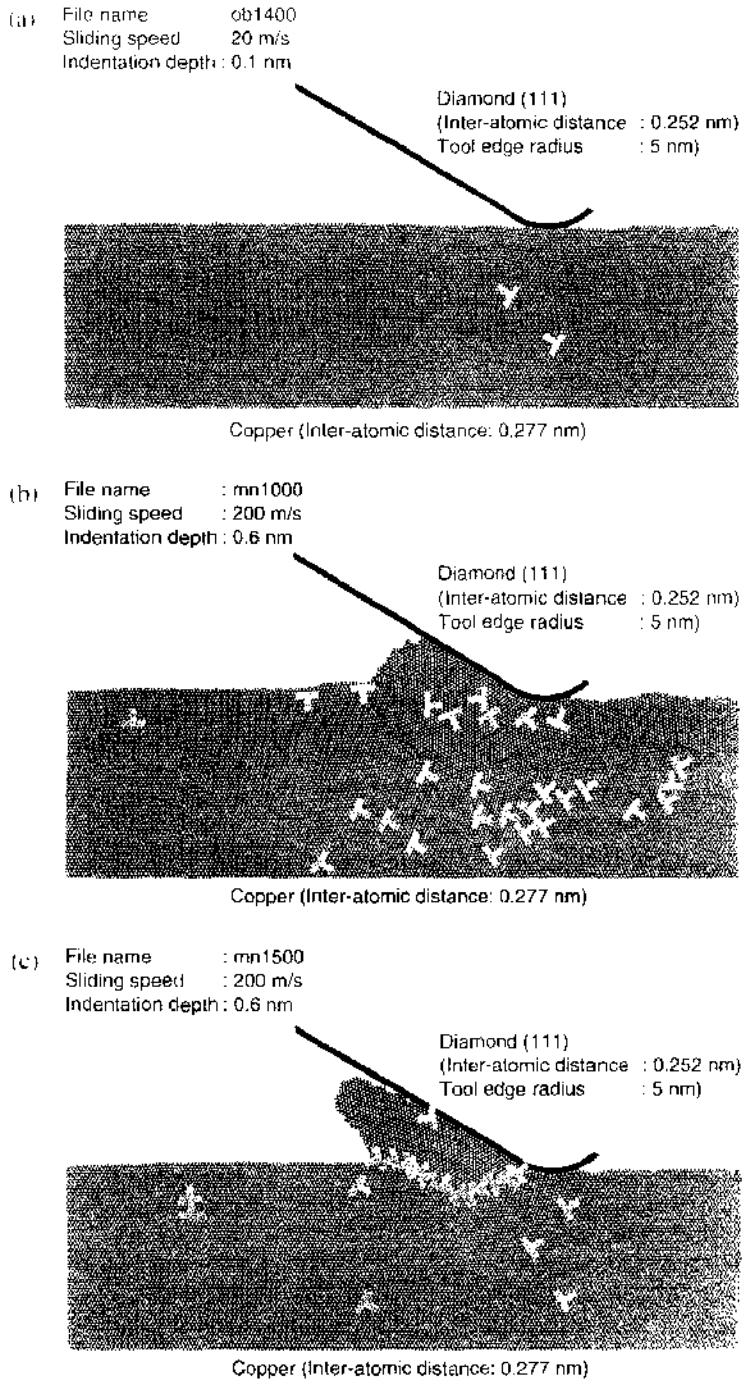


Figure 35. Distribution of dislocations in the subsurface of copper specimen. (The diamond slides from the right to the left.) (a) Adhering regime; (b) ploughing regime; (c) cutting regime. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997). © 1997, Elsevier Ltd.

The formation of various deformation regimes and their transition can be elucidated by the variation of temperature distribution and dislocation motion in the atomic lattice, (see Figs. 35 and 37). For instance, a larger indentation depth or a higher sliding speed indicates a higher input sliding energy, greater temperature rise, and severe plastic deformation. This in turn means a higher density of dislocations with more complicated interactions in the deformed atomic lattice. When the nondimensional indentation depth δ is small, a smaller number of atoms has high temperatures, as shown in Figs. 37a and 37b. Sliding does not cause any considerable temperature rise in the vicinity of the contact zone. With the increase of δ, the number of high-temperature atoms increases quickly (Figs. 37c and 37d). However, compared with the cutting regime (Fig. 37d), the high-temperature atoms in the ploughing

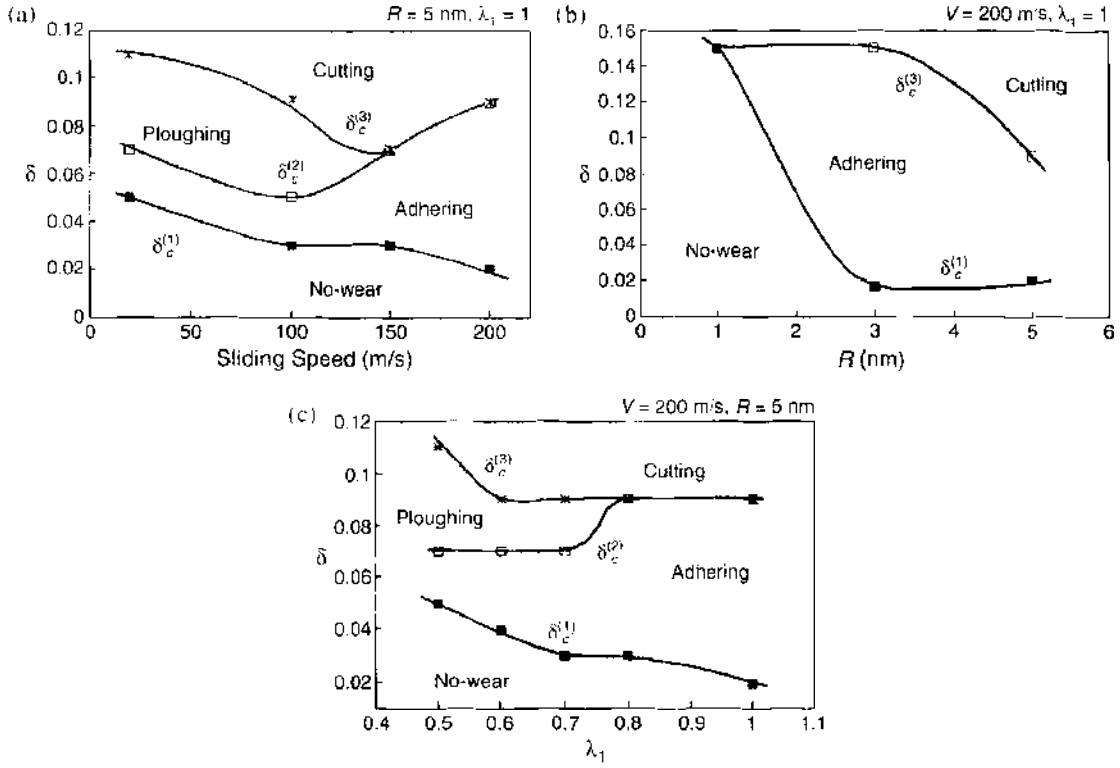


Figure 36. Regime transition under specific sliding conditions. (a) Nondimensional indentation depth versus sliding speed; (b) nondimensional indentation depth versus tip radius; (c) nondimensional indentation depth versus lubrication/contamination. Reprinted with permission from [2]. L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997). © 1997, Elsevier Ltd.

regime (Fig. 37c) distribute in a much wider area. This explains why a ploughing regime has a greater deformation zone and less localized distribution of dislocations (Fig. 35).

4.3.3. Friction

With the above deformation mechanisms in mind, now let us examine the frictional behavior of the system. Figure 38 shows the variation of the conventional friction coefficient, $\mu = |F_x/F_y|$, with the change of δ, where F_x and F_y are, respectively, the frictional force and normal indentation force during sliding. It is clear that in the cutting regime, F_x is proportional to F_y. In other regimes, however, the behavior of F_x is complex. Particularly, μ becomes singular at a specific δ in the no-wear regime.

The singularity of μ is understandable if we examine the sliding forces when δ changes. On the atomic scale, as shown in Fig. 39, the normal sliding force F_y always varies from attractive to repulsive. Thus, at the transition point ($F_y = 0$), μ is infinite. This also explains why μ varies sharply under different sliding conditions, as reported by Refs. [62–68]. All these conditions clearly indicate that the concept of the conventional friction coefficient is no longer meaningful in no-wear, adhering, and ploughing regimes.

In noncontact sliding, the only appropriate way to calculate the frictional force is to use Eq. (12). In contact sliding, however, empirical expressions in terms of the contact area can be developed. For example, the following simple formula can be obtained according to the present theoretical analysis

$$F_x = \begin{cases} -\sum_{n=1}^{N_a}\sum_{m=1}^{N_c} \dfrac{\partial}{\partial x_n}\phi(r_{mn}), & \text{for } L_c = 0 \\ \zeta_1^{(I)} L_c w_a + \zeta_2^{(I)}, & \text{for } L_c^{(1)} < L_c \le L_c^{(2)} \\ \zeta_1^{(II)} L_c w_a + \zeta_2^{(II)}, & \text{for } L_c > L_c^{(2)} \end{cases} \tag{13}$$

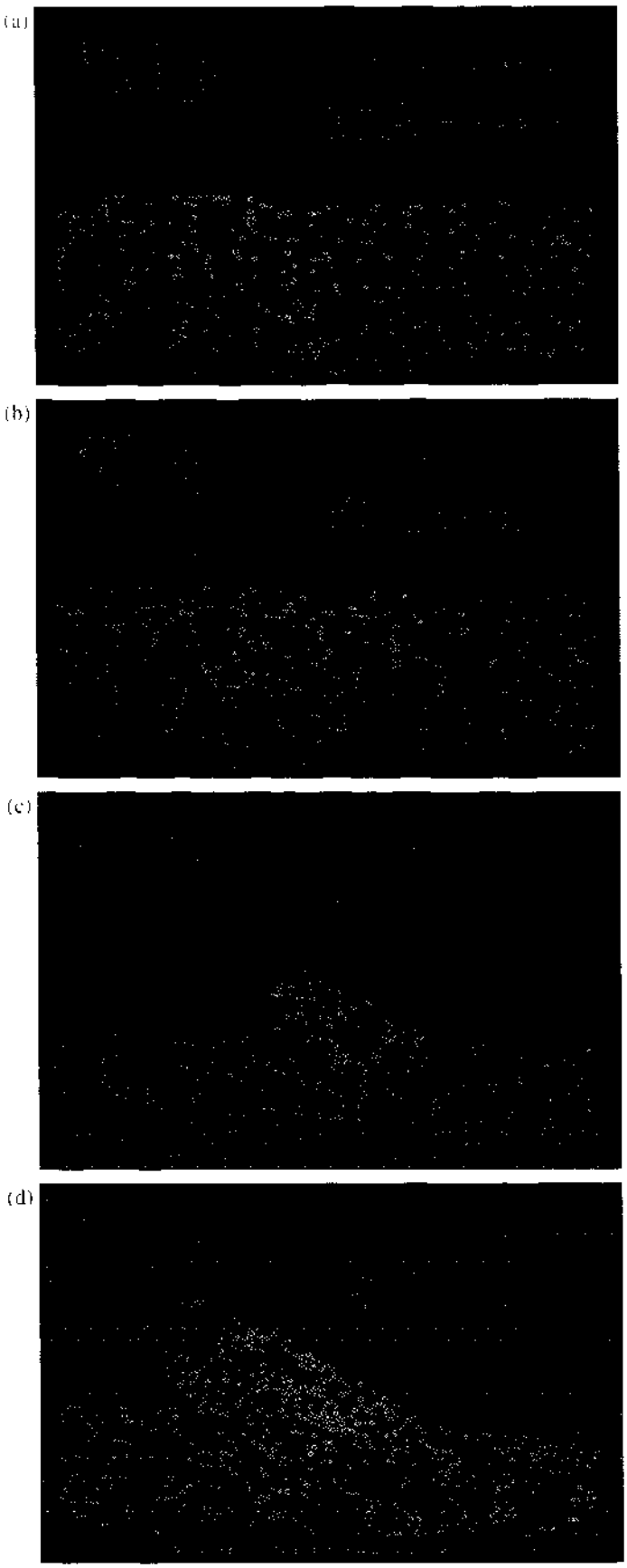


Figure 37. Temperature distribution in the atomic lattice of copper. (The diamond slides from the right to the left.) (a) No-wear regime; (b) adhering regime; (c) ploughing regime; (d) cutting regime. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997), © 1997, Elsevier Ltd.

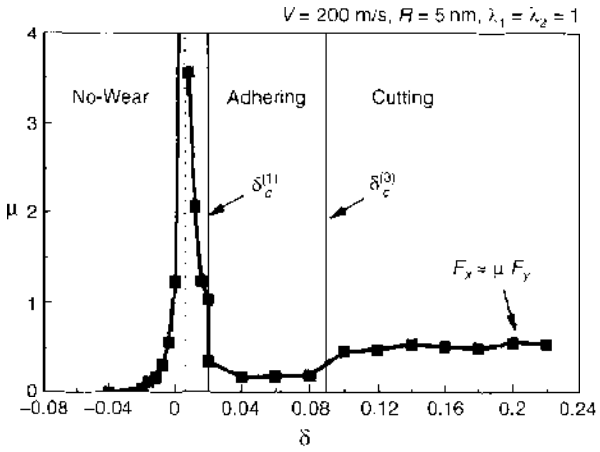


Figure 38. Variation of friction coefficient. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997), © 1997, Elsevier Ltd.

where $\zeta_1^{II} = 409$ MPa, $\zeta_2^{II} = 1.807 \times 10^{-8}$ nN, $\zeta_1^{III} = 4.20$ GPa, $\zeta_2^{III} = -1.899$ nN are constants, N_c is the total number of copper atoms in the model, N_d is the number of diamond atoms, L_c is the atomic contact length, and $w_a = 0.226$ nm is the width of an atomic layer of copper in the direction perpendicular to its (111). Equation (13a) can be derived directly from Eq. (12) by considering that F_x is the resultant force of the atomic forces on all the diamond atoms in x direction. Equations (13b) and (13c) are empirical expressions made by fitting the molecular dynamics simulation data in Fig. 40. The physical meaning of product $L_c w_a$ in Eq. (13) is the atomic contact area in the present sliding system. As shown by Eq. (13) and Fig. 40, we can see that the frictional behavior of an atomic sliding system cannot be described by a single formula. There exist two distinct contact sliding zones, zone II ($L_c^{(1)} \leq L_c < L_c^{(2)}$) and zone III ($L_c \geq L_c^{(2)}$), where $L_c^{(2)} = 2.216$ nm is the transition boundary from Zones II to III, and $L_c^{(1)} = 0.277$ nm is the minimum contact length, defined as the distance between two copper atoms in its (111) plane. The transition from noncontact to contact sliding is a sudden change because L_c does not exist below $L_c^{(1)}$. In Fig. 40, zone II reflects the frictional behavior of the system in the no-wear contact sliding, and zone III shows the behavior in the adhering and ploughing regimes. Thus, $L_c^{(2)}$ can be interpreted physically as critical contact length at which wear takes place. It is easy to obtain according to Fig. 40, so that the shearing stress of the present sliding system in the adhering regime, $\tau = F_x/(L_c w_a)$, is in the range of 0.7 to 1.5 GPa. With so few dislocations activated in the atomic lattice, it is reasonable that τ is close to the theoretical shearing stress of a perfect copper monocrystal, which is $G/2\pi (\approx 7.32 \text{ GPa}) \geq \tau_{\text{theoretical}} \geq G/30 (\approx 1.53 \text{ GPa})$ [45], where $G \approx 46$ GPa is the shear modulus of copper.

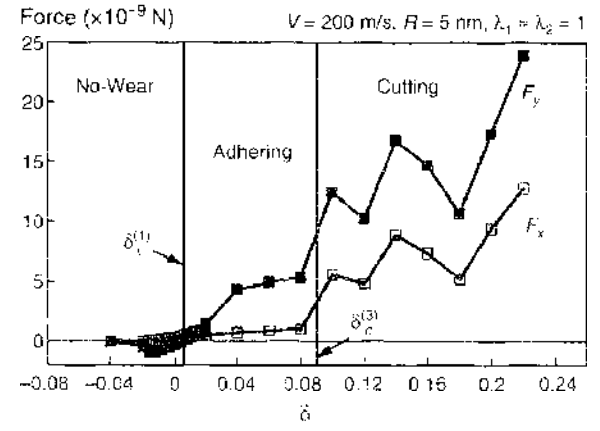


Figure 39. Variation of sliding forces. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997), © 1997, Elsevier Ltd.

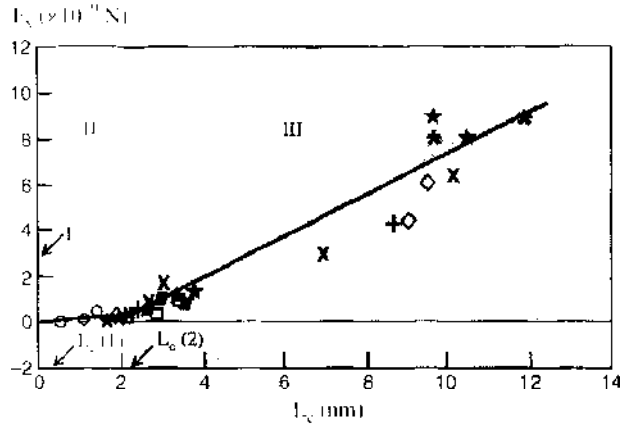


Figure 40. Relationship between the frictional force and contact length: ×: R = 5 nm, V = 20 m/s, λ_1 = 1; *: R = 5 nm, V = 100 m/s, λ_1 = 1; ★: R = 5 nm, V = 200 m/s, λ_1 = 1; ○: R = 1 nm, V = 200 m/s, λ_1 = 1; ◇: R = 5 nm, V = 200 m/s, λ_1 = 0.5; ■: R = 5 nm, V = 200 m/s, λ_1 = 0.6; +: R = 5 nm, V = 200 m/s, λ_1 = 0.7; □: R = 5 nm, V = 200 m/s, λ_1 = 0.8. Reprinted with permission from [2], L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997). © 1997, Elsevier Ltd.

The linear equations, Eqs. (13b) and (13c) are only the rough fittings to the theoretical results. The data scattering in Fig. 40 indicates that other variables, such as the tip radius of asperity R and sliding speed V, also contribute greatly to friction. In other words, F_x should be a function of not only the contact area $L_c w_a$ but also of V, R, and so on.

It is worth to note, from Fig. 39, that the frictional force F_x does not vanish under a stable condition even if the system is in noncontact sliding. It is understandable because the interactions between the diamond and copper atoms always exist. In contact sliding with a small indentation depth, F_x remains small until a ploughing or a cutting is achieved. Hence, the frictionless sliding mentioned by Belak and Stowers [69] is doubtable. Indeed, in their simulation the indentation motion stopped after 2000 time steps but the "frictionless motion" ended at time step 1000. In that period of time, sliding was still at the very initial transient state. Thus, the variation of F_x with the time step of molecular dynamics simulation does not make sense in terms of frictional or frictionless sliding. In addition, the normal force variation they reported is inconsistent with the indentation motion.

In summary, the molecular dynamics analysis has acquired the following important understandings of nanowear and nanofriction on diamond-copper sliding systems. First, there exist four regimes of deformation in general in an atomic sliding system. They are the no-wear regime, which is defect free; the adhering regime, in which surface atom exchange occurs; the ploughing regime, which is characterized by a moving triangular atom-cluster; and the cutting regime, in which material removal takes place.

Second, in the cutting regime, the frictional force follows the simple proportional rule of $F_x = \mu F_y$. In all the other regimes, the formulae in Eq. (13) apply.

Third, the transitions between different deformation regimes are governed by indentation depth, sliding speed, asperity geometry, and surface lubrication conditions. A better lubrication, a smaller tip radius, or a smaller sliding speed can bring about a greater no-wear regime, and no-wear designs may be achieved in practice.

4.4. Scale Effect of Contact Size on Friction Transition

4.4.1. Introduction

The investigation by Zhang and Tanaka [2], described earlier, focused on the friction and wear mechanisms when the radius of the asperity R is a constant while the depth of asperity indentation δ increases. Wear and plastic deformation consequently occur when δ reaches a critical value.

Recently, on the basis of certain experimental observations [74, 75], Hurtado and Kim [76] proposed a micromechanical dislocation model of frictional slip, predicting that when the contact size is small, the friction stress is constant and on the order of the theoretical shear

strength. However, at a critical contact size, there is a transition beyond which the frictional stress decreases with increasing contact size, until it reaches a second transition, where the friction stress gradually becomes independent of the contact size. Hence, the mechanisms of slip are size dependent, or in other words, there exists a scale effect. Before the first transition, the constant friction is associated with a concurrent slip of the atoms without the aid of dislocation motion. The first transition corresponds to the minimum contact size at which a single dislocation loop is nucleated and sweeps through the whole contact interface, resulting in a single-dislocation-assisted slip. This mechanism is predicted to prevail for a wide range of contact sizes, from 10 nm to 10 μm, in radius for typical dry adhesive contacts; however, there are no available experimental data in this size range. The second transition occurs for contact sizes larger than 10 μm, beyond which friction stress is once again constant because of the cooperative glide of dislocations within dislocation pileups. The above dislocation model excludes wear or plastic deformation of the sliding parts.

To clarify this issue, Zhang, Johnson, and Cheong [12] carried out a nanotribology analysis using molecular dynamics by varying the asperity radius from 5 to 30 nm and keeping the indentation depth unchanged. The model consists of a single cylindrical asperity (rigid diamond) of various radii, sliding across a copper (111) plane with a speed of 5 m/s. The indentation depth, d, was 0.46 and -0.14 nm (0.14 nm above the workpiece), respectively, where d is the distance between the surfaces of the asperity and the specimen, defined by the envelopes at the theoretical radii of their surface atoms. As usual, two layers of thermostat atoms are arranged around the Newtonian copper atoms of the specimen to ensure that the heat generated during sliding can conduct out of the control volume properly. The velocities of the atoms in the initial configuration of the model follow the Maxwell distribution. The modified Morse potential, Eq. (9), was applied to describe the interactions between the atoms. It must be noted that the molecular dynamics simulation cannot capture the second transition because it will require too long a computation time to analyze a model on the order of micrometers.

4.4.2. Friction Transition

Figure 41 shows snapshots of the simulation with different asperity sizes. It is clear that the depths of indentation in the simulations are small enough that there are no dislocations created within the copper, corresponding to the no-wear regime described by Zhang and Tanaka [2].

In the case in which the radius of the diamond asperity is less than 12 nm, the carbon atoms slide across the copper atoms in close contact. The surface of the copper workpiece conforms closely to the shape of the asperity tip in contact with the copper (Fig. 42a). There is also strong indication of atomic stick-slip between the atoms of the asperity and the workpiece (Fig. 42b). This implies that the sliding mechanism involved is similar to the ideal slip of two atomic planes in a perfect dislocation-free crystal. Hurtado and Kim [76] referred to this sliding mechanism as concurrent slip. In addition, the friction stress averages around a constant value of 5 GPa, regardless of the contact width (Fig. 43).

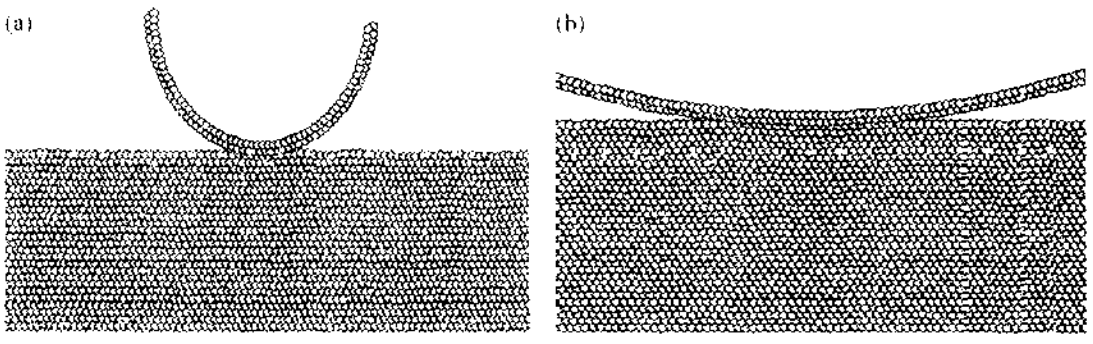


Figure 41. Snapshots of sliding simulation with diamond asperities of different tip radius. (a) Radius = 5 nm; (b) radius = 30 nm. Reprinted with permission from [12]. J. C. Zhang et al., *Tribol. Lett.* 10, 23 (2001). © 2001, Kluwer Academic Publishers.

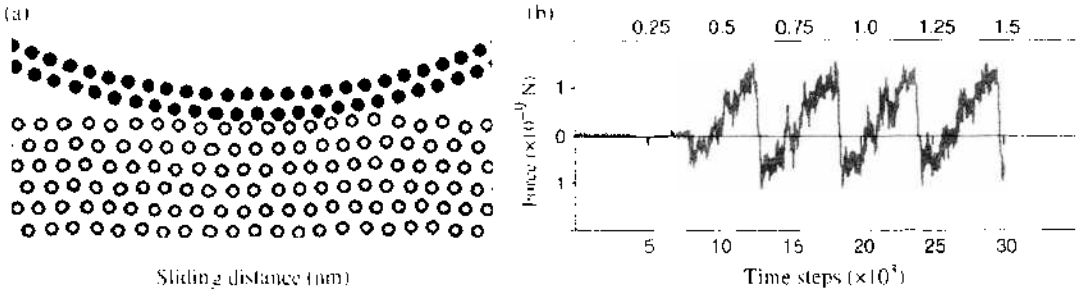


Figure 42. Diamond asperity of radius 8 nm sliding on a monocrystalline copper surface. (a) Surface of copper workpiece conforms closely to shape of asperity with good contact; (b) stick-slip phenomenon in sliding. Reprinted with permission from [12], L. C. Zhang et al., *Tribol. Lett.* 10, 23 (2001). © 2001, Kluwer Academic Publishers.

When the asperity radius exceeds 12 nm, there are considerable differences in the sliding mechanism involved. The surface of the copper workpiece does not conform closely to the shape of the carbon asperity (Fig. 44a), and there is little atomic stick-slip between the atoms of the asperity and the workpiece (Fig. 44b). In addition to that, the frictional stress now decreases with an increase in the contact width (Fig. 43). Hence, the friction stress is constant before the first transition, after which it decreases with the increasing contact width (by increasing the asperity radius). This clearly indicates a change in the mechanism of sliding.

When the depth of indentation is increased to 0.46 nm, a somewhat similar relationship between friction stress and contact width is obtained. Figure 43 compares the variation of the friction stress and the critical contact width at the first transition when the indentation depth changes. It is clear that the indentation depth influences both the critical contact size and the rate of friction reduction after the transition. At this greater indentation depth, however, Figure 45 shows that permanent damage and wear are occurring. Dislocation lines indicating plastic deformation within the body of the solid are visible. This behavior is similar to the adhering regime described by Zhang and Tanaka [2].

4.4.3. Contact Width

The contact width between the asperity and workpiece obtained by the above MD simulation can be compared with the predictions of the JKR theory [13, 32], which shows, for the present configuration of a circular cylinder in contact with a half space (plane-strain), that the indentation load per unit width on the asperity, P, and the contact width, $2a$, follows the relationship of

$$P = \frac{\pi E^* a^2}{4R} - \sqrt{2\pi E^* a w} \tag{14}$$

where R is the radius of the asperity, E^* is the effective modulus of the contact system [13], and w is the work of adhesion and can be determined by a nanoindentation simulation using molecular dynamics analysis. It is found that for the present diamond–copper (C–Cu)

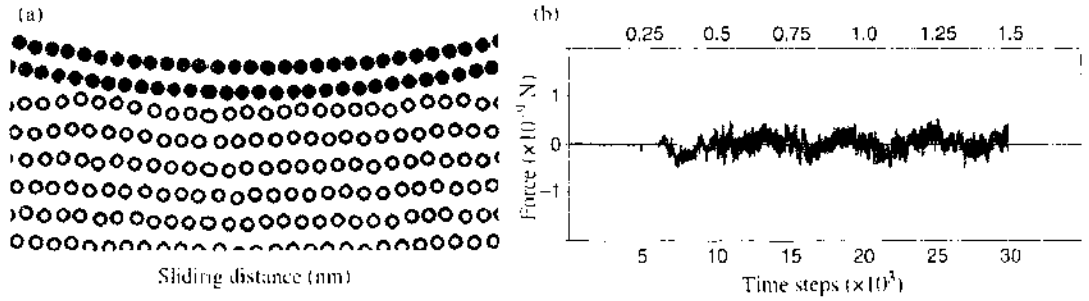


Figure 43. Frictional stress versus contact width for indentation depths of −0.14 nm and 0.46 nm. Reprinted with permission from [12], L. C. Zhang et al., *Tribol. Lett.* 10, 23 (2001). © 2001, Kluwer Academic Publishers.

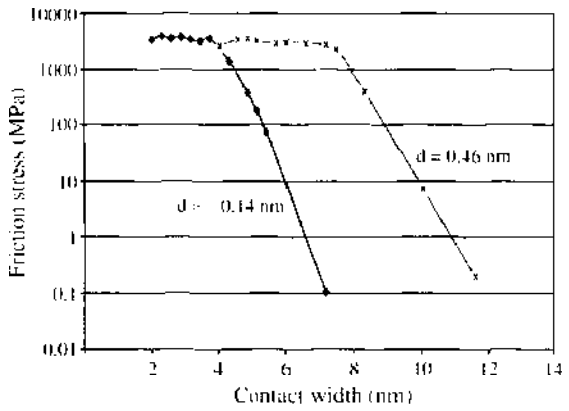


Figure 44. Diamond asperity of radius 30 nm sliding on a monocrystalline copper surface. (a) Surface of copper workpiece does not conform closely to the shape of the asperity; (b) no stick-slip phenomenon in sliding. Reprinted with permission from [12]. L. C. Zhang et al., *Tribol. Lett.* 10, 23 (2001). © 2001, Kluwer Academic Publishers.

system, $w_{C\text{-}Cu} = 1.476$ J/m^2. Because the diamond asperity is assumed to be a rigid body, the E' in Eq. (12) becomes 125.36 GPa by taking $E_C = \infty$, $E_{Cu} = 110$ GPa, and $\nu_{Cu} = 0.35$ [77].

Table 5 compares the contact widths from the molecular dynamics simulation, the JKR theory of Eq. (14), and the Hertzian contact theory under various conditions. The values from the JKR and simulation are different, although the deformation of the copper workpiece at $d = -0.14$ nm was purely elastic, and that at $d = 0.46$ nm was almost purely elastic. A possible cause is that the contact width of the molecular dynamics simulation contains the effect of sliding, whereas Eq. (14) does not. It is also worth noting that the predictions of the JKR theory compared to the predictions of the Hertzian contact theory is much closer to the molecular dynamics results. This indicates that the effect of normal adhesion is considerable.

In calculating the contact width above using Eq. (14), the force P used is from the corresponding molecular dynamics simulation, as listed in the table. In their paper [12], Zhang, Johnson, and Cheong also discussed the sliding on copper by a copper asperity.

4.4.4. Mechanism—An Open Question

Apparently, the above result of molecular dynamics simulation is in agreement with the phenomenon predicted by the dislocation model [76], but the mechanisms are different. A quantitative comparison between the predictions of the two modeling methods is worthwhile, although the specimen materials are different and the dislocation model is three-dimensional, whereas this molecular dynamics simulation is two-dimensional. Table 6 shows the results of the dislocation model and those of the molecular dynamics simulation.

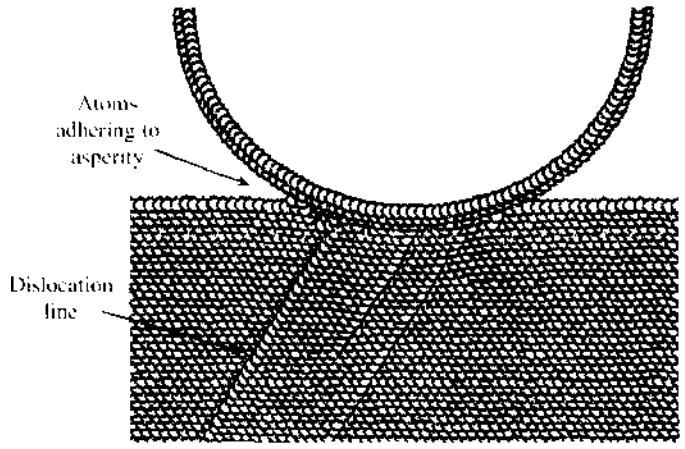


Figure 45. Sliding simulation with indentation depth 0.46 nm and asperity radius 5 nm. Reprinted with permission from [12]. L. C. Zhang et al., *Tribol. Lett.* 10, 23 (2001). © 2001, Kluwer Academic Publishers.

Table 5. Contact lengths by the JKR and Molecular Dynamics analyses for the case of diamond–copper interactions.

		Contact Length 2a (nm)	
		$d = -0.14$ nm, $P = 0.625$ N/m	$d = 0.46$ nm, $P = 22.969$ N/m
$R = 5$ nm	JKR	2.914	3.764
	Molecular Dynamics	2.870	4.120
	Hertz	0	2.160
$R = 8$ nm			
		$d = -0.14$ nm, $P = 0.824$ N/m	$d = 0.46$ nm, $P = 27.34$ N/m
	JKR	3.99	5.152
	Molecular Dynamics	3.731	5.740
	Hertz	0	2.980

The value of τ/μ obtained from the molecular dynamics simulation is close, particularly in the case of $d = 0.46$ nm, to that of the dislocation model, where τ is the shear stress and μ is the bulk shear modulus defined by Hurtado and Kim [76] as $\mu = 2G_1G_2/(G_1 + G_2)$, in which G_1 is the shear modulus of the asperity and G_2 is that of the copper specimen. In this molecular dynamics simulation, the asperity is rigid so that $G_1 = \infty$ and $\mu = 2G_2 = 81.48$ GPa. The ratio a/b at the transition varies considerably with the indentation depth, where a is half of the critical contact width and b is the Burgers vector. The ratio for the case of $d = 0.46$ nm is closer to that of the dislocation model. However, it is interesting to note that in the transition zone the rates of friction reduction with the contact size, $\xi = d\tau/da$, are very different. The molecular dynamics analysis gives a much greater rate. A better understanding is needed to be gained, using a three-dimensional model.

The above analysis concludes that there does exist a critical contact size on the nanometer scale, below which the stick-slip and sliding occurs via concurrent slip of all the atoms in the contact at a friction of the theoretical shear strength of the solid. When the contact size is beyound the critical value, a friction transition takes place to a much lower value. The critical contact size varies with the degree of penetration of the asperity. In addition, the study further confirms that the Hertz theory fails to predict the contact size in vacuum at the nanometer scale when surface energy plays a key role.

4.5. Diamond–Silicon Sliding Systems

4.5.1. Modeling

Let us now consider the two-body and three-body contact sliding problems defined in Fig. 31. In the former, asperities are fixed on the sliding surfaces. To understand the fundamental deformation mechanism in a component induced by the penetration of asperities, the molecular dynamics model illustrated in Fig. 31a is developed [4], where the shape of a hard asperity, which should be irregular in reality, has been simplified to a hemispherical diamond tip of radius R moving with a constant speed V_c. Because a diamond can be considered to be a rigid body compared with silicon, the model enables us to concentrate on the understanding of the deformation of silicon. In a three-body contact sliding, the model shown in Fig. 31b [4] can be used, where the motion of a foreign particle between the two surfaces possesses both a translation and a self-rotation. To facilitate understanding, a single particle is considered for the time being and is approximated by a diamond ball of

Table 6. Comparisons of dimensionless shear stress (τ/μ) and contact size (a/b) at transition.

	Two-Dimensional Molecular Dynamics Model		Three-Dimensional Theoretical Model
	$d = -0.14$ nm	$d = 0.46$ nm	
τ/μ	0.061	0.037	0.023
a/b	6.53	13.06	30
ξ (Pa/m)	2.90×10^{18}	1.46×10^{18}	1.88×10^{13}

radius R, moving horizontally (translation) with a speed V_t, and in the meantime rotating about its center independently with a peripheral speed V_r. When $V_r = 0$, the three-body contact sliding reduces to a two-body sliding. When $V_t = 0$ or $V_r = V_t$, however, it becomes a pure rolling process. To avoid the boundary effect, the dimension of the moving control volume of the silicon specimen was taken as 9.23 × 13.57 × 4.34 nm, containing 28,773 silicon atoms.

4.5.2. Inelastic Deformation

The molecular dynamics simulation showed that there always exists a thin layer of amorphous silicon in a specimen subsurface subjected to a two-body contact sliding, as shown in Fig. 46a. This is in agreement with the experimental findings by Zhang and Zarudi [78] (Fig. 46b). The thickness of the layer decreases with the of decreasing the penetration depth of asperity (Fig. 46c), δ. If δ is large (e.g., 76.3 nm), dislocations can be developed in the crystal silicon below the amorphous layer. When δ becomes smaller, dislocations cannot be activated but the amorphous layer still appears. This means that on the nanometer scale, an inelastic deformation via amorphous phase transformation is a more energetically favorable mechanism. In the case of three-body contact sliding, the mechanism of inelastic deformation is the same: that is, via amorphous phase transformation. However, because of the kinetic difference in the two-body and three-body sliding motions, the extent of subsurface damage is different. In general, a two-body contact sliding introduces a thinner layer of amorphous silicon. A three-body contact sliding, however, may leave a perfect crystal

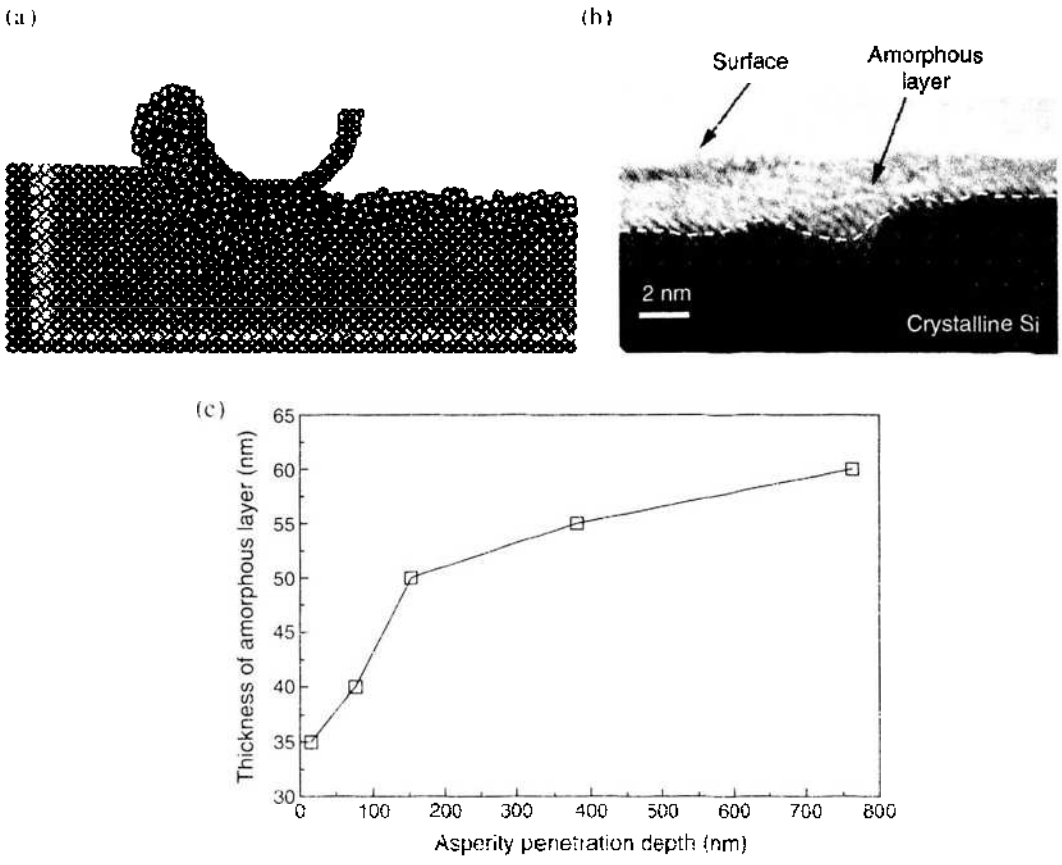


Figure 46. The subsurface microstructure of silicon monocrystals after a two-body contact sliding. The amorphous phase transformation has been predicted. (a) Cross-sectional view of the deformed subsurface of the specimen (V_t = 200 m/s, R = 2.1 nm, δ = 0.99 nm, sliding in [100] direction). Reprinted with permission from [4], L. C. Zhang et al., *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd. (b) Experimental result of the subsurface damage induced. Note the top amorphous layer (V_t = 23.95 m/s, R = 1 μm, δ = 15.2 nm, sliding in [100] direction). Here, each spot represents a silicon atom. Reprinted with permission from [78], L. C. Zhang and I. Zarudi, *Int. J. Mech. Sci.* 43, 1985 (2001). © 2001, Elsevier Ltd. (c) Experimentally measured thickness variation of amorphous layer with the indentation depth of the asperity (V_t = 23.95 m/s, R = 1 μm, sliding in [100] direction).

structure after sliding, although wear has happened, depending on the penetration depth of the particle, its speed ratio of self-rotation to translation, and the variation of atomic bonding strength affected by surface contamination. It was also found that the variation of sliding velocity from 20 to 200 m/s does not change the deformation mechanisms described above.

Figure 47a is a molecular dynamics pediction on the deformation caused by a three-body contact sliding. It is clear that at location ⟨2–4⟩, specified in the figure, the shape of the crystal–amorphous boundary is irregular; but at location ⟨1–3⟩, the boundary is mostly regular along the (111) atomic lattice orientation. This coincides very well with the experimental findings reported by Zarudi et al. [79], shown in Fig. 47b–d. According to the theory of contact mechanics, the stress fields at locations ⟨2–4⟩ and ⟨1–3⟩ are different. At the former location, the hydrostatic stress component dominates so that the boundary does not go along a single atomic lattice orientation. At the latter location, however, the shear stresses play a central role, and hence the phase change terminates in a particular orientation. The above observations are aligned with the stress criteria for various phase transformation events in silicon proposed [3, 80].

4.5.3. Wear Regimes

Similar to the wear mechanisms for the diamond–copper sliding system [2] discussed previously, the wear regimes of the current diamond–silicon system also depend on sliding conditions, as shown by the mechanism diagram, Fig. 48. In a two-body contact sliding with a given sliding speed, the deformation of a silicon monocrystal falls into the no-wear, adhering, ploughing, or cutting regime when the asperity penetration depth varies (see the left half of the figure). Deformation without wear happens only under an extremely small penetration depth, when the atomic lattice of silicon deforms purely elastically. With an increase in the penetration depth, adhering takes place (Fig. 49a), in which some surface atoms stick to the

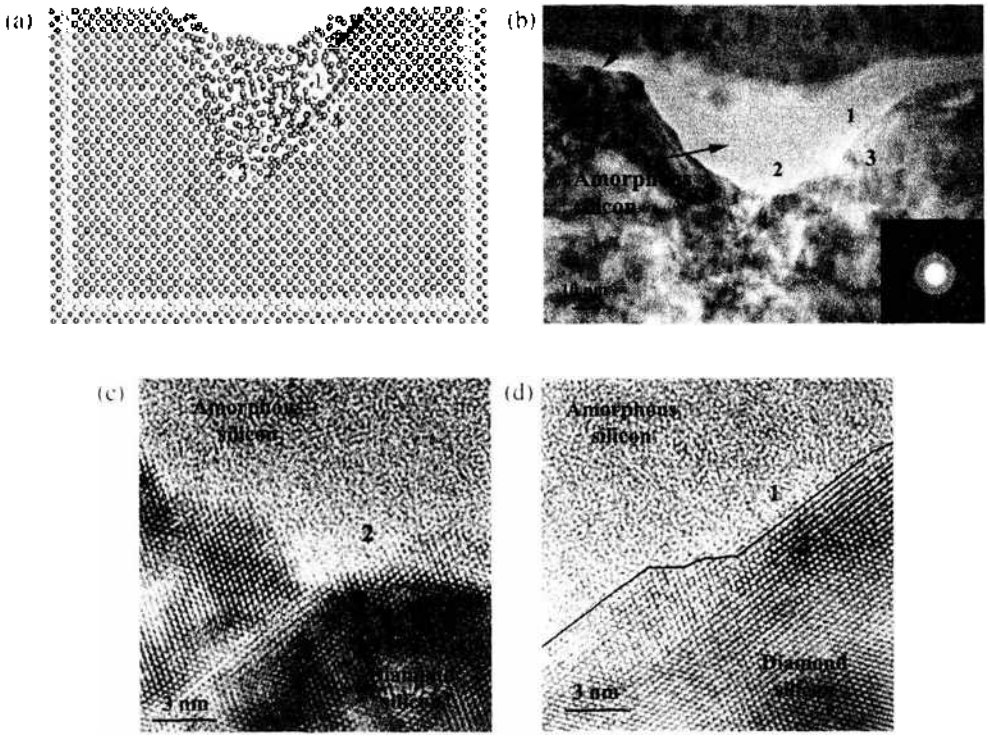


Figure 47. Comparison of theoretical prediction and experimental observation. In (c) and (d), each spot represents a silicon atom observed in experiment. (a) Molecular dynamics prediction of the deformation in the scratch generated by nanopolishing (a cross-sectional view). (b) Cross-sectional view of the deformation in a scratch generated by nanopolishing (experiment). (c) Atomic orientations at the boundary (location ⟨2–4⟩). (d) Atomic orientations at the boundary (location ⟨1–3⟩). Reprinted with permission from [79]. I. Zarudi et al., *Nanotechnology* 15, 104 (2004) © 2004, Institute of Physics Publishing.

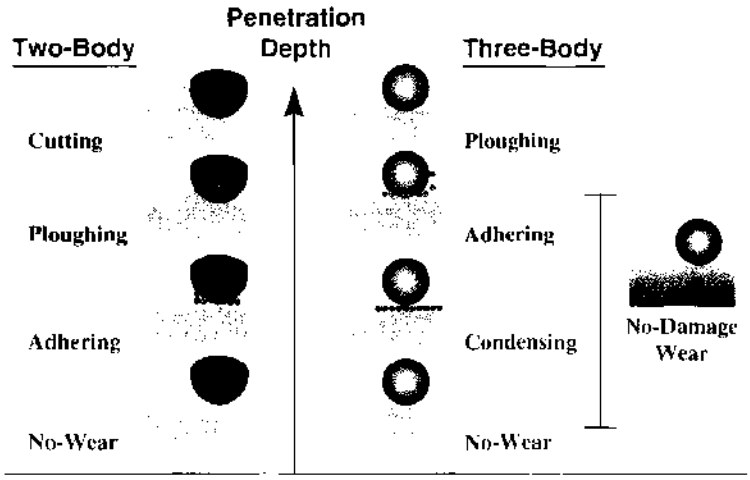


Figure 48. Wear diagram. Diamond asperities/particles move from the right to the left. Rotation of particles is anticlockwise. Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

asperity surface and move together with it to cause wear. However, these atoms may return to the silicon substrate during sliding if the specimen surface has not been contaminated. When the penetration depth increases further, a new wear state, ploughing, characterized by an atomic cluster being pushed to move with the asperity, will appear (Fig. 49b). A further increase of the penetration depth leads to a continuous cutting process.

In a three-body contact sliding, however, silicon will experience different wear regimes. They are the no-wear, condensing, adhering, and ploughing regimes, as shown in the right half of Fig. 48. After the pure elastic deformation in the no-wear regime, the amorphous phase under the particle will experience a remarkable condensing locally without material removal. In other words, because the density of the surface silicon atoms under particle indentation becomes higher, condensing creates a sliding mark on the specimen surface (see Fig. 50a). Thus, condensing is a special wear process without material removal. A further particle penetration will lead to adhering and ploughing (Fig. 50b). These regimes are similar to the corresponding ones in the two-body contact sliding. Cutting rarely happens in three-body sliding processes, but it is possible if the particle penetration depth becomes sufficiently large and the self-rotation speed becomes small.

Another interesting phenomenon associated with the three-body contact sliding is the existence of a regime of no-damage wear. Under certain sliding conditions, the atomic bonding strength among surface silicon atoms can be weakened chemically. When this happens, these atoms can be removed via adhesion because the diamond–silicon attraction is still strong, as shown in Fig. 51. Because of the recrystallization behind the particle, a worn specimen may appear as damage free in the majority of its subsurface, with only little distortion within one or two surface atomic layers. In conjunction with the phenomenon happening in the condensing regime discussed above, it becomes obvious that a perfect subsurface after a three-body contact sliding does not necessarily indicate a no-wear process.

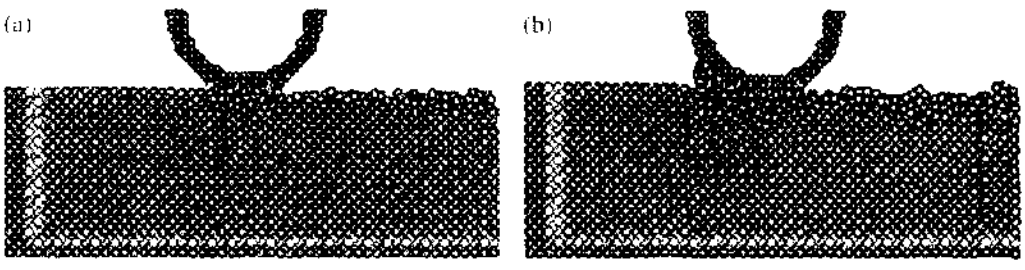


Figure 49. Characteristics of various deformation regimes in two-body contact sliding processes, a cross-sectional view (V_s = 200 m/s, R = 2.1 nm, sliding in [100] direction). (a) Adhering (δ = 0.19 nm); (b) ploughing (δ = 0.39 nm). Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

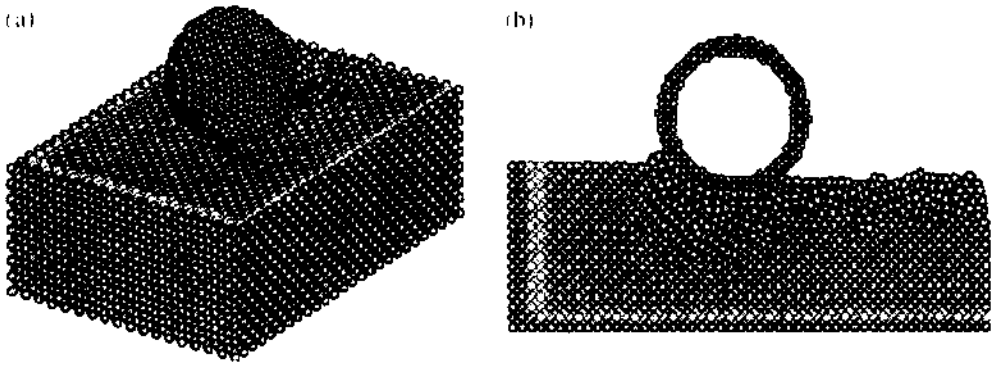


Figure 50. Deformation regimes associated with three-body contact sliding, a cross-sectional view (V_c = 200 m/s, V_j = 100 m/s, R = 2.1 nm, sliding in [100] direction). (a) Condensing (δ = 0.19 nm); (b) ploughing (δ = 0.99 nm). Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

4.6. Multiasperity Sliding

4.6.1. Modeling

Previous study of a single sliding asperity has provided us with important knowledge on the deformation mechanisms of the friction and wear of monocrystalline materials. In a real sliding system, however, a counterpart material is actually subjected to multiasperity interactions, as illustrated in the inserts of Figs. 31a and 31b. When the first asperity has created a damaged zone, the material may deform differently under subsequent sliding interactions. Cheong and Zhang thus discussed the effects of the sliding by multiasperities [81].

The mechanics model consists of three spherical diamond asperities, A, B and C, sliding on an atomically smooth silicon surface, as illustrated in Fig. 52. Their relative positions and orientations are defined by their distances, L_{AB} and L_{AC}, and angles with respect to the sliding direction, α and θ. Three configurations are of special interest; (I) $\alpha = \theta = 0°$, with $L_{AB} < L_{AC}$, representing a repeated single-asperity sliding so that the effect of residual subsurface damage can be understood; (II) $\alpha = 0°$ and $\theta = 90°$, with $L_{AB} = L_{AC}$, standing for the interaction of two parallel asperities; and (III) $\alpha = 90°$ and $\theta = 0°$, indicating the case with parallel sliding asperities coupled with an interaction from a third asperity. Again, because diamond is much harder than silicon, the asperities are modeled as rigid spheres. These spheres slide across the silicon surface at a specified velocity of 40 m/s. The maximum depth of asperity penetration is 1.0 nm.

4.6.2. Configurations II and III

In the cases of configurations II and III, the asperities do not retrace the damaged zones. At the depth of asperity penetration of 1.0 nm, the wear mechanism observed is that of cutting.

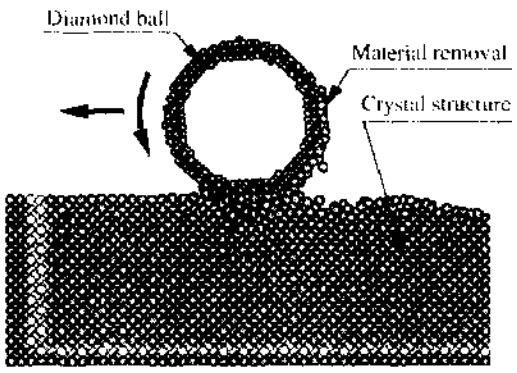


Figure 51. Material removal via adhering in a three-body sliding process when the C–Si interaction is three times stronger than that of Si–Si. The particle translation is from the right to the left, and its rotation is anticlockwise. (V_c = 200 m/s, V_j = 100 m/s, R = 2.1 nm, and δ = 0.19 nm). Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

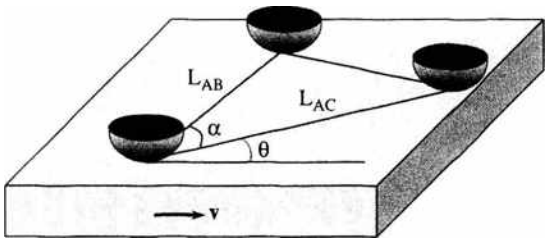


Figure 52. Mechanics model for multiasperity contact sliding. Reprinted with permission from [81], W. C. D. Cheong and L. C. Zhang, *Int. J. Mater. Prod. Tech.* 18, 398 (2003). © 2003, Interscience Enterprises Inc.

The plastic deformation resulting from the sliding asperities is very localized. Figure 53 shows a cross section of the silicon workpiece through the center of the asperities. It can be seen that there is almost no subsurface damage to the silicon workpiece between the two asperities. Plastic deformation is very localized and occurs beneath the asperities.

As in nanoindentation [33], the sliding asperities A, B, and C also create trails of a subsurface amorphous layer in the damaged zones. Phase transformation of silicon occurs at the leading edge beneath the sliding asperity, resulting in the formation of amorphous chips and an amorphous trail along the path traversed by the asperities. Dislocations are absent at this particular depth of asperity penetration, which indicates that the plastic deformation is solely caused by phase transformation.

4.6.3. Configuration I

In this case, the second and third asperities, B and C, retrace the damaged path caused by asperity A. Therefore, the cutting mechanism involved in the first and the following two asperities are very different. Asperity A cuts the silicon workpiece in the same fashion as in the case of a single sliding asperity, causing phase transformation of the original diamond cubic silicon. Asperity B, however, ploughs through the residual amorphous layer in the wake of asperity A. No further phase transformation is found, but the amorphous silicon atoms are pushed away as the asperity ploughs through.

4.6.4. Phase Transformation

As shown above, something apparent in the silicon workpiece in all the three configurations is the formation of an amorphous subsurface layer in regions traversed by the diamond asperities. This is analogous to the amorphous damaged zone resulting from nanoindentation. Hence, it is worthwhile to draw comparisons and to predict the process of phase transformation on the basis of results obtained from nanoindentation [3, 33].

With configurations II and III, wear occurring via cutting is achieved by chipping the amorphous silicon. By considering the exact coordinates of the silicon atoms, it is found that the transformation mechanism is similar to that of nanoindentation [33]. Diamond cubic silicon first transforms into its β-tin phase and then, on the removal of stresses, transforms into an amorphous phase. This explains the trail of subsurface amorphous silicon in the damaged zone behind each sliding asperity. The mechanism of phase transformation is reflected in the coordination numbers of atoms near the regions beneath the asperities.

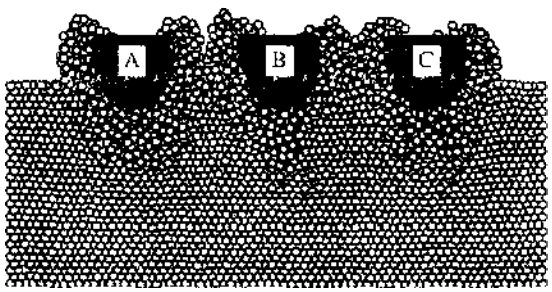


Figure 53. Cross-section of the silicon workpiece through the center of asperities. Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

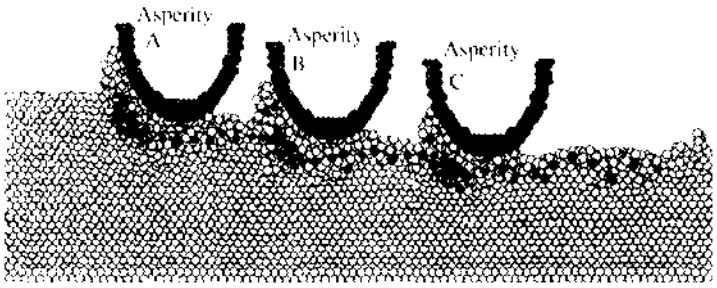


Figure 54. Cross-sectional view of asperities A, B, and C in configuration I. A light-colored circle represents an atom with coordination number 4, and a dark circle represents an atom with coordination number 6. Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

It can be seen that a cluster of six-coordinated body-centered tetragonal β-tin silicon atoms indicated by darkened circles (Fig. 54) forms near the leading edge beneath each asperity. Four-coordinated diamond cubic silicon transforms to its six-coordinated β-tin form as a result of the stresses induced by the asperity. When the stresses are removed as the asperity slides past, the atoms transform into an energetically more favorable amorphous form. The coordination number of the silicon atoms in the trailing amorphous region is mostly equal to four. Therefore, as the asperity slides across the silicon workpiece, diamond cubic silicon continuously transforms into β-tin silicon beneath the asperity and then transforms into amorphous silicon when the asperity passes, leaving a layer of subsurface amorphous silicon in its wake.

Figure 55 shows the number of β-tin silicon atoms formed during the cutting process with asperity configurations II and III. The number of β-tin silicon in configurations II and III is three times that of a single asperity cutting process. This indicates that the formation of β-tin silicon is highly stress-state dependent and occurs only beneath the three asperities. Although regions between the asperities are compressed because of the proximity of the asperities, it is likely that β-tin silicon does not form because of the absence of the required stress states.

With configuration I, asperity A cuts the silicon workpiece in very much the same way as in the other two configurations described above. The asperity cuts through the diamond cubic silicon, leaving behind a trail of subsurface amorphous silicon. However, there is a vast difference in the silicon phase transformation involved because of asperities B and C, which represent subsequent cuts into the amorphous zone. In this case, the six-coordinated atoms also form beneath asperities B and C (Fig. 54). This implies that the β-tin silicon phase is recoverable from the amorphous phase, provided that the required stress field is

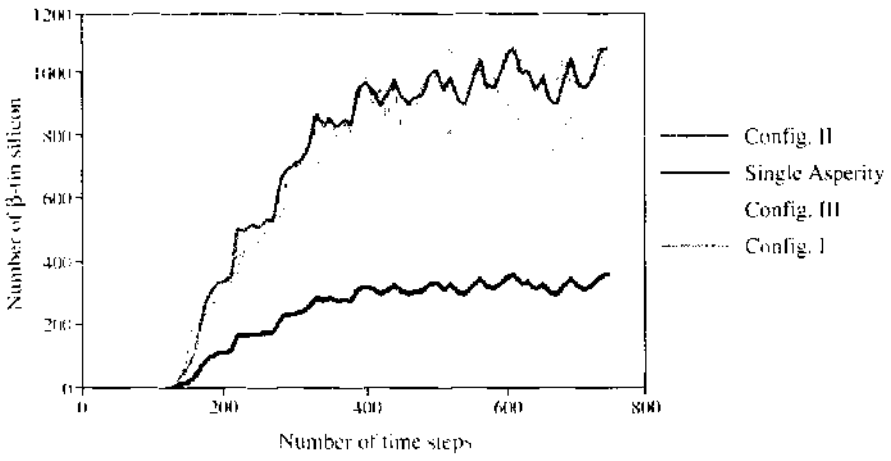


Figure 55. Number of six-coordinated β-tin silicon atoms in configurations I, II, and III compared to that with a single asperity. Reprinted with permission from [4], L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998). © 1998, Elsevier Ltd.

achieved. It must be noted that only some of the β-tin silicon is recovered as asperities B and C retrace the amorphous damaged zone. That is why the number of β-tin silicon atoms in configuration I is slightly less than in configurations II and II (Fig. 55).

5. CHARACTERIZATION OF CARBON NANOTUBES

5.1. Introduction

Carbon nanotubes have attracted tremendous attention since their discovery in 1991 [82]. Experimentally, they have been observed as single-walled nanotubes (SWNT) [83, 84], multiwalled nanotubes (MWNT) [82, 85, 86], bundles [87, 88] and nanoropes [89, 90]. They have remarkable electrical and mechanical properties. For example, they have tensile strength 20 times that of high-strength steel alloys and have a current carrying capacity 1000 times that of copper [91]. As such, nanotubes are expected to have a variety of applications, such as in memory chips, sensors, probes, tips, and reinforcing phase in composite materials.

With the rapidly growing interest in nanotubes and the difficulties in direct measurement resulting from its nanoscale dimension, molecular dynamics simulation has been widely used in characterizing the mechanical properties and understanding the mechanism of deformation. However, such simulations have to be done carefully to represent the real situation. Similar to the nanoindentation and nanotribology problems discussed previously, to have a reliable molecular dynamics simulation of carbon nanotubes, it is also important to select an appropriate interaction potential so that it describes the true deformation of a nanotube correctly and effectively. Second, during a loading process, improper treatment of the temperature rise can lead to incorrect simulation results. In molecular dynamics, as has been extensively discussed in previous sections, heat conduction is accomplished via thermostat atoms and various thermostatting methods. Adiabatic relaxation method, isokinetic thermostattting, Andersen stochastic thermostatting, and Nose–Hoover feedback thermostatting were all reported in the literature [92] for temperature conversion. The adiabatic method is quite noisy. In the isokinetic thermostatting, the temperature is maintained in different ways. For example, in the Berendsen thermostat scheme with velocity scaling, the velocities of the thermostat atoms are scaled to fix the total kinetic energy. In the Gaussian feedback or Evans–Hoover scheme with force scaling, however, the kinetic energy is monitored and information is fed back into the equations of motion so that the kinetic energy is kept constant to dissipate heat by controlling the thermostatting force. The velocity scaling has been used in general because it is somewhat a simpler scheme to implement. For small time steps, the Gaussian isokinetic method and velocity scaling method are identical [92]. However, a very small time step will give an unusually high elongation speed. In contrast, a small displacement step with a small time step will be computationally expensive. The flaw in the isokinetic-thermosatting method is that it is impossible to separate the effects of thermostatting on rate processes. The other two schemes also have this limitation to a certain extent. Third, the system has to be relaxed initially as well as during the simulation so that the velocities of the Newtonian and thermostat atoms reach equilibrium at the specified temperature of the simulation; an appropriate time step and displacement step has to be selected to get a reasonable elongation speed.

It is unfortunate that no comparison or clarification is available in the literature as to the key issues discussed above for a reliable molecular dynamics simulation of carbon nanotubes. For example, researchers have used either the Tersoff–Brenner (T-B) potential or the tight-binding potential in their molecular dynamics calculation on nanotubes. Those who used the potential claimed that the potential was selected because it could avoid the overbinding of radicals. However, the overbinding of radicals in a carbon nanotube is possible only when the bonds are broken. Sinnot and coworkers [93, 94] used the first few rows of atoms on both ends as boundary atoms and the next few rows as thermostat atoms applying the Langevin heat baths. Zhou and Shi [95] used the first two rows of atoms on both ends as boundary atoms but treated all the other atoms as thermostat atoms. No information is available on the method of temperature conversion. Moreover, various time steps, ranging from 0.15 to 15 fs, have been used in the simulation of carbon nanotubes. A natural question is, therefore: Which simulation scheme will be reliable and effective?

In this section, we will discuss various critical issues for a reliable simulation by the molecular dynamics method for analyzing carbon nanotubes and their mechanical properties. In particular, we will address the problems in selecting potentials, number of thermostat atoms to control the simulation temperature, type of thermostats for the conversion of heat, time and displacement steps, and number of relaxation steps to reach the dynamic equilibrium. On the basis of these variables, the structural changes of armchair and zigzag nanotubes and their mechanical properties, including Young's modulus and Poisson's ratio, are also investigated.

5.2. Modeling

The interatomic forces were described by the Tersoff potential [6–7] and the empirical bond order potential formulated by Brenner based on Tersoff potential and known as the T-B potential [96, 97]. The simulations were carried out at 300 K with a time step of 0.5 fs. Open single-walled armchair nanotubes (10, 10) with 100 repeat units along the axial direction and zigzag nanotubes (17, 0) with 58 repeat units along the axial direction, both having a length of about 245 Å, were examined. The first two layers of atoms on both ends were held rigid. The next four layers were taken as thermostat atoms, and the remaining atoms were treated as Newtonian atoms. First, the tubes were annealed at the simulation temperature for 5000 time steps. Then the rigid atoms on both ends were pulled along the axial direction at an increment of 0.05 Å unless otherwise stated in the discussion. Each displacement step was followed by 1000 relaxation steps to dissipate the effect of the preceding displacement step over the entire length of the tube. In another simulation, all the atoms except the boundary atoms that were held rigid were treated as thermostat atoms, as in Ref. [95]. In this case, each displacement step was followed by 50 relaxation steps (in Section 3.4, a different number of relaxation steps were used to examine its effect). The stress was calculated by dividing the axial force by the cross-sectional ring area of the nanotube. The equivalent wall thickness of a carbon nanotube is taken as 0.617 Å [98], which gives the area as 2.582×10^{-19} m^2.

5.3. Deformation Characteristics

For the armchair (10, 10) and zigzag (17, 0) tubes under tensile loading, Fig. 56 compares the axial stress as a function of strain with both Tersoff and T-B potentials using Berendsen and Evans–Hoover thermostats.

The stress–strain relation has four clear stages. Initially it is linear (stage 1). It then becomes nonlinear (stage 2) and reaches a plateau. After this the stress increases sharply (stage 3), reaches the maximum, and then drops down to zero or close to zero (stage 4). The Young's moduli of the tubes were evaluated from the linear region (stage 1), which gave a value of 3.96 TPa for the armchair tube and 4.88 TPa for the zigzag tube. As pointed out by Vodenitcharova and Zhang [98], the value of the Young's modulus of a nanotube depends on the thickness of the tube. Here we use the equivalent thickness of 0.617 Å, as mentioned before. The Poisson's ratios, $(\Delta r/r)/(\Delta l/l)$, were found to be 0.15 for the armchair tube and 0.19 for the zigzag tube, where Δr and Δl are the changes in tube radius and tube length, respectively.

5.4. Potential

At the initial stages of loading, all the stress–strain curves overlapped with one another, and it is almost linear. Thereafter, the Tersoff potential curve lies slightly above the T-B potential curves, but it has the same trend up to a strain of 0.34 and 0.2 for the armchair and zigzag tubes, respectively. The armchair tube has a maximum stress of about 1357 GPa around a strain of 0.4, and the zigzag tube has the maximum stress of 754 GPa around a strain of 0.22. After this, a large cross-sectional necking happens. If the tubes are unloaded from a point little before the maximum stress, the stress–strain curve at unloading overlaps with the loading curve, showing that the tube deformation up to this stage is completely elastic.

With both potentials at various stages of the loading process, the structural changes of the tube were examined by unrolling the tube. Figs. 57a–c and 58a–c give the two-dimensional view of a portion of the unrolled armchair and zigzag tubes at stages 1, 2, and 3 when

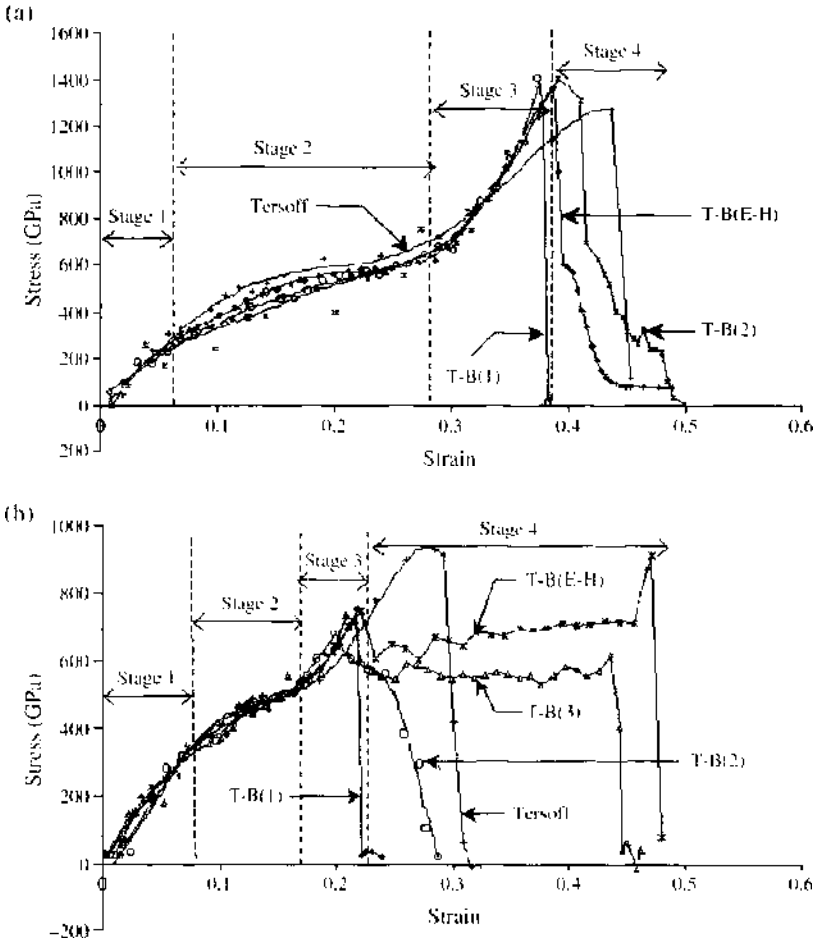


Figure 56. Stress–strain curves with Tersoff and Tersoff–Brenner potentials using Berendsen and Evans-Hoover thermostats. [T-B(1) is the calculation with the first two rows of rigid atoms and next four rows of thermostat atoms on both ends. T-B(2) is the calculation with the first two rows of rigid atoms and all other atoms as thermostat atoms. T-B(3) is the calculation as in T-B(2) but with a smaller displacement step of 0.008 Å.]; (a) (10, 10) armchair SWNT; (b) (17, 0) zigzag SWNT.

using the T-B potential. They show that during loading, both the angles and lengths of the bonds change. In the zigzag tube, on the application of axial stress, both types of bonds are stretched because of the way that they are oriented.

For an armchair tube, Figs. 59a and 59b, compares the variation of these geometrical parameters with the two potentials. With the T-B potential, the change in bond angles is more, but with the Tersoff potential the change in bond lengths is more. At about maximum

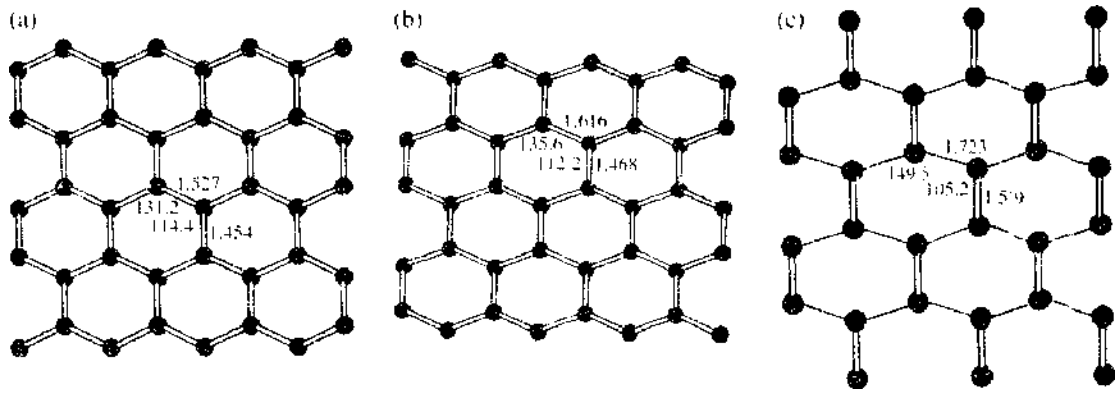


Figure 57. Structural changes of an armchair tube (10,10) in different stages. (a) In deformation stage 1, indicated in Fig. 56a; (b) in deformation stage 2, indicated in Fig. 56a; (c) in deformation stage 3, indicated in Fig. 56a.

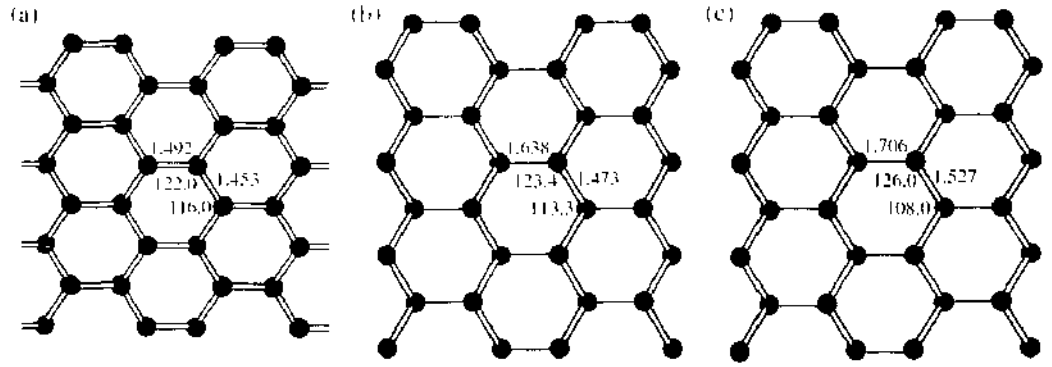


Figure 58. Structural changes of a zigzag tube (17,0) in different stages. (a) In deformation stage 1, indicated in Fig. 56b; (b) in deformation stage 2, indicated in Fig. 56b; (c) in deformation stage 3, indicated in Fig. 56b.

loading, some of the stretched bonds are broken and the Tersoff potential curve starts to deviate from the T-B curve. This seems to indicate that when bonds are broken, radicals will form, and hence the overbinding effect that was not considered in the development of Tersoff potential would become important. The T-B potential results showed the necking of the tube followed by the formation of a one-atom chain, as shown in Figs. 60 and 61. The use of the Tersoff potential, however, did not bring about any necking or formation of the one-atom chain, and the tube broke suddenly after reaching the maximum stress.

The above comparison and discussion show that the T-B potential describes the whole process reliably. However, the Tersoff and T-B potential curves overlap at the initial stages, indicating that the Tersoff potential can be used for the calculation of the mechanical properties of a carbon nanotube such as the Young's modulus and Poisson's ratio.

5.5. Number of Thermostat Atoms

It is necessary to clarify the issue because in the literature the number of thermostat atoms was selected and used without a rational reason. Our comparison was done with the T-B potential using the Berendsen thermostat.

The first method is calculation with the first two rows of rigid atoms and next four rows of thermostat atoms on both ends. This method showed little fluctuation in the temperature up to the third stage. When stage 4 starts, the tube falls into pieces and the temperature increases suddenly and sharply. This is because when some bonds break at the same time, a huge amount of energy is released and leads to a sudden increase in temperature. Moreover, with such an arrangement of thermostat atoms, no significant necking will take place.

The second method is calculation with two rows of rigid atoms and all other atoms as thermostat atoms. With this method, a remarkable necking occurs in stage 4. The armchair tube started necking at a strain of 0.39, and the zigzag tube started its necking at a strain of

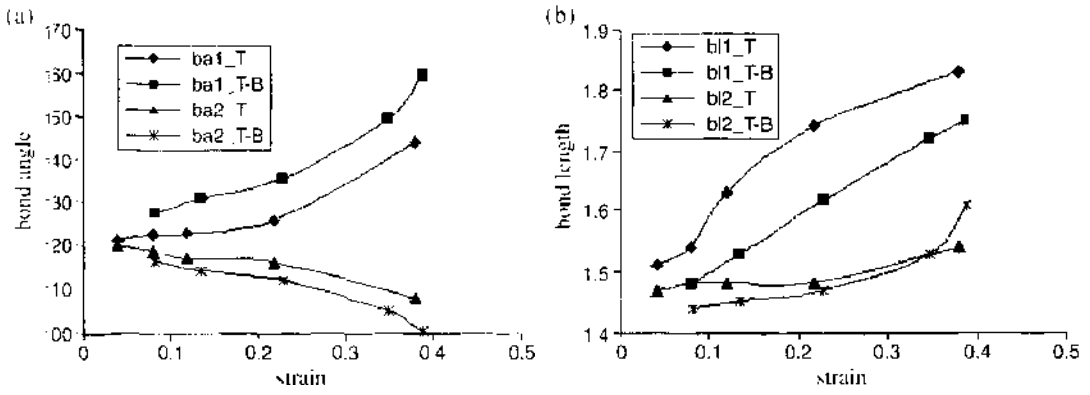


Figure 59. Comparison of the changes in bond angles and bond lengths of an armchair carbon nanotube with the Tersoff and Tersoff–Brenner potentials. (a) Bond angles (ba1 and ba2); (b) Bond lengths (bl1 and bl2).

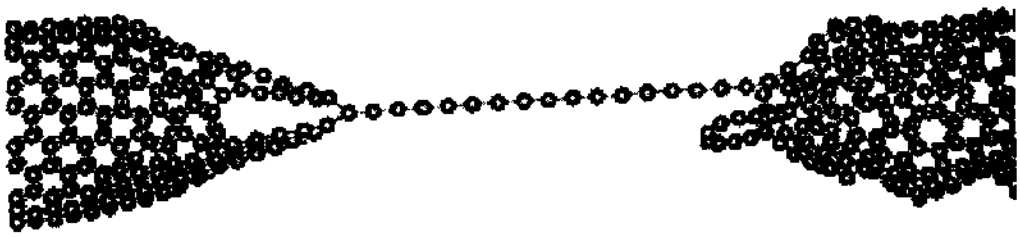


Figure 60. Atomic chain of an armchair tube when using Berendsen thermostat.

0.22. The temperature increase in this stage is only 20°–30°C in a short period. On further application of the tensile force, a one-atom chain formed, as shown in Figs. 60 and 61 and grew with the applied force before the tube breaks, as reported in the literature [95, 99]. Within the chain, the strain is constant with a C–C distance of 1.7 Å. In other words, on further pulling, the C–C distance in the atomic chain did not change, but more and more C atoms joined the chain. This is known as carbon chain unravelling and was observed in an experiment when a capped nanotube was opened by the force of an electric field [100]. However, analysis of the current molecular dynamics results showed that the electronic structure of this chain is close neither to the cumelenic form nor to the bond alternate polyyne, as suggested in Ref. [100]. It is believed to be a pure unravelling process that originates from a place at which a bond is broken. Although in chemical terms the valency of carbon is not satisfied, it may be possible to have such a chain under stress for a short period. Once the growing chain was detached from the tube, the C–C distance decreased to 1.33 Å, which is equivalent to a C–C double bond, as in the cumelenic form. During necking, some bonds are broken and some new bonds are formed to facilitate the closure of the ends at the breakage. As a result, the temperature did not go up suddenly. The slight increase in temperature can be attributed to the activation energy of the unravelling process.

The above comparison and analysis clearly show that the second method is the right one to use in molecular dynamics simulation of carbon nanotubes. Note that the loading rate in the molecular dynamics calculations is much higher compared to that in an experiment. As such, in the experiment not much heat will be generated. In the simulations, because of the computational cost and numerical accuracy, the loading rate cannot be as small as in the experiment. Hence, one has to find an effective way to conduct the heat that is produced as a result of the higher loading rate. This could be achieved either by relaxing the atoms for a long time between each step of pulling, which is computationally expensive, or by treating all the atoms as thermostat atoms, as in the second method. The first method is not reliable in this sense because the Newtonian atoms are not fully surrounded by the thermostat atoms; that is, they are exposed to the environment. As such, the heat conduction in the simulation cannot reflect correctly the real deformation process of a nanotube under experimental conditions.

5.6. Influence of Thermostat Schemes

Heat conversion is a central component in a correct molecular dynamics simulation [72]. An inappropriate conversion scheme will result in incorrect atomic motion and deformation. To examine the effect of thermostat schemes, we use T-B potential, as it is reliable for nanotubes. For the armchair tube, both the Berendsen and Evans–Hoover thermostats gave

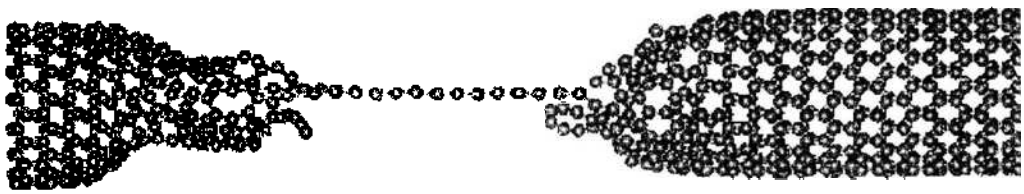


Figure 61. Atomic chain of a zigzag tube when using Berendsen thermostat.

similar results until maximum stress. After this point, with the Berendsen thermostat (first method arrangment) the stress dropped down to zero, the temperature went up, and the tube broke into pieces without the formation of an atomic chain; with the Evans–Hoover thermostat and the Berendsen thermostat (second method arrangment), the stress dropped down to a value close to zero, the tube necked and formed an atomic chain and then broke into two pieces with closed ends.

For a zigzag tube, the Evans–Hoover thermostatting scheme showed an entirely different behavior than the armchair one once the maximum stress was reached. The zigzag tube started to neck at both ends at a strain of 0.23 as with the Berendsen scheme, but on further tension, the necking propagated, as shown in Fig. 62a until the whole tube became narrow, which happened at a strain of 0.475. During this period the stress was almost constant. This could be because in the zigzag tube both types of bonds can get stretched, as shown in Fig. 62b and 62c. As such, the above necking and its propagation is possible, whereas in the armchair tube, the bonds that are normal to the direction of pulling cannot get stretched. As such, the stretched bonds will break once they reach the maximum stress. On further application of the tensile force, again the tube necked at both ends, formed a one-atom chain, and broke within a short period. It is interesting to note that the ultimate failure of the tube happens around a strain of 0.47, which is closer to that of an armchair tube.

The Evans–Hoover scheme can be understood more easily if the discussion in Section 1 is recalled. In this scheme, the force scaling is done on all the atoms, irrespective of the number of thermosat atoms used in the calculation, and the kinetic energy is kept constant. Thus, the heat conduction problem is avoided and a very smooth stress–strain curve was obtained. However, with this scheme, one has to work with small time steps. This means that to have a reasonable loading rate, the displacement step must be small, which would increase significantly the overall computational time. Hence, if the Berendsen scheme with the second method of thermostat atom arrangement is used, one can minimize the heat conduction problem and improve computational efficiency.

5.7. Integral Time Step, Displacement Step, and Relaxation Step

In molecular dynamics simulations, the time step has to be selected to reduce the round-off error and truncation error. A suitable time step should be less than 10% of the vibration

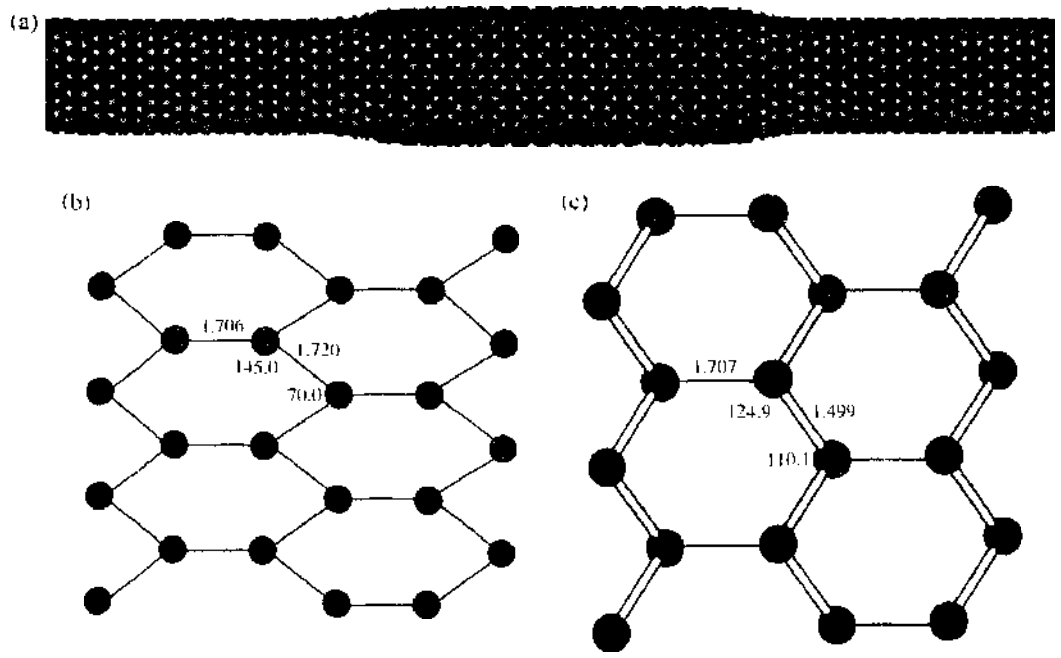


Figure 62. Deformation of a zigzag tube. (a) Showing the necking that propagates along the tube; (b) showing the structure on unrolling the tube at the narrower part; (c) showing the structure on unrolling the tube at the wider part.

period of an atom and, accordingly, for diamond, a time step of 0.5–0.8 fs will provide good results. Hence, in our calculations, we have used a time step of 0.5 fs.

Displacement step is usually chosen according to the time step used. For the armchair tube, variation in the displacement step did not show any significant difference in the stress–strain relationship. However, for the zigzag tube, the Berendsen scheme with the second method arrangement of thermostat atoms and a displacement step of 0.05 Å showed failure at a strain of 0.23. When the displacement step was reduced to 0.008 Å, it brought about necking and its propagation over the entire length of the tube and the formation of the one atom chain and then the ultimate failure at the strain of 0.44, similar to the application of the Evans–Hoover thermostatting scheme. In other words, a little lower elongation speed showed a more reasonable results. Moreover, on unloading the tube from a point during the necking propagation showed that the necking is elastic, although the unloading curve took a different path until a strain of 0.182. This is because on unloading, the bond lengths and bond angles do not change in the same way as on loading.

Initially, the tube is relaxed to its dynamic equilibrium status at the specified simulation temperature. We find that to reach this equilibrium about 4000 molecular dynamics steps are required. As a result, in all our calculations the initial relaxation is done for 5000 molecular dynamics steps. After the initial relaxation, tensile loading is done via small displacement steps. The effect of the displacement is dissipated over the entire length of the tube, and subsequently, a dynamic equilibrium is achieved by relaxing the system after each displacement. The required number of relaxation steps varies with the number of thermostat atoms used. The first method's arrangement required nearly 1000 relaxation steps to reach the dynamic equilibrium. Even so, the necking and the one-atom chain formation were not observed. In contrast, the second method's arrangement, with 50 relaxation steps, showed the necking and the one-atom chain formation for both armchair and zigzag nanotubes.

These results clearly show that even with 20 times the relaxation steps used in the second method, the first method fails to show the necking once the maximum stress is reached. For the zigzag tube, a smaller displacement step of 0.008 Å is required to get the correct behavior, where the ultimate failure of the tube happens around a strain of 0.44.

5.8. Summary

In summarizing the above discussion, we can conclude that the simulation using the T-B potential and the Berendsen thermostat with all atoms as thermostat atoms (except the rigid ones) with 50 relaxation steps following each displacement of 0.008 Å is the correct, reliable, and cost-effective method. Following the reliable simulation technique identified in this study, it is quantified that the Young's modulus and Poisson's ratio of the armchair tube are 3.96 TPa and 0.15, respectively, and the zigzag tube are 4.88 TPa and 0.19, respectively. The armchair tube has a higher tensile stress compared to the zigzag tube. In both armchair and zigzag nanotubes, under tensile loading, the carbon chain unravelling and the one-atom chain were observed at a strain of around 0.4.

6. DEFORMATION OF COPPER NANOWHISKERS

6.1. Introduction

A complete figure of the stress–strain behavior of a nanoscale specimen under tensile loading is of primary importance to the development of nanotechnology. However, the information about the mechanical properties of nanoscopic specimens is lacking, and the potential problems of testing such tiny specimens have not been explored, although there have been some investigations on the deformation of microscopic specimens. This is because adequate testing techniques are unavailable. On the nanometer scale, specimen preparation becomes very difficult, and a mechanical testing must be well controlled with a high level of stability in terms of temperature, strain rate, crystal orientation, and shape of specimen.

In this section, we will use the molecular dynamics method to investigate the deformation of copper nanowhiskers subjected to uniaxial tension, based on the study carried out by Zhang, Tanaka, and Gupta [101]. Some important aspects, such as the effects of the shape

and atomic orientation of tesing samples, will be discussed in detail to provide the theoretical basis for potential experimentation.

6.2. Modeling

As already discussed earlier in the chapter, the Morse potential is reliable for investigating the deformation of monocrystalline copper. For increasing the computational efficiency, a basket of neighboring atoms with 14 units was maintained for each atom, whereas the Morse potential cutoff radius was taken as nine units. The basket was updated adequately within a certain time interval by considering both the accuracy and efficiency. A criterion used for updating was that an atom not in the basket does not enter the Morse potential cutoff radius, or an atom that was within the cutoff radius does not leave the basket before a further updating. The motion of an atom depends on the strain rate of testing, and thus a higher strain rate needs a larger basket.

To properly maintain the temperature at 298 K during the molecular dynamics simulation, thermostat atoms were arranged at the two loading ends of the specimen, which absorb the rise in velocity momentum of Newtonian atoms. Figure 63 shows the initial atomic configuration of a specimen with a hexagonal cross section. The reason for using such a special cross section shape will be presented in the next section.

6.3. Effects of Atomic Orientation and Specimen Shape

When a simple square model with a (100) lattice in the direction of loading is subjected to free relaxation, a shrinkage of about 23% occurs. The surface atoms tend to restructure to (111) to maximize their atomic density, whereas those in the core of the specimen remain in a stable FCC configuration of (100). As a result, a shrinkage stress is generated before the uniaxial tensile loading. A series of testing results shown in Table 7 demonstrate that the shrinkage stress is sensitive to the dimensions of nanowhiskers, and smaller specimens have higher shrinkage stresses. This is because on the nanometer scale a specimen always tends to deform in a manner that may minimize its net potential energy.

In an FCC structure, atomic lattices in different directions have different atomic densities. It is therefore not difficult to understand that specimens with different shapes and surface orientations would have different dimensional stability and shrinkage stress. The above observation suggests that the shape of a specimen may be stabilized and the shrinkage stress be minimized by maximizing the surface atomic density. A systematic investigation shows that a specimen with a uniform square cross section of two (111) and two (112) surfaces has a lower shrinkage stress. A specimen, oblique in shape, with all its four surfaces as (111) planes shows even less distortion during relaxation. The best specimen shape is with a hexagonal section, which has four (111) and two (100) surfaces, as shown in Fig. 63. Such specimens do not have sharp edges and can maximize the gross density of the atoms on the

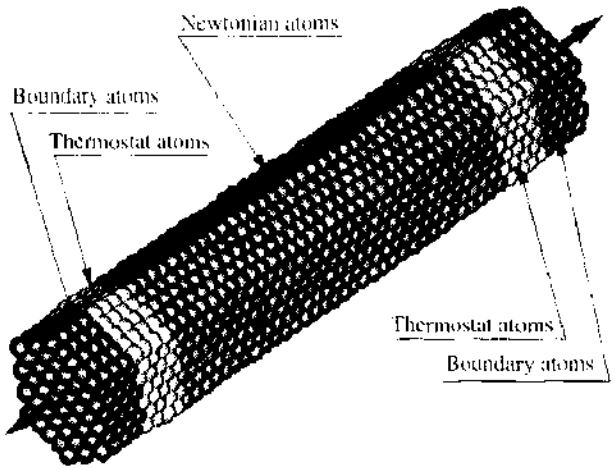


Figure 63. The atomic model of a hexagonal nanowhisker specimen of monocrystalline copper.

Table 7. Effects of size, shape, and surface atomic structure of copper monocrystal nanowhisker.

Specimen Type	Specimen Size	Atom Number	Shrinkage Stresses Generated, MPa	Ultimate Tensile Strength, MPa	Strain at the Onset of Plastic Deformation
	$10 \times 1 \times 1$	263	9000	14000	12.5
Specimen with a	$30 \times 2 \times 2$	763	7500	9000	6.9
rectangular	$30 \times 2 \times 4$	1373	4000	8100	8.0
section and	$30 \times 4 \times 4$	2471	3000	9100	18.2
(100) surfaces	$30 \times 4 \times 8$	4667	1500	8200	15.8
	$30 \times 8 \times 8$	9971	750	6400	7.4
	$34 \times 8 \times 16$	19355	500	—	—
Hexagonal	$48 \times 9 \times 7$	3871	0	6000	5.5

surfaces and minimize the distortion and shrinkage stress. Thus, this type of specimens can produce reliable results.

6.4. Elastic Modulus and Ultimate Tensile Strength

With most materials, there is a gradual transition from elastic to plastic deformation, and the onset point of plastic deformation is difficult to define with precision on the macroscopic or microscopic scale. For a nanospecimen, however, plastic deformation can be said to have occurred when an atom has changed its position permanently. The last column of Table 7 indicates the strain needed for the occurrence of the first predominant slip in different specimens and directly gives rise to the modulus of elasticity. The modulus of elasticity, using the most stable hexagonal specimen, is 120 GPa, which is very close to that of 130 GPa obtained by a macroscopic engineering test. This indicates that macroscopic testing can be used instead of the difficult nanometer-scale experiment, if the measurement of elastic modulus is the only concern.

It is the fact that the ultimate tensile strength (UTS) in a single crystal can be sometimes higher than that of a polycrystal of the same material, tested at the same temperature and strain rate. A distinct rise in UTS has also been observed when the size of specimens reduces to the microscopic scale; for instance, to a microwhisker as small as 6.8 μm in diameter, which shows a UTS of about 1950 MPa as opposed that of 220 MPa measured with macroscopic specimens of polycrystal copper. The present molecular dynamics simulation shows that the UTS of a nanowhisker of monocrystalline copper can go up to 6000 MPa, which is about three times the strength of a microwhisker.

6.5. Shear Banding and Necking

Although the theory of shear-band formation has been well established, little knowledge is available to fully explain the actual mechanism of how a band begins to form.

Figure 64 shows the evolution process of shear banding and necking revealed by the molecular dynamics simulation. Even in the elastic stage, the deformation in the atomic lattice has been nonuniform (see Fig. 64a). Figure 64b and 64c shows that the first and second bands occur when the built-up stress is gradually released. The process of stress build-up and stress release continues in the whole elastic–plastic deformation process when the total strain varies from 5.64% to 11.33%. By this time, all the primary bands have formed. A further straining, as shown in Fig. 64d, brings about an onset of necking at one of the weaker zones between two bands. In conjunction with this necking, stress increases first and then drops quickly when a noticeable necking forms. Here it is interesting to note that the atoms in the first necking zone are quickly restructured by minimizing their potential energy and exhibit a stronger resistance to further deformation. As a result, a second necking takes place between other bands, as shown in Fig. 64e and Fig. 64f. This process continues until a prominent slip occurs, as shown in Fig. 64g, which leads to a stable necking followed by the final breakage of the whisker.

In summary, the above molecular dynamics simulaton brings about some new and interesting understandings of the nanowhisker deformation. We have understood that a hexagonal

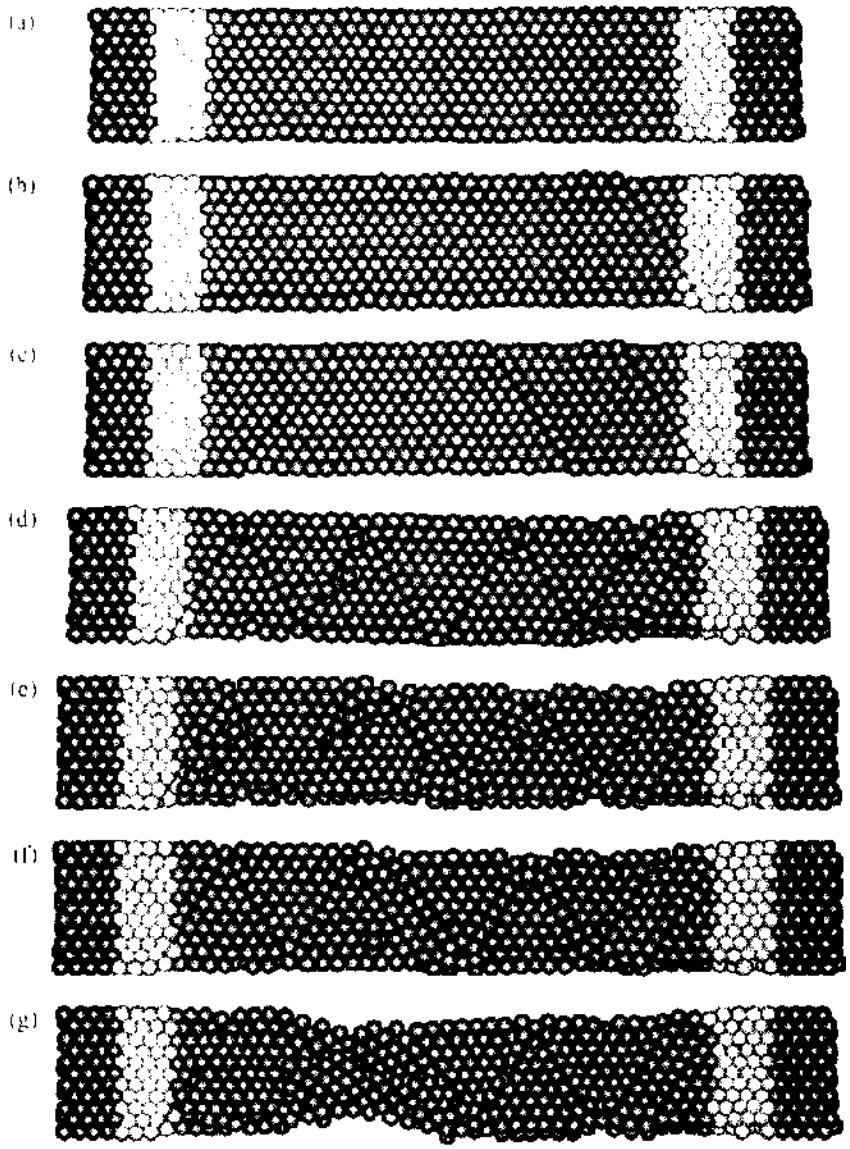


Figure 64. The evolution process of shear banding and necking in a nanowhisker of monocrystal copper under uniaxial tension. These are the cross-sectional views through the center of the specimen. (a) At the strain of 4.61%; (b) at the strain of 5.64%—note the formation of the first shear band; (c) at the strain of 6.36%—the second shear band has formed; (d) at the strain of 11.4%—many shear bands have formed, and necking can also be observed at this stage; (e) at the strain of 14.15%—necking has appeared in a number of places in the specimen; (f) at the strain of 14.51%—necking has developed further; (g) at the strain of 15.59%—the location of necking is stabilized.

nanospecimen with four (111) and two (100) side surfaces should be used for testing to achieve a high level of shape stability and eliminate the effect of shrinkage stress, that the elastic modulus on the nanometer scale is close to that from macroscopic measurement, that the ultimate tensile strength of a monocrystalline copper can be as high as 6 GPa, and that the mechanism of shear banding and necking on the nanometer scale are much more complex than ever thought.

7. BUCKLING OF DIAMOND THIN FILMS

7.1. Background

Thin diamond films have promising applications in protective coatings and microelectronic devices as a result of their high density, extreme hardness, high thermal conductivity, chemical inertness, and infrared transparency. However, the films often suffer from adhesion problems—partially or totally delaminated at the interface or in the substrate material

because of high stresses [102–107] . The total stress is composed of a thermal stress and an intrinsic stress. The former is a result of the difference in the thermal expansion coefficients of the film and substrate materials, and the latter is a result of the accumulating effect of the crystallographic flaws that are built in during deposition. According to Thornton and Hoffman [108], the parameter T/T_m, where T is the substrate temperature and T_m is the coating material melting point, is important in describing the stress-related behavior. For materials with high T/T_m, bulk diffusion becomes very important, and this relaxes the intrinsic stresses and prevents their accumulation. As carbon is a harder, higher-melting point material ($T_m \sim 3700$ K), under typical deposition conditions T/T_m is low (~0.08), and the intrinsic stresses dominate over thermal stresses.

A variety of stress relief patterns such as telephone-cord, sinusoidal wave, sinusoidal shape with extra branches, and a strings-of-beads pattern have been reported in the literature. Gioia and Ortiz [102] have compiled most of these patterns in their paper on delamination of compressed thin films. Studies on carbon films [103–105, 109] have indicated intrinsic stress levels, which are normally compressive, in the range of −0.3 to −12.5 GPa, depending on the growth conditions. According to Gupta and Bhushan [109], the cathodic-arc carbon coatings that exhibit the highest hardness have high residual compressive stresses of about 12.5 GPa, whereas all other carbon coatings exhibit comparable residual compressive stresses of about 0.6–2 GPa. Som et al. [103] studied the delamination of CVD diamond films deposited on silicon and estimated that the biaxial compressive stress to initiate the film buckling was 1.19 GPa. From their optical microscopy study of diamond-like carbon film, Iyer et al. [104] reported a new stress relief pattern having a sinusoidal shape with two extra branches at every peak position. They found the internal compressive stress in the film to be on the order of 1 GPa and dependent on the film thickness. Peng et al. [105] reported that the intrinsic stress on diamond-like carbon film varied from 0.3 to 2.0 GPa.

The studies to date have been mainly using experimental techniques and continuum mechanics. Although Pailthorpe et al. [110], McKenzie et al. [111], and Rosenblum et al. [112] used atomistic simulation in their studies on coating, they concentrated mainly on the energy diffusion, the structure of films, and the mismatch-induced residual thermal stresses in film/substrate systems. Because silicon monocrystal is anisotropic in its physical and mechanical properties, it is of interest to investigate the behavior of coating on different crystal orientations of silicon.

In this section, we will present the study carried out by Mylvaganam and Zhang [113], who investigated the buckling patterns of diamond carbon films deposited on different orientations of monocrystalline silicon under different residual stresses, using molecular dynamics analysis.

7.2. Theory and Simulation Method

Bombardment of diamond-like carbon films by energetic species tends to cause high compressive stresses, and as a result the film buckles and delaminates. Thornton [114] pictured

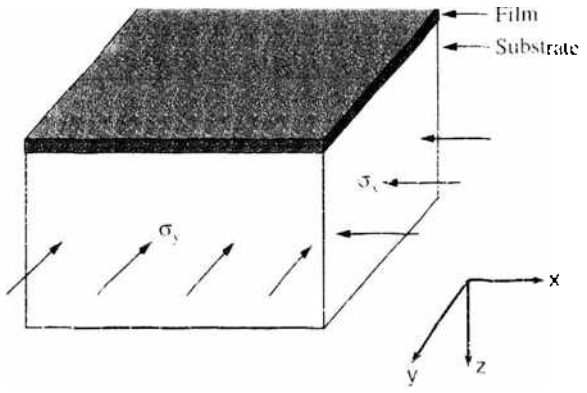


Figure 65. The simulation model for thin film buckling analysis. Reprinted with permission from [113], K. Mylvaganam, L. C. Zhang, *Thin Solid Films* 425, 145 (2003), © 2003, Elsevier Ltd.

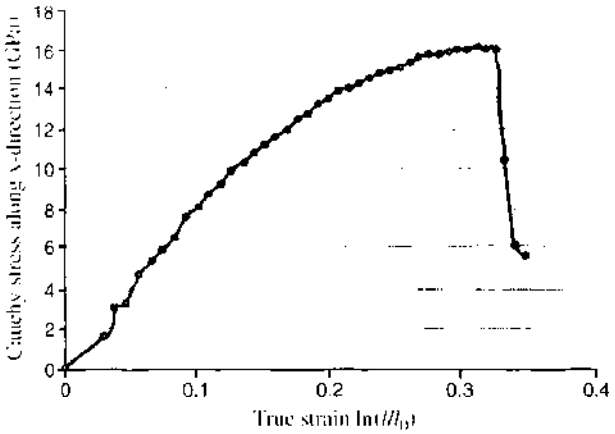


Figure 66. True stress–strain relationship of silicon under a uniaxial tension.

this process as follows: first the incident atoms transfer kinetic energy to the substrate and become loosely bonded "ad atoms." They then diffuse over the surface, exchanging energy with the substrate until they are either desorbed or trapped at low-energy sites of the substrate. Thus the initial kinetic energy transfer to the substrate can induce a tensile stress on the substrate. When the atoms diffuse over the surface, a compressive stress can be induced back to the coated atoms.

The above understanding of the residual stress generation enables us to simulate the thin-film deformation in a convenient way while capturing the nature of the residual stress generation. We apply the tensile stresses on the substrate before placing the coating and then release the stress with the coating on, so that the film will undergo compressions, as experienced in experiment.

We focus on the deformation of a diamond carbon coating on silicon (100) and (110) surfaces under various stress levels with different ratios of compressive stresses in two perpendicular directions, x and y, as shown in Fig. 65. A piece of diamond cubic silicon (100) with the control volume of $10.3 \times 10.3 \times 3.8$ nm (i.e., $19 \times 19 \times 7$ unit cells), containing 21516 atoms, was used as a substrate. The outermost layer of the substrate atoms with the exception of the top (100) surface was thermostated to 300 K by periodic rescaling of its atomic velocities. The tensile stresses were applied by forced atomic displacements at the boundary of the control volume in the following distinct ways to uncover the effect of stress variation in the film plane: uniaxially (along the x-direction), that is, the stress externally applied in y-direction would be zero, and hence the stress ratio is $\sigma_y/\sigma_x = 0$; and biaxially with equal displacements along both x- and y-directions to give the stress ratio of $\sigma_y/\sigma_x \approx 1$. Three atomic layers of diamond coating with dimensions $9.987 \times 9.987 \times 0.1783$ nm, having 4817 atoms, was placed on the stressed substrate. The coating and substrate were relaxed together for 5 ps, and then the stresses on the substrate were released gradually. To investigate the effect of crystal orientation, the above procedure was repeated by placing the coating on Si (110) surface.

Similar to the many successful molecular dynamics simulation studies discussed in previous sections, the Tersoff potential [6–7] was used for Si–Si and C–C interactions, and the Morse

Figure 67. Sinusoidal buckling pattern of the top layer of the film in x-direction after releasing the uniaxial stress (top view; $\sigma_y/\sigma_x = 0$). Reprinted with permission from [113]. K. Mylvaganam, L. C. Zhang, *Thin Solid Films* 425, 145 (2003), © 2003, Elsevier Ltd.

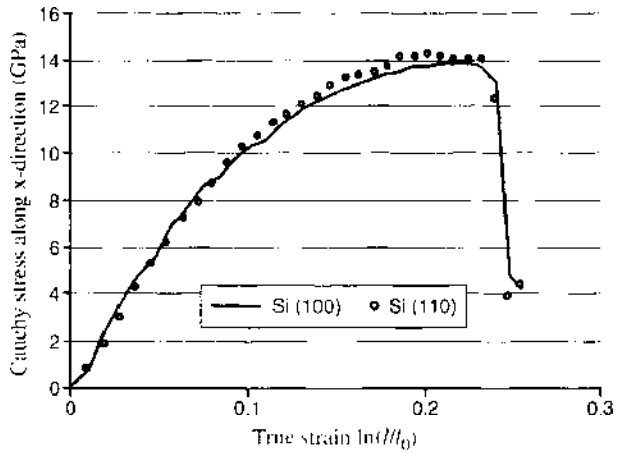


Figure 68. True stress–strain relationship of Si (100) and Si (110) under a biaxial tension.

potential was used for C–Si interactions. The true strain defined by $\ln(l/l_0)$ is calculated from the initial length, l_0 and final length, l, of the sample. The lateral normal stresses (σ_x and σ_y) were calculated by summing forces across the outermost yz- and xz-planes and dividing it by their current area to give rise to the Cauchy stress, or true stress. To facilitate the observation of the buckling deformation, we apply sufficiently large stresses, much higher than the critical buckling stress measured in relevant experiments [103–104], provided that they are within the elastic limit of silicon.

7.3. Uniaxial Stress

The true stress–strain relationship under uniaxial tension is shown in Fig. 66, which is clearly nonlinear. The loading and unloading curve coincides with when the stress was released from a point little before the maximum. This shows that the substrate is fully elastic and that it deforms plastically at a Cauchy stress (along x-direction) of 16 GPa under uniaxial tension.

The film placed on the substrate pretensioned to various levels showed that the film buckled when the substrate was pretensioned to about 2.5 GPa. However, to facilitate the observation of the buckling pattern, we placed the film on the substrate pretensioned to 12.5 GPa, a stress well below the maximum stress at which the substrate deforms plastically.

The film showed sinusoidal buckling and delaminated at the interface. The top layer of the film viewed from the top, at a tilted angle, is shown in Fig. 67. Sinusoidal wrinkles also formed in the x-direction.

7.4. Biaxial Stresses

The true stress–strain relationships of Si (100) and Si (110) under biaxial tension are shown in Fig. 68, which are, again, nonlinear. The stress release behavior shows that both are elastic

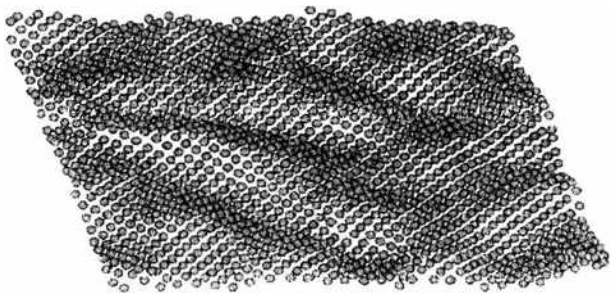


Figure 69. The telephone-cord-like buckling pattern of the top layer of the film after releasing the biaxial stresses on Si (100) ($\sigma_x/\sigma_y = 1$). Reprinted with permission from [113], K. Mylvaganam, L. C. Zhang, *Thin Solid Films* 425, 145 (2003). © 2003, Elsevier Ltd.

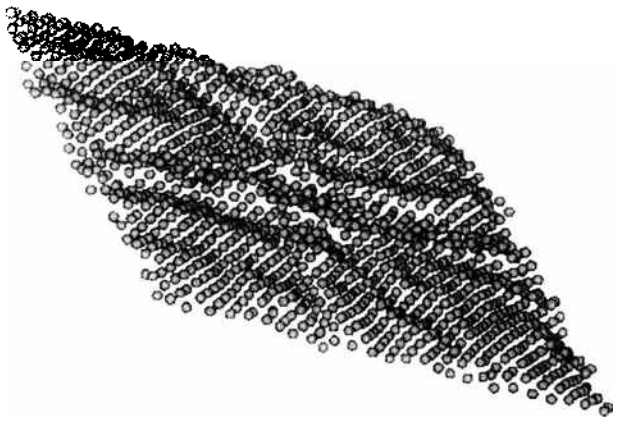


Figure 70. The buckling pattern of the top layer of the film after releasing the biaxial stresses on Si (110) ($\sigma_y/\sigma_x = 1$).

until they reach the maximum Cauchy stress of 13.9 Gpa in the case of Si (100) and of 14.25 GPa in the case of Si (110).

The film placed on the Si (100) substrate showed that the wrinkles formed when the substrate was pretensioned to a stress value of as low as 1.04 GPa. Again, to observe the buckling pattern, we placed the film on the substrate pretensioned to a much higher stress of 13.4 GPa in both x- and y- directions, but still within the elastic limit.

On releasing the stress, the film buckled and delaminated. The top-most layer of the film is shown in Fig. 69. Now it is interesting to note that the size of the wrinkles becomes larger than those in the uniaxial stressing, and the overall buckling pattern approaches a telephone-cord-like structure, as observed in many experiments [102].

To see the anisotropic effect on the buckling pattern, we placed the film on the Si (110) substrate pretensioned to 13.5 GPa, which is similar to the value used for Si (100). On releasing the stress, we find that the film buckled, but with very little delamination. In addition, we note that the wrinkle sizes were not uniform and that the buckling pattern varied significantly. The top most layer of the film is shown in Fig. 70.

The above investigation leads to the following major understandings about the mechanisms of the buckling pattern formation of diamond thin films. The buckling pattern varies with the orientation of the crystal and the residual stress ratio, σ_y/σ_x. Under biaxial stress, when both the film and substrate have the same crystal orientation, a telephone-cord-like buckling pattern will appear.

8. CONCLUDING REMARKS

Through a comprehensive dicussion on many examples of mechanical systems under various loading conditions, we have demonstrated in this chapter that molecular dynamics simulation offers an effective way to characterize the deformation of materials at the atomistic and nanometric scales. In some cases, none of the other methods, such as quantum mechanics and experimentation, can play the same role.

In the example of the nanoindentation of monocrystalline silicon, we have revealed the complex phase transformation mehanisms in relation to stress variations, investigated the environmental effects such as those of oxygen and water, and compared with experiment. In studying the contact sliding systems of diamond–copper and diamond–silicon, we have explored the nanofriction and nanowear mechanisms and have concluded that there exists a scale effect of contact size on nanofriction transition. We have also discussed in detail the modeling and deformation of carbon nanotubes, monocrystalline copper nanowhiskers, and diamond nanofilms, and we have demonstrated their deformation characteristics, shear banding and necking processes, and buckling patterns.

We have particularly emphasized, from the very beginning of the chapter throughout all the examples, that to carry out a reliable molecular dynamics analysis, it is essential to look into every step of a simulation, such as the establishment of the initial model, the relaxation,

the temperature conversion, the control volume size, the potential function, the thermostat technique, the time step, and the stress analysis. Any inappropriate setting of the fundamentals can bring about misleading results.

ACKNOWLEDGMENTS

The author appreciates the contributions of his colleagues, collaborators, and students over the years, particularly Prof K. L. Johnson (Cambridge), H. Tanaka (Osaka), T. X. Yu (HKUST), M. V. Swain, I. Zarudi, W. C. D. Cheong, T. Vodenitcharova, K. Mylvaganam, and C. Y. Tang (University of Sydney). The research presented in this chapter has been supported by many research grants, including the ARC (Large, IREX, Fellowship, Discovery) and the University of Sydney Research Grants.

REFERENCES

1. J. M. Haile, "Molecular Dynamics Simulation," Wiley, New York, 1992.
2. L. C. Zhang and H. Tanaka, *Wear* 211, 44 (1997).
3. L. C. Zhang and H. Tanaka, *JSME Int. J.* A42, 546 (1999).
4. L. C. Zhang and H. Tanaka, *Tribol. Int.* 31, 425 (1998).
5. L. C. Zhang, D. Cheong, and H. Tanaka, in "Abrasive Technology: Current Development and Applications" (J. Wang, W. Scott, L. C. Zhang, Eds.), p. 407, World Scientific, Singapore, 1999.
6. J. Tersoff, *Phys. Rev. Lett.* 56, 632 (1986).
7. J. Tersoff, *Phys. Rev.* B39, 5566 (1989).
8. L. C. Zhang and H. Tanaka, in "Advances in Abrasive Technology," (L. C. Zhang and N. Yasunaga, Eds.), p. 43, World Scientific, Singapore, 1997.
9. C. Kittel, "Introduction to Solid State Physics," 7th Edn, Wiley, New York, 1996.
10. L. C. Zhang, "Solid Mechanics for Engineers," Palgrave, Malaysia, 2001.
11. J. E. Inglesfield, "Potentials in Metals, Computer Simulation of Solids," p. 115, Springer, New York, 1982.
12. L. C. Zhang, K. L. Johnson, and W. C. D. Cheong, *Tribol. Lett.* 10, 23 (2001).
13. K. L. Johnson, "Contact Mechanics," Cambridge Univ. Press, Cambridge, 1985.
14. D. L. Callahan and C. M. Jonanathan, *J. Mater. Res.* 7, 1614 (1992).
15. F. J. Bundy, *J. Phys. Chem. Solids* 41, 3809 (1964).
16. J. Z. Wittig, *Phys. Lett.* 17, 187 (1965).
17. J. Z. Wittig, *J. Z. Phys.* 195, 215 (1966).
18. T. I. Dyuzheva, S. S. Kabalkina, and V. P. Novichkov, *Zh. Eksp. Teor. Fiz.* 74, 1784 (1978).
19. M. C. Gupta and A. L. Ruoff, *J. Appl. Phys.* 51, 1072 (1980).
20. J. C. Jamieson, *Science* 139, 762 (1963).
21. A. Werner, J. A. Sanjurjo, and M. Cardona, *Solid State Commun.* 44, 155 (1982).
22. H. Olijnyk, S. K. Sikka, and W. B. Holzapfel, *Phys. Lett.* 103A (1984).
23. J. Z. Hu and I. L. Spain, *Solid State Commun.* 51, 263 (1984).
24. I. L. Spain, D. R. Black, L. D. Merkle, J. Z. Hu, and C. S. Menoni, High Temp.-High Pressures 16, 507 (1984).
25. J. S. Kasper and S. M. Richards, *Acta Crytallogr.* 77, 752 (1964).
26. B. Welber, C. K. Kim, M. Cardona, and S. Rodriguez, *Solid State Commun.* 17, 1021 (1975).
27. G. J. Piermarini, *Phys. Rev.* B12, 1171 (1975).
28. I. Zarudi and L. C. Zhang, *Tribol. Int.* 32, 701 (1999).
29. L. C. Zhang, H. Tanaka, and I. Zarudi, in "Proceedings of the 12th ASPE Annual Conference," p. 4n2, American Society of Precision Engineering, Raleigh, NC, 1997.
30. L. C. Zhang and I. Zarudi, *Wear* 225, 669 (1999).
31. E. R. Weppelman, J. S. Field, and M. V. Swain, *J. Mater. Res.* 8, 830 (1993).
32. K. L. Johnson, K. R. Kendall, and A. D. Roberts, *Proc. R. Soc. Lond. A* 324, 301 (1971).
33. W. C. D. Cheong and L. C. Zhang, *Nanotechnology* 11, 173 (2000).
34. J. Donohue, "The Structures of the Elements," p. 262, Wiley, New York, 1974.
35. I. V. Gridneva, Y. V. Milman, and V. I. Trefilov, *Phys. Status Solidi* (a)14, 177 (1972).
36. A. P. Gerk and D. Tabor, *Nature* 271, 732 (1978).
37. D. R. Clarke, M. C. Kroll, P. D. Kirchbner, R. F. Cook, and B. J. Hockey, *Phys. Rev. Lett.* 21, 2156 (1988).
38. G. M. Pharr, W. C. Oliver, and D. R. Clarke, *Scripta Metall.* 23, 1949 (1989).
39. G. M. Pharr, W. C. Oliver, and D. R. Clarke, *J. Elec. Mater.* 19, 881 (1990).
40. M. C. Gupta and A. L. Ruoff, *J. Appl. Phys.* 51, 1072 (1980).
41. S. Minomura, "Localization and Metal-Insulator Transitions," p. 63, Plenum Press, New York, 1985.
42. W. C. D. Cheong and L. C. Zhang, *J. Mater. Sci. Lett.* 19, 439 (2000).
43. D. Ge, V. Domnich, and Y. Gogotsi, *J. Appl. Phys.* 93, 2418 (2003).
44. U. Durig and A. Stalder, "Physics of Sliding Friction," Kluwer, Dordrecht, 1996.
45. T. H. Courtney, "Mechanical Behaviour of Materials," McGraw-Hill, Singapore, 1990.
46. K. Kato, T. Uda, and K. Terakura, *Phys. Rev. Lett.* 80, 2000 (1998).

47. H. Bu and J. W. Rabalais, *Surface Sci.* 301, 285 (1994).
48. A. Pasquarello, M. S. Hybertsen, and R. Car, *Nature* 396, 58 (1998).
49. H. Ikeda, K. Hotta, T. Yamada, S. Zaima, H. Iwano, and Y. Yesuda, *J. Appl. Phys.* 77, 5125 (1995).
50. T. Uchiyama, T. Uda, and K. Terakura, *Surface Sci.* 433, 896 (1999).
51. Y. Miyamoto, A. Oshiyama, and A. Ishitani, *Solid State Commun.* 74, 343 (1990).
52. K. Mylvaganam and L. C. Zhang, *Nanotechnology* 13, 623 (2002).
53. I. Zarudi and L. C. Zhang, *J. Mater. Processing Technol.* 84, 149 (1998).
54. I. Zarudi and L. C. Zhang, *J. Mater. Sci. Lett.* 15, 586 (1996).
55. T. Ono, *J. Elec. Soc.* 146, 3807 (1999).
56. I. Zarudi, L. C. Zhang, and M. Swain, *Key Eng. Mater.* 233, 609 (2003).
57. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* 79, 926 (1983).
58. D. C. Rapaport, "The Art of Molecular Dynamics Simulation." Cambridge Univ. Press, Cambridge, 1995.
59. A. R. Leach, "Molecular Modelling: Principles and Applications," 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey, 2001.
60. M. P. Allen and D. J. Tildesley, "Computer Simulation of Liquids." Oxford University Press, New York, 1987.
61. S. Yang and H. Kim, *Bull. Kor. Chem. Soc.* 16, 1028 (1995).
62. C. M. Mate, G. M. McClelland, R. Erlandsson, and S. Chiang, *Phys. Rev. Lett.* 59, 1942 (1987).
63. R. Kaneko, K. Nonaka, and K. Yasuda, *J. Vac. Sci. Tech.* A6, 291 (1988).
64. P. Bak, *Res. Prog. Phys.* 45, 587 (1982).
65. W. Zhong and D. Tomanek, *Phy. Rev. Lett.* 64, 3054 (1990).
66. G. M. McClelland, in "Adhesion and Friction," (H. J. Kreuzer and M. Grunze, Eds.), p. 1. Springer, Berlin, 1989.
67. J. B. Sokoloff, *Surface Sci.* 144, 267 (1984).
68. U. Landman, W. D. Luedtke, and M. W. Ribarsky, *J. Vac. Sci. Tech. A* 47, 2829 (1989).
69. J. Belak and I. F. Stowers, in "Fundamentals of Friction: Macroscopic and Microscopic Processes," (I. L. Singer and H. M. Polloek, Eds.), p. 511. Kluwer Academic, Dordrecht, 1992.
70. A. M. Homola, *Wear* 136, 65 (1990).
71. F. P. Bowden and D. Tabor, "Friction and Lubrication." Methuen, London, 1967.
72. W. C. D. Cheong, L. C. Zhang, and H. Tanaka, *Key Eng. Mater.* 196, 31 (2001).
73. H. Tanaka and L. C. Zhang, in "Progress of Cutting and Grinding," (N. Narutaki, C. Dingchang, Y. Yamane, and A. Ochi, Eds.), p. 262. Japan Society for Precision Engineering, Osaka, 1996.
74. R. W. Carpick, N. Agrait, D. F. Ogletree, and M. Salmeron, *Langmuir* 12, 505 (1996).
75. M. A. Lantz, S. L. O'Shea, M. E. Welland, and K. L. Johnson, *Phys. Rev.* B55, 10776 (1997).
76. J. A Hurtado and K.-S. Kim, *Proc. R. Soc. Lond.* A455, 3363 (1999).
77. W. D. Callister, Jr., "Materials Science and Engineering–An Introduction," 3rd Edn., p. 767. Wiley, New York, 1995.
78. L. C. Zhang and I. Zarudi, *Int. J. Mech. Sci.* 43, 1985 (2001).
79. I. Zarudi, W. C. D. Cheong, J. Zou, and L. C. Zhang, *Nanotechnology* 15, 104 (2004).
80. W. C. D. Cheong and L. C. Zhang, *Key Eng. Mater.* 233, 603 (2003).
81. W. C. D. Cheong and L. C. Zhang, *Int. J. Mater. Product Technol.* 18, 398 (2003).
82. S. Iijima, *Nature* 354, 56 (1991).
83. S. Iijima and T. Ichihashi, *Nature* 363, 603 (1993).
84. D. S. Bethune, C. H. Klang, M. S. Devries, G. Gorman, R. Savoy, J. Vazquez, R. Beyers, *Nature* 363, 605 (1993).
85. T. W. Ebbesen and P. M. Ajayan, *Nature* 358, 220 (1992).
86. X. Zhao, M. Ohkohchi, M. Wang, S. Iijima, T. Ichihashi, and Y. Ando, *Carbon* 35, 775 (1997).
87. J. Qiu, Y. Li, Y. Wang, T. Wang, Z. Zhao, Y. Zhou, F. Li, and H. Cheng, *Carbon* 41, 2170 (2003).
88. H. M. Cheng, F. Li, G. Su, H. Y. Pan, L. L. He, X. Sun, and M. S. Dresselhaus, *Appl. Phys. Lett.* 72, 3282 (1998).
89. A. Thess, R. Lee, P. Nikalaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tomaneck, J. E. Fischer, and R. E. Smalley, *Science* 273, 483 (1996).
90. H. M. Cheng, F. Li, X. Sun, S. D. M. Brown, M. A. Pimenta, A. Marucci, G. Dresselhaus, and M. S. Dresselhaus, *Chem. Phys. Lett.* 289, 602 (1998).
91. P. G. Collins and P. Avouris, *Sci. Am.* 283, 62 (2000).
92. G. Ciccotti and W. G. Hoover, "Molecular-Dynamics Simulation of Statistical-Mechanical Systems." North-Holland, Amsterdam, The Netherlands, 1986.
93. A. Garg and S. B. Sinnott, *Phys. Rev.* B60, 13786 (1999).
94. B. Ni, S. B. Sinnott, P. T. Mikulski, and J. A. Harrison, *Phys. Rev. Lett.* 88, 205505 (2002).
95. L. G. Zhou and S. Q. Shi, *Comput. Mater. Sci.* 23, 166 (2002).
96. D. W. Brenner, *Phys. Rev.* B42, 9458 (1990).
97. D. W Brenner, O. A. Shenderova, J. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, *Condensed Matter* 14, 783 (2002).
98. T. Vodenitcharova and L. C. Zhang, *Phys. Rev.* B68, 165401 (2003).
99. B. I. Yakobson, M. P. Campbell, C. J. Brabec, and J. Bernholc, *Comput. Mater. Sci.* 8, 341 (1997).
100. A. G. Rinzler, J. H. Hafner, P. Nikolaev, L. Lou, S. G. Kim, D. Tomanek, P. Nordlander, D. T. Colbert, and R. E. Smalley, *Science* 269, 1550 (1995).

101. L. C. Zhang, H. Tanaka, and P. Gupta, in "Proceedings of the International Symposium on Processing, Designing and Properties of Advanced Engineering Materials" (ISAEM-97) (T. Kobayashi, M. Umemoto, and M. Morinaga, Eds.), p. 727. Japan Society for the Promotion of Science, Toyohashi, 1998.
102. G. Gioia and M. Ortiz, *Adv. Appl. Mechan.* 33, 120 (1997).
103. T. Som, S. Bhargava, M. Malhotra, H. D. Bist, V. N. Kulkarni, and S. Kumar, *Appl. Phys. Lett.* 72, 3014 (1998).
104. S. B. Iyer, K. S. Harshavardhan, and V. Kumar, *Thin Solid Films* 256, 94 (1995).
105. X. L. Peng and T. W. Clyne, *Thin Solid Films* 312, 207 (1998).
106. K. M. Crosby and R. M. Bradley, *Phys. Rev.* E59, R2542 (1999).
107. P. Peyla, *Phys. Rev.* E62, 1501 (2000).
108. J. A. Thornton and D. W. Hoffman, in "Proceedings of the 2nd Annual Technical Conference of the Society of Vacuum Coaters," Society of Vacuum Coaters, New Orleans, LA, 1979.
109. B. K. Gupta and B. Bhushan, *Thin Solid Films* 270, 391 (1995).
110. B. A. Pailthorpe, D. Mitchell, and N. S. Bordes, *Thin Solid Films* 332, 109 (1998).
111. D. R. McKenzie, D. Muller, and B. A. Pailthorpe, *Phys. Rev. Lett.* 67, 773 (1991).
112. I. Rosenblum, J. Adler, S. Brandon, and A. Hoffman, *Phy. Rev.* B62, 2920 (2000).
113. K. Mylvaganam and L. C. Zhang, *Thin Solid Films* 425, 145 (2003).
114. J. A. Thornton, *Annu. Rev. Mater. Sci.* 7, 239 (1977).

CHAPTER 9

Isomeric Fullerenes and Endofullerenes: Stability Computations on Promising Nanoscience Agents

Zdeněk Slanina

Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Japan, and Institute of Chemistry, Academia Sinica, Nankang, Taiwan, Republic of China

Kaoru Kobayashi, Shigeru Nagase

Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Japan

CONTENTS

1. Introduction 458
2. Outline of Quantum-Chemical and Statistical-Mechanical Calculations 460
3. Energetics and Thermodynamics of Carbon Clusters 463
   3.1. Basic Topology 463
   3.2. Isolated Pentagon Rule Energetics 464
   3.3. Stabilities of Clusters of Different Dimensions 465
4. Small Carbon Clusters 468
5. Topological Generations of Cages 470
6. Smaller Fullerenes 472
7. Higher Fullerenes 475
   7.1. C_{72} Fullerenes 475
   7.2. C_{74} Fullerenes 475
   7.3. C_{76} Isolated Pentagon Rule Set 478
   7.4. C_{78} Isolated Pentagon Rule Set 479
   7.5. C_{80} Isolated Pentagon Rule Set 480
   7.6. C_{82} Isolated Pentagon Rule Set 482
   7.7. C_{84} Isolated Pentagon Rule Set 482
   7.8. C_{86} Isolated Pentagon Rule Set 483

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (457–505)

7.9. C_{88} Isolated Pentagon Rule Set 483
7.10. C_{90} Isolated Pentagon Rule Set 485
7.11. C_{92}–C_{98} Isolated Pentagon Rule Sets 487
8. Endohedral Metallofullerenes 488
9. Fullerene Derivatives 493
10. Conclusions 496
References 496

1. INTRODUCTION

Cage compounds built exclusively from carbon atoms, now commonly known as fullerenes, have entered many research and application fields (for a recent survey, see e.g., Ref. [1]). Though fullerenes were for the first time observed [2], in the gas phase less than 20 years ago and prepared [3] in crystalline form less than 15 years ago, they have attracted an enormous research interest in chemistry, molecular physics, materials science, and medicine.

Not surprisingly, the new systems have also been evaluated as possible agents for nanotechnologies including molecular electronics. One concept for fullerene-based molecular memories was suggested by Gimzewski [4] and belongs to the category of memories operating between two isomeric configurations of a system (i.e., mobile isomeric scheme [5], which has been treated in molecular electronics from its early days). A fullerene cage with a metal encapsulated inside is considered [4] that possesses two possible location sites of the occluded atom inside the cage, the cage being immobilized on a support. As long as the two locations are distinguishable, they can in principle serve for coding the numbers 0 and 1 (Fig. 1). The switching between the positions and reading could be handled by scanning tunneling microscope. Storage density of the arrangement is estimated [4] to be approximately 1000 Gbit/mm^2.

Another concept of quantum computing aims [6] at a usage of spin states of a nitrogen atom encapsulated in the C_{60} cage, $N@C_{60}$. The state of the electron spin in an external magnetic field could be read by nuclear magnetic resonance (NMR) or electron spin resonance (ESR) pulse (Fig. 2). However, the lifetime of the states is relatively short even at very low temperatures, and single spin read-out is not possible. Hence, a larger identical assembly should be organized and read. Recently, yet another candidate species has been added [7] to the list; namely, $CH_2@C_{70}$ with an encapsulated simple polyatomic system. Molecular transistors based on fullerenes [8–10] represent another interesting option for molecular electronics. In addition to spheroidal fullerene cages, fullerene science also deals with elongated cylindrical bodies known as nanotubes, prepared [11] soon after mastering the fullerene synthesis. Nanotubes can also serve [12, 13] as molecular transistors. Nano-memory devices could also be based [14] on nanotubes filled with fullerenes, known as peapods.

Further development of this future nanotechnological application potential requires reliable knowledge of various aspects of the fullerenic and endohedral systems. Calculations can

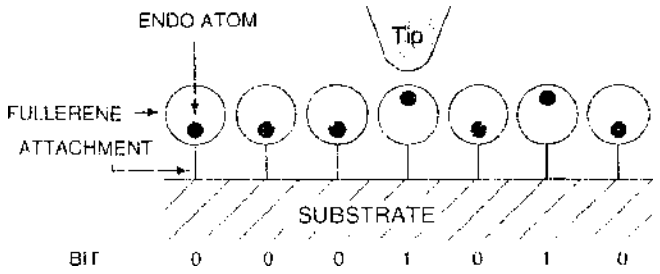


Figure 1. Concept of a molecular memory based on an endohedral fullerene. Each cage in the array contains a metal atom that can be switched between two positions. Reprinted with permission from [4]. J. K. Gimzewski, in "The Chemical Physics of Fullerenes 10 (and 5) Years Later" (W. Andreoni, Ed.), p. 117. Kluwer, Dordrecht, The Netherlands, 1996. © 1996, Kluwer.

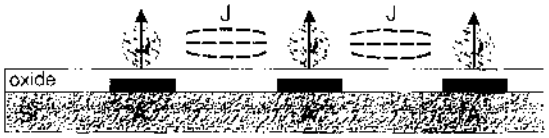


Figure 2. Concept of a quantum computer with information storage in the spin states of N atom encapsulated in C_{60}, $N@C_{60}$, exposed to an external magnetic field. The memory elements are addressable via A-gates. Reprinted with permission from [6], W. Harneit et al., in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish, et al., Eds.), p. 358. Electrochemical Society, Pennington, NJ, 2001. © 2001, Electrochemical Society.

help in a search for such promising candidates, evaluations of their interesting functional characteristics, and also optimizing conditions for their preparation and separation.

Fullerenes and metallofullerenes have represented objects of considerable interest for both experiment [15] as well as computations and theory [16, 17]. Their experimental studies have especially been based on liquid chromatography [18], ^{13}C NMR spectroscopy [19], and more recently also ^{3}He NMR spectroscopy [20]. Fullerenes are commonly defined as polyhedral cages containing only carbon atoms arranged into five- and six-membered rings. More generally, *quasi*-fullerenes have also been recognized [21], namely as cages containing not only the traditional five- and six-membered rings but also other types of cycles like (topological) squares and heptagons.

The very early history [22–26] of carbon clusters (i.e., before fullerenes) goes back to a mass-spectrometric observation by Hahn and his coworkers [27] in the 1940s (though their upper observation limit was only C_{15}). The research interest in the carbon clusters had continued through the 1950s and 1960s when the experiments [28–30] could already reach up to C_{33}. At about that time, simple computations had been applied [31–33] to the problem, in particular by Pitzer and Clementi [31], and later on by Hoffmann [33]. The highest computational level then applied was represented [34, 35] by the semiempirical MINDO/2 (modified INDO, where INDO means intermediate neglect of differential overlap) method. Nevertheless, according to current knowledge, the computed relative energies for small carbon clusters were essentially correct. On the other hand, much simpler Hückel-type calculations of aromatic systems, most common in the chemical computations of the 1960s, missed suggestions [36–38] of carbon-based polyhedrons. So, only in the 1970s were the first Hückel calculations of C_{60} performed [39, 40]; however, they failed to attract any experimental interest.

In contrast to the extremely modest tools available to computational chemistry from the 1950s till the beginning of the 1970s, quantum chemistry could act at much more advanced levels in the time of the C_{60} discovery by Curl, Kroto, Smalley, and others [2]. Consequently, computations have received widespread appreciation and indeed have consistently supported fullerene [16, 17] and metallofullerene [41] research at every important stage of development. Thus, the computations have supplied not only a solid support but even guidance to experiments. For example, this guiding role can be documented by the fact that the four IR bands computationally predicted [42–47] for C_{60} were indeed used [48] by Huffman and Krätschmer in their historical carbon-arc C_{60} synthesis [3]. Let us mention that vibrational analysis using semiempirical quantum-chemical methods for C_{60} and, in particular, for C_{70} still represented a demanding task in the mid 1980s. Fortunately enough, the fullerene research avalanche [49, 50] from the 1990s on has coincided with further considerable progress in computer technology. Consequently, such large-scale semiempirical calculations like those by Bakowies and Thiel [51], or even *ab initio* correlated treatments as those by Häser et al. [52] could be performed.

Computational chemistry has at present been acting as a real contributing partner in fullerene research. Hence, theory and experiment can be viewed as complementing tools in the exponentially growing [49, 50] field. Although there are survey papers [53–80] dealing with numerous aspects of fullerene science, quantum-chemical computations of fullerenes and metallofullerenes are reviewed relatively rarely [16, 17, 41, 81–90]. Moreover, in spite of the fact that fullerenes are prepared at temperatures that are among the highest ever used in chemical synthesis, statistical-mechanical portion of the computations is typically left apart so that the picture is essentially incomplete. Hence, this report supplies such a review of

the computations, covering both quantum-chemical and statistical-mechanical description in order to get a complete and consistent insight. The computations are linked to the existing observations and projected on the new, topical nanoscience aspects.

The time can be considered ripe for such a survey, as there are already well over 20 stable higher fullerenes C_n known from observations [19, 91–93] (n ranging from 60 to 96; higher refers to the fullerenes bigger than C_{60}). Elucidation of their structures has entirely been based on the so-called isolated pentagon rule (IPR) that claims [94, 95] that especially stable fullerenes should have all pentagons surrounded just by hexagons. Bonding patterns for such cages can be generated by various topological schemes [96–99], and in fact the cages can be quite numerous. Sometimes, like for $n = 76$, the IPR cage isomerism can be eliminated; that is, reduced down to just one species by energetics (in the C_{76} case with the help of Jahn-Teller effect [100]). Nevertheless, beyond $n = 76$, isomerism of the IPR structures should generally play a role in observations. Several such mixtures of empty fullerene isomers have indeed been computed and an agreement with experiments found: C_{76} (e.g., Refs. [100–110]), C_{78} (e.g., Refs. [111–114]), C_{80} (e.g., Refs. [115–119]), C_{82} (e.g., Refs. [120–125]), C_{84} (e.g., Refs. [126–132]), C_{86} (e.g., Refs. [91, 133–136]), C_{88} (e.g., Refs. [91, 134–136]), C_{90} (e.g., Refs. [91, 137, 138]), C_{92} (e.g., Ref. [139]), C_{94} (e.g., Ref. [140]), and C_{96} (e.g., Ref. [141]), whereas there is still only a computational prediction [142] for C_{98}. The combined quantum-chemical and statistical-mechanical computations have clearly shown that temperature effects can be important in understanding higher fullerenes. Moreover, a similar interplay has also been described for smaller fullerene systems [143–146] like C_{32} or C_{36}. In addition, various metallofullerenes can coexist [41] in several isomeric forms as well, for example $Ca@C_{72}$ (e.g., Refs. [147–150]), $Mg@C_{72}$ (e.g., Refs. [151, 152]), $Ca@C_{74}$ (e.g., Refs. [147, 149, 153, 154]), $Ca@C_{82}$ (e.g., Refs. [155–161]), $La@C_{82}$ (e.g., Refs. [162–166]), $Tm@C_{82}$ (e.g., Ref. [167]), $Ti_2@C_{80}$ (e.g., Ref. [168]), or $Ti_2@C_{82}$ (e.g., Refs. [169]). Metallofullerene isomerism is primarily based on encapsulation into different isomeric cages, however, different locations of the encapsulate in a given cage were also observed [170] (such species can actually be better suited for molecular-electronic applications). Endohedral fullerenes encapsulated in nanotubes also possess [171] an application potential as molecular memories. Still another interesting area is offered by isomerism of derivatives of fullerenes in the conventional chemical sense like $C_{60}H_{36}$ (e.g., Ref. [172]), $C_{60}F_{36}$ (e.g., Ref. [173, 174]), or $C_{60}F_{48}$ (e.g., Ref. [175]). The derivatized cages obviously also can host encapsulates [176].

In spite of the structural diversity, the physicochemical picture is uniform. Although the interisomeric separation energies themselves are important, they alone cannot predict the relative stabilities of the isomers at the very high temperatures where the fullerene synthesis happens. Owing to the unusually high temperatures, entropy contributions can even overcompensate the enthalpy terms. Hence, the enthalpy–entropy interplay represents an essential background on which various aspects of fullerenes and metallofullerenes can be treated and interpreted.

2. OUTLINE OF QUANTUM-CHEMICAL AND STATISTICAL-MECHANICAL CALCULATIONS

The current quantum-chemical calculations of fullerenes almost always deal with the optimized geometries, either at semiempirical, *ab initio* Hartree–Fock Self-Consistent-Field (HF SCF) or density functional theory (DFT) levels whereas *ab initio*–correlated treatments are very rare. The quantum-chemical calculations can, however, start from preliminary molecular-mechanical geometry optimization [177, 178]. Molecular mechanics can be misleading in special situations like Jahn-Teller cases that are relatively common with fullerenes. Semiempirical quantum-chemical geometry optimizations represent [179, 180] a useful treatment, either as such or as a prerequisite for following *ab initio* treatments [181]. The semiempirical optimizations supply a reliable guess of the second-derivative matrix (known also as force-constant matrix, or Hessian matrix) that strongly influences performance of any geometry optimization. The most common semiempirical methods [182–184] MNDO (modified neglect of differential overlap), AM1 (Austin model 1), and PM3 (parametric method 3) have recently been complemented with a new tool, the SAM1 method [185]

(semi-*ab-initio* model 1). The SAM1 method is implemented in newer AMPAC (Austin model package) program versions [186], whereas the other semiempirical methods are also available in the more common MOPAC (molecular orbital package) program package [187, 188]. Semiempirical methods and a full spectrum of *ab initio* procedures are implemented in the Gaussian program series [189, 190] such as G98 or G03 and also in the streamlined Spartan package [191].

At present, the geometry optimizations are routinely carried out with the analytically constructed potential energy gradient. This approach, in contrast to the older techniques based on numerical gradients, assures a relative speed and accuracy of the stationary-point search. In fact, the primary interest is focused on local energy minima, and the geometry search, unless specifically forced, rarely yields saddle points relevant in kinetic studies. The geometry optimizations are increasingly performed in the Cartesian coordinates (i.e., without any symmetry constraints). The symmetry constraints can be imposed, or in other words an exact symmetry produced, if internal coordinates are used. The explicit implementation of the symmetry is still important as it can scale down the computational time considerably. However, only some portion of the fullerene structures exhibits some higher symmetry, and nonsymmetric species are rather common. Although the semiempirical geometries are considered relatively reliable, the computed energetics is always to be checked with more sophisticated techniques (though frequently in the geometries optimized only at lower levels). There are several options available, such as *ab initio* HF SCF treatment in the standard 6-31G* or 6-311G* basis sets (HF/6-31G*, HF/6-311G*) or DFT approach with Becke's three-parameter functional with the nonlocal Lee–Yang–Parr correlation functional in the standard 6-31G* basis set (B3LYP/6-31G*). For larger systems like fullerenes, an issue of the SCF wavefunction stability can become important [192, 193]. Still, it frequently happens that the relative energetics from various quantum-chemical sources agree quite well and are mutually consistent in sets of fullerene isomers.

Although the optimization procedures mostly produce local energy minima, the types of the located stationary points must be checked accordingly. This step is essentially based on the harmonic vibrational analysis, that is, on the force-constant matrix, constructed numerically (numerical differentiation of the analytical gradient) or analytically. If the vibrational analysis is performed in the Cartesian coordinates, there are six zero (or near zero) frequencies always present that physically represent overall translations and rotations. The vibrational analysis reveals if there is any imaginary vibrational frequency—manifestation of a saddle point. If all the computed vibrational frequencies are real, we deal with a local energy minimum, and the structure should be considered in the thermodynamic stability reasoning. The computed vibrational frequencies can also be used in simulation of the vibrational spectra and in construction of the vibrational partition functions. For the spectra simulation, the vibrational frequencies are to be combined with the computed infrared (IR) intensities or Raman activities. Frequency scaling would be relevant for the spectral simulations though it is not important for high-temperature partition functions [194].

As already mentioned, the symmetries of the optimized structures can be an issue. In some relatively rare cases, they may not be properly recognized by the symmetry searchers built in the quantum-chemical programs, though the symmetries are exactly present in the structures. Other problem appears if the Cartesian coordinates are used, as the symmetry may not be recognized owing to numerical inaccuracies. The symmetries of the optimized structures can still be determined by a procedure [117] that represents a precision of the computed coordinates through a variable parameter, ϵ. The origin of the coordinate system is placed in the center of charge—the point is only candidate for possible center of symmetry. Then, candidates for C_2 axes are either lines connecting any two nuclei or perpendicular bisectors of the distance between any two nuclei of the same kind. Further on, C_n axes with $n > 2$, S_n and S_{2n} axes, and planes of symmetry are investigated in a similar way. The symmetry operations identified through this procedure have to create one of the known point groups of symmetry. For each symmetry operation considered, coordinates of the interrelated atoms before and after symmetry operation are checked with respect to ϵ. If the coordinates are identical within this accuracy ϵ (i.e., the largest difference is smaller than ϵ), a symmetry

element has been found at the accuracy level ϵ. This approach is more appropriate than the standard symmetry search without any flexible accuracy measure.

For a given carbon content (i.e., for a family of fullerene isomers), the geometry optimizations will end with a set of m isomeric structures. Their relative concentrations can be expressed as their mole fractions, w_i, using the isomeric partition functions q_i. In the terms of q_i and the ground-state energy changes $\Delta H^\circ_{0,i}$ the mole fractions are given [195–197]:

$$w_i = \frac{q_i \exp[-\Delta H^\circ_{0,i}/(RT)]}{\sum_{j=1}^{m} q_j \exp[-\Delta H^\circ_{0,j}/(RT)]} \tag{1}$$

where R stands for the gas constant and T for the absolute temperature. Equation (1) is an exact formula that can be directly derived [198] from the standard Gibbs energies of the isomers, supposing the conditions of the interisomeric thermodynamic equilibrium. Although the partition functions are to be constructed within the rigid-rotor and harmonic-oscillator (RRHO) approximation, Eq. (1) itself is essentially exact. In fact, there is only one presumption behind Eq. (1)—the presumption of noninteracting particles (or, more specifically, the ideal gas behavior) and the condition of the interisomeric thermodynamic equilibrium.

Strictly speaking, neither semiempirical nor nonempirical methods offer the $\Delta H^\circ_{0,i}$ terms as their primary computational outputs. The semiempirical quantum-chemical methods are parametrized for room temperature; that is, they produce the conventional heats of formation at room temperature $\Delta H^\circ_{f,298}$ (or the related separation or relative terms $\Delta H^\circ_{f,298,r}$). Thus, one has to convert the primary terms to the heats of formation at the absolute zero

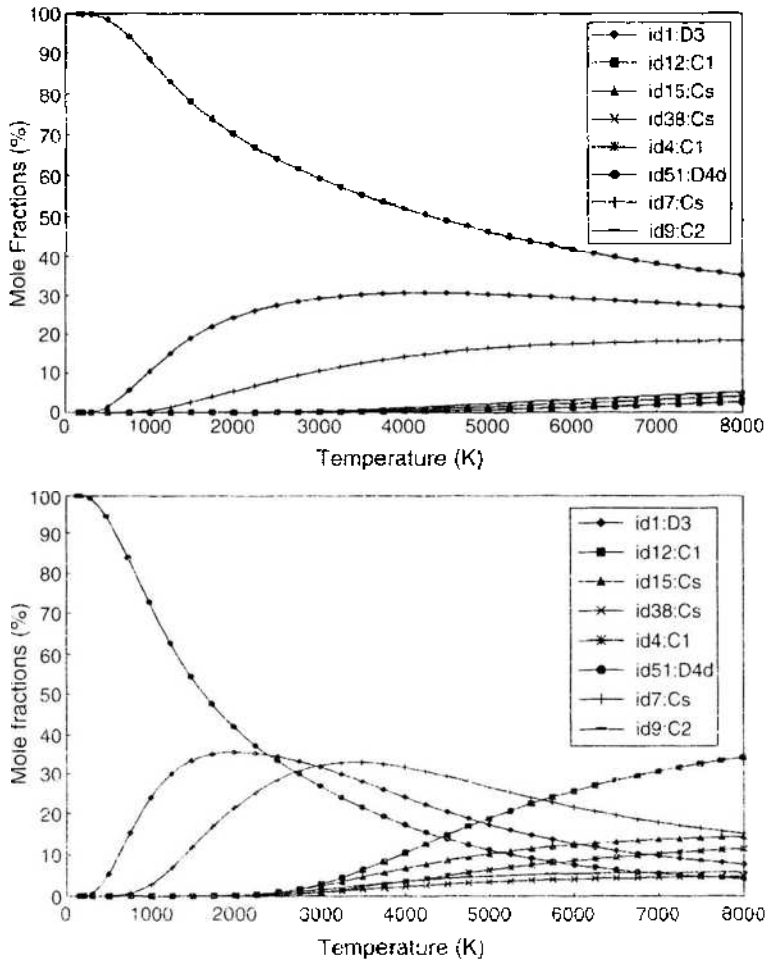


Figure 3. The computed [143] relative concentrations for C_{72}: top, simple Boltzmann factors, Eq. (2); bottom, the rigorous w_i terms from Eq. (1). Reprinted with permission from [143]: X. Zhao et al., *Fullerene Sci. Technol.* 8, 595 (2000). © 2000, Marcel Dekker.

temperature $\Delta H^{\circ}_{T,0}$; that is, the terms that truly appear in Eq. (1). The vibrational zero-point energy can be extracted, and in this case the relative potential energies ΔE_r result from the treatment. This level has in fact exactly the nature of the energy terms from *ab initio* computations; in other words, the relative potential energies ΔE_r from nonempirical treatments are to be corrected for the vibrational zero-point energies in order to serve in Eq. (1). For methodological reasons, we can also mention simple Boltzmann factors:

$$w'_i = \frac{\exp[-\Delta E^{\circ}_{r,i}/(RT)]}{\sum_{j=1}^{m} \exp[-\Delta E^{\circ}_{r,j}/(RT)]} \tag{2}$$

entirely based only on the potential-energy terms with no reference to entropy contributions. It is important to realize that the simple Boltzmann factors can never cross with a temperature change, and thus, they cannot really represent a complicated system—for an illustration, see Fig. 3.

There is still another interesting parameter that should be discussed. Chirality contributions are to be considered in Eq. (1). Clearly enough, there is no asymmetric carbon atom in the conventional sense in the fullerene cages (with sp^2 three-coordinated carbon atoms). Nevertheless, some of the structures are still chiral (i.e., they are not superimposable upon their mirror image). This structural dissymmetry can be directly read from the point group of symmetry, as presence of no reflection symmetry, that is, absence of rotation-reflection axes S_n, is required: only the C_n, D_n, T, O, and I groups obey the requirement. For an enantiomeric pair, its partition function q_i in Eq. (1) has to be doubled (if we assume the presence of both optical isomers, which can be considered well justified under fullerene synthesis conditions). It should be noted that thermodynamic packages in quantum-chemical programs entirely ignore the chirality contribution [199].

3. ENERGETICS AND THERMODYNAMICS OF CARBON CLUSTERS

3.1. Basic Topology

The conventional fullerenes are understood [21, 97] as cages built from three-coordinated (sp^2) carbon atoms arranged into two types of rings: pentagons and hexagons. As any other polyhedra, fullerenes have to obey Euler's polyhedral theorem [200, 201] (more precisely, the theorem for convex polyhedra):

$$V + F = E + 2 \tag{3}$$

where V denotes number of vertexes (atoms), F number of faces (rings), and E number of edges (bonds); for example, for C_{60}: $V = 60$, $F = 32$, $E = 90$. As only pentagons and hexagons are allowed, their numbers n_5 and n_6 give the total count of faces:

$$F = n_5 + n_6. \tag{4}$$

As all atoms are three-coordinated, it must hold for the number of edges:

$$E = \frac{3 * V}{2} \tag{5}$$

as each bond is accounted twice. One can also count the edges through rings:

$$E = \frac{5 * n_5 + 6 * n_6}{2}. \tag{6}$$

Combining Eqs. (5) and (6) yields:

$$V = \frac{5 * n_5 + 6 * n_6}{3} \tag{7}$$

Equation (5) can be placed in Euler's theorem:

$$V + F = \frac{3 * V}{2} + 2 \tag{8}$$

and combination with Eqs. (7) and (4) gives:

$$n_5 + n_6 = \frac{5 * n_5 + 6 * n_6}{6} + 2 \tag{9}$$

which can immediately be reduced to:

$$n_5 = 12 \tag{10}$$

In other words, in any conventional fullerene C_n, the number of five-membered rings must equal 12. The number of six-membered rings is variable, however, Eq. (7) gives a relationship to the number of carbon atoms:

$$n_6 = \frac{n - 20}{2} \tag{11}$$

According to Eq. (11), the smallest possible fullerene has the stoichiometry C_{20} (12 pentagons, no hexagons). Among all possible fullerenes, a special position have the IPR cages [94, 95] (that have all the pentagons surrounded just by hexagons), as the arrangements should be especially stable. In other words, there are no pentagon–pentagon junctions in the IPR fullerenes. As there are always 12 pentagons, then the smallest cage that could be of the IPR type should have $12 * 5$ carbon atoms. Indeed, 60 carbon atoms can create an IPR cage (which is geometrically truncated icosahedron), and the structure is the well-known buckminsterfullerene C_{60}.

3.2. Isolated Pentagon Rule Energetics

The very first MINDO/2 computations [34, 202] of one, two-, and three-dimensional small carbon clusters C_n pointed out a simple, smooth dependency of the relative heats of formation $\Delta H^\circ_{f,298}/n$ on the number of carbons n. Of course, that time virtually nothing was known over n of about 8, the computational upper limit of semiempirical quantum chemistry of those days. After two decades, the curve was extended into the fullerene domain, for example with *ab initio* HF approach in the standard STO-3G basis set (HF/STO-3G) computation [203, 204] or with the MNDO method [51, 205] or with simple semiempirical tight-binding computations [86, 206]. The qualitative picture is all the time the same. Although a shallow minimum could possibly be seen in some of the curves for C_{60}, basically they are still smooth decreasing dependencies (Fig. 4). Xu and Seuseria [206] could find such behavior even with carbon clusters up to C_{8640}.

There is a simple way to rationalize the finding. Let us limit our reasoning to the IPR fullerenes. Then, we deal with two types of bonds, frequently called the 5/6 (between pentagons and hexagons) and 6/6 (shared by two hexagons bonds). Let us suppose moreover that those two types of bonds can be represented [207] by some uniform dissociation energies, $H_{5/6}$ and $H_{6/6}$. In a general IPR fullerene C_n we have always 60 5/6 bonds, whereas the number of the 6/6 bonds is variable, $3n/2 - 60$. Now, we can readily write for the atomization heat of the considered C_n:

$$\Delta H_{at} = 60 \times H_{5/6} + \left(\frac{3n}{2} - 60\right) \times H_{6/6} \tag{12}$$

The atomization and formation heats for carbon aggregates are linked by the heat of vaporization of carbon, ΔH°_{vap}:

$$\frac{\Delta H^\circ_f}{n} = -\frac{\Delta H_{at}}{n} + \Delta H^\circ_{vap} \tag{13}$$

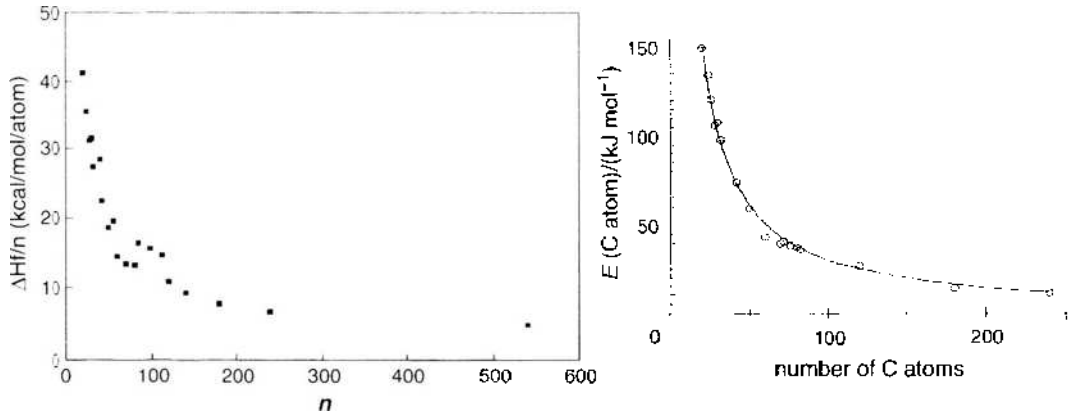


Figure 4. The MNDO heats of formation per atom [51, 205] (left) and *ab initio* HF/STO-3G relative energies [203, 204] (right) for selected fullerenes. Reprinted with permission from [203], R. F. Curl, *Phil. Trans. Roy. Soc. London A* 343, 19 (1993), © 1993, The Royal Society; and from [205], M.-L. Sun et al., *Chem. Phys. Lett.* 233, 279 (1995), © 1995, Elsevier.

or simply in the terms of the bond energies:

$$\frac{\Delta H_f^\circ}{n} = -\frac{3}{2}H_{6,6} + \Delta H_{vap}^\circ + \frac{60}{n} \times (H_{6/6} - H_{5/6}) \tag{14}$$

This functional dependency can formally be expressed by:

$$\frac{\Delta H_f^\circ}{n} = A + \frac{B}{n} \tag{15}$$

where B is a positive constant and A is a relatively small number. Hence, Eq. (15) is indeed a smoothly decreasing curve of the type shown in Fig. 4. From such a curve, one cannot read a particular stability of C_{60} or C_{70}. Just the opposite—it might suggest a straightforward rule that the stability would increase with dimension, which is, however, not observed. Clearly enough, the shape could have been predicted even before any computations, however, it could have been counter-productive as an indication that there had been nothing special in addition to diamond/graphite. Anyhow, in order to get a realistic picture, we have to consider entropy and/or kinetic effects instead, this being still under development.

3.3. Stabilities of Clusters of Different Dimensions

In its most general form, the problem could, in principle, be treated as a complex kinetic scheme described by a huge number of kinetic differential equations. The scheme can be solved if the values of the rate constants for all reaction channels are known. The equilibrium composition comes as the limiting case for infinite time. If we treat the problem from a thermodynamic point of view, we should realize that the conventional standard pressure of 1 atm is considerably different from the actual fullerene-synthesis conditions. Apparently, we should expect lower cluster pressures in the carbon-arc synthesis. The actual entropy and Gibbs free energy change with pressure as can be demonstrated simply on the C_{60} and C_{70} cases based on quantum chemical calculations. For example, the equilibrium constant $K_{60/70}$ for an interconversion between the two clusters, expressed in partial pressures p, offers a deeper insight into the problem [208–212]:

$$K_{60\text{-}70} = \frac{p_{70}^{6/7}}{p_{60}} = \frac{(1 - x_{60})^{6/7}}{x_{60}} P^{1/7} \tag{16}$$

where P stands for the total pressure of the two clusters and x_{60} is the mole fraction of buckminsterfullerene. At higher pressures, C_{70} is more populated than C_{60}, but at the conditions of a saturated carbon vapor (simulated simply as a mixture of seven clusters at present), the stability order is reversed in favor of C_{60} so that agreement with experiment is obtained.

The pressure effects can be studied [212] based on the MNDO, AM1, PM3, and SAM1 methods or using the available experimental data [213]. The results [212] also point out a need for a reliable evaluation of the absolute values of the heats of formation and other input information before the pressure-dependent scheme can be applied to other higher fullerenes. Still, it is already clear that the C_{60}/C_{70} population ratio is modulated by the carbon-vapor pressure.

More generally, one should consider all the clusters C_n (i.e., all the values for n) and for each n value all the isomers. At present, however, just seven clusters have been considered [212]: $n = 1, 2, 3, 4, 5, 60$, and 70. Expansion to any larger set of cluster is just a technical, not a conceptual, problem. Let us suppose that the thermodynamic equilibrium between the carbon gas phase and graphite is established:

$$nC(s) = C_n(g) \tag{17}$$

The partial chemical equilibria (17) are described by equilibrium constants:

$$K_{p,n} = p_n \tag{18}$$

where p_n denotes the partial pressure of C_n (it is actually expressed in atm; 1 atm = 101,325 Pa). If all the partial equilibria (17) are established, we reach the state of the saturated carbon vapor with a unique total saturated pressure p^* (for a given temperature T). In our model truncated set of carbon aggregates, it means:

$$p^* = p_1 + p_2 + p_3 + p_4 + p_5 + p_{60} + p_{70} \tag{19}$$

There is also an alternative and more general formulation of the problem. The alternative formulation is related to the monoatomic carbon cluster $C(g)$:

$$nC(g) = C_n(g) \tag{20}$$

and is described by another set of equilibrium constants, $K'_{p,n}$:

$$K'_{p,n} = \frac{p_n}{p_1^n} \tag{21}$$

The equilibrium constants $K'_{p,n}$ are related to the starting equilibrium constants $K_{p,n}$ by:

$$K'_{p,n} = \frac{K_{p,n}}{K_{p,1}^n} \tag{22}$$

Moreover, we can consider any total pressure p, not just the saturated p^* but also any undersaturated or supersaturated pressure. However, it remains true that:

$$p = p_1 + p_2 + p_3 + p_4 + p_5 + p_{60} + p_{70} \tag{23}$$

In order to simplify the scheme, we can introduce a variable x_1 with a meaning of the mole fraction of the monoatomic cluster $C(g)$. This substitution converts Eq. (23) into the following algebraic equation of the 70th order:

$$1 = x_1 + \sum_{i=2}^{70} K'_{p,i} x_1^i p^{i-1}, \qquad (i = 2-5, 60, 70) \tag{24}$$

One can solve [208–212] Eq. (24) for a selected temperature T and pressure p numerically with any required precision. Once the term x_1 is known, other mole fractions are simply given:

$$x_n = K'_{p,n} x_1^n p^{n-1} \tag{25}$$

The equilibrium constants $K'_{p,n}$ can be derived from quantum-chemical and statistical-mechanical calculations; the constant $K_{p,1}$ comes from experiment.

The scheme is pressure dependent—the pressure does not reflect the He pressure but the carbon-vapor pressure itself (never measured; after all, it is always covered by the He pressure). At pressures around the saturated carbon-vapor pressure, the scheme gives the correct order of the relative populations of C_{60} and C_{70}. If we increase the pressure to very high, unrealistic values like 1 atm, we can reverse the order. The scheme is quite sensitive to the heats of formation. Thus, before the treatment is applied on a large scale, one has to be able to produce accurate values of the total heats of formation (i.e., not only the relative energetics like for the isomeric problem). Moreover, some representative values of temperature and pressure for fullerene synthesis are to be selected. As there is a considerable temperature gradient in the chamber, if we select one pressure value, we shall in a substantial volume of the chamber actually deal with a supersaturated carbon vapor. Clearly enough, the isomeric stability problem is much better understood at present than the relative stabilities of nonisomeric carbon clusters, partly because the former has been been studied [198, 214–216] for a longer time.

There is still another interesting result of the computational studies [208–212]: temperature increase of the clustering degree under the saturation conditions. This somewhat surprising result can in fact be easily rationalized [217]. Whereas the equilibrium constants for cluster formation decrease with temperature, the saturated pressure increases. It is just the competition between these two terms that decides the final temperature behavior. The finding actually has a more general validity as it was also reported for other vapors [218–221].

Let us consider a substance A that exists in both condensed (liquid or solid) and gas phases, and those two phases are in thermodynamic gas–liquid or gas–solid equilibrium. In the gas phase clusters A_i of any dimension are present:

$$iA(g) = A_i(g), \qquad (i = 2, \ldots) \tag{26}$$

and their formation is described by equilibrium constants:

$$K_{p,i} = \frac{p_{A_i}}{p_A^i}, \qquad (i = 2, \ldots) \tag{27}$$

Temperature dependency of the oligomerization equilibrium constant $K_{p,i}$ is given by the van't Hoff equation:

$$\frac{d\ln K_{p,i}}{dT} = \frac{\Delta H_i^\circ}{RT^2}, \qquad (i = 2, \ldots) \tag{28}$$

where ΔH_i° stands for the (negative) standard change of enthalpy upon the respective i-merization.

The total pressure P will be the sum of all the partial pressures derived from Eq. (27):

$$P = p_A + \sum_{i=2}^{\infty} p_A^i K_{p,i} \tag{29}$$

We can again work with the monomeric molar fraction x_1:

$$1 = x_1 + \sum_{i=2}^{\infty} x_1^i P^{i-1} K_{p,i} \tag{30}$$

Generally speaking, this algebraic equation cannot be solved in a closed form. In fact, we do not need the algebraic solution—it is just enough for our purpose to express the temperature derivative dx_1/dT in an applicable form. Moreover, we are not interested in a general pressure but in the saturated pressure. Let us suppose that the saturated pressure simply obeys the Clausius–Clapeyron equation:

$$\frac{d\ln P}{dT} = \frac{\Delta H_{\text{vap}}}{RT^2} \tag{31}$$

where ΔH_{vap} symbolizes the (positive) enthalpy of vaporization. After some algebraic manipulations, one can express the temperature derivative dx_1/dT as follows:

$$-\frac{dx_1}{dT}\left(1+\sum_{i=2}^{\infty} i\xi_i\right)=\sum_{i=2}^{\infty} x_1\xi_i\frac{\Delta H_i^\circ+(i-1)\Delta H_{vap}}{RT^2} \tag{32}$$

where the ξ_i terms are given:

$$\xi_i = x_1^{i-1}P^{i-1}K_{p,i} \tag{33}$$

Equation (32) suggests a simple sufficient (though not necessary) condition for the temperature decrease of x_1 in the saturated vapor (in other words, for temperature increase of the clustering degree $1-x_1$). As long as $\Delta H_{vap} > |\Delta H_i^\circ|/(i-1)$ for every $i \geq 2$, the temperature derivative dx_1/dT is negative, and thus, the clustering degree increases with temperature in the saturated vapor. In all cases tested in computations [217], the simple rule derived from Eq. (32) was satisfied. This finding also gives still another perspective to the high temperatures in fullerene synthesis. They are needed not only in order to evaporate graphite, but also in order to get a sufficient clustering in the vapor phase.

4. SMALL CARBON CLUSTERS

Although the small carbon clusters had been computed [31–35] well before the fullerene discovery, their studies have certainly gained from the discovery. Their relative stabilities can be treated in a similar way as those of fullerenes, and thus, the small clusters represent an interesting model system. They have been treated by both theoretical [222–240] and experimental [241–264] techniques. The basic structural features of the systems were established by Raghavachari and Binkley [223]: (i) linear isomers exist for both even- and odd-numbered species; however, the even species being in the triplet electronic state, and (ii) planar cyclic isomers exist for the even species. Their computations were performed at the HF/6-31G* level only, and thus, a further check with a correlated treatment had been necessary.

On the experiment side, ion-chromatography observations by von Helden et al. [261, 263, 264] suggested that most carbon clusters had more than one stable form. Figure 5 shows [264] that starting from about $n = 7$, cyclic rings exist in addition to the linear forms (though there is a charge dependency). The first observed cyclic case [264] (i.e., C_7) was indeed at the MP2/6-31G* level computed [236] only about 2 kcal/mol above the linear form (though, 17 kcal/mole at the advanced MP4/6-31G* level). On the other hand, the HF/6-31G* treatment does not produce such low-energy species. The cycles are not necessarily planar as a nonplanar structure was computed [266] for C_9 in the MP2/6-31G* approximation (6 kcal/mol above the linear form). Moreover, the C_8 cycle is also computed [237] nonplanar at the MP2/6-31G* level. The nonplanarity of C_8 allows for an inversion motion with an inversion barrier [237, 267] of 13 kcal/mol. However, the nonplanar C_8 cycle was not confirmed in the more sophisticated coupled-cluster computations of Martin and Taylor [240]. Hence, the MP2/6-31G* applicability to carbon clusters still needs additional critical tests.

In spite of the open question of their applicability, there are [268] the MP2/6-31G* calculations also for C_{10}, C_{11}, C_{12}, and C_{13}, with the cyclic form located 131, 48, 102, and 23 kcal/mol, respectively, below the linear isomer. The separation energy for C_{10} is, however, considerably higher than reported by Parasuk and Almlöf [235]. Density functional computations agree [269] with the MP2 results for C_{13} (so that the observations of Giesen et al. [270] can be rationalized). Qualitatively speaking [271], the available computations agree with the findings of ion chromatography [261, 263, 264]. C_{11} should be the first (or, at least first odd-numbered) species for which the cyclic structure becomes dominant. Switch from cyclic to polyhedral species is still not well-known but is expected [272] around $n = 45$.

The linear and cyclic (rhombic, bicyclic) C_4 isomers were computed several times. The cyclic versus linear difference varied considerably [53] from -16 to 1 kcal/mol. Usually, calculations placed the singlet cyclic form at the lower energy. It was suggested [53] to estimate the rhombus energy preference by about 5 kcal/mol. The unclear computational picture for C_4

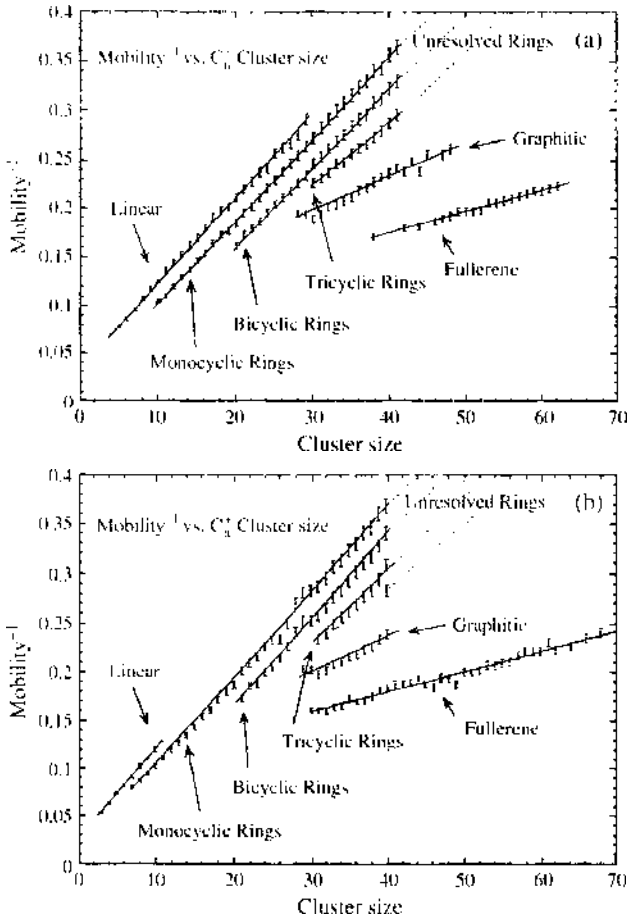


Figure 5. Ion chromatography of carbon clusters performed by Gotts et al. [264] with indication of several structural families. Reprinted with permission from [264], N. G. Gotts et al., *Int. J. Mass Spectr. Ion Process* 149/150, 217 (1995). © 1995, Elsevier.

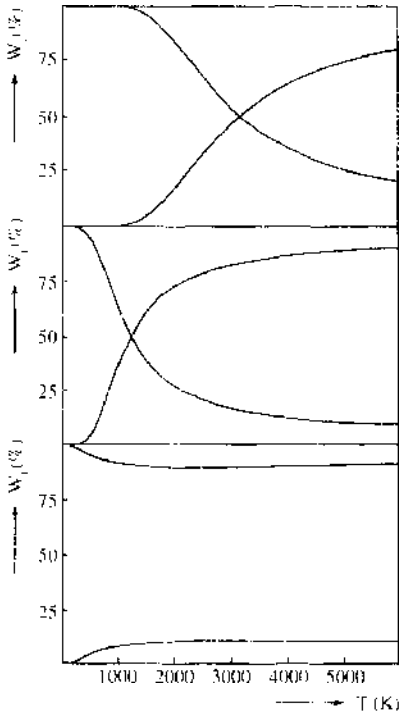


Figure 6. Relative concentrations [232, 278, 279] of the linear (the higher curve at high temperatures) and rhombic structures of C_4 for three separation energetics [53, 273, 280] (bottom: thermodynamic fit [273]). Reprinted with permission from [232], Z. Slanina et al., *Z. Phys. D* 19, 431 (1991). © 1991, Springer-Verlag.

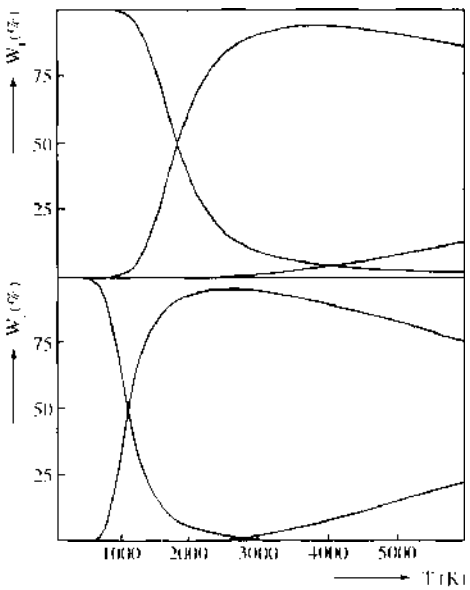


Figure 7. Relative concentrations [232, 214, 281] of the linear (the highest curve at high temperatures), cyclic (the lowest at high temperatures), and another linear structure of C_6 for two separation energetics [222]. Reprinted with permission from [232]. Z. Slanina et al., *Z. Phys. D* 19, 431 (1991). © 1991, Springer-Verlag.

prompted an approach [273] based on available [29, 274] experimental equilibrium constants for C_4 formation (two values for two different temperatures). Using the computed entropy, the cyclic versus linear difference was evaluated [273] as 2.4 kcal/mol (or 2.9 kcal/mol; i.e., the linear form was slightly lower). Incidentally, an extended computation of Nygren and Pettersson [275] placed the linear structure 3.6 kcal/mol below the cycle. Density functional theory may yield a numerical artifact [276] in this case (the rhombic structure higher by 20.6 kcal/mol). At the G1 and G2 levels of theory, the rhombus form is predicted [239] to be more stable by 3.4 and 5.1 kcal/mol, respectively, though the G2 value is probably too large. Still another advanced calculation by Watts et al. [277] finds the cyclic isomer to be more stable by about 1 kcal/mol.

The energy difference between the C_4 isomers is relatively small, however, the entropy contributions can still produce one dominant species at higher temperatures [245, 278, 279]. Figure 6 presents the temperature evolution of the relative concentrations of the C_4 isomers evaluated for three selected values of the separation energy [53, 273, 280]. For example, if the thermodynamic fit [273] is used, the highest relative concentration of the cycle is obtained at around 2700 K, however, only 5–10%. Even if the cycle represents the ground state, the linear isomer still dominates at high temperatures. This conclusion agrees with the experimental finding of only one (and not two) peaks in ion chromatography [261].

There is still another example of a small carbon-cluster system with an important role of entropy, viz. C_6. Raghavachari et al. found [222] an energy separation of -11 kcal/mol between the cyclic and linear C_6 forms, with the cycle representing the ground state. At nonzero temperatures, the picture is, however, considerably different, as shown [232] in Fig. 7. At a temperature of observation [29] (2500 K), the linear species is computed [214, 281] to form more than 95% of the equilibrium mixture. Martin and Taylor confirmed [240] the finding. This thermodynamic reasoning again explains why only one peak was recorded [261] in the ion chromatography of C_6. The system continues to attract a vigorous research interest [282, 283].

5. TOPOLOGICAL GENERATIONS OF CAGES

Enumerations and generations of fullerene topologies represent an essential step before any search on a potential hypersurface. The first systematic studies of the problem were carried out by Fowler et al. [96, 97, 284–302], though various topological aspects of fullerenes have been studied on a wide front [95, 98, 99, 303–331]. Exhaustive generation of all closed

cages built from pentagons and hexagons, and of the IPR cages among them, represents a particularly important task. As long as we deal only with pentagons and hexagons, the cage isomerism first appears [97] with C_{28} (however, the IPR structures start from C_{60}).

Enumerations of fullerene cages have developed together with other algebraic and combinatorial aspects of fullerenes, like the concept of Goldberg polyhedra applied [286] to icosahedral cages, leapfrog transformation [296], or the concept of topological duals [306]. The Stone–Wales transformation [304] represents another useful concept based on topological relationships. It was introduced as a process for rearrangements of the rings in fullerene cages (i.e., for fullerene isomerizations). Such a pairwise interchange of two pentagons and two hexagons is also called the pyracylene transformation [296]. It can be further generalized [332] through inclusion of other rings, different from pentagons and hexagons. It is visualized as a movement of two atoms in which two bonds are broken. It is not necessarily a convenient, feasible kinetic process—it is thermally (but not photochemically) forbidden. Nevertheless, it still represents a useful topological concept.

Recently, a way was suggested how to activate the kinetic process—its feasibility can be enabled [333–335] by a catalysis or even autocatalysis by free atoms present in the reaction mixture at very high temperatures. Hence, catalysis seems to be of crucial importance not only for production of nanotubes [11] but also of fullerenes themselves. However, the mechanism of fullerene synthesis is still virtually unknown. There are tremendous experimental difficulties in elucidation of the mechanism, and computational support is essential. There are some generally supposed steps like production of small linear, cyclic, and polycyclic species, and their combination into cages. Finally, multiple isomerizations should transform general cages (containing several types of rings) into a few most stable IPR structures. Although kinetics in electronically excited states could bring some reduction of the barriers, a catalytic action is also a very plausible step. The catalytic action is supposed through an influence of either elemental atoms or their small clusters, in particular carbon, nitrogen, oxygen, and hydrogen atoms. Graphite always contains small amounts of many chemical elements, and N_2 and O_2 also come as an impurity in the inert gas used as the medium in the fullerene synthesis. During the process of graphite evaporation in the electric arc, all the components are atomized. The free atoms can create intermediate complexes with the cages. In fact, even carbon atoms themselves could act this way [333]. However, the computations in particular favor [335] catalytic activity in the complexes containing nitrogen atoms.

Although the kinetic aspects of the fullerene synthesis are of tremendous importance, the Stone–Wales transformation can be a useful concept even if all the energetics is disregarded. Then, it works as formal topological structure generator [132]. This application represents another exhaustive or near-exhaustive generation of topologically possible cages and was recently applied [143], to the cage-isomerism problem of C_{32}. Moreover, the Stone–Wales transformations can give some kinetic insight even without considering kinetic barriers. They were for example applied to an analysis [296] of interconnection links for the C_{84} IPR structures. The analysis demonstrated that the set of transformations would decompose into two disjoint families.

A versatile technique for the enumerations of all fullerene cages and of the IPR cages was systematically developed by Fowler and Manolopoulos [96, 97, 308]. They introduced a concept of two-dimensional representation of fullerenes by their ring spiral, and the concept has been developed into a key tool in the fullerene enumerations. They noticed that fullerenes, at least below some dimension threshold, could be peeled like an orange—each face, after the first, borders its immediate predecessor, so that the rings come off the cage in a single continuous spiral. It is now known that the simple procedure cannot be applied to any cage—a counterexample is given [302] by a fullerene with 380 atoms. It is the smallest tetrahedral fullerene without a spiral (T symmetry). It is not clear, however, whether this is the smallest possible unspirable fullerene of any symmetry. However, the spiral algorithm seems to fail only for such n values that are not relevant to our current considerations.

Within the spiral algorithm the bonding topology is reconstructed from the sequence of rings in the spiral. Thus, all possible C_n fullerene graphs can be generated by considering all

the ways in which 12 pentagons and $(\frac{n}{2} - 10)$ hexagons can be combined into a spiral:

$$S_n = \frac{(\frac{n}{2} + 2)!}{12! \times (\frac{n}{2} - 10)!} \tag{34}$$

The number of the spirals, S_n, is higher than the resulting number of related fullerenes and it is subjected to a selection and reduction procedure. The uniqueness test is based on the eigenvalues of the adjacency matrix (hence, we deal with the spectrally distinct cages).

The total number of spectrally distinct C_n fullerene isomers increases rapidly with the carbon content n. In particular, it was found [308–313] that the number of spectrally distinct C_{60} isomers is 1812. For example, for $n = 50, 40, 30$, and 20, the computed [96, 97] numbers of all isomers are 271, 40, 3, and 1, respectively. Let us stress that these counts consider only pentagon/hexagon rings, with no other cycles allowed, and a chiral pair is considered essentially as one isomer. Odd-numbered cages are not usually considered in the enumerations. They are also known [314, 336, 337] but they cannot simply be built from three-coordinated atoms only. For example, if we consider just one two-coordinated carbon in the odd-numbered cages, and pentagons/hexagons only, the number of pentagons is reduced [314] to 10. If we allow for two tetra-coordinated and one two-coordinated carbons, the number of pentagons is 14, and so forth.

Among all the possible isomers, the IPR structures are supposed to be a particularly important species (i.e., high-stability candidates). The requirement of isolated pentagons can readily be implemented into the spiral algorithm. As already mentioned, there cannot be an IPR structure for $n < 60$. This fact is obvious as there are always 12 pentagons, and 12 isolated pentagons represent $12 \times 5 = 60$ carbon atoms. The second smallest IPR fullerene comes for $n = 70$. It is still a unique structure like those for $n = 72$ and 74. For any higher carbon atom, we already meet IPR isomers; for example [96, 97], for $n = 76, 78, 80, 82, 84, 86, 88$, and 90, there are 2, 5, 7, 9, 24, 19, 35, and 46 IPR structures, respectively.

Other enumeration algorithms have been developed, for example, the net-drawing method [98, 99], ring-by-ring approach of Liu et al. [309, 310], circumscribing technique by Dias [320], or topological treatment based on the Stone–Wales transformations [132]. There are also interesting applications of Pólya's enumeration theorem, known as the most general result in chemical enumerations [198, 233]. Pólya's theorem was for example applied [321, 324, 330, 331] to enumerations of isomers for various substituted fullerenes.

6. SMALLER FULLERENES

Although fullerene science is focused at $n = 60$ and above, there has been a vigorous interest in a lower bound for observable fullerenes C_n. Von Helden et al. [261, 272] reported fullerene-like structures at and above C_{30} using gas-phase ion chromatography technique. The fraction of fullerene isomers was observed close to 5% for C_{32}, about 1% for C_{30}, and virtually no cages below $n = 30$. Moreover, synthetic routes even toward the smallest fullerene have indeed been developed [339–342] and the system computed (e.g., Refs. [216, 343–346]). Similarly, C_{36} fullerene was also isolated as a solid material [347] and computed [144, 145, 348, 349]. Narrow nanotubes related to C_{36} and even to C_{20} are also known [350–353]. Cage isomerism was computed [143–145] for C_{32} and C_{36}.

There are six conventional C_{32} cages built from pentagons and hexagons. They were computed by Murry et al. [348], and a D_3 structure was pointed out as the system ground state. No IPR structures are of course possible for C_{32}. Hence, pentagon/pentagon junctions will always be present. However, one can reduce the number of pentagon–pentagon junctions by introduction of four-membered and/or seven-membered rings. Then, Euler's network closure requirement has the following form:

$$2n_4 + n_5 - n_7 = 12, \tag{35}$$

where n_i denotes the number of rings with i vertices. For example, this equation says that a removal of two pentagons will increase the number of four-membered rings by one. An exhaustive generation of all C_{32} cages built from four-, five-, six-, and/or seven-membered

rings was performed [143] using the topological treatment based on the generalized Stone–Wales transformations [132]. In order to reduce the amount of necessary quantum-chemical computations, an additional limitation, viz. $2n_4 + n_7 \leq 4$, was imposed. With this limitation, the topological search generates 199 cage structures for C_{32}. Six of them are the conventional fullerenes and 193 *quasi*-fullerenes.

All the 199 structures were computed at the AM1, PM3, and SAM1 levels [143] and partly also at *ab initio* levels. All the computations point out a structure with two four-membered rings (and no heptagons) as the ground state in the 199-membered set (the second lowest structure is, however, a conventional fullerene). Moreover, if we consider five structures lowest in energy (Fig. 8), only two of them are conventional fullerenes, the other cages contain at least one square, though no heptagon yet. The first structure with a heptagon (Fig. 8) appears only about 230 kJ/mol above the ground state. The first two structures with a nonzero n_7 contain one square and one heptagon each. The third lowest structure with a nonzero n_7 contains no square. The three heptagon-containing structures are lower in energy than the four conventional fullerenes not included among the lowest five cages.

In order to gain a deeper insight into the relative-stability problem, temperature development of the equilibrium mole fractions in the C_{32} system was computed [143], too (Fig. 3). If we completely ignore the entropic terms [the simple Boltzmann factors from Eq. (2)], the resulting temperature evolution is rather simple and uniform. However, the entropy contribution can reverse the stability order in some cases, and this indeed happens in the C_{32} isomeric set. The structure lowest in energy (coded 51; a D_{4d} symmetry) is the most populated species only till a temperature of about 2500 K. Beyond that point it is surpassed by another structure, the conventional fullerene labeled by 1. However, at still higher temperatures a third species becomes a leading term, the *quasi*-fullerene structure coded by 7. The equimolarity points are located rather high in temperature, however, in a still relevant temperature region. The computational results suggest that the C_{32} system, if isolated, should definitely represent an interesting structural case.

For C_{36}, there are already 15 conventional fullerene cages, of course none of them being an IPR structure. However, one can again reduce the number of pentagon–pentagon junctions by introduction of four-membered and/or seven-membered rings. The topological search [144], based on the generalized Stone–Wales rearrangements [132], was limited to conventional fullerenes and to *quasi*-fullerenes with one or two squares (no heptagon), one heptagon (no square), or one square and one heptagon. The topological search supplied

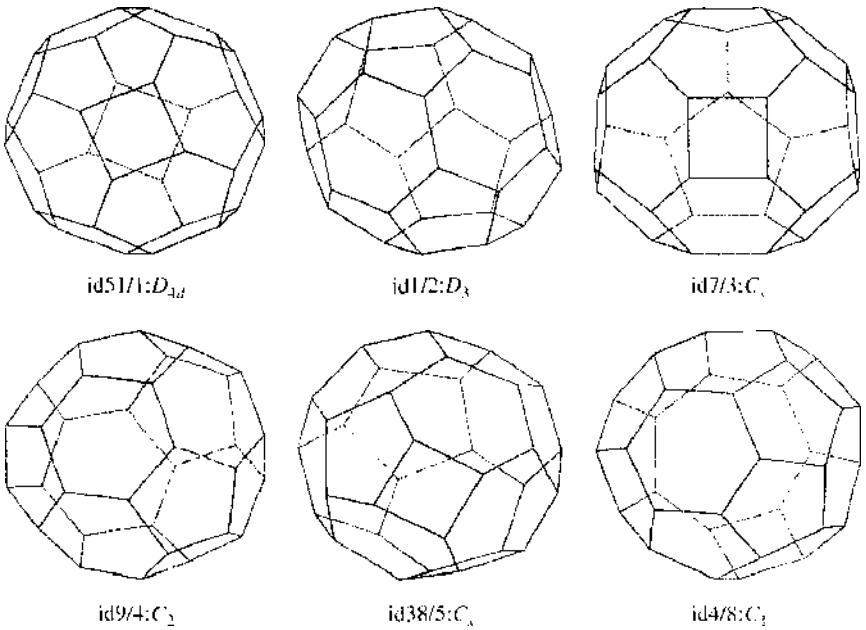


Figure 8. Six computed [143] important structures of C_{32}. Reprinted with permission from [143], X. Zhao et al., *Fullerene Sci. Technol.* 8, 595 (2000). © 2000, Marcel Dekker.

598 cages. At present, the completeness of the search is not completely understood yet, however, it is certainly a comprehensive enough search for the first approach. All 598 structures were optimized at the SAM1 level, and the 10 lowest structures selected (Fig. 9); among those 10 cages, 6 are the conventional fullerenes. We can characterize the cages by a ring index, a vector containing the counts n_i of the rings considered: (n_4, n_5, n_6, n_7). Then, three low-energy *quasi*-fullerenes are of the (1, 10, 9, 0) type (one square) and one is of the (1, 11, 7, 1) type. For the three lowest structures, their separation energetics was further checked by the HF/4-31G and B3LYP/6-31G* computations [144, 145] in the optimized SAM1 geometries. It turns out that the SAM1 and B3LYP/6-31G* data agree very well, whereas the HF/4-31G (free of any correlation energy) values seem to be somewhat underestimated.

In contrast to the above C_{32} system with the ground state containing two squares, in C_{36} the ground state is already represented by a conventional fullerene. The ground state exhibits a D_{2d} symmetry. The second lowest species is also a conventional fullerene, with a C_{2v} symmetry and 36 kJ/mol higher on the SAM1 scale. The highest member of the

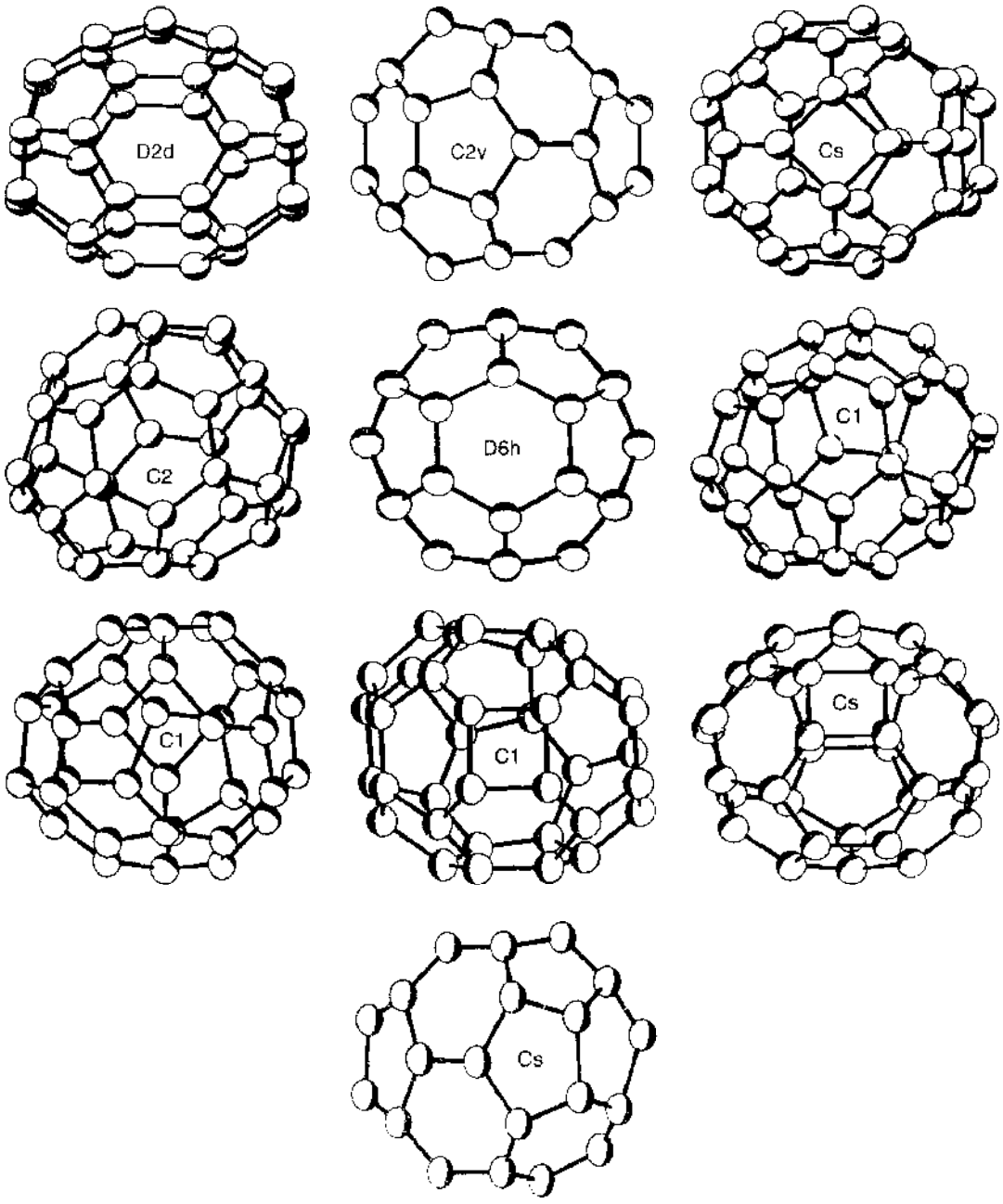


Figure 9. The SAM1 optimized structures [144] of C_{36}: 10 lowest isomers out of a set of 598 cages. Reprinted with permission from [144], Z. Slanina et al., *Chem. Phys. Lett.* 290, 311 (1998). © 1998, Elsevier.

10-membered set is a C_s conventional fullerene and is located about 162 kJ/mol above the ground state in the SAM1 approach. However, the third lowest structure contains one square, it has a C_s symmetry, and in the SAM1 terms it is located some 77 kJ/mol above the ground state. There is one interesting high-symmetry conventional fullerene. The structure exhibits a D_{6h} topological symmetry; however, flexible symmetry diagnostic tool [117] gives only C_{6v} symmetry at the SAM1 level [144] and C_{2h} symmetry after the B3LYP/6-31G* geometry optimization [145]. A closer analysis reveals that we are actually dealing with a Jahn–Teller distortion owing to the degenerate frontier orbitals; however, the finding is method dependent [144–146]. Even after the symmetry relaxation, the structure is located above the D_{2d} ground state. The temperature dependencies of the relative concentrations of the selected C_{36} low-energy isomers were also evaluated [144, 145]. Even single and multireference MP2 perturbation theory was applied recently to the system by Gordon et al. [146] giving still another perspective to its description.

7. HIGHER FULLERENES

7.1. C_{72} Fullerenes

C_{72} is the last member among higher fullerenes [19, 92] C_{60}–C_{96} that has not been isolated in condensed phase. C_{74} has also been counted as a "missing fullerene," however, some access to C_{74} is already possible [354, 355]. A low solubility [356] in conventional solvents may be one reason for the difficulties. Hence, as C_{72} could only be recorded in gas phase [357, 358], its structure is not known. Still, C_{72} can act as a host cage for some metallofullerenes [41, 147, 148] like Ca@C_{72} or La_2@C_{72}.

There is just one IPR satisfying structure for C_{72}, namely with D_{6d} symmetry [97]. Nevertheless, it was pointed out in conjunction with the Ca@C_{72} computations [148] that a non-IPR (i.e., IPR-violating) structure with one pentagon–pentagon junction is by a few kcal/mol lower in energy than the IPR cage. Moreover, it was demostrated [359] on the Si_{60} case that the IPR/non-IPR stability order can be reversed by the entropy factor. Hence, the C_{72} system was investigated at semiempirical [360] and DFT levels [361]. At the DFT level, the geometry optimizations were carried out [361] using the B3LYP/3-21G treatment. In the optimized B3LYP/3-21G geometries, the harmonic vibrational analysis was carried out and also the B3LYP/6-31G* separation energies were evaluated. The electronic excitation energies were evaluated by means of the ZINDO (Zermers INDO) method [362, 363].

The following structures were considered [361]: the IPR cage (a), two non-IPR cages [148] with one pentagon-pentagon junction (b) and (c), a structure [148] with one heptagon (d), a cage [149] with two heptagons (e), and two structures [356] each with two pentagon–pentagon junctions (f), (g). Figure 10 presents the B3LYP/3-21G optimized structures of the seven C_{72} isomers. The (c) structure of C_{2v} symmetry with just one pentagon/pentagon fusion represents the lowest energy isomer, being followed by the IPR structure (a). The other structure with just one pentagon–pentagon junction, (b) isomer of C_2 symmetry, comes as the third lowest species. The remaining four cages are located more than 30 kcal/mol above the lowest one.

Figure 11 presents the temperature development of the relative concentrations of the seven C_{72} isomers in a high-temperature region. The lowest energy structure (c) is the most populated species at any temperature. On the other hand, the IPR structure (a) is always negligible. Its elimination cannot be ascribe only to the potential energy. In fact, several structures higher in potential energy than the (a) isomer are more populated at high temperatures. This is in particular true for the (b) structure with one pentagon–pentagon junction and the (e) structure with two heptagons [the (e) cage is located in the B3LYP/6-31G* scale about 20 kcal/mol above the (a) isomer]. Clearly enough, the IPR structure is suppressed not only by its energy but also by its unfavorable entropy term.

7.2. C_{74} Fullerenes

Although C_{74} has not been isolated in condensed phase yet, Shinohara et al. [355] recorded electronic spectrum of C_{74} anion and suggested that the cage could have D_{3h} symmetry

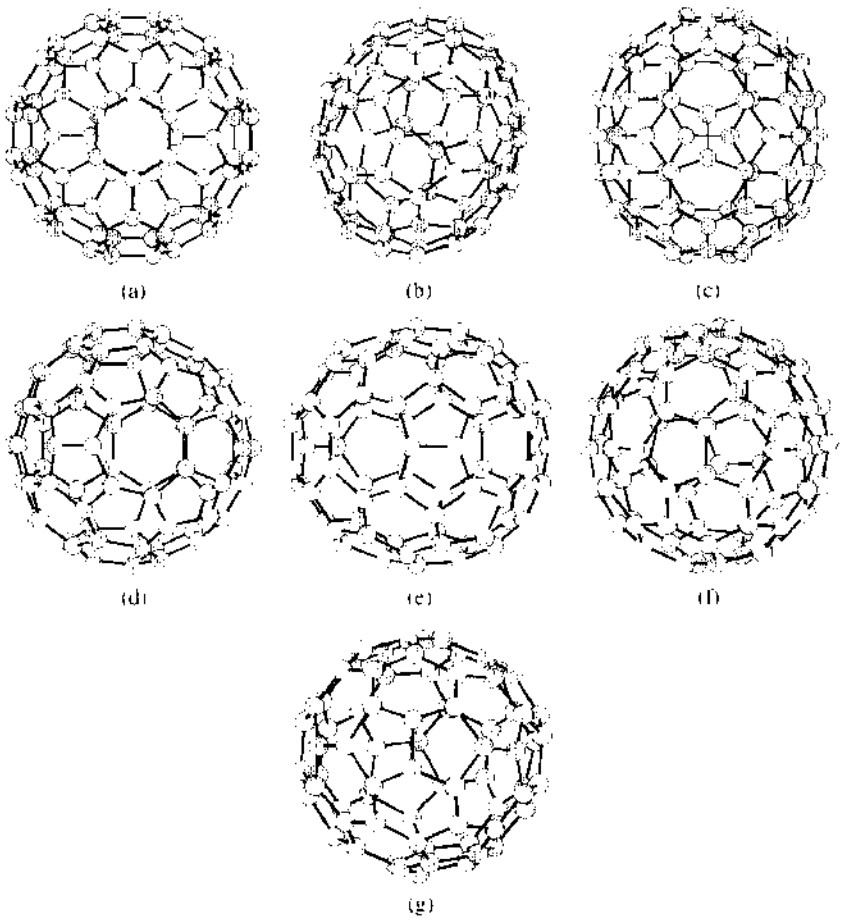


Figure 10. B3LYP/3-21G optimized structures [361] of C_{72} isomers: (a) IPR, (b) 5/5 pair, (c) 5/5 pair, (d) 7-membered ring, (e) two 7-membered rings, (f) two 5/5 pairs, (g) two 5/5 pairs. Reprinted with permission from [361], Z. Slanina et al., *Chem. Phys. Lett.* 384, 114 (2004). © 2004, Elsevier.

(i.e., the only one available [97] IPR structure). This interesting finding prompted DFT computations [364] that treated a set of six isomers (Fig. 12), five of them being non-IPR species. The set consists of the three isomers considered previously [147, 149] in connection with $Ca@C_{74}$ and of three additional cages that also exhibited some non-negligible populations in semiempirical evaluations [365]. In the evaluations [365], the cages were labeled by some code numbers that are also used here, combined with the symmetry of the complexes: $1/D_{3h}$ (IPR), $4/C_1$ (5/5 fusion), $52/C_2$ (5/5 fusion), $103/C_1$ (5/5 fusion), $368/C_1$ (5/5 fusion and 7-ring), and $463/C_1$ (5/5 fusion and 7-ring). The computational methodology was the same as in the C_{72} case [361].

Figure 13 presents the temperature development in the C_{74} set. The IPR structure is the lowest in energy and prevails at any relevant temperature. The computations thus agree with the experimental finding [355]. There is, however, an interesting point regarding the electronic partition function. Furche and Ahlrichs [366] recently pointed out a very low location of the first triplet states for some C_{80} isomers (actually even with a possibility of a triplet ground state). It can be expected that the triplet species will polymerize and form insoluble solids. There are also other fullerenic systems [314, 367] with computed low triplet states, and the same feature was also recently reported [368] for the IPR C_{74}. Although the computations [368] were carried out with the unrestricted Hartree–Fock method that gives lower energies than the corresponding restricted open-shell Hartree-Fock treatment [314, 367], the aspect was also found [364] in the ZINDO and time-dependent [369] DFT treatments.

In principle, every excited electronic state could be treated within Eq. (1) as an individual isomer with its own rotational–vibrational partition function. This approach would, however, considerably increase the computational demands. Consequently, one has to follow the usual approach in which the rotational and vibrational levels of any excited electronic state are put

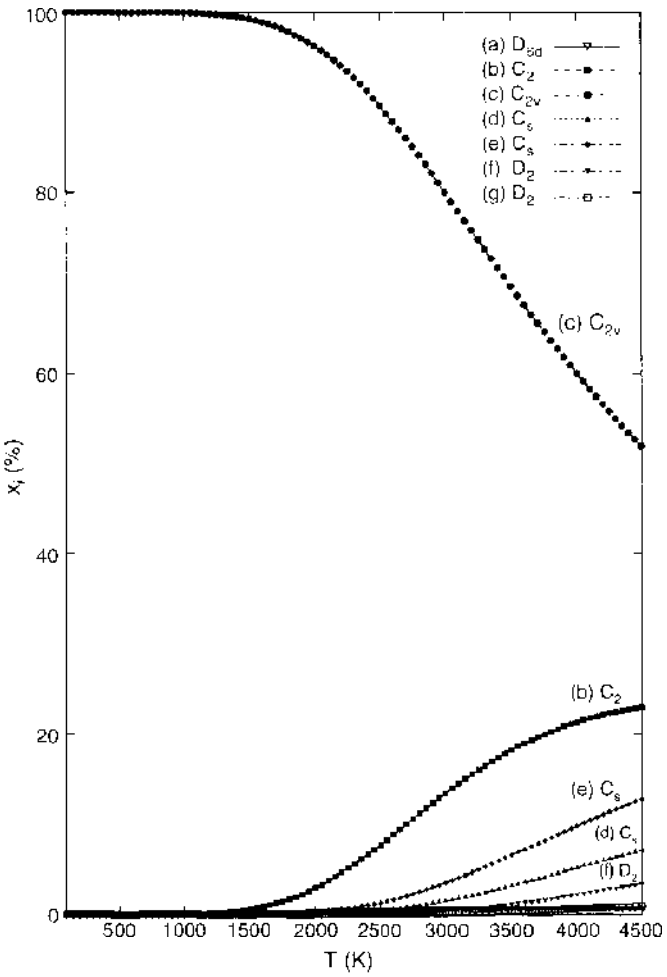


Figure 11. Relative concentrations [361] of the C_{72} isomers based on the B3LYP/6-31G* energetics and the B3LYP/3-21G and ZINDO entropy. Reprinted with permission from [361], Z. Slanina et al., *Chem. Phys. Lett.* 384, 114 (2004). © 2004, Elsevier.

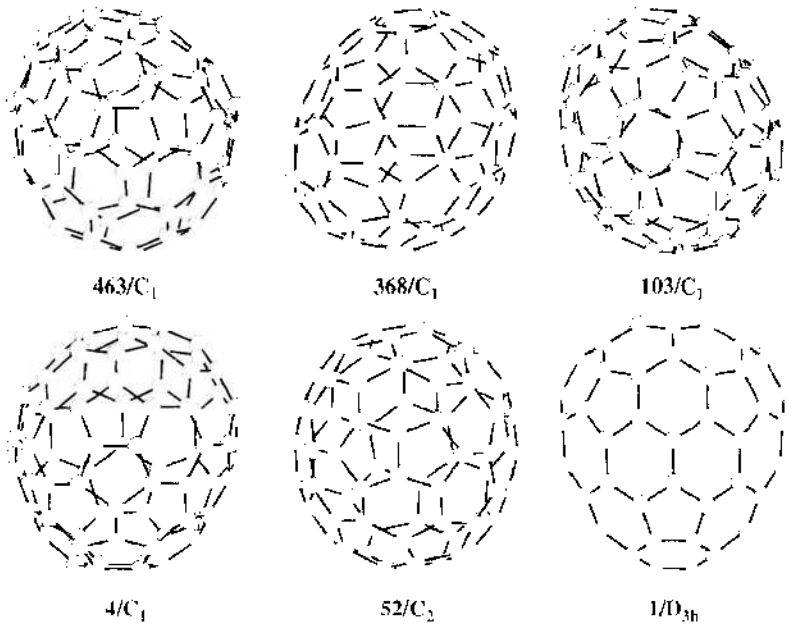


Figure 12. B3LYP/3-21G optimized structures [364] of C_{74} isomers: 463/C_1 5/5 fusion and 7-ring, 368/C_1 5/5 fusion and 7-ring, 103/C_1 5/5 fusion, 4/C_1 5/5 fusion, 52/C_2 5/5 fusion, 1/D_{3h} IPR.

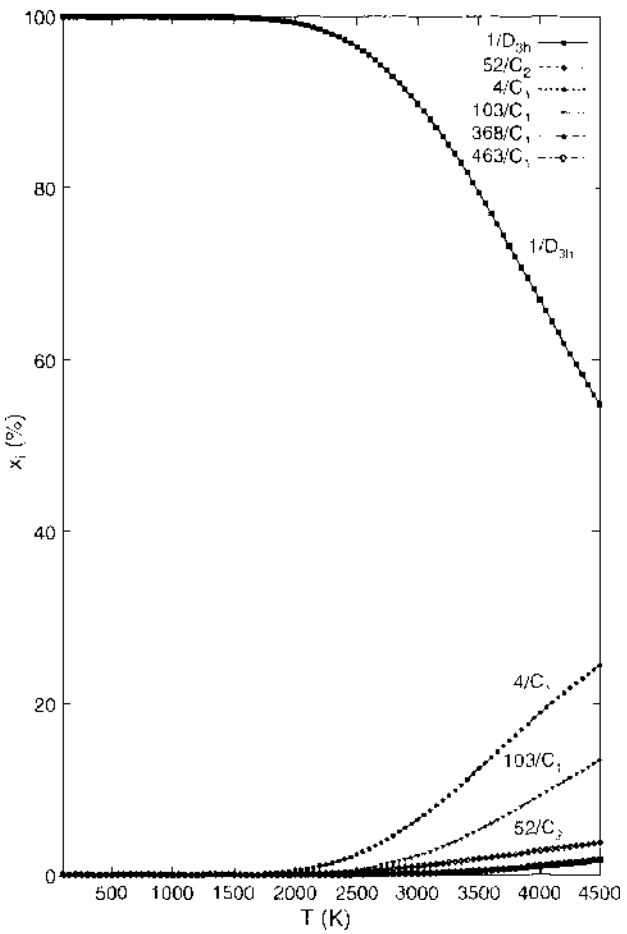


Figure 13. Relative concentrations [364] of the C_{74} isomers based on the B3LYP/6-31G* energetics and the B3LYP/3-21G and ZINDO entropy; both singlet and triplet electronic excited states are included in the partition functions.

equal to those of the respective electronic ground state. There is, however, one interesting point. Let us suppose that the triplet states from the gas phase end up as polymers after condensation. Then, it can be shown [370] that their explicit consideration as individual isomers in Eq. (1) would not influence the ratios of x_i, and thus the comparison with the observed data (as long as we deal only with the x_i ratios). Anyhow, the topic is in fact not particularly important for C_{74}, as even if the triplet states are excluded from the partition functions, the results are not qualitatively different from those in Fig. 13. However, the presence of low-lying triplet states would suggest [370] reduced production yields as they are scaled down owing to the part of the cages that were converted into the polymeric material.

7.3. C_{76} Isolated Pentagon Rule Set

The system was firstly isolated by Ettl et al. [101] and treated in further studies [100, 102–110]. C_{76} is actually the smallest fullerene that allows for cage isomerism of its IPR structures, though modest. There are just two [97] different IPR structures, namely with the topological symmetries D_2 and T_d. However, the T_d topological symmetry undergoes a relaxation as the structure exhibits degenerate, partially filled frontier orbitals. Thus, it has to be distorted according to the Jahn–Teller theorem, lowering both symmetry and energy. In the AM1 calculations [100, 371], this relaxation process ends in a D_{2d} symmetry (though other quantum-chemical methods could predict other symmetry subgroups). Moreover, the D_{2d} isomer is still located about 108 kJ/mol above the D_2 ground state on the AM1 potential hypersurface. The separation energy derived directly from the heat of formation at room temperature $\Delta H^{\circ}_{f,298}$ is 103 kJ/mol. Interestingly enough, the geometrical distortion is quite small. This distortion can be measured by the rotational constants (identical for a spherical top like

tetrahedral T_d species). Even after the symmetry relaxation, the rotational constants remain almost identical for the D_{2d} structure: 0.00173, 0.00173, and 0.00174 cm^{-1}. For the other species, D_2, the rotational constants are, however, distinctly different: 0.00194, 0.00171, and 0.00156 cm^{-1}. As all the computed vibrational frequencies of both structures are real [100, 371], one really deals with local energy minima after the symmetry relaxation.

From the relative-stability point of view, the D_2 structure is actually in a triply convenient situation. It is lower in energy, it is enhanced by its chirality factor of 2, and its low vibrational frequencies are lower than those of the D_{2d} isomer. Owing to this triple coincidence, it must be prevailing even at very high temperatures. For example, even at a temperature of 4000 K, the ground-state structure represents [100, 371] more than 95% of the two-isomer equilibrium mixture. Thus, the computations clearly point out that the D_{2d} C_{76} species can hardly be observed.

7.4. C_{78} Isolated Pentagon Rule Set

The situation is very much different if we move from C_{76} to C_{78}. The IPR cage isomerism of C_{78} has been frequently studied [111–114, 293, 373] not only by computations but also in experiments, however, the picture is still not yet complete. In experiment, at first only two isomers were observed by Diederich et al. [111]: D_3 and C_{2v} symmetry in a ratio 1:5; the latter species is also labeled as C_{2v}(I). Later on, Kikuchi et al. [112], however, reported three isomers of C_{78} with the symmetries D_3, C_{2v}(I), and C_{2v}(II) and in a ratio 2:2:5. The challenging observations were soon completed with computations [113, 114] of the relative stabilities. In fact, it was only the third case of such computations in the fullerene field, following two model computations without an available experimental counterpart [215, 343].

The semiempirical stability computations on the C_{78} IPR system were performed [113, 114, 373] with the semiempirical MNDO, AM1, and PM3 methods. There are five [97] IPR isomers for C_{78}, and each of the topologically possible structures indeed leads to a local energy minimum in the geometry optimizations. Their symmetries derived from the MNDO optimizations are [113] D_{3h}(I), D_{3h}(II), D_3, C_{2v}(I), and C_{2v}(II). In all three semiempirical methods employed, the C_{2v}(II) structure represents the lowest (global) energy minimum (in fact, in agreement with the MM3 force-field results).

According to the semiempirical computations [114], the global minimum represents more than 50% of the equilibrium mixture even at a high temperature of 4000 K. On the other hand, the relative population of the D_{3h}(I) structure is always practically negligible. The remaining three structures exhibit comparable relative stabilities, and in the high-temperature limit this three-membered group represents nearly 50% of the equilibrium isomeric mixture. The picture is basically the same for the three semiempirical energetics considered. Fowler et al. [293] conclude that the structures D_{3h}(I), D_{3h}(II), D_3, C_{2v}(I), and C_2(II) should exhibit 8, 8, 13, 21, and 22 lines, respectively, in their ^{13}C NMR spectra. Diederich et al. [111] isolated two C_{78} isomers that exhibited 13 and 21 NMR lines, respectively, and were therefore identified as the D_3 and C_{2v}(I) structures. The presence of only two isomers (rather than of five) was rationalized [111] by kinetics. Two product channels were considered, the minor one leading to the D_3 species and the major one to the C_{2v}(I) species. A kinetic control, if present, would of course reduce applicability of the equilibrium thermodynamical treatment.

It was not however the end of the story as later on Kikuchi et al. [112] reported their observaton of the D_3, C_{2v}(I), and C_{2v}(II) isomers of C_{78} in a 2:2:5 ratio. This second experiment is actually qualitatively more consistent with the computations than the 1:5:0 ratio [111]. However, in order to get even a quantitative agreement, we have to go to relatively very high temperatures. Still, it should be realized that a relatively small change in the observed ratio can scale down the anticipated temperatures considerably. Moreover, there is certainly room for further computational activities; for example, on the system energetics. In addition to hat computational need, there have also been additional experimental results reported recently [374–376]. In particular, the independent ^3He NMR spectroscopy data supplied by Saunders et al. [20, 125, 377] could support a more ample isomerism. This possibility is indeed consistent with the recent experimental work of Yamamoto, who reported [378] isolaton of a fourth isomer of C_{78}, namely the D'_{3h} species. The observation prompted a new

computational evaluation [372, 379] with the B3LYP/6-31G* energetics and ZINDO electronic excitations. Figure 14 presents [372] the outcome (including a visualization of the role of the electronic partition function). The computed results [372, 379] agree reasonably well with the experimental findings [112, 378] on the C_{78} system.

7.5. C_{80} Isolated Pentagon Rule Set

This system of seven [97] topologically possible IPR structures represents an instructive example of the thermodynamic isomeric interplay. The isomers have in the literature been conventionally coded [115–119, 136] by letters rather than by a numerical code, **A–G**. Of the seven IPR C_{80} structures, the species **B** exhibits a high topological symmetry, I_h, the same as seen in the icosahedral C_{60}. However, according to the concept [286, 380, 381] of the Goldberg polyhedra based on the topological duals, it has to be an electronic open shell. Consequently, it has to undergo a Jahn–Teller distortion toward lower energy and lower symmetry. Similarly to the distortions of the topologically tetrahedral T_d C_{76} case, we can also check the Jahn–Teller deformation on the rotational constants. The distortions are again quite small, as the rotational constants of the AM1 fully optimized species **B** read [117] 0.00155, 0.00156, and 0.00157 cm^{-1} (i.e., a near-spherical top). The symmetry of the **B** structure is lowered to D_2 in the relaxation process. Nevertheless, the particular species still exhibits the highest energy in the computed IPR set. The computations (both at room and at the absolute zero temperature) point out the **C** isomer (D_{5d} symmetry) as the system ground state, being followed by the **A** species of a D_2 symmetry.

Figure 15 presents [136, 382] the temperature dependencies of the relative concentrations, w_i, for the seven-membered mixture under the thermodynamic-equilibrium conditions.

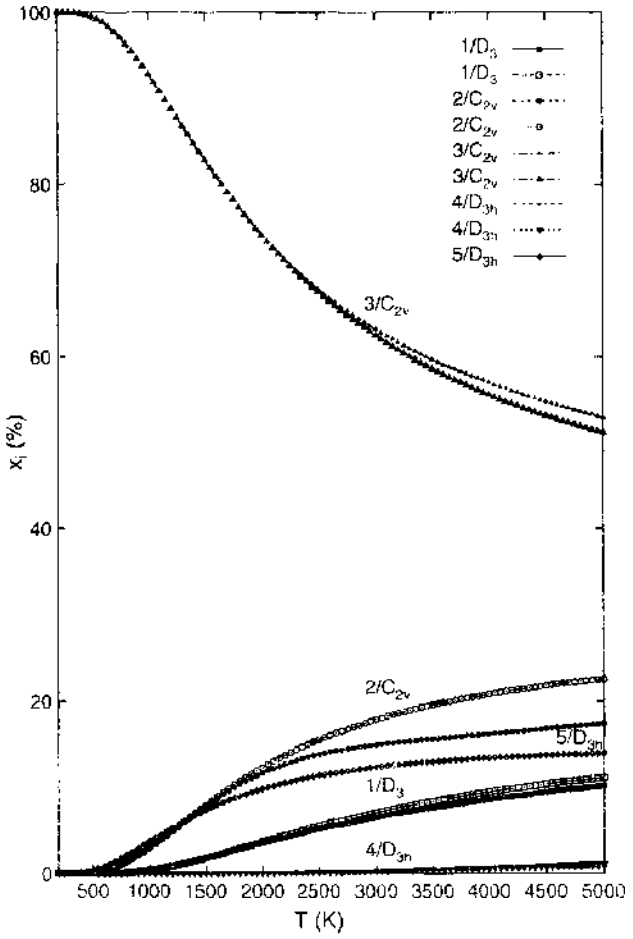


Figure 14. Mole fractions [372] x_i of five IPR isomers of C_{78} (the full and empty symbols refer to the calculations with and without q_{vib}, respectively). Reprinted with permission from [372], Z. Slanina et al., *Int. J. Quantum Chem.* 96 (2003), © 2003, John Wiley and Sons.

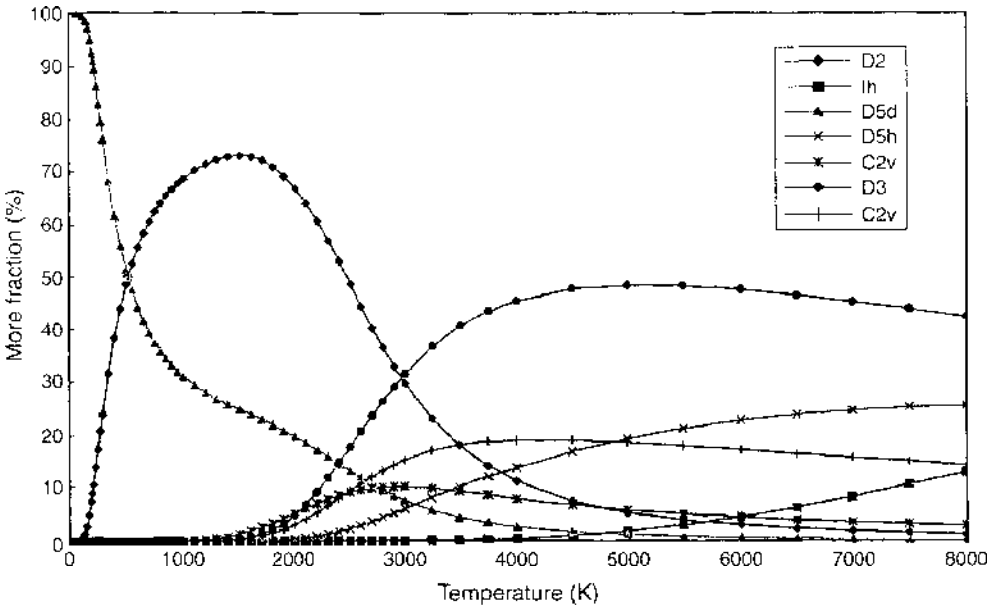


Figure 15. The HF/4-31G computed [136, 382] relative concentrations of C_{80} (topological symmetries indicated). Reprinted with permission from [382], Z. Slanina et al., *Advan. Strain. Inter. Org. Mol.* 7, 185 (1999). © 1999, Elsevier.

The dependencies are reported for the HF/4-31G energetics (Fig. 15 refers to the topological symmetries of the isomers). As always at very low temperatures, the ground-state structure, **C** D_{5d} symmetry, has to be prevailing, and the relative concentrations of other structures have to simply obey the $\Delta H^{\circ}_{f,0,\mathrm{rel}}$ order. This low-temperature region, however, has no relevancy to the fullerene synthesis. At higher temperatures, the preexponential factors in Eq. (1) become increasingly important (whereas the importance of the exponential part is gradually decreasing at the same time). In fact, at a relatively low temperature, the **A** (D_2 symmetry) species reaches equimolarity with the **C** species. Beyond this temperature point of the two-component equimolarity, the **A** species of D_2 symmetry is always more populated (though not being the system ground state). Interestingly enough, at still considerably higher temperatures, yet another species becomes the most populated.

The C_{80} separation energies were computed [117, 136] in the SAM1, AM1, HF/STO-3G, HF/3-21G, and HF/4-31G approaches. The results agree reasonably well so that the agreement between various types of computations supports reliability of the predicted relative concentrations. In fact, the computational prediction [117] could soon be critically tested against observations. NMR spectra of a C_{80} isomer were recorded [118] and they were indeed consistent with the **A** (D_2 symmetry) species. According to the HF/4-31G computations [136, 383], the two-component equimolarity point between the important **A** and **C** structures is reached already at 517 K. At a temperature of some 1500 K (which probably represents a reasonable lower bound of temperatures for the fullerene synthesis), the **A** structure of D_2 symmetry forms about 73% of the equilibrium isomeric mixture. In fact, the **A** isomer reaches its relative concentration maximum at a temperature of 1497 K in the HF/4-31G energetics. Thus, the computations and observations single out as the most populated species a structure that is not the ground state of the system, but it is still preferred by the Gibbs function. This interesting thermodynamical event has been known with other chemical systems [198, 384], too, but it is relatively frequent for isomeric fullerenes owing to the extremely high temperatures needed for their preparations. Recently, Shinohara et al. [119] isolated also a minor C_{80} isomer, the D_{5d} species.

Interestingly enough, once a metal is encapsulated, the electronic structure is influenced considerably. There is a substantial charge transfer to the cage, and this can in some convenient cases eliminate the Jahn–Teller situation otherwise present with the empty cage. This interesting event happens with the I_h C_{80} cage, for example, if two La atoms are encapsulated [385–387]. In fact, for La_2C_{80} the I_h-related encapsulate is the lowest energy isomer [385].

7.6. C_{82} Isolated Pentagon Rule Set

^{13}C NMR spectra for C_{82} were reported by Kikuchi et al. [112], and that time three symmetry species (C_2, C_{2v}, C_{3v}) were concluded with a ratio of 8:1:1. Later on, however, the original interpretation was modified [19, 91, 92]; in the newer interpretation, only C_2 symmetry is considered with the C_{82} species (while the two former minor components are supposed to be other fullerenes). In fact, it is even suggested [91] that C_{82} has at least two isomers, and both of them have C_2 symmetry. Several related papers are available dealing with the C_{82} system [120–125].

There are nine C_{82} cage topologies [97] that satisfy the IPR condition. Their relative stabilities at high temperatures were evaluated [124] with the AM1 quantum-chemical semiempirical method. The cage separation energies are rather similar, giving the following energy order at the absolute zero temperature: C_2, C_2, C_s, C_s, C_2, C_s, C_1, C_s, and C_s. In particular, the two lowest structures are of C_2 symmetry and are separated by about 17 kJ/mol only.

At very low temperatures, the ground-state structure of a C_2 symmetry is of course dominant. At moderate temperatures, the other low-energy C_2 isomer represents the second most populated structure. However, its molar fraction in fact exhibits a temperature maximum of about 20% and then decreases. At very high temperatures, some other structures show significant populations, including the third C_2 isomer. For a more quantitative correlation of computations and observations, some information on the temperature history of the particular sample would be needed. Any information on a representative reaction temperature during the fullerene synthesis is still virtually missing. However, the computations reveal that the group of C_2 species always represents more than 50% of the equilibrium isomeric mixture, for example 84% at 1500 K. This computational finding is consistent with the updated interpretation [19, 91, 92] of the available NMR spectra as it allows for both one or two C_2 structures. Incidentally, with some difficulty, the computations could to some extent also support the previous NMR interpretation (C_2, C_{2v}, C_{3v} in a ratio 8:1:1).

Although isomerism of the empty C_{82} cages is not important, the situation is quite different with some C_{82}-based metallofullerenes. For example $Ca@C_{82}$ consists of four isomers [155–161], $La@C_{82}$ of two isomers [162–166], or $Tm@C_{82}$ of three isomers [167, 388–390].

7.7. C_{84} Isolated Pentagon Rule Set

C_{84} fullerenes [126–132] are rather special as they are relatively abundant, and thus, they were among the very first fullerene species observed [126] immediately after the synthesis of C_{60} and C_{70}. Their observations have frequently been complemented with theoretical studies [51, 96, 129, 130, 292, 296, 391–396], and, in turn, there is now a larger set of identified C_{84} isomers. However, a still more sophisticated wave of the C_{84} computations is to be expected before the topic is entirely settled. In fact, for C_{84}, altogether 24 IPR structures are to be considered [97], and this isomeric count can be quite demanding on computational resources.

Kikuchi et al. [112] concluded from ^{13}C NMR spectra the presence of two major isomers of D_2 and D_{2d} symmetry. The two major isomers are now quite well characterized [397–399]. Saunders et al. [20, 125, 377] could, however, see with their ^3He NMR technique even up to nine C_{84} isomers. The two major isomers were separated by Dennis et al. [131], and more recently they could recognize several minor isomers [400–402].

Bakowies et al. [394] computed all the C_{84} IPR isomers at the MNDO semiempirical level. Their study was soon completed [130, 132, 395, 396] with computational evaluation of the temperature dependencies of the relative concentrations in the isomeric set. In fact, the C_{84} isomeric system is also the interesting case where the lowest-energy cage is not the most populated species at the conditions of the fullerene synthesis. The two species reported by Kikuchi et al. [112] are conventionally labeled [394] **22** (D_2 symmetry) and **23** (D_{2d} symmetry). In the MNDO computations [394], the lowest isomer **23** is located only about 2 kJ/mol below structure **22**. On the other hand, the structure highest in energy (also with a D_2 symmetry) is located about 240 kJ/mol above the lowest minimum.

The lowest energy minimum, **23**, is actually the most populated species only at very low temperatures [130], below room temperature. The **22**/**23** two-component equimolarity point

is found at a temperature of 276 K only. After this relative stability interchange, structure **22** (in spit of being slightly higher in the potential energy) becomes the most populated isomer over a wide temperature interval. Moreover, the **22**/**23** ratio is almost constant, being roughly 2:1 anywhere above 500 K; for example, the ratio of the two isomers is 58.2:41.7, 41.4:20.6, and 19.0:8.84 at temperatures 500, 2000, and 4000 K, respectively. One could interpret [199, 216, 298] this 2:1 ratio in the terms of the chirality partition function from Eq. (1). There are several other species exhibiting a significant or non-negligible mole fraction at very high temperatures. This computational prediction [130] was after several years indeed confirmed in the observations of Saunders et al. [20, 125, 377] and Dennis et al. [400–402]. For example, if we select a threshold w_i value of 10%, the computations can point three additional isomers of this relatively high stability.

The agreement of the computational predictions with some of the newer experiments is comprehensively analyzed in our study [132]. Some non-IPR structures of C_{84} were also evaluated [365]. Competition between some of the species creates an interesting temperature dependency with a local maximum (as also seen for C_{80} in Fig. 15). In fact, even in very complex fullerene mixtures, we should be able, at least for some specific cases, to suggest an optimal region of temperatures where the production yield should be maximized. Although any temperature history of the sample used for the NMR spectroscopy by Kikuchi et al. [112] is virtually unknown, the computed data can match their observed ratio 2:1 in a very wide temperature interval. Overall, the C_{84} system represents another instructive example of a basic theory–experiment agreement, though further developments can be expected.

7.8. C_{86} Isolated Pentagon Rule Set

As soon as the first isomeric fullerene systems such as C_{76}, C_{78}, C_{80}, C_{82}, or C_{84} were successfully computed, it had become rather clear that the interisomeric thermodynamic equilibrium presumption was applicable. This conceptual finding has been somewhat surprising and could be interpreted through a cooperative cancellation of several effects. Anyhow, still higher fullerenic systems should be submitted for computations as long as some observations are available. At present, the observations extend [19, 91, 92, 93, 374] till C_{96} or even till [374] C_{102} (and also include C_{110}). C_{86} is then the next system in the row to be computed [133–136]. In contrast to the 24 IPR isomers for C_{84}, there are only 19 topologically different C_{86} IPR structures. As for the relative stabilities, they were computed [133–136] primarily with the SAM1 method [185] (for other semiempirical treatments, see Ref. [403]). However, also the AM1, HF/STO-3G, and STO/3-21G treatments consistently produce [133–136] a C_2 species as the ground state of the C_{86} IPR set.

Figure 16 shows the temperature development of the computed [133, 382] relative stabilities in the C_{86} IPR set. At very low temperatures, the ground-state structure is of course prevailing, however, its mole fraction decreases relatively fast. At first, a C_s species reaches close to 20% (it has its population maximum at a temperature of 1503 K). There is a third significant species (C_3 symmetry). Although it is relatively high in energy, its entropy contribution still allows for an equimolarity (about 29% each) with the ground state at a temperature of 1739 K. Let us mention, however, that at this temperature, two additional species already exhibit relative concentrations of about 10%. Achiba et al. [91] found from ^{13}C NMR spectra the presence of two C_2 isomers in their C_{86} sample; the relative abundance of the isomers was about 4:1. The computations [133–136] suggest C_2 and C_s structures around 1500 K. At higher temperatures, the computations predict a coexistence of more than two isomers. It is, however, true that one computed symmetry is C_s, not C_2, and this minor difference between theory and experiment has not been clarified yet.

7.9. C_{88} Isolated Pentagon Rule Set

After the small drop in the number of the IPR structures when going from C_{84} to C_{86}, the C_{88} IPR system is already quite extended. There are [97] 35 topologically different C_{88} IPR structures. Nevertheless, also this system has been computed and reasonably understood [134–136, 404, 405], again primarily using the SAM1 semiempirical method.

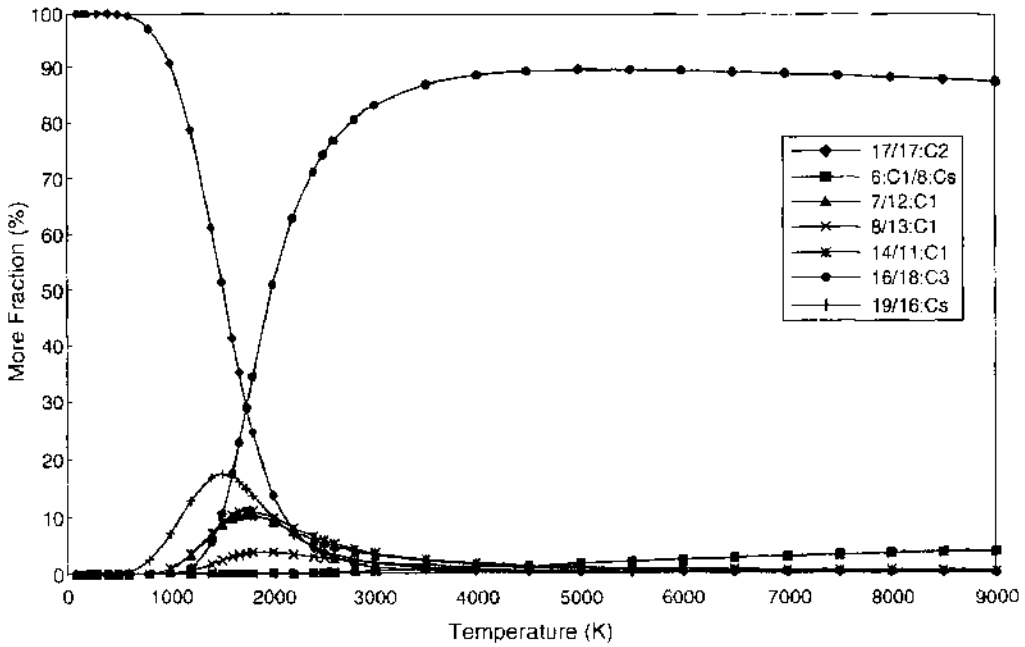


Figure 16. The SAM1 computed [133, 382] relative concentrations of C_{88} isomers (numbering according to Refs. [97, 133]). Reprinted with permission from [382], Z. Slanina et al., *Advan. Strain. Inter. Org. Mol.* 7, 185 (1999), © 1999, Elsevier.

As already mentioned (and demonstrated with, e.g., the topologically icosahedral structure for C_{80}), the original topological symmetries represent a kind of symmetry upper bound. The true symmetries after quantum-chemical geometry optimizations can be the same or lower (not higher). In particular, during quantum-chemical calculations, the topological symmetry can be lowered owing to Jahn–Teller and pseudo Jahn–Teller effects. In addition to those specific physical reasons, it can also be reduced simply owing to general energy reasons. Hence, only after quantum-chemical geometry optimizations can the symmetry properties be clarified. Moreover, such symmetry reductions are not rare for fullerene and other cages and can frequently happen even without a Jahn–Teller effect. For example, the icosahedral B_{32} cage (a topological boron analogy of C_{60}) actually exhibits [406] imaginary frequencies when computed *ab initio*. Similarly, hexa-anion of C_{60} may seem to be icosahedral (as the degenerate LUMO is just filled), but actually its symmetry relaxes with a considerable energy gain [406] of some 60–150 kJ/mol. With the examples in mind, it is not surprising that the C_{88} system at the SAM1 computational level in five cases falls below the topological upper symmetry limit. In four cases, the starting topological symmetry itself is so low that it does not allow for degenerate representations. Hence, we deal with only one pure Jahn–Teller distortion (namely from T to D_2).

The system was computed [404] at four different levels of theory: SAM1, HF/STO-3G, HF/3-21G, HF/4-31G. The computations consistently predict a species of C_1 symmetry as the ground state of the C_{88} IPR set. The IPR cage highest in energy is located about 300 kJ/mol above this system ground state. Owing to the temperature interplay governed by Eq. (1), the species, which is the second lowest in energy and has a C_2 symmetry, exhibits [404, 407] a fast relative-stability increase with a temperature maximum (Fig. 17). The maximum molar fraction is about 23% and happens at a temperature of about 1470 K in the SAM1 computations (or already at a temperature of 1270 K, and amounts to about 30%, in the HF/4-31G approach). The third lowest species, again having a C_2 symmetry, has a temperature profile with a maximum, too, though rather modest. There is, however, a structure with a steady increase that eventually becomes dominant—a C_1 symmetry, in fact reduced from a C_2 topological expectation. The isomer is relatively rich in potential energy—over 100 kJ/mol above the ground state. It has an equimolarity point with the ground-state structure that comes in the HF/4-31G computations at a temperature of about 1990 K with 29% for each species (see Fig. 17). In the SAM1 energetics, the equimolarity is reached at somewhat higher

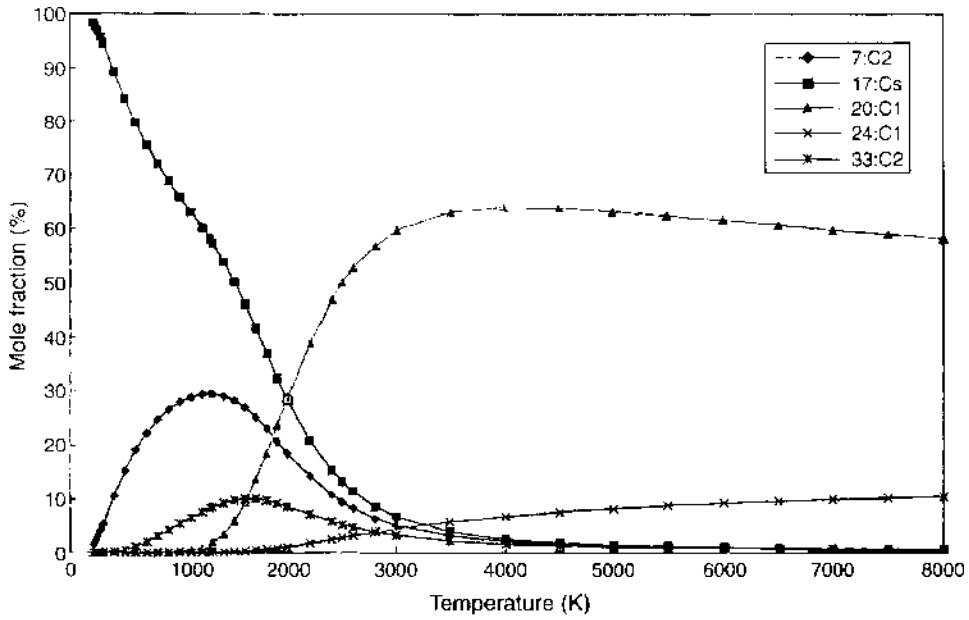


Figure 17. The computed [404, 407] (HF/4-31G energetics) relative concentrations of C_{88}. Reprinted with permission from [407], Z. Slanina et al., *Scripta Mater.* 43, 733 (2000). © 2000, Elsevier.

temperature, at about 2240 K with about 26% for each of the two isomers. For example, at a temperature of 3000 K, the C_1 relative concentration amounts already to 59% and 47% in the HF/4-31G and SAM1 energetics, respectively.

^{13}C NMR spectrum for C_{88} in solution is available from Achiba et al. [91] and is consistent with an isomer of C_2 symmetry as a major structure. Although our computations point out even two C_2 structures, their temperature maxima appear at rather low temperatures when the ground-state structure still has a strong population and should be seen in the experiment, too. However, if we consider temperatures above 2240 and 1990 K according to the SAM1 and HF/4-31G predictions, respectively, we shall deal with the fast increasing C_1 structure as the most populated and eventually the dominant one.

As already mentioned, the structure exhibits a C_2 topological symmetry, which is reduced to C_1 only in the quantum-chemical optimization. Such symmetry reductions can bring rather small coordinate distortions. For example, the topologically icosahedral C_{80} cage exhibits [117] the Jahn–Teller geometry distortions of the order 10^{-3} or 10^{-2} Å. Hence, we can well speak on a near-icosahedral structure in this case. The C_1 C_{88} structure exhibits in our search a C_2 symmetry for $\epsilon > 4 \times 10^{-2}$ Å, too. However, the vibrational amplitudes of C_{60} from electron diffraction, for example, have still larger values. The same is true, for example, for the vibrational amplitudes of kekulene at room temperature [404]. The vibrational amplitude data suggest that we can actually consider the C_1 structure as having a near C_2 symmetry. This argument would produce an agreement between computation and experiment. Moreover, the theory–experiment comparison also suggests that the measured sample originated at temperatures somewhat over 2000 K.

7.10. C_{90} Isolated Pentagon Rule Set

C_{90} is another system where the treatment based on the interisomeric thermodynamic equilibrium presumption reveals rather unexpected complex internal relationships. The computational details are given in Refs. [137, 138]. The C_{90} system consists [97] of 46 topologically different C_{90} IPR structures. The computations consistently point out a structure with C_2 symmetry (labeled **38** and **45** according to numbering systems in Ref. [138] and Ref. [97], respectively) as the system ground state. Interestingly enough, there is an exception at the HF/3-21G level. The D_{5h} structure (only one such high symmetry species in the whole set) comes as practically isoenergetic with the **38**/**45** structure at semiempirical levels, but this event is likely to be just a computational artifact [138, 408].

According to the preliminary experimental results reported by Achiba et al. [91], altogether five C_{90} species were identified from ^{13}C NMR spectra. The observed C_{90} species were in fact distributed in three HPLC fractions: one C_{2v}, three C_2, and one C_1. However, more advanced interpretations reported later on in Refs. [19, 92, 409] change the assignment into the following quintet [409]: two C_2, one C_{2v}, one C_s, and one C_1. The symmetries of the five SAM1 lowest-energy structures read [138] C_2 **(38/45)**, C_{2v} **(29/46)**, C_s **(43/35)**, D_{5h} **(19/1)**, C_1 **(22/30)**. In other words, at least two of the structures suggested by the potential energy screening do not appear in the original interpretation [91]. The computation–observation agreement, however, becomes still even worse if we look into the NMR pattern of the computed C_{2v} structure. The last mentioned structure gives 24 lines, 3 of them weaker. The C_{2v} species considered in the experiment is different and exhibits 25 lines, 5 of them weaker.

At best, we could speak on a partial agreement with the experiment at this intermediate stage. In order to get the final picture, however, we really have to investigate possible thermal effects on the relative stabilities in this large isomeric set in spite of the already considerable dimension of the set. After applying the combined treatment, it turns out that there are just five structures that exhibit a significant population in the high-temperature region, relevant to the fullerene synthesis (Fig. 18). In addition to the three structures lowest in the potential energy (C_2 **38/45**, C_{2v} **29/46**, C_s **43/35**), two high-energy species are also exhibiting significant populations: C_2 **16/18** (ΔE_r = 130 kJ/mol) and C_1 **12/9** (ΔE_r = 202 kJ/mol). As always, at very low temperatures, the ground-state structure has of course to be the dominant species; however, its decrease of relative stability with increasing temperature found for this system is considerable. According to the SAM1 computations [137, 138], the ground-state isomer drops to two-component equimolarity with the **16/18** structure of C_2 symmetry at a temperature of 2012 K. Two other structures, C_s **43/35** and C_{2v} **29/46**, show moderate maxima close

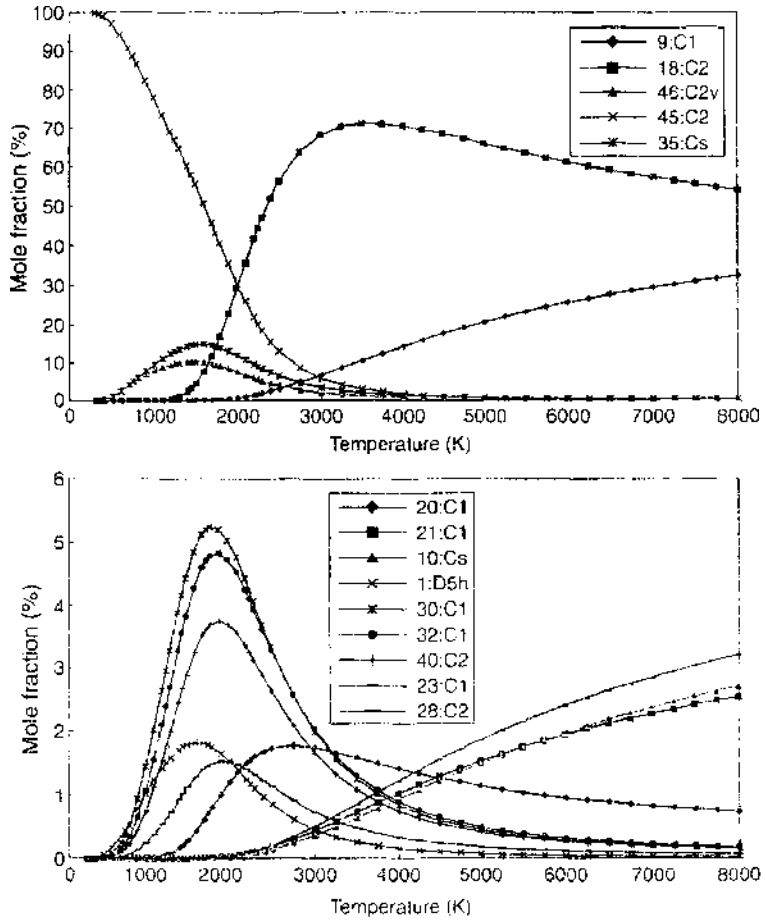


Figure 18. The SAM1 computed [138] relative concentrations of C_{90}: top, major isomers; bottom, some less-populated species (numbering according to Ref. [97]). Reprinted with permission from [138]. Z. Slanina et al., *Chem. Phys.* 219, 193 (1997), © 1997, Elsevier.

to 1500 K. The last structure of Fig. 18, C_1 **12/9**, becomes quite important at very high temperatures, though its population is still found too low around 1000 K. Figure 18 also screens several less-populated structures—it is of interest that the species with the highest temperature maximum in this second group, **22/30**, also has again a C_1 symmetry, and the same is even true for the second most populated species in the second group, **32/32**. On the other hand, the high symmetry D_{5h} isomer is actually not favored by entropy and turns out to be unimportant even if it is almost isoenergetic with the **38/45** ground-state structure. Let us also mention that the results [138] from the SAM1 and HF/4-31G energetics are not particularly different.

We do not know what could be a temperature interval in which the C_{90} sample, examined in the observations [91], was actually synthesized. We may expect the temperatures somewhere beyond 800–1000 K. Our computations predict that at elevated temperatures, we primarily deal with the five structures of the following symmetries: C_2, C_2, C_s, C_{2v}, C_1. This SAM1 high-temperature set compares now better with the original conclusion [91] from the ^{13}C NMR spectra: C_{2v}, three times C_2, C_1. However, we have to remember that in the computations and experiment, we in fact deal with two different C_{2v} isomers with different NMR patterns. The original experimental conclusion is based [91] on an NMR spectrum consisting of exactly 70 lines, five of them weaker—those are the primarily observed facts. Any IPR C_{90} isomer of C_2 symmetry exhibits always 45 lines (none weaker). There are C_{2v} C_{90} structures with exactly 25 lines (5 of them weaker). This was the essence of the interpretation [91] of the observed facts. However, in the computations, we deal with a different C_{2v} structure—only 24 NMR lines (3 of them weaker). Moreover, we have to consider still another species, not mentioned in the original experimental interpretation [91], a C_s structure with 46 lines (2 of them weaker). If we now combine together the NMR patterns of the two structures predicted by the computations, C_{2v} and C_s, we also get exactly 70 lines (and exactly 5 of them weaker). With this computation-supported analysis of the NMR spectra, we actually reach a complete agreement between the computations and the more advanced experimental report [409]. Let us note for the completeness that the C_s species is somewhat more populated than the C_{2v} one, which actually should help to equalize the intensities of their NMR lines.

7.11. C_{92}–C_{98} Isolated Pentagon Rule Sets

The complete set of 86 IPR isomers of C_{92} was described [139] by the SAM1 quantum-chemical method, and their energetics is also checked at the B3LYP/6-31G* level. Although the lowest-energy cage is not identical in both approaches, it still exhibits a D_2 symmetry in either case. The lowest energy structure is not the most populated isomer at higher temperatures—it is replaced by a D_3 structure. Further stability interchanges are possible at very high temperatures, when C_3 and C_1 structures are also important. Achiba et al. [409] reported preliminary results on their ^{13}C NMR investigations of C_{92}. They list [409] seven structures: two D_2, C_2, C_3, D_3, and two C_1. Hence, there is a reasonable correspondence between the computations [139] and the preliminary observed data [409].

The combined computations were also carried [140] out at semiempirical level for the complete set of 134 IPR isomers of C_{94}. Selected low-energy cage structures were recomputed at *ab initio* level. All the methods point out a C_2 species as the lowest energy isomer. In the temperature range 700–2000 K, the SAM1 calculations point out three structures with a comparable stability: C_2, C_s, and C_1. There are also two minor isomers of C_s and C_1 symmetries. Preliminary results of Achiba et al. [409] can distinguish up to three C_2 isomers and one C_s structure. The results of Howard et al. [410, 411] suggest three isomers.

There are also preliminary experimental results available [93, 409] for C_{96}. The isomeric abundance should for the top 10 species decrease along a sequence: C_1 (four isomers), C_2 (three isomers), C_s, D_2, and D_{3d}. Computations were performed [141] for the complete set of 187 IPR isomers of C_{96}, using four semiempirical quantum-chemical methods: MNDO, AM1, PM3, and SAM1. The relative concentration of the lowest energy species (D_2) decreases rather fast, especially in the SAM1 method. There are always some C_1 species among the most populated isomers. On the other hand, populations of high-symmetry isomers are not

impressive (with a D_{6d} exception in the PM3 method). Overall, the computations and experiment are in a reasonable agreement at this stage.

Finally, computational predictions [142] are available for the complete set of 259 IPR isomers of C_{98} at the SAM1, PM3, AM1, and MNDO levels. All the applied methods point out a C_2 species as the lowest energy structure. Interesting stability interchanges in the isomeric set are found so that not only the C_2 ground-state structure but also other three structures (C_1, C_s, and C_2) are significantly populated at elevated temperatures.

8. ENDOHEDRAL METALLOFULLERENES

The combined stability computations have been applied also to a few metallofullerenes, especially to $Ca@C_{72}$, $Ca@C_{74}$, and $Ca@C_{82}$. $Ca@C_{72}$ was isolated [147], though its observed structure is not yet available. It follows from its very first computations [148, 149] that there are four isomers especially low in potential energy. In fact, C_{72} has only one [97] IPR structure. The endohedral $Ca@C_{72}$ species created by putting Ca inside the sole IPR cage has been labeled [148] by (a). The other three $Ca@C_{72}$ isomers considered in Ref. [148] are related to two non-IPR C_{72} cages (b) and (c) and to a C_{72} structure with one heptagon (d) (see Fig. 19).

The extended computations [150] started from the four optimized structures [148] derived using *ab initio* HF treatment with a combined basis set: 3-21G basis for C atoms and a dz basis set [412] with the effective core potential (ECP) on Ca (for the sake of simplicity, the treatment is coded by HF/3-21G~dz). The structures [148] were reoptimized at the B3LYP/3-21G~dz level. In the optimized B3LYP/3-21G~dz geometries, the harmonic vibrational analysis was carried out with the analytical force-constant matrix. In the same geometries, single-point energy calculations were also performed at the B3LYP/6-31G* level. The electronic excitation energies were evaluated by means of time-dependent DFT response theory [369] at the B3LYP/3-21G~dz level.

Figure 20 presents [150] the temperature development of the relative concentrations of the four $Ca@C_{72}$ isomers in a high-temperature region. At very low temperatures

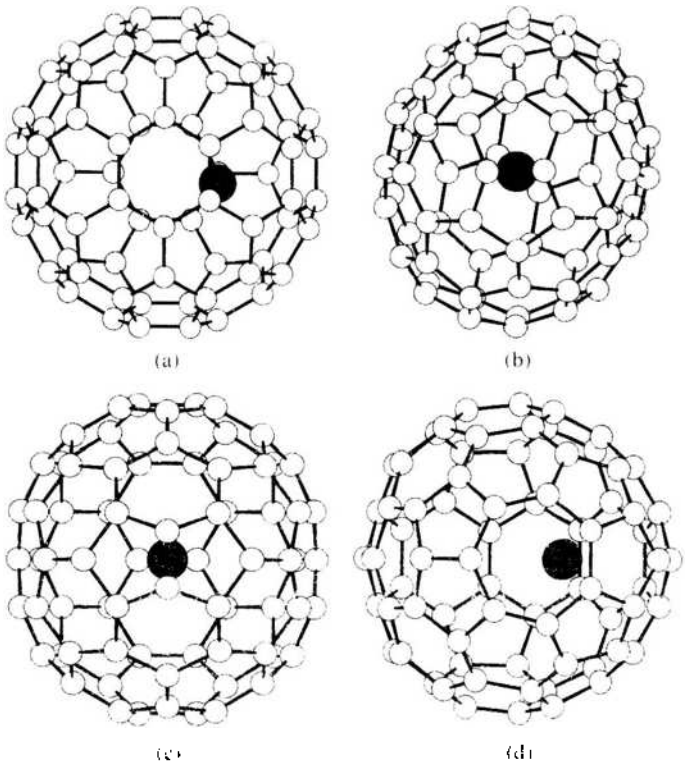


Figure 19. B3LYP/3-21G~dz optimized structures [150] of $Ca@C_{72}$ isomers: (a) IPR, (b) 5/5 pair, (c) 5/5 pair, (d) 7-membered ring. Reprinted with permission from [150], Z. Slanina et al., *Chem. Phys. Lett.* 372, 810 (2003), © 2003, Elsevier.

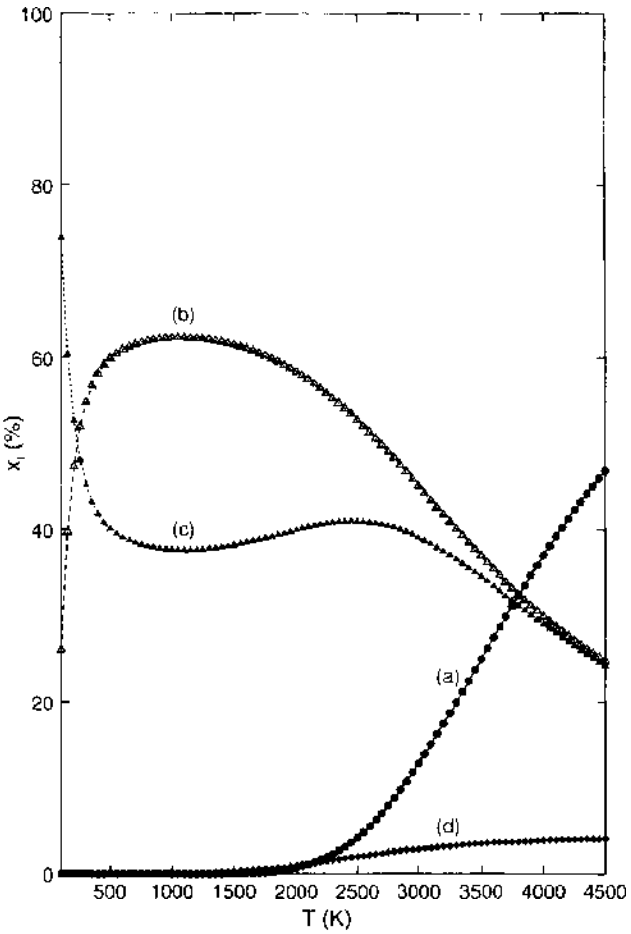


Figure 20. Relative concentrations of the Ca@C_{72} isomers based [150] on the B3LYP/6-31G* energetics and the B3LYP/3-21G~dz entropy. Reprinted with permission from [150], Z. Slanina et al., *Chem. Phys. Lett.* 372, 810 (2003). © 2003, Elsevier.

(not shown in Fig. 20), the structure lowest in the $\Delta H^{\circ}_{0,i}$ scale must be prevailing. However, already at a temperature of 226 K (that has no practical meaning), the relative concentrations of the (c) and (b) structures are interchanged, and beyond the point, the (b) structure is always somewhat more populated. Even more interesting is the behavior of the IPR-satisfying (a) structure. As the structure is the highest in potential energy, it must be the least populated species at low temperatures. However, later on the entropy contributions (low symmetry, some lower vibrational frequencies, and some lower electronic excitation energies) elevate the (a) isomer into the status of a minor isomer that could also be observed. On the other hand, the (d) isomer has the least chance to be detected. Interestingly enough, the concentration order at high temperatures for Ca@C_{72} is quite similar to that previously computed [151] for Mg@C_{72}.

In contrast to Ca@C_{72}, Ca@C_{74} was not only isolated [147] but even its structure was determined [153]. According to the ^{13}C NMR spectra recorded by Achiba et al. [153], Ca@C_{74} exhibits D_{3h} symmetry of its cage. There is only one IPR structure possible [97] for the empty C_{74}. The sole C_{74} IPR cage has D_{3h} symmetry. The Ca location in the C_{74} IPR cage was computed [41, 147, 149] to be off the three-fold axis. The metal atom is instead placed on an orthogonal two-fold axis. The two carbon atoms nearest to Ca are connected by a bond between two cage hexagons so that the system has C_{2v} symmetry. Two non-IPR C_{74} cages (one pair of connected pentagons in each cage) were also computed [149] for Ca encapsulation, but the resulting endohedrals are higher in energy. This indicates a difference from the Ca@C_{72} system [150] with two non-IPR C_{72} cages especially convenient for the Ca encapsulation.

A set of six Ca@C_{74} isomers was subjected to the stability computations [154] along the same lines [150] as Ca@C_{72} (though the electronic excitation energies were evaluated by

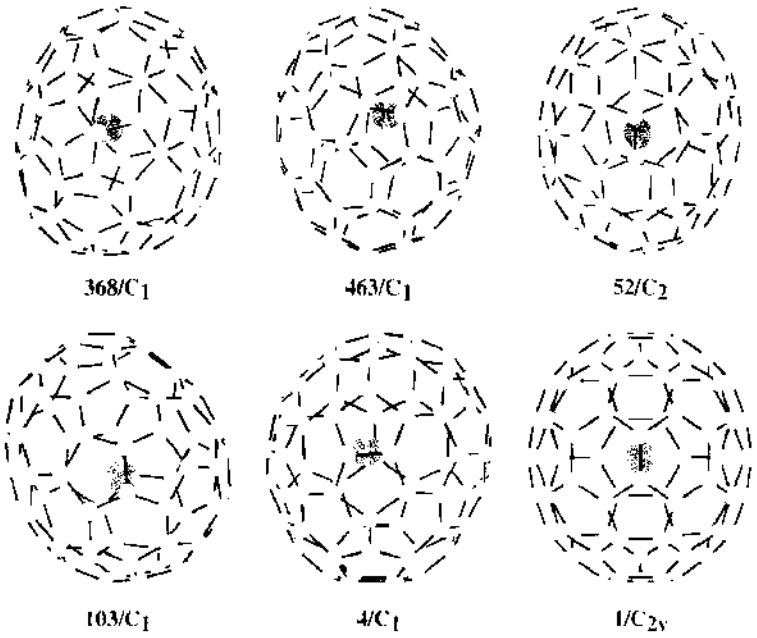


Figure 21. B3LYP/3-21G~dz optimized structures [154] of Ca@C_{74} isomers. Reprinted with permission from [154], Z. Slanina et al., *Chem. Phys.* 301, 153 (2004). © 2004, Elsevier.

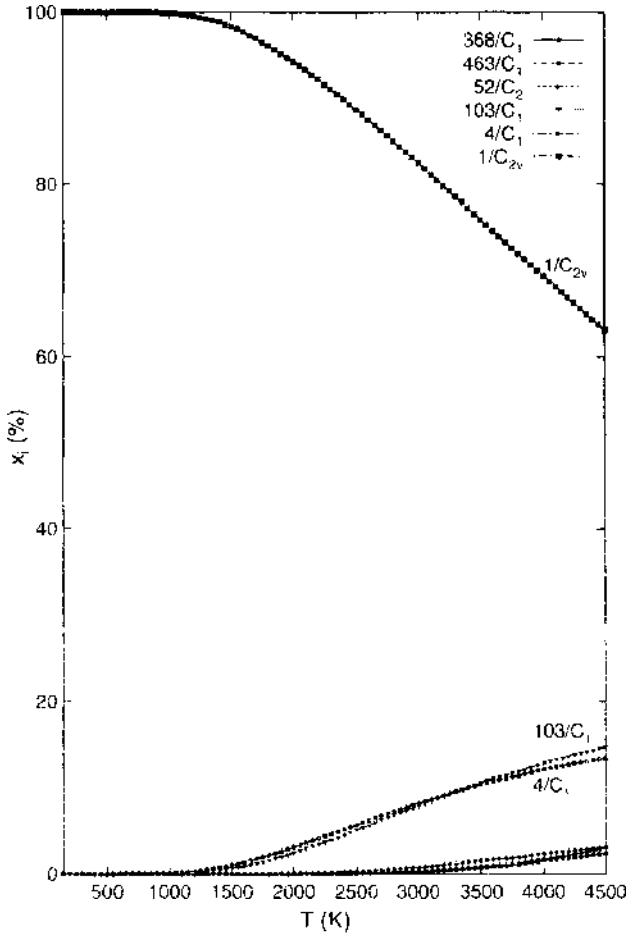


Figure 22. Relative concentrations [154] of the Ca@C_{74} isomers based on the B3LYP/6-31G* energetics and the B3LYP/3-21G ~dz entropy. Reprinted with permission from [154], Z. Slanina et al., *Chem. Phys.* 301, 153 (2004). © 2004, Elsevier.

the ZINDO method [362, 363]). The set consists of the three isomers considered in initial computations [147, 149] to which three cages were added (Fig. 12) that exhibited some non-negligible populations in evaluations [365] of the empty C_{74}. Figure 21 presents [154] the B3LYP/3-21G~dz optimized structures of $Ca@C_{74}$.

Temperature development [154] of the relative concentrations of the six $Ca@C_{74}$ isomers is shown in Fig. 22. It turns out that the $1/C_{2v}$ species has not only favorable enthalpy but also entropy term, and thus all the remaining isomers can act as minor species at best. In the IPR C_{74} structure with the D_{3h} symmetry, there are obviously three equivalent sites for location of Ca, being linked by rotations around the three-fold axis. It makes sense to find a transition state that separates two such equivalent minima. Clearly enough, the structure is reached by a 60° rotation around the three-fold axis. The structure is an activated complex indeed as it exhibits [154] just one imaginary frequency (76i cm $^{-1}$). The calculations give for the potential barrier a low value of 7–9 kcal/mol, thus placing it just between the well-known barriers for the ammonia inversion [413] and bullvalene autoisomerization [414]. Such low kinetic barrier should enable the $Ca@C_{74}$ autoisomerizations at moderate temperatures; the rearrangements can actually be seen in the NMR spectra [153]. The motion produces an averaging and thus the D_{3h} symmetry deduced from the ^{13}C NMR spectra.

Whereas $Ca@C_{74}$ is represented by just one structure, the third illustrative system, $Ca@C_{82}$, exhibits the richest isomerism among the Ca endohedrals [147, 155, 157, 415–417]. Shinohara et al. [155] isolated four isomers of $Ca@C_{82}$ and labeled the isomers by (I), (II), (III), and (IV). Dennis and Shinohara concluded [158, 418] from the ^{13}C NMR spectra of

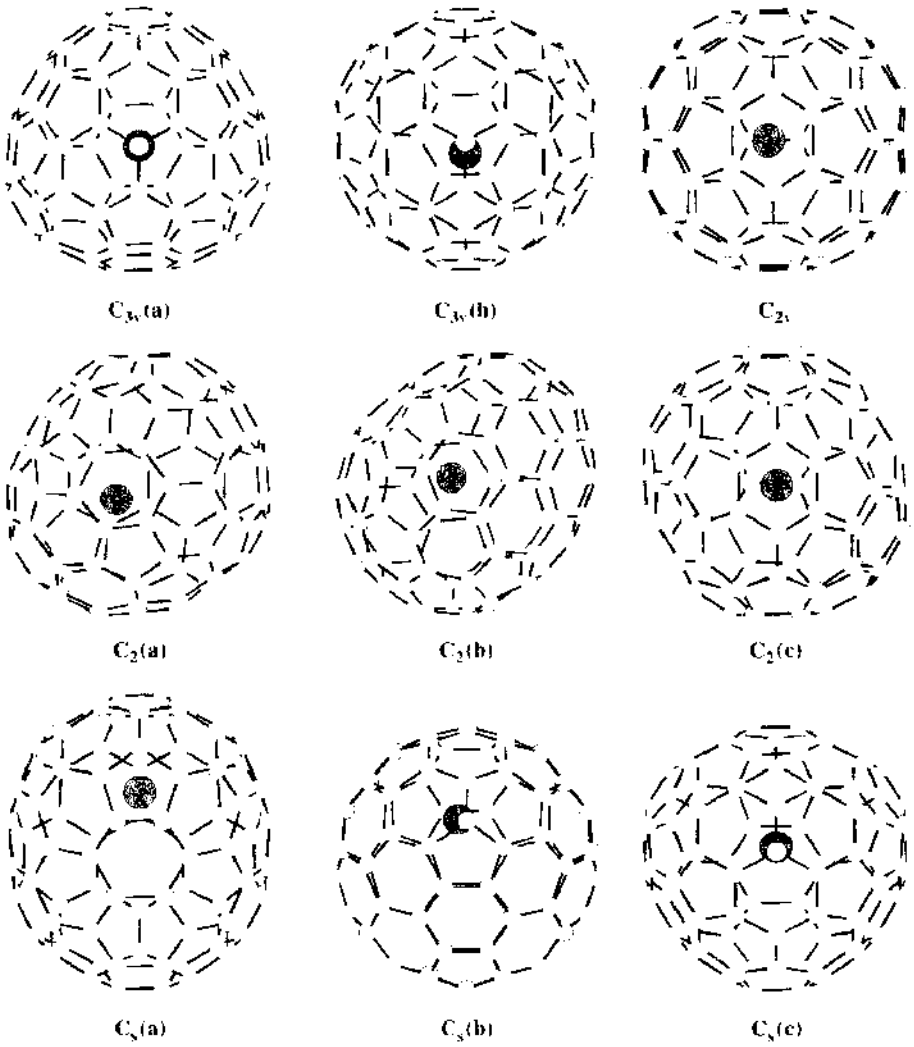


Figure 23. B3LYP/3-21G~dz optimized [161] structures of $Ca@C_{82}$ isomers. Reprinted with permission from [161]. Z. Slanina et al., *J. Chem. Phys.* 120, 3397 (2004). © 2004, American Institute of Physics.

Ca@C_{82}(III) its symmetry as C_2. The ultraviolet photoelectron spectra measured by Hino et al. [159] support the finding; a similarity with Tm@C_{82}(II) was also noted [167]. Very recently, Achiba et al. measured [160] the ^{13}C NMR spectra of all four isomers and assigned the symmetry of isomers (I), (II), (III), and (IV) as C_s, C_{3v}, C_2, and C_{2v}, respectively.

The Ca@C_{82} structure–energetics relationships were also computed [156], and a qualitative agreement with the experiment was found [41]. The computations were performed at the HF and DFT levels, and in both cases the C_{2v} structure was the lowest isomer in potential energy. There were still three other low-energy species: C_s, C_2, and C_{3v}. The combined stability computations are also available [161] for the full set of nine isomers [156] of Ca@C_{82}.

The nine C_{82} IPR structures [97] produce nine Ca@C_{82} cages with the following symmetries recognized [156] at the HF level: C_{3v}(a), C_{3v}(b), C_{2v}, C_2(a), C_2(b), C_2(c), C_s(a), C_s(b), and C_s(c). Figure 23 presents the structures of the nine Ca@C_{82} isomers reoptimized [161] at the B3LYP/3-21G~dz level. It has turned out that in five cases, the original HF structures after the DFT reoptimizations within the same symmetry lead to saddle points with imaginary vibrational frequencies, not to the required local energy minima. When the five saddle points are relaxed and reoptimized, the following local minima are obtained: C_{3v}(b) → C_s, C_{2v} → C_s, C_2(a) → C_1, C_2(b) → C_1, C_s(b) → C_1.

Figure 24 presents the temperature development [161] of the relative concentrations of the nine Ca@C_{82} isomers in a wide temperature region. The enthalpy part of the Gibbs energy is taken from the B3LYP/6-31G*//B3LYP/3-21G~dz calculations; the entropy part is evaluated at the B3LYP/3-21G~dz level. At very low temperatures, the structure lowest in the $\Delta H^\circ_{0,i}$ scale must be prevailing. However, at a temperature of 1700 K, the relative concentrations of the C_{2v} → C_s and C_s(c) structures are interchanged, and beyond the point, the C_s(c) structure is always somewhat more populated. The C_s(c) isomer and also C_2(c)

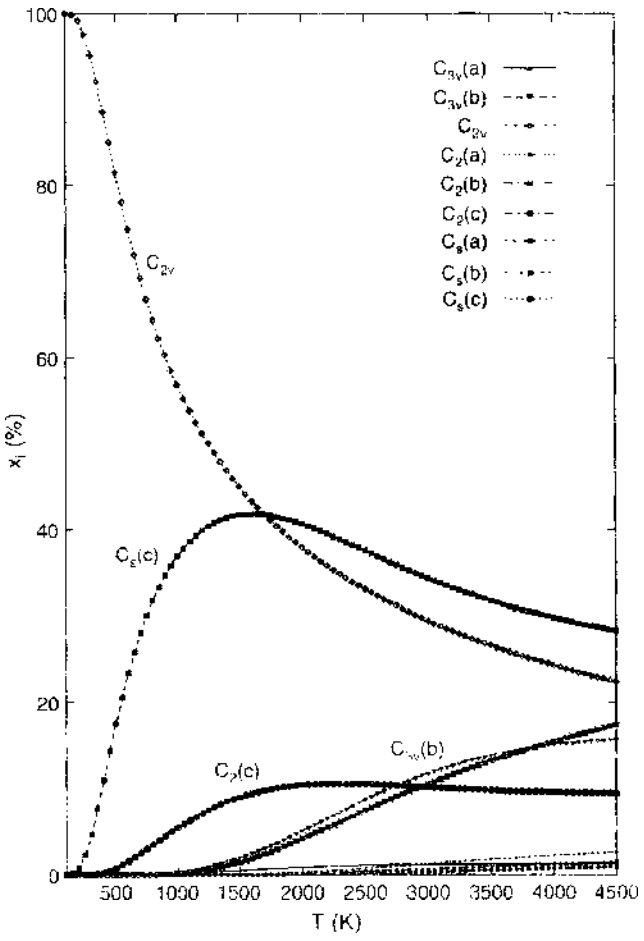


Figure 24. Relative concentrations [161] of the Ca@C_{82} isomers based on the B3LYP/6-31G* energetics and the B3LYP/3-21G~dz entropy. Reprinted with permission from [161], Z. Slanina et al., *J. Chem. Phys.* 120, 3397 (2004), © 2004, American Institute of Physics.

exhibit a temperature maximum. Then, there are still two other structures with significant populations at high temperatures: $C_{3v}(b) \rightarrow C_s$ and $C_s(b) \rightarrow C_1$. Although the former species is a bit more populated, their concentrations are rather close. In contrast to the five species with significant populations at least in some temperature regions, the remaining four isomers are computed to be negligible at any temperature. The $C_{3v}(a)$ structure is remarkable as it is lower in the potential energy than $C_{3v}(b) \rightarrow C_s$; however, it is actually suppressed by entropy. The observed yields [160] of the isomers were nearly equal except for the considerably less produced C_{3v} species (though the HPLC chromatograms [155] could indicate somewhat larger production differences). Figure 24 is in reasonable agreement with the qualitative population information in a relatively wide temperature interval. However, the fifth isomer also pointed out by the computations, $C_s(b) \rightarrow C_1$, has not been observed. The fifth species could be hidden in a chromatographic fraction as a minor component.

9. FULLERENE DERIVATIVES

Halogenated (especially fluorinated) and hydrogenated fullerenes represent [419] prominent families of fullerene derivatives. As their preparation can be carried out at higher temperatures, entropy term should have a significant role. An instructive example is supplied by $C_{60}F_{36}$ that was prepared [420] by heating C_{60} and MnF_3 at a temperatue of 623 K. Two isomers were recognized [420]: C_3 and T. The structure of the C_3 isomer has only recently been clarified [421] by ^{19}F NMR, whereas the T isomer was characterized [422] by x-ray. Moreover, a third isomer with a C_1 symmetry has very recently been isolated [423] (for Schlegel diagrams, see Refs. [420–424] and also Fig. 25). The C_1 species undergoes a

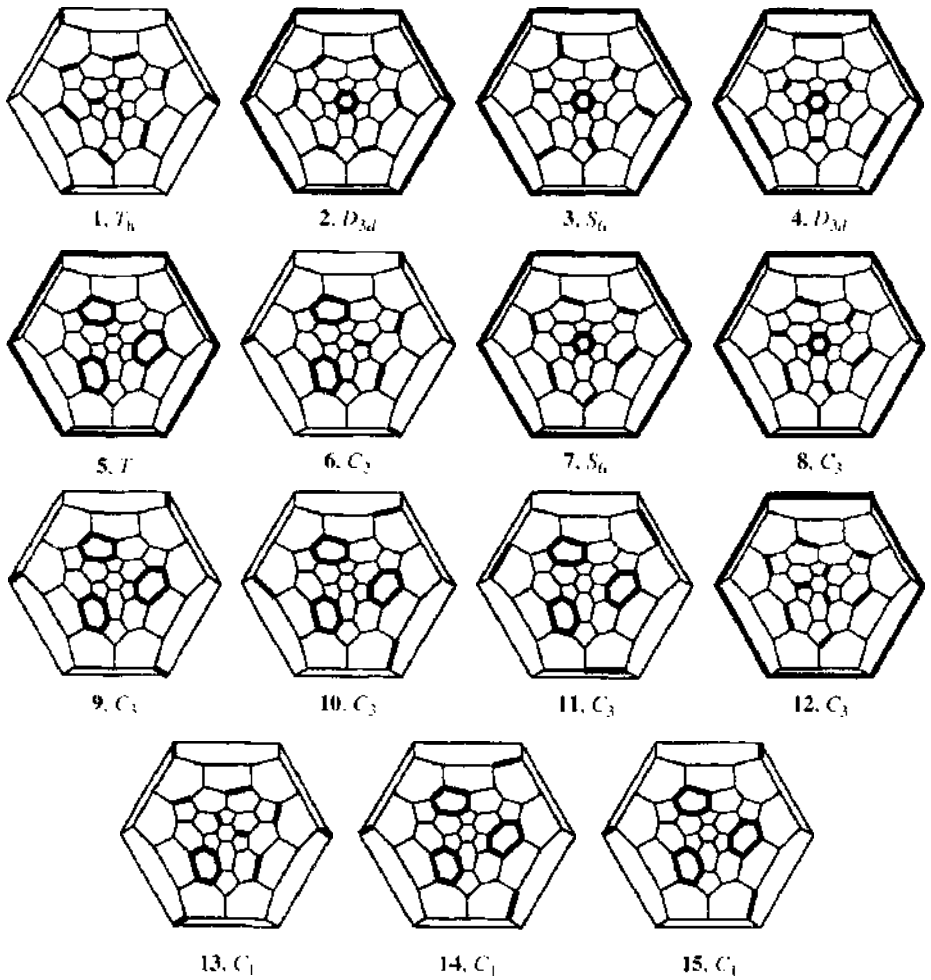


Figure 25. Schlegel diagrams of 15 isomers of $C_{60}X_{36}$; heavy lines indicate C=C bonds or benzene rings. Reprinted with permission from [424]. J. Nossal et al., *J. Am. Chem. Soc.* 123, 8482 (2001). © 2001, American Chemical Society.

remarkable room-temperature rearrangement and establishes equilibrium with the C_3 isomer in solution [425].

$C_{60}F_{36}$ isomers have also been computed with a conclusion [426–428] that there are four symmetric low-energy isomers for $C_{60}F_{36}$: C_3, T, S_6, and C_2 symmetry, and then one non-symmetric C_1 isomer [423]. However, the computed energetics somewhat depends on the applied methods [429, 430], and the final stability picture is influenced [173] by entropy. Further computations were thus carried out [174] at the B3LYP/6-31G* level. The PM3 and B3LYP/6-31G* separation energies are somewhat different, as it was also found [423, 427] for the AM1 method. While in the AM1 and PM3 methods, the C_3 structure is the lowest cage; at the B3LYP/6-31G* level the T isomer comes as the species lowest in the potential energy.

There are two observed concentration ratios that can help to resolve the issue. The C_3:T concentration ratio [421] for a temperature of 623 K is 12:1 and the C_3:C_1 ratio [425] at room temperature is 6:1. The computed entropies can be used within an inverse problem, that is, deduction [273] of the separation energetics from the observed concentrations and computed partition functions. For example [174], the C_3:C_1 ratio of 6:1 at room temperature requires the C_1–C_3 potential-energy separation of 1.68 kcal/mol. As it is known [425] that the C_1 isomer is mostly converted to the C_3 isomer in solution, we can actually require that (C_3 + C_1):T is equal to 12:1 (if we neglect the residual, nonconverted amount of C_1). Figure 26 presents the temperature development based on the energy fit. The S_6 and C_2 isomers should always be negligible (the S_6 structure is obviously suppressed by entropy), and indeed, they have never been observed.

$C_{60}H_{36}$ is historically the first prepared [431] derivative of C_{60}. Although $C_{60}H_{36}$ was originally produced by the Birch reduction, the system can also be prepared [424, 432–434] by hydrogenation at higher temperatures up to 700 K. $C_{60}H_{36}$ and other hydrides have

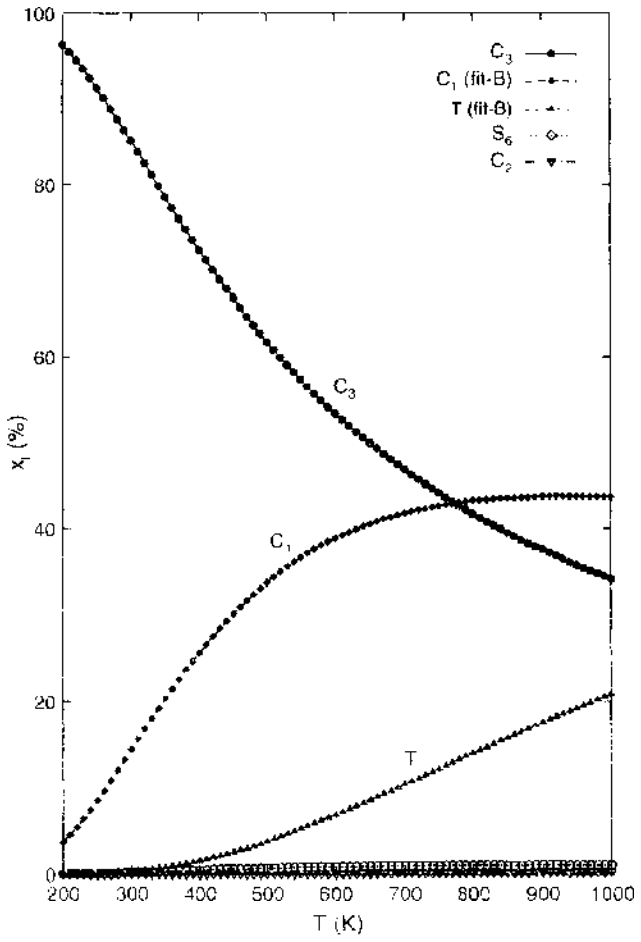


Figure 26. Relative concentrations of the five $C_{60}F_{36}$ isomers based on the energetics refined [174] to the observed C_3:C_1 and C_3:T ratios

extensively been calculated [424, 426–428, 435–438], too. Recently, 14 $C_{60}H_{36}$ isomers were computed [424] at the B3LYP/6-31G* level in order to assist interpretation of NMR spectra. Although the degree of parallelism between fluorination and hydrogenation is not yet clear, it makes sense to apply [172] the new C_1 $C_{60}F_{36}$ topology [174, 423] also to the $C_{60}H_{36}$ case.

The original 14-membered $C_{60}H_{36}$ isomeric set [424] was limited [172] to 8 structures that have low energies either in the AM1 or B3LYP/6-31G* method or in both. Then, the ninth isomer was added—the C_1 $C_{60}H_{36}$ species derived from the observed C_1 $C_{60}F_{36}$ topology [423] (label F/C_1 in Fig. 27).

Similar to the $C_{60}F_{36}$ case [174], the PM3 and B3LYP/6-31G* energetics do not agree throughout. The PM3 method points [172] out the 7/S_6 isomer as the lowest energy species in the set. However, this semiempirical result does not agree with the prediction from the B3LYP/6-31G* treatment [424] where the 5/T structure falls 15 kcal/mol below the 7/S_6 isomer.

The PM3 relative energies together with the PM3 evaluated partition functions produce a temperature development [172] of the relative concentrations in which the F/C_1 structure becomes the second most populated at high temperatures (Fig. 27). The F/C_1 isomer is actually only the fourth lowest in the PM3 potential energy, but its less favorable energy term is compensated by its entropy. Clearly enough, if the PM3 energetics is replaced by the B3LYP/6-31G* data, the picture will be dominated [172] by the 5/T structure. More experimental data are needed before the complex system can be fully understood. The search for low-energy $C_{60}H_{36}$ isomers should also continue further, given the enormous number of possible topologies [321]—there are 600, 873, 146, 368, 170 topological isomers of $C_{60}X_{36}$.

$C_{60}F_{48}$ represents another fullerene derivative subjected to the combined stability treatment. $C_{60}F_{48}$ has been known [429, 439–442] to exist in two isomeric forms of D_3 and

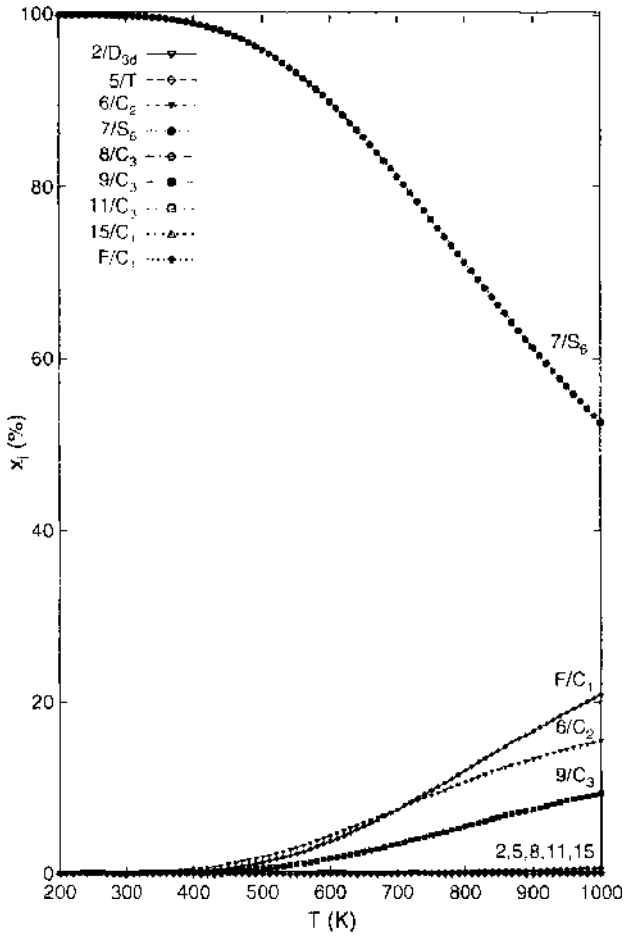


Figure 27. Relative concentrations [172] of the nine $C_{60}H_{36}$ isomers based on the PM3-computed energy and entropy. Reprinted with permission from [172], Z. Slanina, K. Kobayashi, and S. Nagase, *Chem. Phys. Lett.* 382, 211 (2003), © 2003, Elsevier.

S_6 symmetries. However, the quantum-chemical calculations [175, 429, 441, 442] have not agreed on their stability order, though a near-isoenergetic picture is otherwise always encountered. In order to clarify the situation, the entropy effects were evaluated [175] for synthetic temperatures of about 500 K. The PM3 entropy-based inverse problem [273] suggests that the D_3 isomer should be more stable in the potential energy by 2.05–2.55 kcal/mol (to which term the *ab initio* data are closer [175] than the semiempirical ones).

10. CONCLUSIONS

This survey shows the ongoing productive interaction between theory and experiment in fullerene research that is building a frame for nanotechnological applications of fullerenes and nanotubes. For higher fullerenes, the mutual interaction is indeed essential, given the complexity of the systems. The considerable thermal effects on the relative isomeric populations revealed by the quantum-chemical computations result from a complex interplay between rotational, vibrational, electronic, potential-energy terms, chirality factors, and so forth. Such effects would never be seen if only energetics is considered and entropy neglected. The treatment is, however, built on the presumption of the interisomeric thermodynamic equilibrium. We do not know yet to which degree this presumption is satisfied in reality. We can only acknowledge that the thermodynamic-equilibrium treatment has already produced a reasonable computation–observation agreement for the isomeric systems from C_{76} to C_{96}, with no serious failure. This relatively large tested set supports the belief in a still wider applicability of the equilibrium treatment. There are also other aspects of the combined treatment that should further be studied—especially anharmonicity of vibrations, motions of encapsulates in cages, and reliability of interisomeric energetics. The non-IPR structures have not sufficiently been explored yet (though they may be behind the observation [443] of possible isomers of C_{60} and C_{70}). Heterofullerenes, like cages [205, 444] based on boron nitride, should also attract more interest. Once the thermodynamic treatment is completely exposed, the interest will gradually shift toward complex kinetic schemes and to fullerene sets with variable stoichiometry. The steadily growing family of fullerene and metallofullerene cages will further attract interest of more applied branches of nanoscience and nanotechnology.

ACKNOWLEDGMENTS

The reported research was supported by a grant-in-aid for NAREGI Nanoscience Project, Scientific Research on Priority Area (A), and Scientific Research (B), from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and also by the Japan Society for the Promotion of Science. The authors also wish to thank the following organizations for kindly permitting the reprinting of copyrighted material: American Institute of Physics; Elsevier Scientific Publishing Company; JAI Press, Inc. Kluwer Academic Publishers; Marcel Dekker, Inc.; The American Chemical Society; The Electrochemical Society, Inc.; The Royal Society; and John Wiley & Sons, Inc.

REFERENCES

1. M. Buhl and A. Hirsch, *Chem. Rev.* 101, 1153 (2001).
2. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* 318, 162 (1985).
3. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, *Nature* 347, 354 (1990).
4. J. K. Ginzewski, in "The Chemical Physics of Fullerenes 10 (and 5) Years Later" (W. Andreoni, Ed.), p. 117, Kluwer, Dordrecht, The Netherlands, 1996.
5. Z. Slanina, *Int. J. Quantum Chem.* 42, 1625 (1992).
6. W. Harneit, M. Waiblinger, C. Meyer, K. Lips, and A. Weidinger, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish, P. V. Kamat, and D. Guldi, Eds.), Fullerenes for the New Millennium, Vol. 11, p. 358, Electrochemical Society, Pennington, NJ, 2001.
7. L. Dunsch, P. Georgi, M. Krause, C.-R. Wang, T. Kai, T. Tomiyama, T. Okazaki, and H. Shinohara, in "The 201st ECS Meeting" (P. Kamat, D. Guldi, and K. Kadish, Eds.), Abstract 955, Electrochemical Society, Philadelphia, 2002.
8. C. Joachim, *Superlat. Microstruct.* 28, 305 (2000).
9. S. Ami and C. Joachim, *Nanotechnol.* 12, 44 (2001).

10. R. Stadler, S. Ami, M. Forshaw, and C. Joachim, *Nanotechnol.* 12, 350 (2001).
11. S. Iijima, *Nature* 354, 56 (1991).
12. A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker, *Science* 294, 1317 (2001).
13. A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker, *Physica E* 16, 42 (2003).
14. C. H. Lee, K. T. Kang, K. S. Park, M. S. Kim, H. S. Kim, H. G. Kim, J. E. Fischer, and A. T. Johnson, *Jpn. J. Appl. Phys. 1* 42, 5392 (2003).
15. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes," Academic Press, San Diego, CA, 1996.
16. J. Cioslowski, "Electronic Structure Calculations on Fullerenes and Their Derivatives," Oxford University Press, Oxford, 1995.
17. G. E. Scuseria, in "Modern Electronic Structure Theory, Part" (D. R. Yarkony, Ed.), p. 279, World Scientific, Singapore, 1995.
18. K. Jinno and Y. Saito, *Advan. Chromat.* 36, 65 (1996).
19. Y. Achiba, K. Kikuchi, Y. Aihara, T. Wakabayashi, Y. Miyake, and M. Kainosho, in "The Chemical Physics of Fullerenes 10 (and 5) Years Later" (W. Andreoni, Ed.), p. 139, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
20. M. Saunders, R. J. Cross, H. A. Jiménez-Vázquez, R. Shimshi, and A. Khong, *Science* 271, 1693 (1996).
21. R. Taylor, in "The Chemistry of Fullerenes" (R. Taylor, Ed.), p. 1, World Scientific, Singapore, 1995.
22. K. Heinzinger and Z. Slanina, *MPG-Spiegel* 2, 11 (1993).
23. J. Baggott, "Perfect Symmetry, The Accidental Discovery of Buckminsterfullerene," Oxford University Press, Oxford, 1994.
24. H. Aldersey-Williams, "The Most Beautiful Molecule, An Adventure in Chemistry," Aurum Press, London, 1995.
25. Z. Slanina, *Chem. Intell.* 4, 52 (1998).
26. Z. Slanina, *Int. J. Hist. Eth. Natur. Sci. Technol. Med. NTM* 9, 41 (2001).
27. J. Mattauch, H. Ewald, O. Hahn, and F. Strassmann, *Z. Phys.* 20, 598 (1943).
28. R. E. Honig, *J. Chem. Phys.* 22, 126 (1954).
29. J. Drowart, R. P. Burns, G. DeMaria, and M. G. Inghram, *J. Chem. Phys.* 31, 1131 (1959).
30. H. Hintenberger, J. Franzen, and K. D. Schuy, *Z. Naturforsch.* 18a, 1236 (1963).
31. K. S. Pitzer and E. Clementi, *J. Am. Chem. Soc.* 81, 4477 (1959).
32. S. J. Strickler and K. S. Pitzer, in "Molecular Orbitals in Chemistry, Physics, and Biology" (B. Pullman and P.-O. Löwdin, Eds.), p. 281, Academic Press, New York, 1964.
33. R. Hoffmann, *Tetrahedron* 22, 521 (1966).
34. Z. Slanina, *Radiochem. Radioanal. Lett.* 22, 291 (1975).
35. Z. Slanina and R. Zahradník, *J. Phys. Chem.* 81, 2252 (1977).
36. H. P. Schultz, *J. Org. Chem.* 30, 1361 (1965).
37. D. E. H. Jones, *New Scientist* 32, 245 (1966).
38. E. Ōsawa, *Kagaku* 25, 854 (1970); *Chem. Abstr.* 74, 75698v (1971).
39. D. A. Bochvar and E. G. Gal'pern, *Dokl. Akad. Nauk SSSR* 209, 610 (1973).
40. R. A. Davidson, *Theor. Chim. Acta* 58, 193 (1981).
41. K. Kobayashi and S. Nagase, in "Endofullerenes—A New Family of Carbon Clusters (T. Akasaka and S. Nagase, Eds.), p. 155, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
42. Z. C. Wu, D. A. Jelski, and T. F. George, *Chem. Phys. Lett.* 137, 291 (1987).
43. V. Elser and R. C. Haddon, *Nature* 325, 792 (1987).
44. R. E. Stanton and M. D. Newton, *J. Phys. Chem.* 92, 2141 (1988).
45. D. E. Weeks and W. G. Harter, *Chem. Phys. Lett.* 144 (1988) 366.
46. D. E. Weeks and W. G. Harter, *J. Chem. Phys.* 90, 4744 (1989).
47. Z. Slanina, J. M. Rudziński, M. Togasi, and E. Ōsawa, *J. Mol. Struct. (Theochem)* 61, 169 (1989).
48. D. R. Huffman and W. Krätschmer, *Materials Res. Soc. Proc.* 206, 601 (1991).
49. T Braun, *Angew. Chem., Int. Ed. Eng.* 31, 588 (1992).
50. T. Braun, A. Schubert, B. Maczelka, and L. Vasvári, "Fullerene Research 1985–1993," World Scientific, Singapore, 1995.
51. D. Bakowies and W. Thiel, *J. Am. Chem. Soc.* 113, 3704 (1991).
52. M. Häser, J. Almlöf, and G. E. Scuseria, *Chem. Phys. Lett.* 181, 497 (1991).
53. W. Weltner Jr. and R. J. van Zee, *Chem. Rev.* 89, 1713 (1989).
54. R. F. Curl and R. E. Smalley, *Science* 242, 1017 (1988).
55. H. W. Kroto, *Science* 242, 1139 (1988).
56. R. C. Haddon, *Accounts Chem. Res.* 21, 243 (1988).
57. D. A. Jelski and T. F. George, *J. Chem. Educ.* 65, 879 (1988).
58. D. J. Klein and T. G. Schmalz, in "Quasicrystals, Networks, and Molecules of Fivefold Symmetry" (I. Hargittai, Ed.), p. 239, VCH Publishers, New York, 1990.
59. E. Ōsawa, *Kagaku* 45, 552 (1990).
60. D. R. Huffman, *Phys. Today* 44, 22 (1991).
61. H. W. Kroto, A. W. Allaf, and S. P. Balm, *Chem. Rev.* 91, 1213 (1991).
62. H. W. Kroto, *Angew. Chem., Int. Ed. Engl.* 31, 111 (1992).
63. J. R. Heath, in "ACS Symposium Series No 481—Fullerenes, Synthesis, Properties, and Chemistry of Large Carbon Clusters" (G. S. Hammond and V. J. Kuck, Eds.), p. 1, ACS, Washington, DC, 1992.

64. J. E. Fischer, P. A. Heiney, D. E. Luzzi, and D. E. Cox, in "ACS Symposium Series No 481—Fullerenes. Synthesis, Properties, and Chemistry of Large Carbon Clusters" (G. S. Hammond and V. J. Kuck, Eds.). ACS, Washington, DC, 1992.
65. Z. Slanina, *Chem. Listy* 86, 327 (1992).
66. R. E. Smalley, *Accounts Chem. Res.* 25, 98 (1992).
67. J. P. Hare and H. W. Kroto, *Accounts Chem. Res.* 25, 106 (1992).
68. J. E. Fischer, P. A. Heiney, and A. B. Smith III, *Accounts Chem. Res.* 25, 112 (1992).
69. F. Diederich and R. L. Whetten, *Accounts Chem. Res.* 25, 119 (1992).
70. R. C. Haddon, *Accounts Chem. Res.* 25, 127 (1992).
71. P. J. Fagan, J. C. Calabrese, and B. Malone, *Accounts Chem. Res.* 25, 134 (1992).
72. J. H. Weaver, *Accounts Chem. Res.* 25, 143 (1992).
73. J. M. Hawkins, *Accounts Chem. Res.* 25, 150 (1992).
74. F. Wudl, *Accounts Chem. Res.* 25, 157 (1992).
75. S. W. McElvany, M. M. Ross, and J. H. Callahan, *Accounts Chem. Res.* 25, 162 (1992).
76. R. D. Johnson, D. S. Bethune, and C. S. Yannoni, *Accounts Chem. Res.* 25, 169 (1992).
77. D. Koruga, S. Hameroff, J. Withers, R. Loutfy, and M. Sundareshan, "Fullerene C_{60}—History, Physics, Nanobiology, Nanotechnology," Elsevier, Amsterdam, The Netherlands, 1993.
78. M. L. Cohen and V. H. Crespi, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 197, VCH Publishers, New York, 1993.
79. S. C. Erwin, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 217, VCH Publishers, New York, 1993.
80. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *J. Mater. Res.* 8, 2054 (1993).
81. T. G. Schmalz, and D. J. Klein, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 83, VCH Publishers, New York, 1993.
82. G. E. Scuseria, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 103, VCH Publishers, New York, 1993.
83. C. T. White, J. W. Mintmire, R. C. Mowrey, D. W. Brenner, D. H. Robertson, J. A. Harrison, and B. I. Dunlap, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 125, VCH Publishers, New York, 1993.
84. R. C. Haddon and K. Raghavachari, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), p. 185, VCH Publishers, New York, 1993.
85. J. Cioslowski, *Rev. Comput. Chem.* 4, 1 (1993).
86. C. Z. Wang, B. L. Zhang, K. M. Ho, and X. Q. Wang, *Int. J. Mod. Phys. B* 7, 4305 (1993).
87. J. González, F. F. Guinea, and M. A. H. Vozmediano, *Int. J. Mod. Phys. B* 7, 4331 (1993).
88. W. Andreoni, in "Electronic Properties of Fullerenes (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), p. 85, Springer-Verlag, Berlin, 1993.
89. Z. Slanina, S.-L. Lee, and C.-H. Yu, *Rev. Comput. Chem.* 8, 1 (1996).
90. J. Cioslowski, N. Rao, and D. Moncrieff, *J. Am. Chem. Soc.* 122, 8265 (2000).
91. Y. Achiba, K. Kikuchi, Y. Aihara, T. Wakabayashi, Y. Miyake, and M. Kainosho, in "Science and Technology of Fullerene Materials" (P. Bernier, D. S. Bethune, L. Y. Chiang, T. W. Ebbesen, R. M. Metzger, and J. W. Mintmire, Eds.), p. 3, Materials Research Society, Pittsburgh, 1995.
92. Y. Achiba, *Kagaku* 52, 15 (1997).
93. R. Mitsumoto, H. Oji, Y. Yamamoto, K. Asato, Y. Ouchi, H. Shinohara, K. Seki, K. Umishita, S. Hino, S. Nagase, K. Kikuchi, and Y. Achiba, *J. Phys. IV* 7, C2-525 (1997).
94. H. W. Kroto, *Nature* 329, 529 (1987).
95. T. G. Schmalz, W. A. Seitz, D. J. Klein, and G. E. Hite, *J. Am. Chem. Soc.* 110, 1113 (1988).
96. D. E. Manolopoulos and P. W. Fowler, *J. Chem. Phys.* 96, 7603 (1992).
97. P. W. Fowler and D. E. Manolopoulos, "An Atlas of Fullerenes," Clarendon Press, Oxford, 1995.
98. M. Yoshida and E. Ōsawa, *Bull. Chem. Soc. Jpn.* 68, 2073 (1995).
99. M. Yoshida and E. Ōsawa, *Bull. Chem. Soc. Jpn.* 68, 2083 (1995).
100. Z. Slanina, M.-L. Sun, S.-L. Lee, and L. Adamowicz, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish and R. S. Ruoff, Eds.), Vol. 2, p. 1138, The Electrochemical Society, Pennington, NJ, 1995.
101. R. Ettl, I. Chao, F. N. Diederich, and R. L. Whetten, *Nature* 353, 149 (1991).
102. D. E. Manolopoulos, *J. Chem. Soc., Faraday Trans.* 87, 2861 (1991).
103. H. P. Cheng and R. L. Whetten, *Chem. Phys. Lett.* 197, 44 (1992).
104. Q. Li, F. Wudl, C. Thilgen, R. L. Whetten, and F. Diederich, *J. Am. Chem. Soc.* 114, 3994 (1992).
105. S. Hino, K. Matsumoto, S. Hasegawa, H. Inokuchi, T. Morikawa, T. Takahashi, K. Seki, K. Kikuchi, S. Suzuki, I. Ikemoto, and Y. Achiba, *Chem. Phys. Lett.* 197, 38 (1992).
106. G. Orlandi, F. Zerbetto, P. W. Fowler, and D. E. Manolopoulos, *Chem. Phys. Lett.* 208, 441 (1993).
107. J. M. Hawkins and A. Meyer, *Science* 260, 1918 (1993).
108. S. J. Austin, P. W. Fowler, G. Orlandi, D. E. Manolopoulos, and F. Zerbetto, *Chem. Phys. Lett.* 226, 219 (1994).
109. R. H. Michel, H. Schreiber, R. Gierden, F. Hennrich, J. Rockenberger, R. D. Beck, M. M. Kappes, C. Lehner, P. Adelmann, and J. F. Armbruster, *Ber. Bunsenges. Phys. Chem.* 98, 975 (1994).
110. R. H. Michel, M. M. Kappes, P. Adelmann, and G. Roth, *Angew. Chem., Int. Ed. Engl.* 33, 1651 (1994).
111. F. Diederich, R. L. Whetten, C. Thilgen, R. Ettl, I. Chao, and M. M. Alvarez, *Science* 254, 1768 (1991).

1112. K. Kikuchi, N. Nakahara, T. Wakabayashi, S. Suzuki, H. Shiromaru, Y. Miyake, K. Saito, I. Ikemoto, M. Kainosho, and Y. Achiba, *Nature* 357, 142 (1992).
1113. D. Bakowies, A. Geleßus, and W. Thiel, *Chem. Phys. Lett.* 197, 324 (1992).
1114. Z. Slanina, J.-P. François, D. Bakowies, and W. Thiel, *J. Mol. Struct. (Theochem)* 279, 213 (1993).
1115. S. J. Woo, E. Kim, and Y. H. Lee, *Phys. Rev. B* 47, 6721 (1993).
1116. K. Nakao, N. Kurita, and M. Fujita, *Phys. Rev. B* 49, 11415 (1994).
1117. M.-L. Sun, Z. Slanina, S.-L. Lee, F. Uhlík, and L. Adamowicz, *Chem. Phys. Lett.* 246, 66 (1995).
1118. F. H. Hennrich, R. H. Michel, A. Fischer, S. Richard-Schneider, S. Gilb, M. M. Kappes, D. Fuchs, M. Bürk, K. Kobayashi, and S. Nagase, *Angew. Chem., Int. Ed. Engl.* 35, 1732 (1996).
1119. C.-R. Wang, T. Sugai, T. Kai, T. Tomiyama, and H. Shinohara, *J. Chem. Soc., Chem. Commun.* 557 (2000).
1120. K. Kikuchi, N. Nakahara, T. Wakabayashi, M. Honda, H. Matsumiya, T. Moriwaki, S. Suzuki, H. Shiromaru, K. Saito, K. Yamauchi, I. Ikemoto, and Y. Achiba, *Chem. Phys. Lett.* 188, 177 (1992).
1121. S. Nagase, K. Kobayashi, T. Kato, and Y. Achiba, *Chem. Phys. Lett.* 201, 475 (1993).
1122. G. Orlandi, F. Zerbetto, and P. W. Fowler, *J. Phys. Chem.* 97, 13575 (1993).
1123. S. Nagase and K. Kobayashi, *Chem. Phys. Lett.* 214, 57 (1993).
1124. Z. Slanina, S.-L. Lee, K. Kobayashi, and S. Nagase, *J. Mol. Struct. (Theochem)* 339, 89 (1995).
1125. M. Saunders, H. A. Jiménez-Vázquez, R. J. Cross, W. E. Billups, C. Gesenberg, A. Gonzalez, W. Luo, R. C. Haddon, F. Diederich, and A. Herrmann, *J. Am. Chem. Soc.* 117, 9305 (1995).
1126. F. Diederich, R. Ettl, Y. Rubin, R. L. Whetten, R. Beck, M. Alvarez, S. Anz, D. Sensharma, F. Wudl, K. C. Khemani, and A. Koch, *Science* 252, 548 (1991).
1127. R. D. Beck, P. S. John, M. M. Alvarez, F. Diederich, and R. L. Whetten, *J. Phys. Chem.* 95, 8402 (1991).
1128. K. Kikuchi, N. Nakahara, M. Honda, S. Suzuki, K. Saito, H. Shiromaru, K. Yamauchi, I. Ikemoto, T. Kuramochi, S. Hino, and Y. Achiba, *Chem. Lett.* 1607 (1991).
1129. K. Raghavachari and C. M. Rohlfing, *J. Phys. Chem.* 95, 5768 (1991).
1130. Z. Slanina, J.-P. François, M. Kolb, D. Bakowies, and W. Thiel, *Fullerene Sci. Technol.* 1, 221 (1993).
1131. T. J. S. Dennis, T. Kai, T. Tomiyama, and H. Shinohara, *Chem. Commun.* 619 (1998).
1132. E. Ōsawa, H. Ueno, M. Yoshida, Z. Slanina, X. Zhao, M. Nishiyama, and H. Saito, *J. Chem. Soc., Perkin Trans. 2* 943 (1998).
1133. Z. Slanina, S.-L. Lee, M. Yoshida, and E. Ōsawa, *Chem. Phys.* 209, 13 (1996).
1134. Z. Slanina, S.-L. Lee, M. Yoshida, and E. Ōsawa, in "Physics and Chemistry of Fullerenes and Their Derivatives" (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), p. 389. World Scientific, Singapore, 1996.
1135. Z. Slanina, S.-L. Lee, M. Yoshida, and E. Ōsawa, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials, Vol. 3" (K. M. Kadish and R. S. Ruoff, Eds.), p. 967. The Electrochemical Society, Pennington, NJ, 1996.
1136. Z. Slanina, S.-L. Lee, and L. Adamowicz, *Int. J. Quantum Chem.* 63, 529 (1997).
1137. Z. Slanina, X. Zhao, S.-L. Lee, and E. Ōsawa, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials, Vol. 4" (K. M. Kadish and R. S. Ruoff, Eds.), p. 680. The Electrochemical Society, Pennington, NJ, 1997.
1138. Z. Slanina, X. Zhao, S.-L. Lee, and E. Ōsawa, *Chem. Phys.* 219, 193 (1997).
1139. Z. Slanina, X. Zhao, P. Deota, and E. Ōsawa, *J. Mol. Model.* 6, 312 (2000).
1140. X. Zhao, Z. Slanina, H. Goto, and E. Ōsawa, *J. Chem. Phys.* 118, 10534 (2003).
1141. X. Zhao and Z. Slanina, in "Fullerenes, Vol. 12: The Exciting World of Nanocages and Nanotubes" (P. Kamat, D. Guldi, and K. Kadish, Eds.), p. 679. The Electrochemical Society, Pennington, NJ, 2002.
1142. X. Zhao and Z. Slanina, *J. Mol. Struct. (Theochem)* 636, 195 (2003).
1143. X. Zhao, Z. Slanina, M. Ozawa, and E. Ōsawa, P. Deota, K. Tanabe, *Fullerene Sci. Technol.* 8, 595 (2000).
1144. Z. Slanina, X. Zhao, and E. Ōsawa, *Chem. Phys. Let.* 290, 311 (1998).
1145. Z. Slanina, F. Uhlík, X. Zhao, and E. Ōsawa, *J. Chem. Phys.* 113, 4933 (2000).
1146. S. A. Varganov, P. V. Avramov, S. G. Ovchinnikov, and M. S. Gordon, *Chem. Phys. Lett.* 362, 380 (2002).
1147. T. S. M. Wan, H. W. Zhang, T. Nakane, Z. D. Xu, M. Inakuma, H. Shinohara, K. Kobayashi, and S. Nagase, *J. Am. Chem. Soc.* 120, 6806 (1998).
1148. K. Kobayashi, S. Nagase, M. Yoshida, and E. Ōsawa, *J. Am. Chem. Soc.* 119, 12693 (1997).
1149. S. Nagase, K. Kobayashi, and T. Akasaka, *J. Mol. Struct. (Theochem)* 462, 97 (1999).
1150. Z. Slanina, K. Kobayashi, and S. Nagase, *Chem. Phys. Lett.* 372, 810 (2003).
1151. Z. Slanina, X. Zhao, X. Grabuleda, M. Ozawa, F. Uhlík, P. M. Ivanov, K. Kobayashi, and S. Nagase, *J. Mol. Graphics Mod.* 19, 252 (2001).
1152. Z. Slanina, F. Uhlík, L. Adamowicz, K. Kobayashi, and S. Nagase, *Int. J. Quantum Chem.* 96, (2003).
1153. T. KoJama, R. Fujii, Y. Miyake, S. Suzuki, H. Nishikawa, I. Ikemoto, K. Kikuchi, and Y. Achiba, in "Fullerenes, Vol. 13: Fullerenes and Nanotubes: The Building Blocks of Next Generation Nanodevices" (D. M. Guldi, P. V. Kamat, and F. D'Souza, Eds.), p. 548. The Electrochemical Society, Pennington, NJ, 2003.
1154. Z. Slanina, K. Kobayashi, and S. Nagase, *Chem. Phys.* 301, 153 (2004).
1155. Z. D. Xu, T. Nakane, and H. Shinohara, *J. Am. Chem. Soc.* 118, 11309 (1996).
1156. K. Kobayashi and S. Nagase, *Chem. Phys. Lett.* 274, 226 (1997).
1157. T. Kimura, T. Sugai, and H. Shinohara, *Int. J. Mass Spectrom.* 188, 225 (1999).
1158. T. J. S. Dennis and H. Shinohara, *Appl. Phys. A* 66, 243 (1998).
1159. S. Hino, K. Umishita, K. Iwasaki, M. Aoki, K. Kobayashi, S. Nagase, T. J. S. Dennis, T. Nakane, and H. Shinohara, *Chem. Phys. Lett.* 337, 65 (2001).

160. T. Kodama, R. Fujii, Y. Miyake, K. Sakaguchi, H. Nishikawa, I. Ikemoto, K. Kikuchi, and Y. Achiba, *Chem. Phys. Lett.* 377, 197 (2003).

161. Z. Slanina, K. Kobayashi, and S. Nagase, *J. Chem. Phys.* 120, 3397 (2004).

162. K. Kobayashi and S. Nagase, *Chem. Phys. Lett.* 282, 325 (1998).

163. T. Akasaka, T. Wakahara, S. Nagase, K. Kobayashi, M. Waelchli, K. Yamamoto, M. Kondo, S. Shirakura, S. Okubo, Y. Maeda, T. Kato, M. Kako, Y. Nakadaira, R. Nagahata, X. Gao, E. van Caemelbecke, and K. M. Kadish, *J. Am. Chem. Soc.* 122, 9316 (2000).

164. T. Akasaka, T. Wakahara, S. Nagase, K. Kobayashi, M. Waelchli, K. Yamamoto, M. Kondo, S. Shirakura, Y. Maeda, T. Kato, M. Kako, Y. Nakadaira, X. Gao, E. van Caemelbecke, and K. M. Kadish, *J. Phys. Chem. B* 105, 2971 (2001).

165. T. Wakahara, S. Okubo, M. Kondo, Y. Maeda, T. Akasaka, M. Waelchli, M. Kako, K. Kobayashi, S. Nagase, T. Kato, K. Yamamoto, X. Gao, E. van Caemelbecke, and K. M. Kadish, *Chem. Phys. Lett.* 360, 235 (2002).

166. K. Kobayashi and S. Nagase, *Mol. Phys.* 101, 249 (2003).

167. T. Kodama, N. Ozawa, Y. Miyake, K. Sakaguchi, H. Nishikawa, I. Ikemoto, K. Kikuchi, and Y. Achiba, *J. Am. Chem. Soc.* 124, 1452 (2002).

168. B. Cao, M. Hasegawa, K. Okada, T. Tomiyama, T. Okazaki, K. Suenaga, and H. Shinohara, *J. Am. Chem. Soc.* 123, 9679 (2001).

169. B. Cao, K. Suenaga, T. Okazaki, and H. Shinohara, *J. Phys. Chem. B* 106, 9295 (2002).

170. C. R. Wang, P. Georgi, L. Dunsch, T. Kai, T. Tomiyama, and H. Shinohara, *Curr. Appl. Phys.* 2, 141 (2002).

171. Y. K. Kwon, D. Tomanek, and S. Iijima, *Phys. Rev. Lett.* 82, 1470 (1999).

172. Z. Slanina, K. Kobayashi, and S. Nagase, *Chem. Phys. Lett.* 382, 211 (2003).

173. Z. Slanina and F. Uhlik, *Chem. Phys. Lett.* 374, 100 (2003).

174. Z. Slanina, O. V. Boltalina, K. Kobayashi, and S. Nagase, *Fullerene Nanotub. Carb. Nanostruct.* 12, 691 (2004).

175. Z. Slanina, F. Uhlik, and L. Adamowicz, *Fullerene Nanotub. Carb. Nanostruct.* 11, 219 (2003).

176. E. Dietel, A. Hirsch, B. Pietzak, M. Waiblinger, K. Lips, A. Weidinger, A. Gruss, and K. P. Dinse, *J. Am. Chem. Soc.* 121, 2432 (1999).

177. T. Clark, "A Handbook of Computational Chemistry. A Practical Guide to Chemical Structure and Energy Calculations," John Wiley, New York, 1985.

178. D. B. Boyd, *Rev. Comput. Chem.* 1, 321 (1990).

179. J. J. P. Stewart, *Rev. Comput. Chem.* 1, 45 (1990).

180. M. C. Zerner, *Rev. Comput. Chem.* 2, 313 (1991).

181. W. J. Hehre, L. Radom, P. R. von Schleyer, and J. A. Pople, "Ab Initio Molecular Orbital Theory," John Wiley, New York, 1986.

182. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* 99, 4899 (1997).

183. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).

184. J. J. P. Stewart, *J. Comput. Chem.* 10, 209 (1989).

185. M. J. S. Dewar, C. Jie, and J. Yu, *Tetrahedron* 49, 5003 (1993).

186. AMPAC 6.0, Semichem, Shawnee, KS (1997).

187. J. J. P. Stewart, MOPAC 5.0, QCPE 455, Indiana University, Bloomington, IN (1990).

188. J. J. P. Stewart, MOPAC 2002, Fujitsu Ltd., Tokyo, Japan (1999).

189. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian 98, Revision A.11.1, Gaussian, Inc., Pittsburgh, 1998.

190. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian 03, Revision B.02, Gaussian, Inc., Pittsburgh, 2003.

191. W. J. Hehre, L. D. Burke, and A. J. Schusterman, Spartan, Wavefunction, Inc., Irvine, CA, 1993.

192. H. B. Schlegel and J. J. W. McDouall, in "Computational Advances in Organic Chemistry" (C. Ögretir and I. G. Csizmadia, Eds.), p. 167, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

193. F. Jensen, "Introduction to Computational Chemistry," John Wiley, Chichester, UK, 1999.

194. Z. Slanina, F. Uhlik, and M. C. Zerner, *Rev. Roum. Chim.* 36, 965 (1991).

195. Z. Slanina, *Int. Rev. Phys. Chem.* 6, 251 (1987).

196. Z. Slanina, *Comput. Chem.* 13, 305 (1989).

197. Z. Slanina, *Theor. Chim. Acta* 83, 257 (1992).
198. Z. Slanina, "Contemporary Theory of Chemical Isomerism." Kluwer, Dordrecht, The Netherlands, 1986.
199. Z. Slanina, F. Uhlík, J.-P. François, and E. Ōsawa, *Croat. Chem. Acta* 73, 1047 (2000).
200. H. S. M. Coxeter, "Regular Polytopes." Macmillan, New York, 1963.
201. H. S. M. Coxeter, "Introduction to Geometry." John Wiley, New York, 1969.
202. Z. Slanina, *Chem. Intell.* 4, 52 (1998).
203. R. F. Curl, *Phil. Trans. Roy. Soc. London A* 343, 19 (1993).
204. G. E. Scuseria, Unpublished data (1993).
205. M.-L. Sun, Z. Slanina, and S.-L. Lee, *Chem. Phys. Lett.* 233, 279 (1995).
206. C. H. Xu and G. E. Scuseria, *Chem. Phys. Lett.* 262, 219 (1996).
207. Z. Slanina and E. Ōsawa, *Fullerene Sci. Technol.* 5, 167 (1997).
208. J. M. Rudziński, Z. Slanina, M. Togasi, F. Ōsawa, and T. Iizuka, *Thermochim. Acta* 125, 155 (1988).
209. Z. Slanina, J. M. Rudziński, and E. Ōsawa, *Carbon* 25, 747 (1987).
210. Z. Slanina, J. M. Rudziński, and E. Ōsawa, *Collect. Czech. Chem. Commun.* 52, 2381 (1987).
211. Z. Slanina, J. M. Rudziński, M. Togasi, and E. Ōsawa, *Thermochim. Acta* 140, 87 (1989).
212. Z. Slanina, X. Zhao, N. Kurita, H. Gotoh, F. Uhlík, J. M. Rudziński, K. H. Lee, and L. Adamowicz, *J. Mol. Graphics Mod.* 19, 216 (2001).
213. T. Kiyobayashi and M. Sakiyama, *Fullerene Sci. Technol.* 1, 269 (1993).
214. Z. Slanina, *Chem. Phys. Lett.* 145, 512 (1987).
215. Z. Slanina, L. Adamowicz, D. Bakowies, and W. Thiel, *Thermochim. Acta* 202, 249 (1992).
216. Z. Slanina and L. Adamowicz, *Thermochim. Acta* 205, 299 (1992).
217. Z. Slanina, *Z. Phys. Chem.* 217, 1119 (2003).
218. Z. Slanina, *Chem. Phys.* 150, 321 (1991).
219. P. S. Dardi and J. S. Dahler, *J. Chem. Phys.* 93, 3562 (1990).
220. Z. Slanina, *Thermochim. Acta* 207, 9 (1992).
221. Z. Slanina, *Thermochim. Acta* 209, 1 (1992).
222. K. Raghavachari, R. A. Whiteside, and J. A. Pople, *J. Chem. Phys.* 85, 6623 (1986).
223. K. Raghavachari and J. S. Binkley, *J. Chem. Phys.* 87, 2191 (1987).
224. J. M. L. Martin, J. P. François, and R. Gijbels, *J. Chem. Phys.* 90, 3403 (1989).
225. V. Parasuk and J. Almlöf, *J. Chem. Phys.* 91, 1137 (1989).
226. C. Liang and H. F. Schaefer III, *J. Chem. Phys.* 93, 8844 (1990).
227. J. M. L. Martin, J. P. François, and R. Gijbels, *J. Chem. Phys.* 93, 8850 (1990).
228. J. Kurtz and L. Adamowicz, *Astrophys. J.* 370, 784 (1991).
229. J. M. L. Martin, J. P. François, and R. Gijbels, *J. Chem. Phys.* 94, 3753 (1991).
230. J. M. L. Martin, J. P. François, and R. Gijbels, *J. Comput. Chem.* 12, 52 (1991).
231. D. W. Ewing, *Z. Phys. D* 19, 419 (1991).
232. Z. Slanina, J. M. Rudziński, and E. Ōsawa, *Z. Phys. D* 19, 431 (1991).
233. J. M. L. Martin, J. P. François, R. Gijbels, and J. Almlöf, *Chem. Phys. Lett.* 187, 367 (1991).
234. J. D. Watts and R. J. Bartlett, *Chem. Phys. Lett.* 190, 19 (1992).
235. V. Parasuk and J. Almlöf, *Theor. Chem. Acta* 83, 227 (1992).
236. Z. Slanina, J. Kurtz, and L. Adamowicz, *Chem. Phys. Lett.* 196, 208 (1992).
237. Z. Slanina, J. Kurtz, and L. Adamowicz, *Mol. Phys.* 76, 387 (1992).
238. L. A. Curtiss and K. Raghavachari, in "Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy" (S. R. Langhoff, Ed.), p. 139. Kluwer, Dordrecht, The Netherlands, 1995.
239. K. Raghavachari and L. A. Curtiss, in "Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy" (S. R. Langhoff, Ed.), p. 173. Kluwer, Dordrecht, The Netherlands, 1995.
240. J. M. L. Martin and P. R. Taylor, *J. Phys. Chem.* 100, 6047 (1996).
241. W. Krätschmer, N. Sorg, and D. R. Huffman, *Surf. Sci.* 156, 814 (1985).
242. S. W. McElvany, W. R. Creasy, and A. O'Keefe, *J. Chem. Phys.* 85, 632 (1986).
243. S. W. McElvany, B. I. Dunlap, and A. O'Keefe, *J. Chem. Phys.* 86, 715 (1987).
244. W. Krätschmer and K. Nachtigall, in "Polycyclic Aromatic Hydrocarbons and Astrophysics" (A. Léger, L. d'Hendecourt, and N. Boccara, Eds.), p. 75. D. Reidel, Dordrecht, The Netherlands, 1987.
245. R. J. van Zee, R. F. Ferrante, K. J. Zeringue, W. Weltner Jr., and D. W. Ewing, *J. Chem. Phys.* 88, 3465 (1988).
246. S. W. McElvany, *J. Chem. Phys.* 89, 2063 (1988).
247. J. R. Heath, A. L. Cooksy, H. W. Gruebele, C. A. Schmuttenmaer, and R. J. Saykally, *Science* 244, 564 (1989).
248. M. Vala, T. M. Chandrasekhar, J. Szczepanski, R. van Zee, and W. Weltner Jr., *J. Chem. Phys.* 90, 595 (1989).
249. M. Vala, T. M. Chandrasekhar, J. Szczepanski, and R. Pellow, *High Temp. Sci.* 27, 19 (1989).
250. L. N. Shen and W. R. M. Graham, *J. Chem. Phys.* 91, 5115 (1989).
251. H. M. Cheung and W. R. M. Graham, *J. Chem. Phys.* 91, 6664 (1989).
252. J. Kurtz and D. R. Huffman, *J. Chem. Phys.* 92, 30 (1990).
253. J. R. Heath, R. A. Sheeks, A. L. Cooksy, and R. J. Saykally, *Science* 249, 895 (1990).
254. J. Szczepanski and M. Vala, *J. Phys. Chem.* 95, 2792 (1991).
255. L. N. Shen, P. A. Withey, and W. R. M. Graham, *J. Chem. Phys.* 94, 2395 (1991).
256. J. R., Heath, A. van Orden, F. Kuo, and R. J. Saykally, *Chem. Phys. Lett.* 182, 17 (1991).
257. J. R. Heath and R. J. Saykally, *J. Chem. Phys.* 94, 1724 (1991).
258. J. R. Heath and R. J. Saykally, *J. Chem. Phys.* 94, 3271 (1991).
259. P. A. Withey, L. N. Shen, and W. R. M. Graham, *J. Chem. Phys.* 95, 820 (1991).

260. Q. Jiang and W. R. M. Graham, *J. Chem. Phys.* 95, 3129 (1991).
261. G. von Helden, M.-T. Hsu, P. R. Kemper, and B. T. Bowers, *J. Chem. Phys.* 95, 3835 (1991).
262. D. Zajfman, H. Feldman, O. Heber, D. Kella, D. Majer, Z. Vager, and R. Naaman, *Science* 258, 1129 (1992).
263. G. von Helden, P. R. Kemper, N. G. Gotts, and M. T. Bowers, *Science* 259, 1300 (1993).
264. N. G. Gotts, G. von Helden, and M. T. Bowers, *Int. J. Mass Spectr. Ion Process.* 149/150, 217 (1995).
265. Z. Slanina, F. Uhlík, and L. Adamowicz, *J. Radioanal. Nucl. Chem.* 219, 69 (1997).
266. Z. Slanina, S.-L. Lee, J.-P. François, J. Kurtz, L. Adamowicz, and M. Smigel, *Mol. Phys.* 81, 1489 (1994).
267. Z. Slanina, S.-L. Lee, J.-P. François, J. Kurtz, and L. Adamowicz, *Chem. Phys. Lett.* 223, 397 (1994).
268. Z. Slanina, S.-L. Lee, M. Smigel, J. Kurtz, and L. Adamowicz, in "Science and Technology of Fullerene Materials" (P. Bernier, D. S. Bethune, L. Y. Chiang, T. W. Ebbesen, R. M. Metzger, and J. W. Mintmire, Eds.), p. 163. Materials Research Society, Pittsburgh, 1995.
269. Z. Slanina, X. Zhao, E. Ōsawa, and L. Adamowicz, *Fullerene Sci. Technol.* 8, 369 (2000).
270. T. F. Giesen, A. van Orden, H. J. Hwang, R. S. Fellers, R. A. Provençal, and R. J. Saykally, *Science* 265, 756 (1994).
271. D. A. Plattner and K. N. Houk, *J. Am. Chem. Soc.* 117, 4405 (1995).
272. G. von Helden, N. G. Gotts, and M. T. Bowers, *Nature* 363, 60 (1993).
273. Z. Slanina, *Chem. Phys. Lett.* 173, 164 (1990).
274. M. W. Chase, Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, *J. Phys. Chem. Ref. Data* 14, Supplement 1, JANAF Thermochemical Tables, Third Edition, Vols. 1, 2 (1985).
275. M. A. Nygren and L. G. M. Pettersson, *Chem. Phys. Lett.* 191, 473 (1992).
276. B. G. Johnson, in "Modern Density Functional Theory: A Tool for Chemistry" (J. M. Seminario and P. Politzer, Eds.), p. 169. Elsevier, Amsterdam, The Netherlands, 1995.
277. J. D. Watts, J. Gauss, J. F. Stanton, and R. J. Bartlett, *J. Chem. Phys.* 97, 8372 (1992).
278. Z. Slanina, *Chem. Phys. Lett.* 161, 265 (1989).
279. Z. Slanina, *Chem. Phys. Lett.* 166, 317 (1990).
280. D. Michalska, H. Chojnacki, B. A. Hess, Jr., and L. J. Schaad, *Chem. Phys. Lett.* 141, 376 (1987).
281. Z. Slanina, *Thermochim. Acta* 127, 237 (1988).
282. A. van Orden and R. J. Saykally, *Chem. Rev.* 98, 2313 (1998).
283. Z. Slanina, *Skept. Inq.* 24, 18 (2000).
284. P. W. Fowler and J. Woolrich, *Chem. Phys. Lett.* 127, 78 (1986).
285. P. W. Fowler, *Chem. Phys. Lett.* 131, 444 (1986).
286. P. W. Fowler and J. I. Steer, *J. Chem. Soc., Chem. Commun.* 1403 (1987).
287. P. W. Fowler, J. E. Cremona, and J. I. Steer, *Theor. Chim. Acta* 73, 1 (1988).
288. A. Ceulemans and P. W. Fowler, *Phys. Rev. A* 39, 481 (1989).
289. A. Ceulemans and P. W. Fowler, *J. Chem. Phys.* 93, 1221 (1990).
290. P. W. Fowler, *J. Chem. Soc., Faraday Trans.* 86, 2073 (1990).
291. P. Fowler, *Nature* 350, 20 (1991).
292. P. W. Fowler, *J. Chem. Soc., Faraday Trans.* 87, 1945 (1991).
293. P. W. Fowler, R. C. Batten, and D. E. Manolopoulos, *J. Chem. Soc., Faraday Trans.* 87, 3103 (1991).
294. A. Ceulemans and P. W. Fowler, *Nature* 353, 52 (1991).
295. P. W. Fowler and D. E. Manolopoulos, *Nature* 355, 428 (1992).
296. P. W. Fowler, D. E. Manolopoulos, and R. P. Ryan, *J. Chem. Soc., Chem. Commun.* 408, (1992).
297. P. W. Fowler, *J. Chem. Soc., Perkin Trans. II* 145, (1992).
298. P. W. Fowler, D. E. Manolopoulos, and R. P. Ryan, *Carbon* 30, 1235 (1992).
299. D. E. Manolopoulos, D. R. Woodall, and P. W. Fowler, *J. Chem. Soc., Faraday Trans.* 88, 2427 (1992).
300. P. W. Fowler and V. Morvan, *J. Chem. Soc., Faraday Trans.* 88, 2631 (1992).
301. P. W. Fowler, D. E. Manolopoulos, D. B. Redmond, and R. P. Ryan, *Chem. Phys. Lett.* 202, 371 (1993).
302. D. E. Manolopoulos and P. W. Fowler, *Chem. Phys. Lett.* 204, 1 (1993).
303. S.-L. Lee, *Theor. Chim. Acta* 81, 185 (1992).
304. A. J. Stone and D. J. Wales, *Chem. Phys. Lett.* 128, 501 (1986).
305. C. Coulombeau, and A. Rassat, *J. Chim. Phys.* 88, 173 (1991).
306. W. N. Lipscomb and L. Massa, *Inorg. Chem.* 31, 2297 (1992).
307. A. C. Tang, Q. S. Li, C. W. Liu, and J. Li, *Chem. Phys. Lett.* 201, 465 (1993).
308. D. E. Manolopoulos, J. C. May, and S. E. Down, *Chem. Phys. Lett.* 181, 105 (1991).
309. X. Liu, D. J. Klein, T. G. Schmalz, and W. A. Seitz, *J. Comput. Chem.* 12, 1252 (1991).
310. X. Liu, D. J. Klein, W. A. Seitz, and T. G. Schmalz, *J. Comput. Chem.* 12, 1265 (1991).
311. X. Liu, T. G. Schmalz, and D. J. Klein, *Chem. Phys. Lett.* 188, 550 (1992).
312. D. E. Manolopoulos, *Chem. Phys. Lett.* 192, 330 (1992).
313. X. Liu, T. G. Schmalz, and D. J. Klein, *Chem. Phys. Lett.* 192, 331 (1992).
314. M.-L. Sun, Z. Slanina, and S.-L. Lee, *Fullerene Sci. Technol.* 3, 627 (1995).
315. P.W. Fowler, C. M. Quinn, and D. B. Redmond, *J. Chem. Phys.* 95, 7678 (1991).
316. S. Wei, Z. Shi, and A. W. Castleman, Jr., *J. Chem. Phys.* 94, 3268 (1991).
317. D. J. Klein, W. A. Seitz, and T. G. Schmalz, *J. Phys. Chem.* 97, 1231 (1993).
318. D. J. Klein, *Chem. Phys. Lett.* 217, 261 (1994).
319. A. T. Balaban, D. J. Klein, and C. A. Folden, *Chem. Phys. Lett.* 217, 266 (1994).
320. J. R. Dias, *Chem. Phys. Lett.* 209, 439 (1993).
321. K. Balasubramanian, *Chem. Phys. Lett.* 182, 257 (1991).

322. K. Balasubramanian, *Chem. Phys. Lett.* 183, 292 (1991).
323. K. Balasubramanian, *J. Chem. Inf. Comput. Sci.* 32, 47 (1992).
324. K. Balasubramanian, *Chem. Phys. Lett.* 198, 577 (1992).
325. K. Balasubramanian, *Chem. Phys. Lett.* 197, 55 (1992).
326. K. Balasubramanian, *J. Mol. Spectr.* 157, 254 (1992).
327. K. Balasubramanian, *Chem. Phys. Lett.* 201, 306 (1993).
328. K. Balasubramanian, *Chem. Phys. Lett.* 206, 210 (1993).
329. K. Balasubramanian, *Chem. Phys. Lett.* 202, 399 (1993).
330. K. Balasubramanian, *J. Phys. Chem.* 97, 6990 (1993).
331. K. Balasubramanian, *J. Phys. Chem.* 97, 4647 (1993).
332. D. Babić, S. Bassoli, M. Casartelli, F. Cataldo, A. Graovac, O. Ori, and B. York, *J. Mol. Simulat.* 14, 395 (1995).
333. B. R. Eggen, M. I. Heggie, G. Jungnickel, C. D. Latham, R. Jones, and P. R. Briddon, *Science* 272, 87 (1996).
334. B. R. Eggen, M. I. Heggie, G. Jungnickel, C. D. Latham, R. Jones, and P. R. Briddon, *Fullerene Sci. Technol.* 5, 727 (1997).
335. Z. Slanina, X. Zhao, F. Uhlík, M. Ozawa, and E. Ôsawa, *J. Organomet. Chem.* 599, 57 (2000).
336. S. W. McElvany, J. H. Callahan, M. M. Ross, L. D. Lamb, and D. R. Huffman, *Science* 260, 1632 (1993).
337. J.-P. Deng, D.-D. Ju, G.-R. Her, C.-Y. Mou, C.-J. Chen, Y.-Y. Lin, and C.-C. Han, *J. Phys. Chem.* 97, 11575 (1993).
338. S. Fujita, "Symmetry and Combinatorial Enumeration in Chemistry," Springer-Verlag, Berlin, 1991.
339. F. Wahl, J. Wörth, and H. Prinzbach, *Angew. Chem., Int. Ed. Engl.* 32, 1722 (1993).
340. H. Prinzbach, A. Weller, P. Landenberger, F. Wahl, J. Worth, L. T. Scott, M. Gelmont, D. Olevano, and B. von Issendorff, *Nature* 407, 60 (2000).
341. Z. Wang, X. Ke, Z. Zhu, F. Zhu, M. Ruan, H. Chen, R. Huang, and L. Zheng, *Phys. Let. A* 280, 351 (2001).
342. Z. Iqbal, Y. Zhang, H. Grebel, S. Vijayalakshmi, A. Lahamer, G. Benedek, M. Bernasconi, J. Cariboni, I. Spagnolatti, R. Sharma, F. J. Owens, M. E. Kozlov, K. V. Rao, and M. Muhammed, *Eur. Phys. J. B* 31, 509 (2003).
343. Z. Slanina and L. Adamowicz, *Fullerene Sci. Technol.* 1, 1 (1993).
344. Z. Slanina and L. Adamowicz, *J. Mol. Struct. (Theochem)* 281, 33 (1993).
345. J. Lu, S. Re, Y. Choe, S. Nagase, Y. S. Zhou, R. S. Han, L. M. Peng, X. W. Zhang, and X. G. Zhao, *Phys. Rev. B* 67, 125415 (2003).
346. Z. X. Zhang, Z. Y. Pan, Y. X. Wang, Z. J. Li, and Q. Wei, *Mod. Phys. Lett.* 17, 877 (2003).
347. C. Piskoti, J. Yarger, and A. Zettl, *Nature* 393, 771 (1998).
348. R. L. Murry, J. R. Colt, and G. E. Scuseria, *J. Phys. Chem.* 97, 4954 (1993).
349. S. G. Louie, *Nature* 384, 612 (1996).
350. L. F. Sun, S. S. Xie, W. Liu, W. Y. Zhou, Z. Q. Liu, D. S. Tang, G. Wang, and L. X. Qian, *Nature* 403, 384 (2000).
351. L. C. Qin, X. L. Zhao, K. Hirahara, Y. Miyamoto, Y. Ando, and S. Iijima, *Nature* 408, 50 (2000).
352. N. Wang, Z. K. Tang, G. D. Li, and J. S. Chen, *Nature* 408, 50 (2000).
353. Z. Slanina, F. Uhlík, and L. Adamowicz, *J. Mol. Graphics Mod.* 21, 517 (2003).
354. M. D. Diener and J. M. Alford, *Nature* 393, 668 (1998).
355. H. Moribe, T. Inoue, H. Kato, A. Taninaka, Y. Ito, T. Okazaki, T. Sugai, R. Bolskar, J. M. Alford, and H. Shinohara, "The 25th Fullerene-Nanotubes Symposium," Awaji, Japan, 2003, paper 1P-1.
356. H. Kato, A. Taninaka, T. Sugai, and H. Shinohara, *J. Am. Chem. Soc.* 125, 7782 (2003).
357. O. V. Boltalina, E. V. Dashkova, and L. N. Sidorov, *Chem. Phys. Lett.* 256, 253 (1996).
358. O. V. Boltalina, I. N. Ioffe, L. N. Sidorov, G. Seifert, and K. Vietze, *J. Am. Chem. Soc.* 122, 9745 (2000).
359. Z. Slanina, S.-L. Lee, K. Kobayashi, and S. Nagase, *J. Mol. Struct. (Theochem)* 312, 175 (1994).
360. Z. Slanina, X. Zhao, F. Uhlík, and E. Ôsawa, in "Electronic Properties of Novel Materials—Science and Technology of Molecular Nanostructures (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), p. 179. AIP, Melville, NY, 1999.
361. Z. Slanina, K. Ishimura, K. Kobayashi, and S. Nagase, *Chem. Phys. Lett.* 384, 114 (2004).
362. D. R. Kanis, M. A. Ratner, T. J. Marks, and M. C. Zerner, *Chem. Mater.* 3, 19 (1991).
363. R. D. Bendale and M. C. Zerner, *J. Phys. Chem.* 99, 13830 (1995).
364. Z. Slanina, K. Kobayashi, and S. Nagase, *Chem. Phys. Lett.* in press (2005).
365. X. Zhao, K. H. Lee, Z. Slanina, and E. Ôsawa, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (P. V. Kamat, D. M. Guldi, and K. M. Kadish, Eds.), Vol. 7, p. 711. The Electrochemical Society, Pennington, NJ, 1999.
366. F. Furche and R. Ahlrichs, *J. Chem. Phys.* 114, 10362 (2001).
367. Z. Slanina and S.-L. Lee, *J. Mol. Struct. (Theochem)* 304, 173 (1994).
368. V. I. Kovalenko and A. R. Khamatgalimov, *Chem. Phys. Lett.* 377, 263 (2003).
369. M. E. Casida, C. Jamorski, K. C. Casida, and D. R. Salahub, *J. Chem. Phys.* 108, 4439 (1998).
370. Z. Slanina, F. Uhlík, K. Kobayashi, and S. Nagase (to be published).
371. Z. Slanina and S.-L. Lee, in "HPC-ASIA '95, Electronic Proceedings." National Center for High-Performance Computing, Hsinchu, Taiwan, ROC, 1995.
372. Z. Slanina, F. Uhlík, L. Adamowicz, K. Kobayashi, and S. Nagase, *Int. J. Quantum Chem.* 100, 610 (2004).
373. D. Bakowies and W. Thiel, *Chem. Phys.* 151, 309 (1991).
374. R. Taylor, G. J. Langley, A. G. Avent, T. J. S. Dennis, H. W. Kroto, and D. R. M. Walton, *J. Chem. Soc., Perkin Trans. 2* 1029 (1993).

375. T. Wakabayashi, K. Kikuchi, S. Suzuki, H. Shiromaru, and Y. Achiba, *J. Phys. Chem.* 98, 3090 (1994).
376. M. Benz, M. Fanti, P. W. Fowler, D. Fuchs, M. M. Kappes, C. Lehner, R. H. Michel, G. Orlandi, and F. Zerbetto, *J. Phys. Chem.* 100, 13399 (1996).
377. R. J. Cross, M. Saunders, A. Khong, R. Shimshi, and S. Uljon, in "Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish and R. S. Ruoff, Eds.), Vol. 4, The Electrochemical Society, Pennington, NJ, 1997.
378. K. Yamamoto, in "The 200th ECS Meeting, San Francisco, 2001," abstract no. 1489, The Electrochemical Society, Pennington, NJ, 1997.
379. F. Uhlík, Z. Slanina, and E. Ōsawa, *Eur. Phys. J. D* 16, 349 (2001).
380. M. Goldberg, *Tohoku Math. J.* 40, 226 (1934).
381. M. Goldberg, *Tohoku Math. J.* 43, 104 (1937).
382. Z. Slanina, X. Zhao, and E. Ōsawa, *Advan. Strain. Inter. Org. Mol.* 7, 185 (1999).
383. S.-L. Lee, M.-L. Sun, and Z. Slanina, *Int. J. Quantum Chem., Quantum Chem. Symp.* 30, 355 (1996).
384. Z. Slanina, *J. Mol. Struct. (Theochem)* 65, 143 (1990).
385. K. Kobayashi, S. Nagase, and T. Akasaka, *Chem. Phys. Lett.* 245, 230 (1995).
386. K. Kobayashi, S. Nagase, and T. Akasaka, *Chem. Phys. Lett.* 261, 502 (1996).
387. K. Kobayashi and S. Nagase, *Chem. Phys. Lett.* 262, 227 (1996).
388. U. Kirbach and L. Dunsch, *Angew. Chem., Int. Ed. Eng.* 35, 2380 (1996).
389. T. Pichler, M. S. Golden, M. Knupfer, J. Fink, U. Kirbach, P. Kuran, and L. Dunsch, *Phys. Rev. Lett.* 79, 3026 (1997).
390. T. Pichler, M. Knupfer, M. S. Golden, T. Boske, J. Fink, U. Kirbach, P. Kuran, L. Dunsch, and C. Jung, *Appl. Phys. A* 66, 281 (1998).
391. F. Negri, G. Orlandi, and F. Zerbetto, *Chem. Phys. Lett.* 189, 495 (1992).
392. K. Raghavachari, *Chem. Phys. Lett.* 190, 397 (1992).
393. X.-Q. Wang, C. Z. Wang, B. L. Zhang, and K. M. Ho, *Phys. Rev. Lett.* 69, 69 (1992).
394. D. Bakowies, M. Kolb, W. Thiel, S. Richard, R. Ahlrichs, and M. M. Kappes, *Chem. Phys. Lett.* 200, 411 (1992).
395. Z. Slanina and S.-L. Lee, *J. Mol. Struct. (Theochem)* 333, 153 (1995).
396. Z. Slanina and S.-L. Lee, *NanoStruct. Mater.* 4, 39 (1994).
397. S. Margadonna, C. M. Brown, T. J. S. Dennis, A. Lappas, P. Pattison, K. Prassides, and H. Shinohara, *Chem. Mater.* 10, 1742 (1998).
398. T. J. S. Dennis, M. Hulman, H. Kuzmany, and H. Shinohara, *J. Phys. Chem. B* 104, 5411 (2000).
399. I. Margiolaki, S. Margadonna, K. Prassides, S. Assimopoulos, K. P. Meletov, G. A. Kourouklis, T. J. S. Dennis, and H. Shinohara, *Physica B* 318, 372 (2002).
400. N. Tagmatarchis, A. G. Avent, K. Prassides, T. J. S. Dennis, and H. Shinohara, *Chem. Commun.* 1023 (1999).
401. T. J. S. Dennis, T. Kai, K. Asato, T. Tomiyama, H. Shinohara, T. Yoshida, Y. Kobayashi, H. Ishiwatari, Y. Miyake, K. Kikuchi, and Y. Achiba, *J. Phys. Chem. A* 103, 8747 (1999).
402. J. A. Azamar-Barrios, T. J. S. Dennis, S. Sadhukan, H. Shinohara, G. E. Scuseria, and A. Penicaud, *J. Phys. Chem. A* 105, 4627 (2001).
403. Z. F. Chen and W. Thiel, *Chem. Phys. Lett.* 367, 15 (2003).
404. Z. Slanina, F. Uhlík, M. Yoshida, and E. Ōsawa, *Fullerene Sci. Technol.* 8, 417 (2000).
405. G. Y. Sun, *Chem. Phys. Lett.* 367, 26 (2003).
406. Z. Slanina, S.-L. Lee, F. Uhlík, and L. Adamowicz, in "Physics and Chemistry of Fullerenes and Their Derivatives" (H. Kuzmany, J. Fink, M. Mehring, and S. Roth, Eds.), p. 385, World Scientific, Singapore, 1996.
407. Z. Slanina, X. Zhao, S.-L. Lee, and E. Ōsawa, *Scripta Mater.* 43, 733 (2000).
408. G. Y. Sun, *Chem. Phys.* 289, 371 (2003).
409. T. Minami, Y. Miyake, K. Kikuchi, and Y. Achiba, in "The 18th Fullerene General Symposium, Okazaki" (E. Ōsawa, Ed.), p. 42, The Fullerene Research Association of Japan, Toyohashi, Japan, 2000.
410. H. Richter, K. Taghizadeh, W. J. Grieco, A. L. Lafleur, and J. B. Howard, *J. Phys. Chem.* 100, 19603 (1996).
411. H. Richter, A. J. Labrocca, W. J. Grieco, K. Taghizadeh, A. L. Lafleur, and J. B. Howard, *J. Phys. Chem. B* 101, 1556 (1997).
412. P. J. Hay and W. R. Wadt, *J. Chem. Phys.* 82, 299 (1985).
413. W. J. Hehre, L. Radom, P. R. von Schleyer, and J. A. Pople, "Ab Initio Molecular Orbital Theory," p. 265, John Wiley, New York, 1986.
414. E. M. Kosower, "An Introduction to Physical Organic Chemistry," p. 251, John Wiley, New York, 1968.
415. L. S. Wang, J. M. Alford, Y. Chai, M. Diener, J. Zhang, S. M. McClure, T. Guo, G. E. Scuseria, and R. E. Smalley, *Chem. Phys. Lett.* 207, 354 (1993).
416. Y. Kubozono, T. Ohta, T. Hayashibara, H. Maeda, H. Ishida, S. Kashino, K. Oshima, H. Yamazaki, S. Ukita, and T. Sogabe, *Chem. Lett.* 457 (1995).
417. F. G. Hopwood, K. J. Fisher, P. Greenhill, G. D. Willett, and R. Zhang, *J. Phys. Chem. B* 101, 10704 (1997).
418. H. Shinohara, in "Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish and R. S. Ruoff, Eds.), Vol. 4, p. 467, The Electrochemical Society, Pennington, NJ, 1997.
419. R. Taylor, "Lecture Notes on Fullerene Chemistry, A Handbook for Chemists," Imperial College Press, London, 1999.
420. O. V. Boltalina, J. M. Street, and R. Taylor, *J. Chem. Soc., Perkin Trans. 2* 649 (1998).
421. A. A. Gakh and A. A. Tuinman, *Tetrahedron Lett.* 42, 7133 (2001).

422. P. B. Hitchcock and R. Taylor, *Chem. Commun.* 2078 (2002).
423. A. G. Avent, B. W. Clare, P. B. Hitchcock, D. L. Kepert, and R. Taylor, *Chem. Commun.* 2370 (2002).
424. J. Nossal, R. K. Saini, A. K. Sadana, H. F. Bettinger, L. B. Alemany, G. E. Scuseria, W. E. Billups, M. Saunders, A. Khong, and R. Weisemann, *J. Am. Chem. Soc.* 123, 8482 (2001).
425. A. G. Avent and R. Taylor, *Chem. Commun.* 2726 (2002).
426. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 315, 71 (1994).
427. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 466, 177 (1999).
428. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 589, 209 (2002)
429. J. Cioslowski, N. Rao, A. Szarecka, and K. Pernal, *Mol. Phys.* 99, 1229 (2001).
430. Z. Slanina, F. Uhlík, O. V. Boltalina, and V. P. Kolesov, *Phys. Solid Stat.* 44, 534 (2002).
431. R. E. Haufler, J. Conceicao, L. P. F. Chibante, Y. Chai, N. E. Byrne, S. Flanagan, M. M. Haley, S. C. O'Brien, C. Pan, Z. Xiao, W. E. Billups, M. A. Ciufolini, R. H. Hauge, J. L. Margrave, L. J. Wilson, R. F. Curl, and R. E. Smalley, *J. Phys. Chem.* 94, 8634 (1990).
432. M. I. Attalla, A. M. Vassallo, B. N. Tattam, and J. V. Hanna, *J. Phys. Chem.* 97, 6329 (1993).
433. C. Ruchardt, M. Gerst, J. Ebenhoch, H. D. Beckhaus, E. E. B. Campbell, R. Tellgmann, H. Schwarz, T. Weiske, and S. Pitter, *Angew. Chem., Int. Ed. Engl.* 32, 584 (1993).
434. M. Gerst, H. D. Beckhaus, C. Rüchardt, E. E. B. Campbell, and R. Tellgmann, *Tetrahedron Lett.* 34, 7729 (1993).
435. B. I. Dunlap, D. W. Brenner, and G. W. Schriver, *J. Phys. Chem.* 98, 1756 (1994).
436. P. W. Fowler, J. P. B. Sandall, and R. Taylor, *J. Chem. Soc., Perkin Trans. 2* 419 (1997).
437. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 589, 195 (2002).
438. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 622, 185 (2003).
439. A. A. Gakh, A. A. Tuinman, J. L. Adcock, R. A. Sachleben, and R. N. Compton, *J. Am. Chem. Soc.* 116, 819 (1994).
440. S. I. Troyanov, P. A. Troshin, O. V. Boltalina, I. N. Ioffe, L. N. Sidorov, and E. Kemnitz, *Angew. Chem. Int. Edt. Engl.* 40, 2285 (2001).
441. S. J. Austin, P. W. Fowler, J. P. B. Sandall, and F. Zerbetto, *J. Chem. Soc., Perkin Trans.* 2, 155 (1996).
442. B. W. Clare and D. L. Kepert, *J. Mol. Struct. (Theochem)* 389, 97 (1997).
443. J. F. Anacleto, H. Perreault, R. K. Boyd, S. Pleasance, M. A. Quilliam, P. G. Sim, J. B. Howard, Y. Makarovsky, and A. L. Lafleur, *Rapid Commun. Mass. Spectr.* 6, 214 (1992).
444. D. Goldberg, Y. Bando, O. Stéphan, and K. Kurashimna, *Appl. Phys. Lett.* 73, 2441 (1998).

CHAPTER 10

Carbon Nanocones

Henning Heiberg-Andersen
Institute for Energy Technology, Kjeller, Norway

CONTENTS

1. Introduction 507
2. Conic Shapes in Graphite: Topology, Geometry, and Stability Considerations 510
   2.1. Carbon Chemistry in Two and Three Dimensions 510
   2.2. Nanocones 511
   2.3. Helical Cone Growth 515
   2.4. Horns, Pipettes, Zipped Structures, and Tubular Graphite Cones 518
3. Atomic and Electronic Structure: Theoretical/Computational Methods and Predictions 519
   3.1. Molecular Mechanics 519
   3.2. Density Functional Theory 520
   3.3. The Tight-Binding Model 522
   3.4. The Effective Mass Theory and Continuum Models 523
   3.5. Predictions 525
4. The Nucleation Puzzle 532
   References 533

1. INTRODUCTION

This chapter reviews the research on carbon nanocones, which started a decade ago. The term *nanocones* here refers to conic graphene shells, closed like fullerenes or open-ended. This implies a distinction from the the various conical whisker structures, the first of which were studied already in 1960 by Bacon [1]. However, due to remarkable theoretical crossing points and recent experimental results [2], a parallel discussion of the related *helical* graphite cones seems indispensable. As is common in the literature, *graphite cones* will in the following refer to both nanocones and topologically different conical carbon structures. Newer members of this family, like nanohorns [3–5] and the recently discovered nanopipettes [6], will be discussed only briefly, as a substantial literature on these structures has not accumulated yet. A discussion of the boron-nitride nanoscale cones [7] is omitted from this review; these

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (507–536)

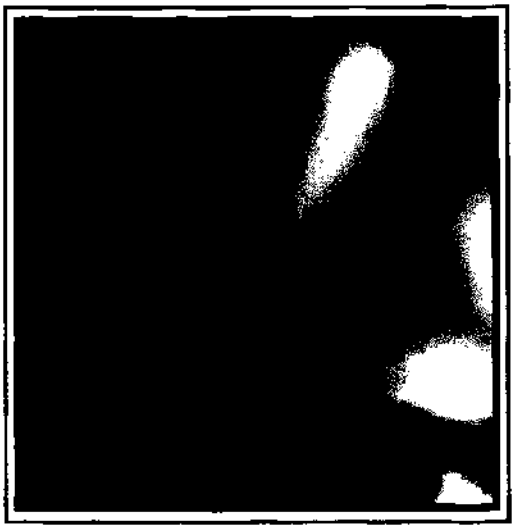


Figure 1. The first STM image of fullerene cones. Reprinted with permission from [22]. M. Ge and K. Sattler. *Chem. Phys. Lett.* 220, 192 (1994). © 1994, Elsevier.

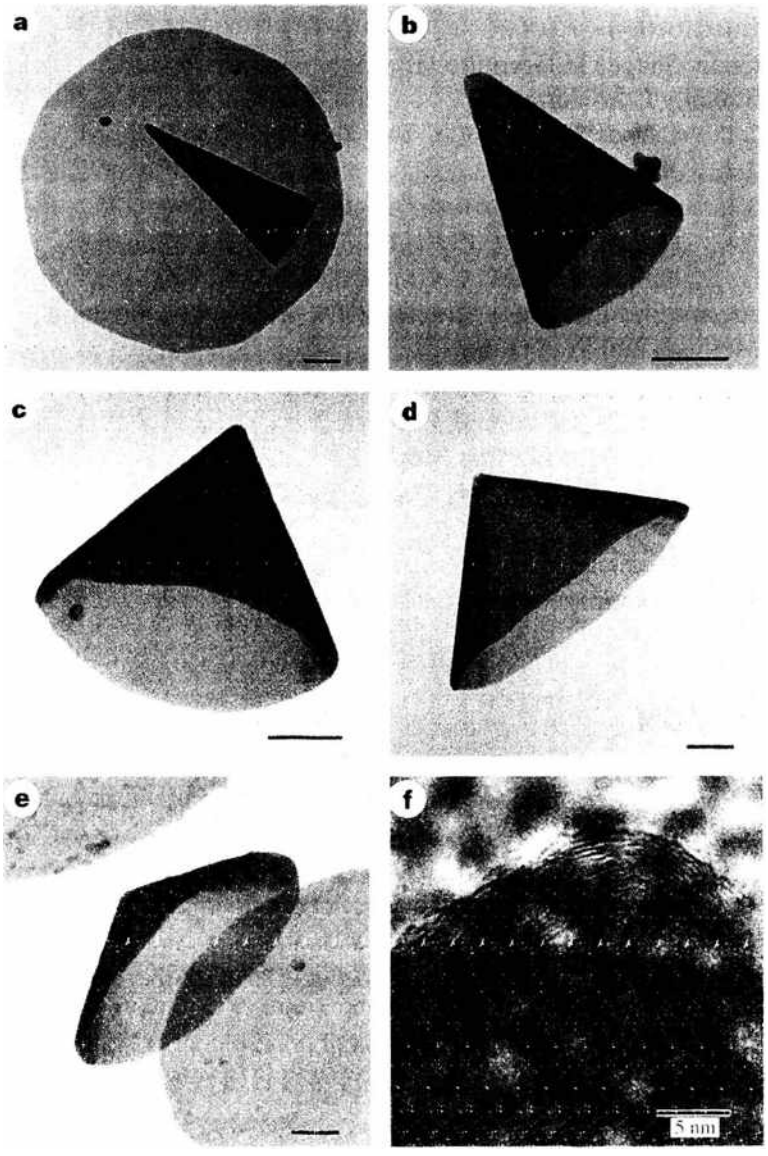


Figure 2. HRTEM images of the five cone shapes synthesized by Krishnan et al. (a–e): multilayered by (f). Reprinted with permission from [25]. A. Krishnan et al., *Nature* 388, 451 (1997). © 1997, Nature Publishing Group.

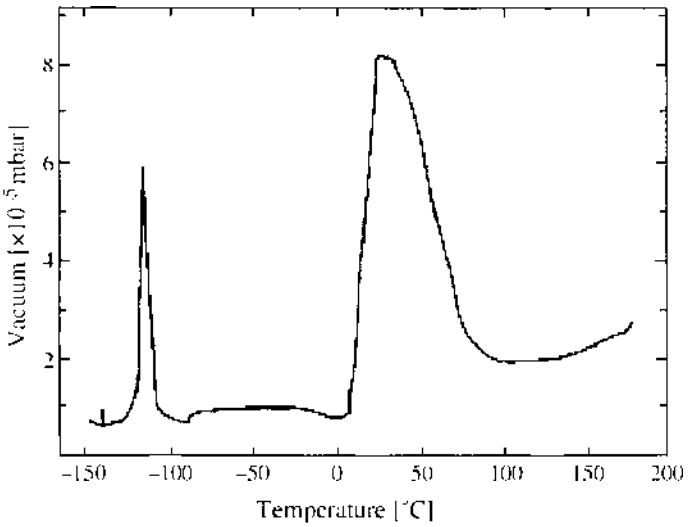


Figure 3. Hydrogen desorption curve from a storage experiment with samples from Kværner's carbon-black process [24] containing nanocones. See patent letter [36] for a description of the method. Reprinted with permission from [36]. A. T. Skjeltorp and A. J. Maeland, U.S. Patent No. 6,290,753 B1, Sept. 18, 2001.

analogous structures are still sufficiently different from carbon nanocones that they should rather be included in a treatment of the emerging field of boron-nitride fullerenes [8–19].

In 1994, independent theoretical predictions [20, 21] of the likely formation of carbon nanocones appeared in the scientific literature, and before the year was over, observation of fullerene cones synthesized in the hot vapor phase were reported by Ge and Sattler [22]. In addition to tubes, open-ended cones were produced in the same experiment. A striking fact is that all the observed cones had apex angles close to 19°, which is consistent with a graphene shell with five pentagonal faces at the tip. Because a tube has six pentagons[1] at a capped end, it was suggested [22, 23] that the tubes and cones originated from a similar seed. Figure 1 shows a scanning tunneling microscopy (STM) image from this experiment. A large open-ended cone is in front of a smaller fullerene cone.

A new trigger for further research came in 1997. Under pyrolysis of heavy oil in a cycle known as *Kværner's carbon-black process* [24], large quanta of open-ended cones with all the five apex angles consistent with Euler's rule were synthesized by Krishnan et al. [25]. Figure 2 shows high-resolution transmission electron microscopy (HRTEM) images of these five basic types of cones observed in the sample. From this point, industrial production rates and commercial applications of carbon nanocones appeared feasible, and in 2001 Terrones et al. obtained high yields of closed, open-ended, and lamp-shade formed (open in both ends) nanocones by pyrolysis of palladium [26]. The partly investigated areas of applications for the carbon nanocones are field emission (FEM), magnetic flux manipulation, and hydrogen storage. The cones, which have one to five pentagons at the tip, can show up a broader range of topology-dependent electronic properties than the tubes, which invariably have six pentagons at the tips. The different aspect ratios also make a factor. A detailed overview of the results achieved on electronic properties of nanocones is given in Section 3. The appearance of the many new carbon structures has renewed the interest in carbon as a light, inexpensive material for hydrogen storage in cars [27–35]. Experiments with samples from Kværner's carbon-black process are very encouraging in this respect, and a U.S. patent on hydrogen storage has been issued [36]. Figure 3 shows a hydrogen desorption curve for a sample from Kværner's carbon-black process. The part played by the cones in these experiments is not understood.

Contrary to the icosahedral fullerenes, the appearance of the various types of nanocones did not attract much attention from pure chemists. The study [37–47] of fluoranthenoids (conjugated hydrocarbons composed of one pentagonal ring among otherwise hexagonal rings) and indacenoids (two fused pentagons surrounded by hexagons) started at the

[1] Other face combinations are possible, but considered less likely, as will be discussed in the next section.

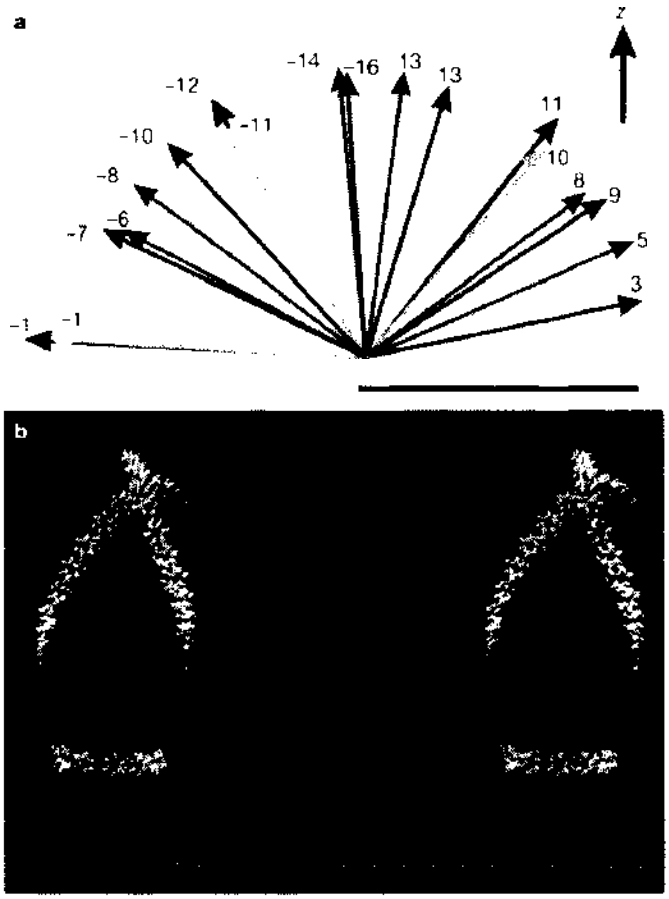


Figure 4. Core shell of the HIV-1 virus: Capsid protein (CA) helical polymorphism and cone formation. (a) Orientations of the unit-cell vectors in six reconstructed CA helical families. See [49] for details. (b) Stereo view of a model of the HIV-1 capsid. Pentagonal defects in the hexagonal network are shown in red. A contiguous line of CA hexagons is highlighted in gold to illustrate reversal of the helical band. Reprinted with permission from [49], S. Li et al., *Nature* 407, 409 (2000). © 2000, Nature Publishing Group.

beginning of the 1990s and has proceeded seemingly independent of the discovery of carbon nanocones, as the focus has been on enumeration of isomers. The observation of fullerene cones [22] inspired, however, an American biochemistry research group [48–50]. They found that both the viral and synthesized cores of the human immunodeficiency virus (HIV) exhibited conic shapes with apex angles of approximately 19°, just like the previously observed fullerene cones. Pursuing this coincidence led to the hypothesis that the capsid protein shell, which surrounds the nucleocapsid protein and genomic RNA, was arranged in a hexagonal–pentagonal network. An image reconstruction that supported this hypothesis is shown in Fig. 4. As for the HIV-cores, the nucleation mechanisms for the nanocones is a great unsolved puzzle, and the pieces will be discussed in Section 4.

2. CONIC SHAPES IN GRAPHITE: TOPOLOGY, GEOMETRY, AND STABILITY CONSIDERATIONS

2.1. Carbon Chemistry in Two and Three Dimensions

Due to the small energy difference between the filled $2s$ and the unfilled $2p$ shell of carbon, the atoms attain their lowest state of energy by promoting the $2s$ electrons into a hybridization with the $2p$ electrons when they form a molecule or a solid. The two hybridizations sp^2 and sp^3 correspond to the periodic lattice forms of carbon, graphite and diamond, respectively. In diamond, all the four electrons take part in the hybridization. The sp^2 hybridization of graphite, on the other hand, involves only two of the three p-orbitals, which defines the hybridization plane. The last p-orbital oriented normal to this plane is occupied by the fourth

electron outside the 1*s* core. The three others take part in bonds in the hybridization plane, which explains the hexagonal arrangement of the atoms in a plane graphite sheet, usually called *graphene*. The hybridized orbitals look like *s*-orbitals when viewed along the relative vector of two atoms in the sheet and are therefore denoted by the Greek letter σ. For the same reason, an electron occupying the last *p*-orbital is called a π-electron. Because this orbital is only half-filled, neighboring atoms can form covalent π-bonds.

For sp^2 hybridized hydrocarbons made up of an even number of carbon atoms, so-called conjugated systems, we can represent a set of such pairwise π-interactions involving all the carbon atoms by double bonds in the graph defined by the σ-bond skeleton. This kind of pictorial structural formulas are called Kekulé structures, after the nineteenth century German chemist. However, for conjugated systems showing some degree of symmetry, one can usually draw two or more equivalent Kekulé structures. This reflects the delocalization of the π-orbitals, which is incorporated in the standard molecular orbital theories by a linear combination of atomic orbitals (LCAO). The simplest of these is the Hückel [51] theory, where each atomic site contributes a single *p*-orbital to the molecular π-orbitals and only interactions, assumed identical, between neighboring atoms are taken into account. The Hamiltonian of the Hückel theory is therefore a matrix[2] that commutes with the adjacency matrix of the combinatorial graph describing the σ-bond network (see, e.g., [52].) These connections between topology, molecular spectra, and bonding structures have resulted in an infinity of graph-theoretical papers on planar hydrocarbons and, later, fullerenes.

The fullerenes, which have attracted attention through two decades, and all the novel carbon materials, imply, however, a bending of the graphene sheets, which in turn implies a modification of the hybridization concept. We will avoid introduction of too many chemical terms in the present discussion, as few of these will be encountered in the main body of work on carbon nanocones. Therefore, we adapt Haddon's [53, 54] recipe for transferring the σ–π separation of orbitals from two to three dimensions. His argument is that the success of the Hückel theory for planar systems is not due to the possible reality of the concept of specific σ- and π-electrons, but rather the orthogonality of the σ- and π orbitals. This orthogonality can be maintained in three dimensions, for which purpose he introduced two methods, both described in [53]: A modified Hückel theory (different from the previous extended Hückel theory [55] which explicitly incorporates both the 2*s* and 2*p* orbitals of each carbon atom) and the π-orbital axis vector (POAV) analysis. These methods have not played any noticeable role in the investigation of carbon nanocones; it is the transferability of the σ–π separation to curved surfaces we want to stress here. In accordance with Haddon's concept, the original Hückel theory, which is purely topological, is found to predict the correct level ordering around the Fermi level for fullerenes [56].

2.2. Nanocones

According to the discussion of the last section, the topology associated with the σ-bonds is relevant also for curved carbon materials. When Euler had proved his famous theorem connecting the faces, edges, and vertices of graphs on various surfaces, he could immediately deduce the sum rule

$$\sum_k (6-k) f_k = 12 \tag{1}$$

for the number f_k of *k*-sided faces in a planar cubic[3] graph, and it is well-known that a three-dimensional polyhedron can be represented by a planar graph by stretching its faces. Because a cubic graph corresponds exactly to the σ-framework of a closed π-conjugated carbon cluster, a fullerene has therefore exactly 12 pentagonal faces and all other faces hexagonal. There are thus five basic types of fullerene cones, with 1, 2, 3, 4, or 5 pentagons at the tip, and, respectively, 11, 10, 9, 8, or 7 at the base end. Because there are no pentagons

[2] In solid-state physics, the Hückel matrix is usually called the "hopping matrix."

[3] *Cubic* means that each vertex of the graph is connected to three neighbors, as are the corners of a cube.

in a flat hexagonal sheet, and placing six pentagons at both ends gives a tube, it is clear that the apex angle decreases with the number of pentagons at the tip, and later in this section the apex angles will be quantified.

The most symmetric closed cluster with one of the mentioned pentagon distributions is the five-fold symmetric UFO shape, where the base is identical to the tip with one pentagon, and the remaining ten pentagons are placed pairwise along the edge, but this is a stretching of the cone concept. The next highest symmetry is the tetrahedron obtained by placing three pentagons at each of the four corners. Then comes members of the C_{mv} point groups, with $m \leq 3$ and two to five pentagons at the (only) tip. For open-ended cones, m ranges from 1 (five pentagons) to 5 (one pentagon). According to the sum rule (1), one may consider many different combinations with triangles, squares, heptagons, and so on. All these will lead to larger deviations from the optimum σ-bond angle of 120°, and their physical realizations are therefore considered to be relatively unlikely.

An open-ended cone can be obtained by cutting one to five sectors with disclination angle $\pi/3$ out of a flat hexagonal sheet and connecting the resulting dangling bonds on each side of the cut, as shown in Fig. 5. Sector angles not divisible by $\pi/3$ will give unmatched dangling bonds if we attempt to construct a single-walled nanocone. We see that cutting out one sector of disclination angle $\pi/3$ angle introduces a pentagon among the hexagons in the sheet, cutting out two such sectors gives two pentagons or one square, and so on, until all the tip configuration possibilities consistent with the sum rule are exhausted. If we assume all bond lengths equal, the apex angle φ for any possible face combination is then found from the circumference ratio between a circular sheet and the cone resulting by cutting out a sector with total disclination angle $n(\pi/3)$ by

$$\sin\left(\frac{\varphi}{2}\right) = \frac{2\pi - n(\pi/3)}{2\pi} \quad (n = 1, 2, \ldots, 5) \tag{2}$$

so the apex angles φ are respectively 112.9°, 83.6°, 60.0°, 38.9°, and 19.2° for 1, 2, 3, 4, and 5 pentagons, or their alternative face combinations allowed by the sum rule. This topological construction, commonly referred to as "the pentagon model," should not be taken as a pathway to cone formation; for the tubes, at least, there is strong theoretical evidence [57] that the pentagons are incorporated at an early stage of the nucleation process.

Open-ended cones with all these five apex angles, together with tubes (six pentagons at the tip) and disks (no pentagons) were synthesized in Kvaerner's carbon-black process [24] by Krishnan et al. [25]. HRTEM images of each geometry are shown in Fig. 2. All the observed fullerene cones, on the other hand, have apex angles close to 19° [22, 23], which corresponds to five pentagons at the tip, and seven at the base. Figure 1 is a STM image of a sample from the first synthesis of fullerene cones by Ge and Sattler [22]. The largest cone in front has a sharp and bright edge, suggesting an open edge with dangling bonds. The smaller cone

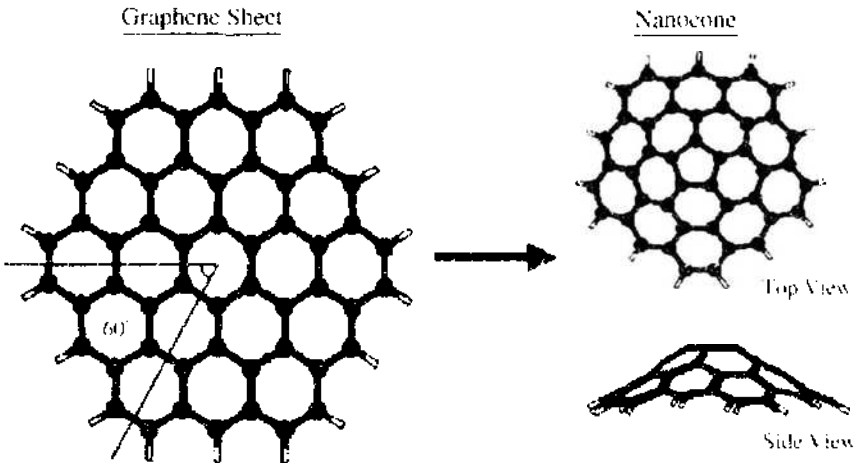


Figure 5. The cut-and-paste procedure for making a nanocone from a flat graphene sheet. Reprinted with permission from [58], O. A. Shenderova et al., *Nanotechnology* 12, 191 (2001). © 2001, Institute of Physics Publishing

partly hidden behind it has a curved less bright end consistent with a cap without dangling bonds.

In 2001, what appears to be a third variant of nanocones were created by pyrolysis of palladium precursors [26]. In the presence of Ar at 850–1000°C, high yields of graphite cones up to 5-μm long were obtained. Many of the observed cones had apex angles a few degrees off those allowed by the pentagon model. These small but significant deviations were attributed to absence of closing pentagons at the tips. According to this hypothesis, a receipt for producing cones open in *both* ends is found.

Next, we will discuss a few theoretical results enabling stability considerations without performing computations. Thereby, it is possible to limit the scope of the expensive simulations to a fraction of the infinitely many possible cone topologies and (resulting) geometries.

2.2.1. Geometrical Stability Considerations

In 1993, before any fullerene cones were actually observed, Tománek, Zhong, and Krastev [59] studied the stability of multishell fullerenes in the form of spheres, tubes or cones, as shown schematically in Fig. 6. Using continuum elasticity theory to calculate the bending energy ΔE_b required to bend a plane graphene sheet into the required shape, they obtained for the three types of single-walled fullerenes

$$\Delta E_b(\text{sphere}) = 4\pi D(1+\alpha) = \epsilon(\text{sphere}) \tag{3}$$

$$\Delta E_b(\text{tube}) = \frac{\pi D L}{R_{\text{out}}} + \epsilon(\text{sphere}) \tag{4}$$

$$\Delta E_b(\text{cone}) = -\frac{\pi D}{\tan(\varphi/2)} \ln\left[1 - \frac{L}{R_{\text{out}}} \tan\left(\frac{\varphi}{2}\right)\right] + \epsilon(\text{sphere}) \tag{5}$$

where D is the flexural rigidity [60], α is a combination of elastic constants, and L, R_{out}, and the apex angle φ are shown in Fig. 6. The logarithmic factor in the bending energy of a cone stems from integration of the cylindrical expression with increasing radius along the cone axis. It is noteworthy that the bending energy of the spheres are constant, independent of the spheres' radii. This result were confirmed in [59] by tight-binding calculations, a method that will discussed in Section 3.3, and by comparison with other energy values for spherical fullerenes published at that time [61–63]. The bending energy of an open-ended cone can thus be obtained by subtracting the half of ϵ(sphere) from (5).

In a previous work [64], the values $D = 1.41$ eV and $\alpha = 0.165$ were suggested for graphite. The cone formula then gives the intuitive result that the bending energy increases with decreasing apex angle α, and plotted as a function of L, the curve will flatten when L becomes large. However, in the first synthesization of fullerene cones by Ge and Sattler published a year later [22], only the fullerene cones with the smallest apex angle of ~19° were observed. This is consistent with the hypothesis put forward in [23], that the cones and the more abundant tubes originated from a similar seed. But it also suggests that (5) must be used with some care in comparison of bending energies of cones with different apex angles: The fullerene cones with ~19° apex angle are realized by placing five pentagons at the tip, and

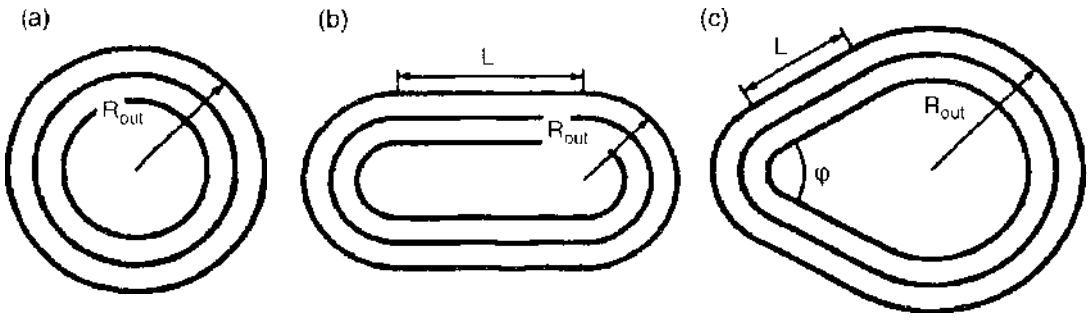


Figure 6. Schematic figures of the three basic fullerene geometries. Reprinted with permission from [59]. D. Tománek et al., *Phys. Rev. B* 48, 15461 (1993). © 1993, American Physical Society.

seven at the base. The pentagon distributions of all other fullerene cones will effect larger geometrical deviations from the ice-cream shape shown in Fig. 6, with larger and more localized stress around the pentagons.

Another central question addressed in [59] was at which number of atoms a multishell fullerene becomes more favorable than a larger single-shell fullerene with the same shape. This number is determined primarily by the increased bending energy required to form two concentric fullerene shells and the stabilizing van der Waals attraction between them. The estimate of [59] was based on the assumption that the van der Waals attraction is proportional to the area of contact between the shells. *Ab initio* calculations of the intershell energy of two concentric nanotubes [65] was taken as support of this assumption. However, the total van der Waals energy doesn't scale linearly with the number of shells; depending on the intershell distance, the inner and outer shells contribute differently [66]. Therefore, the estimate of [59] will not apply to more than two shells. This experimentally verified [67] general feature of multilayer systems is predicted by the macroscopic analysis [66], where each layer is given a macroscopic permeability, and continuity of the solutions of the Laplace equation is required at the surface of each layer. The conic coordinate system is one of the 11 in which the Laplace equation can be solved, but it doesn't relate the surface of a cone to its center of mass, so it is not possible to carry out a strictly correct macroscopic calculation of the van der Waals energy of multilayered cones. Because no other theoretical works on multilayered nanocones are reported to this date, the therefore superfluous term single-walled will not be used in the remainder of this review.

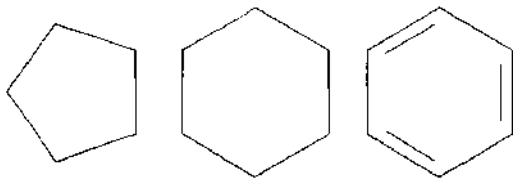
One of the main conclusions of [59] was that the cones, single-shelled or multishelled, were the least stable of the three fullerene shapes. This agrees well with later experimental observations, where the appearance of fullerene cones have been absolutely exceptional. The spherical fullerenes were found more stable than the tubes, in agreement with a previous work of Adams et al. [68], where an empirical formula based on computational results was proposed. Also in that work, the bending energy single-shelled spherical fullerenes were found to be independent of the radii. For the tubes, however, the energies obtained were inversely proportional to the squared radii, in contradiction to (4). More recently, Park et al. [69] parameterized the empirical formulas of [68] to their *ab initio* results while taking the effects of local curvature at the pentagons into account. The conclusion of this work was that nanotubes with diameters about 13 Å were the most stable structures. The conclusion of [59] that the cones were the least stable structures has not been objected by later works, however.

2.2.2. Topological Stability Considerations

For a given apex angle, the question is which of the yet infinitely many possible configurations of the pentagons are likely to be found in real nanocones. The rich literature on fullerenes gives some important guidelines. The simplest concept to practice is the *isolated pentagon rule* (IPR) [70], theoretically justified from different points of view: In the valence bond approach of Taylor [71, 72], the main point is that fused pentagons is one of the many configuration modes where Kekulé structures cannot be drawn without double bonds in the pentagons. The presence of double bonds shortens the already strained pentagonal bonds, producing the Mills-Nixon effect [73, 74]. According to Haddon and Raghavachari, part of the stress is due to re-hybridization [75, 76], and finally fused pentagons introduces anti-aromaticity [71, 77, 78].

A more profound approach is imparted by the important *leapfrog principle*, first applied to fullerenes by Fowler and Steer [79]. On a cubic graph, the leapfrog operation consists of transforming the vertices into hexagons while preserving the original faces. A leapfrog fullerene satisfies automatically the IPR and is guaranteed to have closed Hückel shells [80]. It may appear cumbersome in practice to obtain a leapfrog cone by explicitly performing the operation, so it is useful to know that a fullerene has a *Fries* Kekulé structure if, and only if, it is a leapfrog [81]. This particular Kekulé structure is the one where the maximum fraction the hexagonal faces can be drawn as benzenoid rings, and its name is due to the early organic chemist who proposed a correspondence between the stability and the fraction of benzenoid rings for hydrocarbons. According to the discussion in the last paragraph, the

molecules with Fries Kekulé structures are favorable as the double bonds can be kept outside the pentagons; the three types of faces in the Fries structural formula of a leapfrog fullerene are shown below.



Except for the leapfrogs, which consist of $n = 60 + 6k$ carbon atoms, where k can be any non-negative integer except 1, there are two other series of fullerenes [82], $n = 70 + 30k$ and $n = 84 + 36k$ with all bonding orbitals fully occupied, and thus, as far as the Hückel theory applies, stable against Jahn-Teller distortions. In the two latter series, there are no restrictions on k, except that it must be a non-negative integer. The limitation on k in the leapfrog series is due to the nonexistence of cubic graphs with 12 pentagons and 1 hexagon (22 vertices). Contrary to the leapfrogs, the fullerenes of the last two series have two unbonding orbitals and should therefore be more reactive. Very few fullerenes with closed Hückel shells that do not belong to one of these three series are found. Most of the topologically possible fullerenes have either more than half of the Hückel energy eigenvalues negative (pseudo-closed), or identical HOMO and LUMO energies (open shell). HOMO and LUMO mean, respectively, highest occupied and lowest unoccupied molecular orbital.

If there are dangling bonds at the base end of an open-ended cone, there is no well-defined bonding topology allowing stability considerations. If the open base end is terminated by hydrogen atoms, on the other hand, we can say a lot more. Because the carbon atoms don't make π-bonds with the terminating hydrogens, the molecular graph determining the Hückel spectrum is obtained by deleting the hydrogens and their adjacent σ-bonds. In a recent work it is proved [83] that all even-membered hydrocarbon cones having a Fries Kekulé structure or an induced molecular subgraph consistent with a Fries Kekulé structure have closed Hückel shells. Thus, we have a common recipe for constructing the *tips* of chemically stable fullerene and hydrocarbon cones. The tightest configuration of pentagons at a cone tip consistent with a Fries Kekulé structure is obtained simply by placing the pentagons as close as they can be without breaking the IPR. All wider distributions of the pentagons consistent with a Fries Kekulé are leapfrogs of the tightest one. The relative stability of the latter are not yet investigated. In addition, the optimal cap topologies of the leapfrog fullerene cones have not been identified.

2.3. Helical Cone Growth

Instead of joining the dangling bonds at the opposite edges of the cut in Fig. 5, imagine a continuous helical growth of the edges around a central screw dislocation. The result is a helically wound structure that can be enveloped by a conic, cylindric, or irregular surface. This is the growth form of helical graphite cones. It is different from the whisker growth explored by Bacon in 1959 [1], where the scrolled graphene sheet is aligned with the whisker axis, and the conic shape is realized when the winding axis is not perpendicular to the whisker axis.

A decade after Bacon's work, Double and Hellawell found by optical, scanning, and eletron microscopy that the flake graphite crystals in metal-carbon eutectics [84, 85], where the alloy is constrained to grow along a duplex front, could change orientation during growth by formation of stacking faults involving rotations corresponding to disclination angles of $n(\pi/3) \pm \omega$, with ω equal to 13.2°, 21.8°, and 27.8°. In 1974, the same authors proposed a cone-helix growth mode [86] for graphite consistent with these observations and other contemporary works [87–92] suggesting crystal growth with components normal to the basal graphene plane. In particular, pyrolysis of CO on a SiC substrate at temperatures above 1800°C [91, 92] produced whiskers where the growing ends had the shape of a cone with 140° apex angle. Notice that this angle is far off the largest apex angle consistent with the

pentagon model outlined in Section 2.2. The corresponding disclination angle is 21.8°, and the cone-helix growth conjecture is supported by the fact that this angle gives an optimum coincidence configuration of the carbon atoms in the self-overlapping graphene sheet. Similar "good fit" arrangements occur for the other ω values given above, and exact coincidence is achieved for the disclination angles allowed in the pentagon model.

In 1992, a work [93] inspired by the contemporary speculations on nucleation mechanisms for fullerenes was published by authors apparently unaware of the work of Double and Hellawell. Based on electron microscopy and diffraction studies of the old samples of [92], they formulated a cone-helix growth model that differed from the one just outlined only by a proposal of stacked five-membered carbon rings in the core of the screw dislocation. This implies only two four-connected carbon atoms at each turn of the screw, and the reality of the concept can hardly be pursued by macroscopic geometry considerations.

It would be difficult to omit the helical cones from this review, as there is an interesting theoretical connection between a nonhexagonal face in a nanocone and the screw dislocation in the cone-helix growth model, which will be discussed in Section 3.5.5. In addition, the first-known natural occurrence of large arrays of graphite cones were recently reported by Jaszczak et al. [2]. Surfaces of spherical or spheroidal aggregates of graphite enclosed by calcite were found covered with cones, as shown in the high-quality field emission scanning electron microscopy (FESEM) images of Fig. 7. Cones as long as 40 μm, which is far beyond other reported graphite cone lengths, were observed. The graphite spheres, found in the Central Metasedimentary Belt of the Canadian Grenville province, are presumably deposited by metamorphic fluids [2], and the peak metamorphic temperatures in this area are believed to be below 700°C [94]. About 2% of the examined spheres had surfaces covered with cones, and the presence of a small pyrite crystal was associated with the cone formation.

Figure 8 shows the distribution of measured apex angles for these naturally occuring graphite cones, with the angles favored by the cone-helix and the pentagon model highlighted. Together with the layered growth ripples seen in the scanning electron microscopy (SEM) and FESEM images of Fig. 9, the angle distribution clearly supports the cone-helix growth model. The lower panel of Fig. 10 shows the *renormalized* distribution of measured apex angles (upper panel) for the nanocones synthesized by Krishnan et al. [25]. In that experiment, rather large error bars were assigned to the measured angles, and the

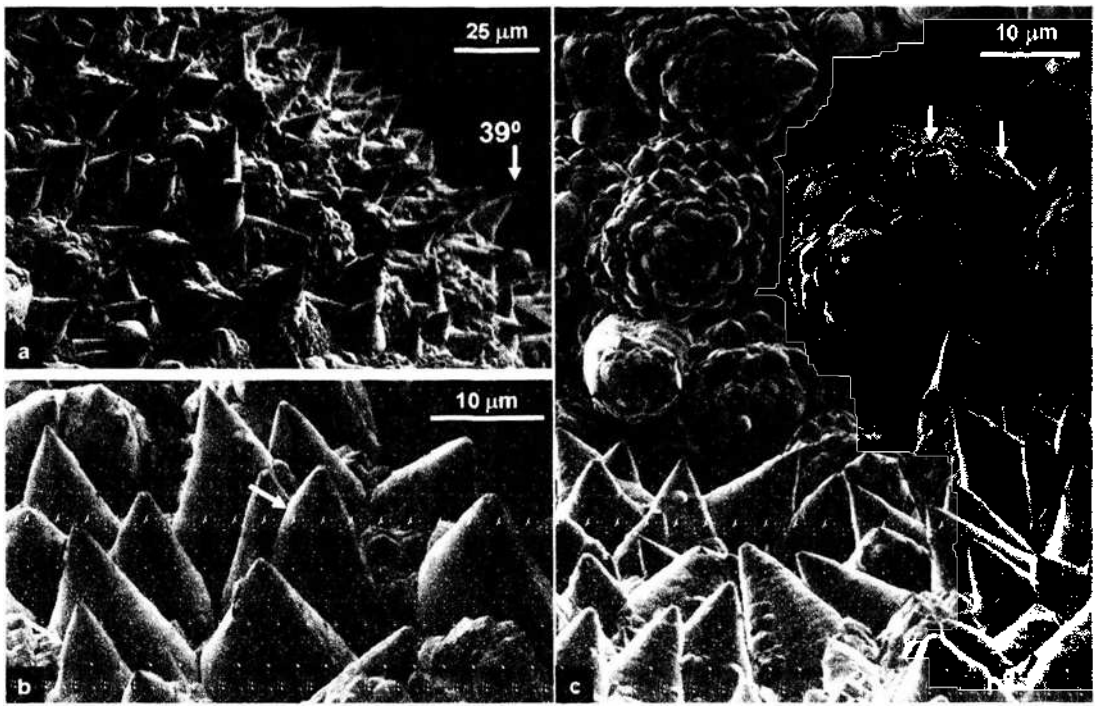


Figure 7. FESEM images of a cone-covered graphite aggregate. In (b) and (c), the arrows show, respectively, change in apex angle and ripples. Reprinted with permission from [2], J. A. Jaszczak et al., *Carbon* 41, 2085 (2003). © 2003 Elsevier

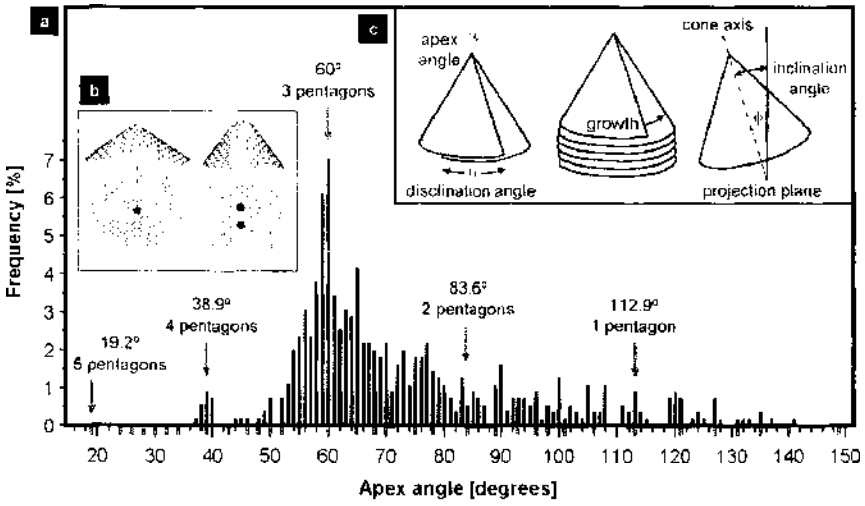


Figure 8. Freqency histogram of the natural occurence of various cone apex angles in a single sample. Arrows shows the apex angles predicted by the pentagon model schematically shown in inset (b). Outlined bars below the horizontal axis show apex angles favored by cone-helix growth model shown schematically in inset (c). Reprinted with permission from [2], J. A. Jaszczak et al., *Carbon* 41, 2085 (2003). © 2003, Elsevier.

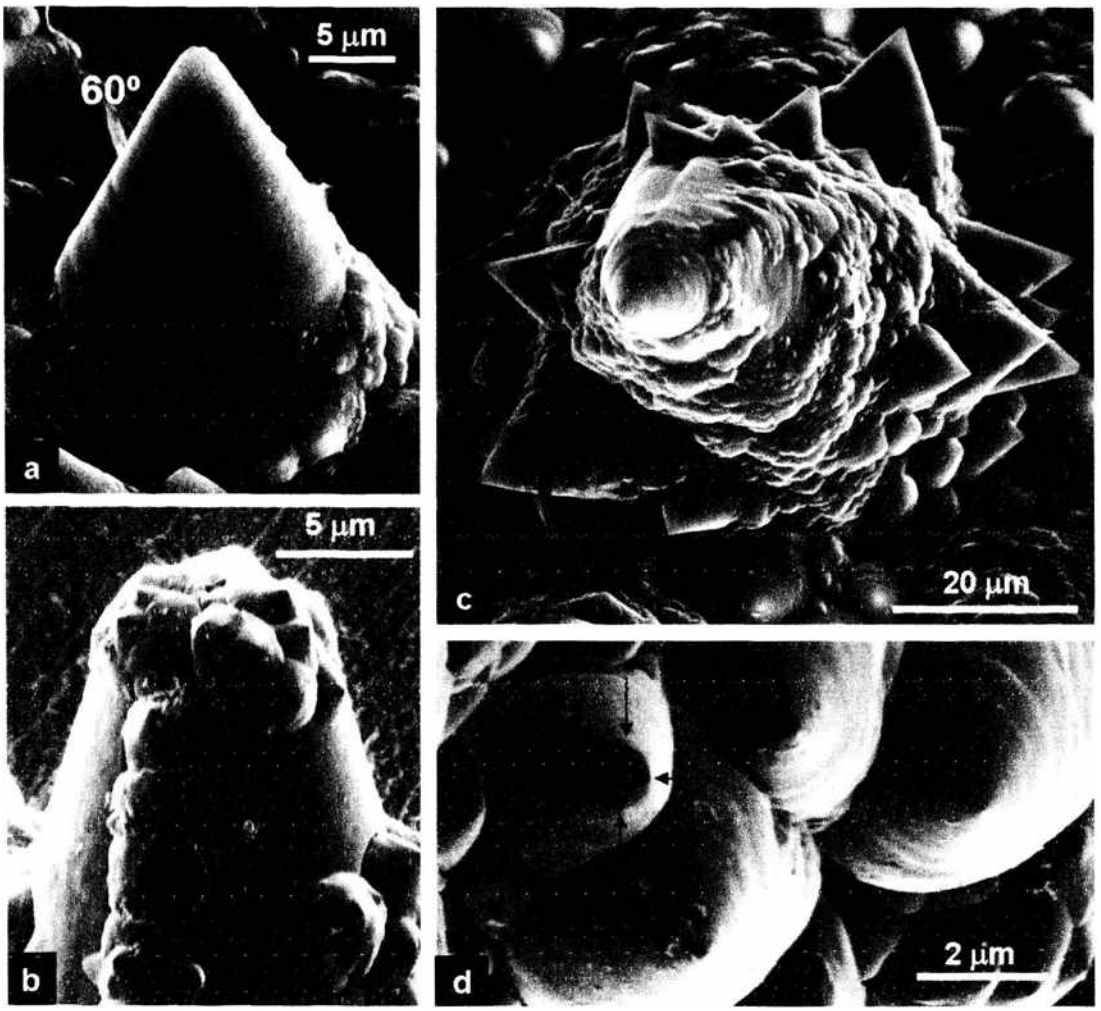


Figure 9. SEM (a and b) and HESEM (c and d) images of naturally occurring graphite cones with the predominant 60° apex angle. The arrows in (c) show mutliple tips. Reprinted with permission from [2], J. A. Jaszczak et al., *Carbon* 41, 2085 (2003). © 2003, Elsevier.

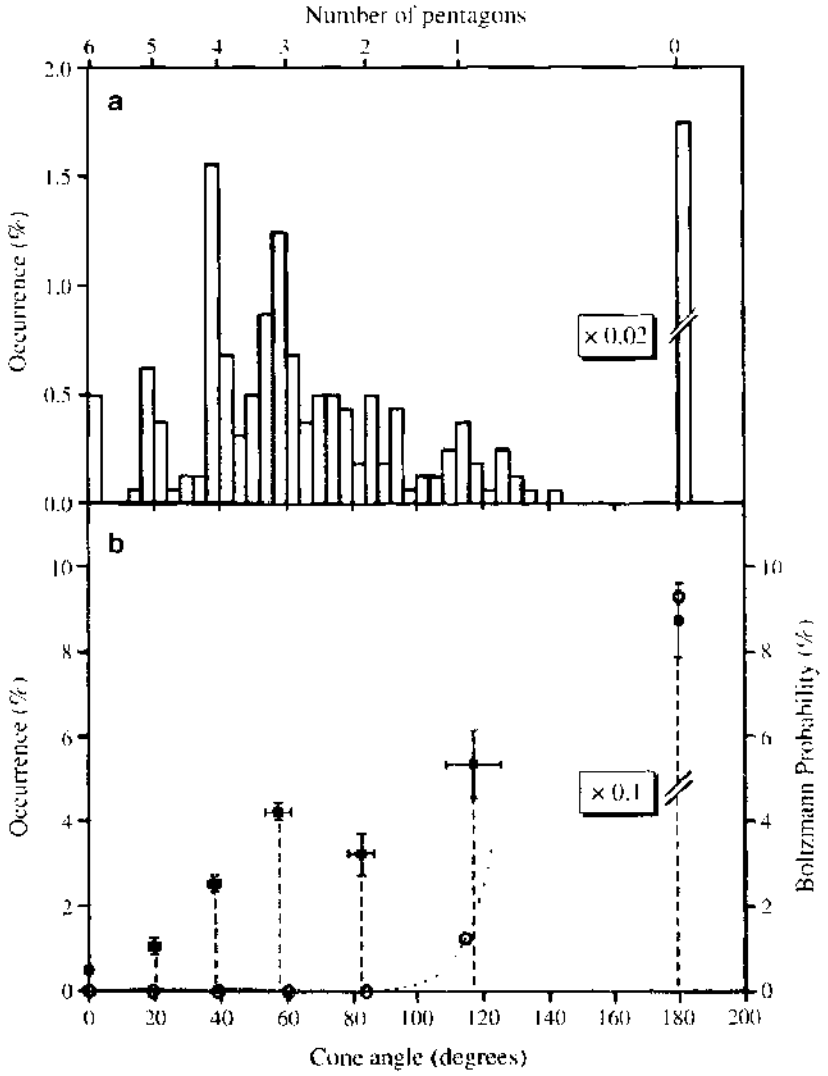


Figure 10. Measured apex angles of cones synthesized in Kværner's carbon-black process. Statistical distribution (a) and averaged distribution centered at the angles allowed by the pentagon model (b). Reprinted with permission from [25], A. Krishnan et al., *Nature* 388, 451 (1997). ©1997, Nature Publishing Group.

renormaliztion consisted of taking the total count at the average angle of each peak, and correcting for underestimation of larger apex angles. The result is the shown discrete distribution in close agreement with the pentagon model. If we accept this procedure, it must be concluded from Figs. 8 and 10 that the 60° apex angle is predominant both for naturally occuring graphite cones and synthesized nanocones. The unavoidable question then arises: Are there considerable amounts of nanocones among the naturally occurring graphite cones? This is a tough nut: The predominance of 60° apex angle seems to be consistent with the cone-helix model, as this corresponds to a screw dislocation giving exact coincidence of the self-overlapping graphene sheet [86] and, as pointed out in [2], smaller apex angles may be disfavored by the resulting higher elastic bending energy. If we regard bending energy as an important parameter for the nanocones, on the other hand, cones with larger apex angles should be more abundant, and the peak at 60° did in fact surprise the authors of [25]. The answer seems to be outside the reach of the theory presented so far in this review. Also, the role of the pyrite crystal and the enclosing calcite layer is not understood.

2.4. Horns, Pipettes, Zipped Structures, and Tubular Graphite Cones

Rearranging the six pentagons at a conventional dome-like nanotube end cap will yield a tube with a conic tip of lower symmetry. Such asymmetric single-walled tubes are called nanohorns, and were synthesized for the first time by Iijima et al. [4] in 1999. An unusual high

production rate, up to 10 g/h, was obtained by CO_2 laser ablation at room temperature in the absence of a metal catalyst. The average apex angle of the tips was about 20°, corresponding to the pentagon configuration of lowest symmetry, namely five pentagons at the tip proximity (which gives 19.2° apex angle according to the pentagon model). The various pentagon distributions consistent with this observation and the resulting geometries and electronic structures were subsequently studied in terms of *ab initio* methods by Berber, Kwon and Tománek [5]. Nanohorns have been considered as a storage material for hydrogen, and an adsorption model was recently proposed [32].

The sharpest cone-shaped carbon structures ever found are the nanopipettes reported by Mani et al. in 2003. It was shown that the pipettes are formed by wrapping a graphene sheet helically around a central nanontube. At each turn, the wrapped sheet falls just short of overlapping itself, such that the surface of the pipette is smoother than the surface of an ordinary conical whisker. Due to the very sharp tips and the hollow cores, the pipettes are being considered for high-precision drug delivery; the standard microneedles fabricated for that purpose are typically wider at the tip [95].

Another new class of graphite cones found in pores of glassy carbon were reported by Dimovski, Libera, and Gogotsi [96] in 2002. The small apex angles varied within a narrow range with distinct peaks at 3°, 7°, and 15°, clearly outside the realm of the pentagon model. Selected-area electron diffraction and high-quality TEM images revealed that these structures were made of thick graphitic walls, up to several hundred layers, around a hollow cone-shaped core. The tips were dome-like, as a result from zipping between adjacent layers. It was suggested that these cones originated from a herring-bone carbon nanotube [97]; deposition of carbon at the terminated planes and extended growth beyond the core of the herring-bone tube can form a conical cavity and zipped layers at the tip. Similar structures, but with a nickel catalyst at the tip, have been grown by Merkulov et al. [98–100].

With the conic structures synthesized by Zhang, Jiang, and Wang [101] in 2003, the geometric possibilities must now be near exhausted. HRTEM images show that these tubular cones have hollow cores with constant radii, and the conic tips result from a gradual shortening of the concentric tubular layers along the cone axis. The distance between the layers is the same as for multiwalled nanotubes. Despite this, the outer surfaces of these cones have helically wound ridges, which isolated could be taken as an indication of whisker structure. This observation was, however, attributed to stress inhomogeneities due to noncircular endings of the individual layers at the conic tip. The tubular cones have wider tips than the nanopipettes, with radius of curvature about 1 μm, which corresponds to the sharpness of fabricated microneedles [95].

3. ATOMIC AND ELECTRONIC STRUCTURE: THEORETICAL/COMPUTATIONAL METHODS AND PREDICTIONS

This part describes the methods that have been central in the investigation of the atomic and electronic structure of carbon nanocones, and all significant results obtained will be discussed. With the exception of a paper of Shenderova et al. [58], no published theoretical or computational work involves the atomic coordinates of carbon nanocones in excited states. In the ground state, the atomic coordinates are understood as the average positions around which the atoms vibrate in the lowest energy mode. According to the Born-Openheimer approximation, the electronic wave function of the ground state depends parametrically on these coordinates. Except in the topological or continuum models to be discussed, the atomic coordinates are thus explicitly needed in order to perform electronic structure calculations. The next three sections describe the most frequently used methods to obtain these coordinates and/or the resulting electronic structures.

3.1. Molecular Mechanics

For covalent systems, this method offers the simplest version of dynamic simulations and the shortest route to an approximate ground state geometry. In the formal derivation of molecular mechanics (MM) (see, e.g., [102], [103], or [104]), the electronic degrees of freedom are

integrated out from the time-dependent Schrödinger equation to obtain the classical equations of nuclear motion in an effective potential, which attempts to incorporate the coupled nuclear-electronic quantum dynamics. MM simulations of the novel carbon materials require a potential where the π-bonding is realistically accounted for. The major developments in this direction are due to Brenner [105], who modified the Tersoff potential [106] into a form suitable for fitting to the large accumulation of carbon data.

The Tersoff potential was in turn formulated in terms of the repulsive and attractive pair interactions used by Abell [107] to explain the universality observed in binding-energy curves by Ferrante, Smith, and Rose [108, 109] in 1983. Expressed as a sum over atomic sites, the Tersoff potential reads [106]

$$E = \sum_i E_i = \frac{1}{2} \sum_{i,\, i \neq j} V_{ij} \tag{6}$$

where the interaction energy between two nuclei, V_{ij}, is given the mentioned pair form

$$V_{ij} = f_c(r_{ij})[A \exp(-\lambda_1 r_{ij}) - B_{ij} \exp(-\lambda_2 r_{ij})] \tag{7}$$

Here, $f_c(r_{ij})$ is an optional cutoff function, A, λ_1, and λ_2 are parameters to be fitted, and r_{ij} is the distance between the two atomic sites. The second term describes bonding, and B_{ij} is a function of r_{ij}, which implicitly includes the bond order. For this reason, potentials developed from the above expression are often referred to as bond-order potentials.

In 1990, Brenner [105] introduced the major modifications of (7) that appeared necessary in order to apply the potential in MM simulations of π-conjugated systems. These include additional terms to account for inherent overbinding of radicals by (7) and replacing B_{ij} by an implicit many-body term to account for nonlocal environmental effects on the bonding. Although the Brenner potential is computationally efficient, it is sufficiently complex that a discussion of the various terms and the 13 parameters must be omitted here. It is found to reproduce a large database of solid-state and molecular properties of carbon, including the in-plane lattice constant, cohesive energy, and elastic properties of graphene sheets [58]. Additional improvements were achieved by Burgos, Halac, and Bonadeo in 1998 [110]. These authors introduced a torsion-like term in the Brenner potential in order to overcome inadequate descriptions of the dynamics of covalent carbon systems. This step reduced the rms deviations for vibrational frequencies significantly, and the re-parameterized potential gave excellent agreement with data on systems not included in the fit. It was pointed out [110] that the torsion-like term introduced did not contribute to the static energies of the crystalline carbon forms.

As in the framework of the stochastic Monte Carlo approach (see, e.g., [111]), simulated annealing [112] can be carried out within the MM scheme. In this process, the energy is dissipated slowly from the system, which will then undergo all metastable states on its way to down to the global minimum.

3.2. Density Functional Theory

Density functional theory (DFT) [113–115] is the only *ab initio* method hitherto applied to carbon nanocones. The process of an *ab initio* simulation of a molecule is as follows: First the geometry of the molecule is optimized by iteration. In contrast to a MM simulation, where the electronic degrees of freedom are integrated out in advance, the electron density is updated at every stage of the *ab initio* iteration. If convergence is reached, hopefully to a geometry close to the the one corresponding to the global energy minimum, one can subsequently calculate a variety of electronic properties. There are several DFT implementations of this procedure available; the Amsterdam density functional (ADF) code [116, 117], CPMD [118], Gaussian [119], NWChem [120], CASTEP [121, 122], and VASP [123] are the most well-known. ADF is a specialized DFT code, while Gaussian and NW-Chem are multipurpose packages. CPMD stands for Car-Parrinello molecular dynamics. This remarkable method allows DFT simulations under user-specified temperature and pressure and can be used, among other things, to find global energy minima by simulated annealing (see previous

section). NWChem, CASTEP, and, of course, CPMD perform Car-Parrinello simulations in addition to regular DFT simulations. Because no Car-Parrinello simulations involving cones have so far been reported, this method will not be discussed further. ADF, Gaussian, and CASTEP are commercial while CPMD, NWChem, and VASP are public codes.

There are more accurate *ab intio* schemes than DFT, like quantum Monte Carlo (see, e.g., [124]), but the working principle of DFT is far more efficient, making the method a natural choice for extensive calculations. However, for large carbon structures, it is, for reasons to be pointed out below, not a trivial task to reach convergence. In addition, implementation of DFT involves an approximate treatment of exchange correlations, although the method has accurately reproduced the true geometry of a large variety of molecules.

The basis of DFT is the theorem of Hohenberg and Kohn [125], stating that the ground state electronic energy is determined completely by the electron density $n(\mathbf{r})$, which in turn depends parametrically on the the fixed nuclear coordinates. An efficient minimization procedure is therefore obtained by expressing the electon density as a sum over the occupied Kohn-Sham [126] (KS) orbitals $\{\phi_i\}$:

$$n(\mathbf{r}) = \sum_i a_i |\phi_i(\mathbf{r})|^2 \tag{8}$$

where a_i is the occupation number for each orbital (which equals 2 for all orbitals for an even number of electrons in the ground state). These orbitals are solutions of the (pseudo-eigenvalue) KS equations

$$H_{KS}\phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \tag{9}$$

and the KS Hamiltonian (in atomic units) is given by

$$H_{KS} = \frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}) \tag{10}$$

where the effective potential is composed of three terms:

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ne}}(\mathbf{r}) + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} + V_{\text{xc}}(\mathbf{r}) \tag{11}$$

where $V_{\text{ne}}(\mathbf{r})$ is the interaction between the nuclei and the electron and the second term is called the Hartree term and describes the mutual interaction between the electrons. Both these terms are simply Coulomb interactions. The last term, on the other hand, is the functional derivative of the unknown exchange correlation energy $E_{\text{xc}}(n)$:

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}(n)}{\delta n(\mathbf{r})} \tag{12}$$

which implicitly contains the remnants of the original many-body problem. In order to carry out calculations, $E_{\text{xc}}(n)$ is approximated by an integral over an exchange correlation energy density ϵ_{xc}, which depends only on the density and its gradient at the actual point in space [127].

$$E_{\text{xc}}(n) \simeq \int d\mathbf{r}\, \epsilon_{\text{xc}}[n(\mathbf{r}); \nabla n(\mathbf{r})] \tag{13}$$

In the simplest approximation, ϵ_{xc} is taken as the exchange correlation energy of an interacting electron gas with homogeneous density $n(\mathbf{r})$ [126, 128]. This is called the local density approximation (LDA), and is known to work remarkably well. To improve from this approximation, it is necessary to consider nonuniform electron gases in the evaluation of (13). The various methods developed along this line [127, 129–132] usually incorporates derivatives of the density, too, and are referred to as generalized gradient approximations (GGA).

For geometry optimization of large systems, like carbon nanocones, the first problem to appear is not the inaccuracies due to the above approximations. Much larger deviations from the ground state geometry can occur as a result of convergence into one of the many geometries corresponding to local energy minima, if the optimization converges at all. It is

well-known that the DFT algorithm tends to assign too low energies to the unoccupied KS orbitals [117]. As a result, the electrons may reorganize themselves between the orbitals at successive geometry iterations, and convergence is hampered. This can be overcome by forcing the algorithm to keep the occupation numbers, but this will in many cases hinder convergence of the self-consistent electron field calculations to be carried at each step in the geometry optimization. The user manual for the ADF code [116, 117] suggests several tricks to overcome these problems.

There are practically two different implementations of DFT. ADF and Gaussian expand the KS orbitals in basis sets of user-specified complexity (and quality). The other codes mentioned in the beginning of this section use plane-wave basis sets and pseudo-potentials. This implementation, descried in detail in [104], requires in addition simulation cells, so the user will have to deal with more strategic parameters in a simulation.

Despite the relative effectiveness of the DFT algorithm compared to other *ab initio* schemes, the required CPU time is sufficiently long that complementary computationally faster methods are widely used in the study of nanocones. These methods, which will be discussed next, necessarily imply simplifying models that have given both insight and controversies.

3.3. The Tight-Binding Model

The tight-binding (TB) model [133, 134] resembles in some aspects the the crude Hückel model and its various extensions discussed in Section 2.1, and its efficiency has made it one of the most widely used electronic structure models. In addition, the model is the basis of a powerful molecular dynamics simulation technique (TBMD) [135–138]. In this scheme, the total energy E_{tot} of the given system is expressed as

$$E_{\text{tot}} = \sum_{\alpha} \frac{p_{\alpha}^2}{2m_{\alpha}} + E_{\text{bs}} + U_{\text{rep}} \tag{14}$$

where the first term sums the kinetic energies of all the atoms constituting the system, and U_{rep} is an effective potential taking account of repulsive effects. It is the middle term, called the *band structure* term,

$$E_{\text{bs}} = \sum_{i} \epsilon_i f(\epsilon_i, T) \tag{15}$$

which extracts most of the CPU time during calculation of the eigenvalues

$$\epsilon_i = \langle \Psi_i | H | \Psi_i \rangle \tag{16}$$

where H is the tight-binding electronic Hamiltonian for the given atomic coordinates, and Ψ_i is the ith eigenfunction of H. The factor $f(\epsilon_i, T)$ is simply the Fermi-Dirac energy distribution at temperature T. The Hellmann-Feynman force [139, 140] is the derivative of E_{bs} with respect to atomic displacements. In its present formulation, the computational cost of a TBMD simulation is in cubic proportion to the number of atoms, and several ways around this bottleneck have been proposed [141–150]. The majority of these efforts are based on the concept of localized orbitals [141–144] calculation of the density matrix [145–149]. The idea of Goedecker and Colombo [150], on the other hand, was to replace the Fermi-Dirac distribution with the matrix

$$F_T = f(H, T) \tag{17}$$

which enables the transformation of (15) into the expression

$$E_{\text{bs}} = \text{Tr}[HF] = \sum_{i\alpha j\alpha'} \langle \phi_{i\alpha} | H | \phi_{j\alpha'} \rangle \langle \phi_{j\alpha'} | F | \phi_{i\alpha} \rangle \tag{18}$$

where $\phi_{i\alpha}$ is the ith TB basis function centered on atomic site number α. According to [150], evalution of the traces scales linearly with the number of atoms, and the sum of the sum

over the matrix products of (18) is well suited for parallel processing and can be computed efficiently.

For systems in the ground state, the calculations of atomic and electronic structure can to a large extent be done separately; the simplest TB Hamiltonian is the Hückel matrix, which only reflects topology, and the atomic coordinates for more advanced TB interactions can be determined by other means, for example one of the MM potentials discussed in Section 3.1.

For quantitative electronic structure calculations, realistic potentials for the TB Hamiltonian have been developed. The environment-dependent binding potential of Tang et al. [151] goes beyond the traditional two-center approximation [152] of the hopping integrals by allowing the parameters and the repulsive energy to depend on the binding energy, thereby improving the transferability between different carbon systems significantly. A comprehensive parameterization of the TB interaction for CH bonds at the ends of hydrocarbons was done previously by Davidson and Pickett [153]. In its simplest form, the TB Hamiltonian is equivalent to the Hückel matrix discussed in Section 2.2, with identical interactions—hopping integrals—between nearest neighboring atoms.

3.4. The Effective Mass Theory and Continuum Models

The "effective mass theory" (EMT) has been applied to band analysis of periodic structures since the beginning of the 1930s [154–161]. The first modification of the EMT for the case of perturbed periodic fields was done by Luttinger and Kohn [162] in 1954, and the developments due to DiVincenzo and Mele 30 years later [163] is the basis for two different continuum models [164, 165] currently applied to the cones [166–170].

For a flat graphene sheet there are two degenerate Bloch states at the K point of the first Brillouin zone, and the EMT approach is equivalent to the $k \cdot p$ expansion [171] of the π-bands about K; the essence of both theories is to replace the graphene bands by conical dispersions at the Fermi level E_F. In atomic units, the EMT equation for the one-electron wave function $\Psi(\mathbf{r})$ then takes the form [163]

$$-iv_F\left(\sigma_1\frac{\partial}{\partial_x}+\sigma_2\frac{\partial}{\partial_y}\right)=E\Psi(\mathbf{r}) \tag{19}$$

where v_F is the Fermi velocity and E is the one-electron energy. The meaning of the Pauli matrices σ_1 and σ_2 will be explained below. As pointed out in [163], the Pauli matrices anticommute, and (19) is thus algebraically identical to a two-dimensional Dirac equation. The difference between a cone and a flat sheet may therefore be incorporated in terms of gauge fields [166–170].

Because the controversy resulting from the two different gauge transformations applied to the cones [166–170] has not converged yet, the LCAO expression

$$\Psi(\mathbf{r})=\sum_{\mathbf{R}}C(\mathbf{R})\phi(\mathbf{r}-\mathbf{R}) \tag{20}$$

as a function of the distance from a pentagonal defect in an otherwise perfect hexagonal graphene sheet has been a valuable reference point. Here, ϕ and $\mathbf{R}$ are the π-orbitals and positions of the atoms, respectively. For a single pentagon, application of the EMT Hamiltonian to (20) and satisfaction of boundary conditions at the atom sites of the pentagons gives [172, 173] the coefficient

$$C(\mathbf{R})\propto f(\mathbf{R})z^{(3m+2)/5} \tag{21}$$

at the Fermi level. Here, m is an integer and $z=Re^{i\theta}$, where R is the distance from the center of the pentagon and θ is the azimuthal angle. The factor $f(\mathbf{R})$ is shown in Fig. 11 together with the K and K' points of the Brillouin zone. At the A and B sites of the pentagon, $f(\mathbf{R})$ is given by

$$f(\mathbf{R}_A)=e^{i\mathbf{K}R_A} \tag{22}$$

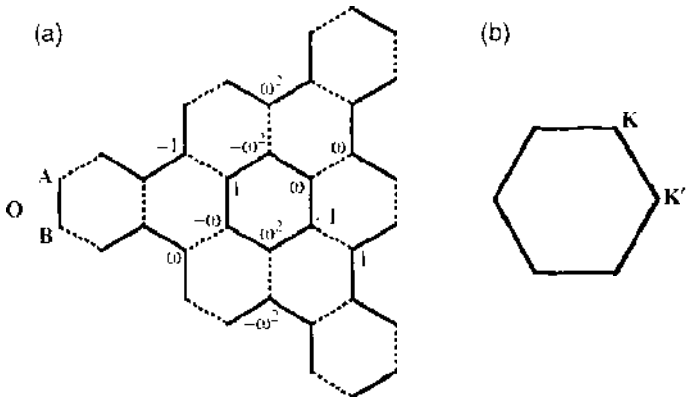


Figure 11. (a) The factor $f(\mathbf{r})$ defined in (22) and (23) on the sites A and B, respectively. Dotted lines show antiphase bonds. (b) The first Brillouin zone with the K and K' points shown. Reprinted with permission from [173], K. Kobayashi, *Phys. Rev. B* 61, 8496 (2000). © 2000, American Physical Society.

and

$$f(\mathbf{R}_B) = e^{i\mathbf{K}'\mathbf{R}_B}(-1)^m\omega^2 \tag{23}$$

where $\omega = e^{2\pi i/3}$. In a 2000 paper, Kobayashi [173] applied this theory to simulation of STM images of open-ended cones terminated by hydrogens, and the results will be discussed in Section 3.5.4.

Of the two gauge-field approaches to the cones, we briefly outline the one of Kochetov and Osipov [164, 167–169]. The cut-and-paste procedure used in Section 2.2 to make a cone from a flat graphene sheet can be formalized as the embedding

$$(r, \theta) \longrightarrow (ar\cos\theta, ar\sin\theta, c) \tag{24}$$

of r and θ defined in the last paragraph. The cone parameters a and c define the apex angle; $\sin(\varphi/2) = a/\sqrt{a^2 + c^2}$, and the convenience parameter $\chi = 1 + c^2/a^2$ is introduced in [167]. As found in [163], the EMT Hamiltonian for two-dimensional graphite is algebraically identical to the Dirac equation. The physical difference is that the two-component spinor represents the two sublattices instead of spin up or down. To avoid a lengthy declaration of the indices in the Hamiltonian derived in [167], we just mention that it commutes with the angular momentum for a two-dimensional Dirac system and that the gauge field accounting for the curvature is Abelian. The solutions have the spinor form

$$\Psi(\mathbf{r}) = \begin{pmatrix} u(r)e^{ij\theta} \\ v(r)e^{i(j+1)\theta} \end{pmatrix} \tag{25}$$

where j is an integer unit of angular momentum. (The spin connection term of the Hamiltonian does not contribute to the Dirac equation in two dimensions [174].)

An essential difference between this theory and the alternative scheme of Lammert and Crespi [166, 170] is captured by the general solution

$$\begin{pmatrix} u(r) \\ v(r) \end{pmatrix} = Ar^{\xi} \begin{pmatrix} J_{\eta}(\sqrt{\chi}aEr) \\ \pm J_{\bar{\eta}}(\sqrt{\chi}aEr) \end{pmatrix} \tag{26}$$

of the reduced equation for $u(r)$ and $v(r)$. In (26), A is a normalization factor, $\xi = (1 - \sqrt{\chi})$ and $J_{\eta}(J_{\bar{\eta}})$ are Bessel functions. Introducing the Frank index $\nu = 1 - \sin(\varphi/2)$, the disclination angle divided by 2π, η and $\bar{\eta}$ can be expressed as $\eta = \pm[\sqrt{\chi}(j + \nu + 1/2) - 1/2]$ and $\bar{\eta} \pm [\sqrt{\chi}(j + \nu + 1/2) + 1/2]$. The omnipresent χ factor, which characterizes the geometry of the cone, has no direct analogue in the solution of the Hamiltonian proposed in [166], where the gauge field instead carries information about the changed boundary

conditions when a graphene sheet is changed into a cone by the cut-and-paste procedure. For comparison, the radial solutions obtained in the Scheme of Lammert and Crespi read [170]

$$\begin{pmatrix} u(r) \\ v(r) \end{pmatrix} = \begin{pmatrix} (i\epsilon)^{-1/2} J_{\epsilon(\mu-1/2)}(kr) \\ \pm(i\epsilon)^{-1/2} J_{\epsilon(\mu+1/2)}(kr) \end{pmatrix} \tag{27}$$

where $\epsilon = \pm 1$, $k = E/\hbar v_F$, v_F is the Fermi velocity, and the the index μ is defined as

$$\mu = \frac{j - \Phi_K}{1 - \Omega} \tag{28}$$

where Φ_K is the normalized fictitious magnetic flux introduced by the gauge field, and Ω is the number of disclinations. Another difference is that the latter gauge fluxes are non-Abelian, which is instrumental for certain predictions to be discussed in Section 3.5.3.

3.5. Predictions

In this section, we review and, when possible, compare the results predicted by the various methods discussed in the last section. Because no firm theoretical predictions on the atomic or electronic structure of fullerene cones are published yet, it is implicitly understood that the term nanocones refers to open-ended cones in the following.

3.5.1. Relative Stability of Isomeric Cone Tips

The question of relative stability of different cone geometries and topologies can clearly be addressed by DFT or other schemes taking into account the interplay between geometry and electronic structure stated by the Hohenberg-Kohn theorem. It will, however, be impossible to restrict the analysis without any guiding or emerging theory. The only investigation of this type was done in 1997 by Han and Jaffe [175]. They calculated various tip topologies in terms of conjugated hydrocarbons to establish some rules beyond IPR for the relative position of the pentagons.

For each tip topology with two to four pentagons, they calculated the bonding energy of the isomers obtained by relocation of the pentagons among the hexagons allowed by the IPR rule. The maximum number of carbon atoms was 56, the least number needed to circumscribe the five pentagons of the ~19° tip with hexagons in the next to tightest configuration. The computations were done in DFT, POAV, and additional methods not mentioned previously in this review. The result for all the tips was the same: The tightest configuration of the pentagons allowed by the IPR rule is the most stable. This finding was attributed to less in-plane and out-of-plane strain of these tips.

What is omitted from the analysis is the fact that these configurations were the only of those investigated consistent with a Fries Kekulé structure. In the valence bond approach discussed in Section 2.2.2, the conclusion would have been the same as the one reached by computing. The question is then about the relevance of the valence bond theory. In [83] it was proved that all even hydrocarbon cones with tips consistent with a Fries Kekulé structure had closed Hückel shells. Calculating the Hückel spectra of the isomers compared to these tips in [175], it is found that, except for the case with two pentagons, they are either open-shell (three pentagons) or pseudo-closed (four and five pentagons). These isomers are likely to be Jahn-Teller distorted into structures with enhanced bond-alternation [176], which in turn makes the valence bond picture more relevant than for the closed shell structures. The good thing with the *ab initio* simulations is that they incorporate all these effects, and their predictions are independent of any theoretical confusion about the parts played by the various effects. The rules established by Han and Jaffe were used to choose the relative position of the three pentagons at the 60° tip studied by DFT and TB in [177]. It remains to investigate the relative stability of wider distributions of the pentagons consistent with Fries Kekulé structures.

3.5.2. Meta-Stable Structures

The 2001 paper of Shenderova et al. [58] is so far the only work involving investigation of the atomic structure when the nanocones are not in the ground state. As pointed out in [58], carbon nanocones with one pentagon at the tip are geometrically compatible with the diamond pentaparticles observed in vapor phase growth of diamond nanoparticles [178] and as five-fold micro-crystals embedded in chemical vapor–deposited diamond films [179]. Atomistic simulations of these structures were done previously by Shenderova and Brenner [180].

In [58], generation of nanocones by cutting layers of the pentagonal diamond crystals were computer-modeled. Simulated annealing (see Section 3.1) via MM gave wavy metastable structures for cones radii larger then 14 Å. Figure 12 shows three of these states together with the conventional shape of the lowest energy configuration. This behavior was attributed to compressive stress in the inner region and tensile stress in the outer region, both stemming from the inherent misfit of the diamond pentaparticle. When the temperature was raised above 300 K in the simulation, the wavy cone shapes converted to the conventional one. The many-body potential used in these MM simulations was that of Brenner [105] discussed in Section 3.1. It is hard to judge if the additional torsion-like term [110] discussed in the same section would have changed the results noticeably.

Another unique feature of this work was the modeling of cone assemblies. It was found that the structure to the right in Fig. 13 possessed the lowest residual stress and system energy.

3.5.3. Local Density of States, Field Emission, and Magnetic Properties

The local density of states (LDOS) at the tip of a cone or the end cap of a tube is directly connected to the field emission (FEM) properties. If there is a peak near the Fermi level, the threshold voltage for emission is decreased, and assemblies of cones have been considered [177] for flat panel displays [181] and templating [58]. Consequently, there are several theoretical papers dealing with the connection between the LDOS and the "topological defects,"

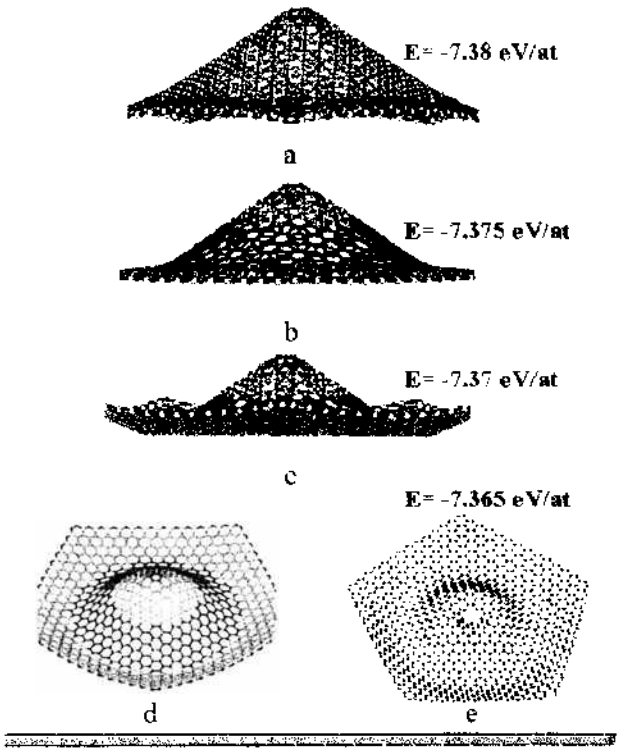


Figure 12. Four configurations for nanocone with radius 32 Å. Configuration (a) is stable, while (b), (c), and (e) are metastable structures. The top view of configuration (e) is shown in (d). The shading illustrates the distribution of hydrostatic stress. Reprinted with permission from [58], O. A. Shenderova et al., *Nanotechnology* 12, 191 (2001) © 2001, Institute of Physics Publishing.

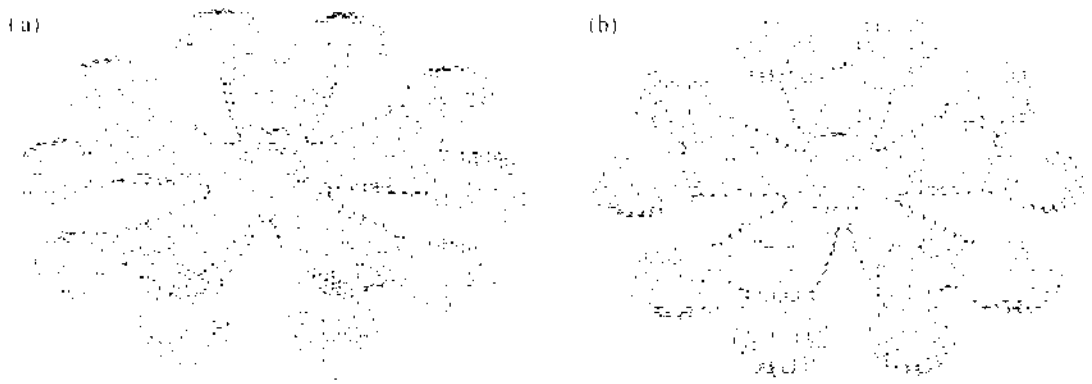


Figure 13. Two-dimensional free-standing structures with (a) different or (b) same type of convex/concave angle for consequent circular groups of cones. Reprinted with permission from [58]. O. A. Shenderova et al., *Nanotechnology* 12, 191 (2001). © 2001, Institute of Physics Publishing.

that is, the nonhexagonal faces of the curved graphene sheets [166–170, 173, 177, 182–185]. Most remarkable is the predicted LDOS of cones with three pentagons at the tip and the divergency of the predictions for cones with two pentagons at the tip.

We start with the most transparent results. In a series of papers in the 1990s [182–184], Tamura and Tsukada (with additional authors in [184]) investigated the LDOS at sites belonging to n-membered rings in the otherwise hexagonal graphene sheet. The aim of these calculations was the shear topological aspects, so the simplest TB Hamiltonian with constant hopping integral T was used (see Section 3.1), and the Fermi level was chosen as the zero point of the energy. For a single n-membered ring, the symmetry of the model sheet can be chosen n-fold, and the local density of states at any of the n sites of the ring is given by [184]

$$N_i^{(n)}(E) = \frac{1}{n}\sum_{j=0}^{n-1} N_i\left(\frac{2\pi}{n}j, E\right) \tag{29}$$

where i refers to any of the sites, and

$$N_i(k, E) = \sum_{\alpha} |\langle i|\alpha, k\rangle|^2 \delta(E - E_\alpha) \tag{30}$$

where $\langle i|\alpha, k\rangle$ is the projection of the state α with symmetry $k = 2\pi/n$ on site i. $N_i(k, E)$ can be expressed as a Green's function

$$N_i(k, E) = \frac{1}{\pi}\mathrm{Im}\langle i|E - H^{(k)}|i\rangle \tag{31}$$

where $H^{(k)}$ is the k-fold symmetric TB Hamiltonian. This latter expression was calculated by a recursion method invented by Haydock, Heine and Kelly [187].

Figure 14 shows the LDOS so obtained for a single n-membered ring in a graphene sheet for $n = 4$ to $n = 8$. The asymmetry with respect to the Fermi level for the odd-membered rings is due to the lack of a sublattice structure [182]. The symmetry seen in the LDOS for the even-membered rings is likely to be disturbed when other than purely topological effects are taken into account.

In 2001, Charlier and Rignanese [177] applied the full TB molecular dynamics scheme outlined in Section 3.3 to model cones consisting of 2000–2500 atoms. These fairly large numbers were chosen to diminish the effects of dangling bonds at the ends, which in this case were not terminated by hydrogen. The TB Hamiltonian incorporated both $2s$ and $2p$ orbitals, and the LDOS for the optimized cone geometries with one to five pentagons at the tip were obtained by the recursion method [187] mentioned above. For the cones with two and four pentaons, two different configurations of the pentagons were investigated. As in the previous work of Tamura and Tsukada [182] it was found that the shape of the LDOS around the Fermi level is very sensitive to the relative position of the pentagons. The LDOS for the different configurations of two pentagons are, however, very different in these two works.

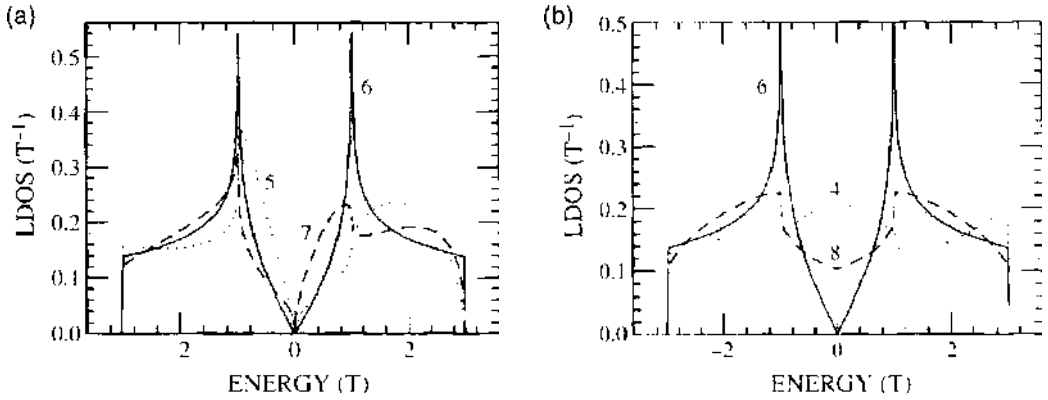


Figure 14. The LDOS at sites belonging to the n-membered rings for (a) $n = 5, 6, 7$ and (b) $n = 4, 6, 8$. The energy unit T is the constant hopping integral in the TB Hamiltonian. Dotted lines shows the LDOS for $n = 4$ and 5, dashed lines represent $n = 7$ and 8. Reprinted with permission from [184]. R. Tamura et al., *Phys. Rev. B* 56, 1404 (1997). © 1997, American Physical Society.

The most significant result of this work [177] was a high sharp peak just above the Fermi level in the LDOS of the cone with three pentagons (60° apex angle) in the configuration favored by the analysis of Han and Jaffe [175]. Charlier and Rignanese supported this result by DFT calculations of the LDOS for smaller cones terminated by hydrogen at the edges and suggested applications of the prominent FEM characteristics resulting from this peak.

The continuum model of Kochetov and Osipov [167–169] also predicts a distinct behavior for nanocones with 60° apex angle. In this model, the configuration of the pentagons cannot be incorporated as the cone tip is viewed simply as the origin of a conical surface. For the same reason, the local density of states diverges at this point, so it must be calculated as the total density of states $D(E, \delta)$ on a small patch $0 < r < \delta$, where r is the distance from the apex. In the leading order, the model of Kochetov and Osipov then gives [167]

$$D(E, \delta) \propto \begin{cases} E\delta^2, & \varphi = 0 \text{ (Flat sheet)} \\ E^{4/5}\delta^{8/5}, & \varphi = 112.9^\circ \\ E^{1/2}\delta, & \varphi = 83.6^\circ \\ E^0\delta^0, & \varphi = 60.0^\circ \end{cases} \tag{32}$$

where φ is the apex angle. Thus, according to this model, only the cones with 60.0° have nonvanishing density of states at the Fermi level, which is chosen to be the zero point of the energy E. This agrees well with the results of Charlier and Rignanese, except for the case with two pentagons configured as close as possible without breaking the IPR rule, where their TB calculation actually gives a peak in the LDOS at the Fermi level. This particular result of Charlier and Rignanese is, however, in turn in opposition to the topological analysis of Tamura and Tsukada [182], where the LDOS had a minimum at the Fermi level unless the pentagons were separated by at least one hexagon. The two-pentagon case was not followed up by *ab initio* calculations in [177].

The alternative continuum model of Lammert and Crespi [165, 166, 170] predicts the same δ-dependence of the LDOS as given in (32), but the energy dependence is entirely different. In their 2000 paper [166], they obtain they obtain analogously

$$D(E, \delta) \propto \begin{cases} E\delta^2, & \text{(Flat sheet)} \\ E^{3/5}\delta^{8/5}, & \text{(One disclination)} \\ \delta, & \text{(Two disclinations)} \end{cases} \tag{33}$$

which implies an enhancement of the LDOS at the apex of cones with two pentagons and is thus in agreement with Charlier and Rignanese, which again is in disagreement with Osipov and Kochetov, Tamura and Tsukada at this particular point. In [183] and [184], it was suggested that the nonzero LDOS at the Fermi level for the case with two pentagons

are due to power-law localized electron states. This understanding was supported by the later work of Kobayashi [173], which will be discussed in the next section. According to the 2000 paper of Lammert and Crespi, on the contrary, these states are extended. It was actually these conflicting points of view that attracted the attention of Osipov and Kochetov to nanocones. At present, this controversy is not resolved. The counter-arguments raised by Osipov and Kochetov in 2001 and 2003 [167–169] are not confronted in the 2004 paper [170] of Lammert and Crespi.

In the 2001 paper of Shenderova et al. [58], the LDOS at the single pentagon of a 45 carbon cone terminated by hydrogen was calculated in the TB model, using the environment-dependent TB potential for carbon [151], discussed in Section 3.3. The TB parameters of Davidson and Pickett [153] for the CH bonds were refined by DFT-based procedure further described in [58]. In contrast to the previously discussed results, these calculations gave a local maximum at the Fermi level. This difference indicates a correlation between the LDOS and the size of the cone. Maybe it can be explained simply in terms of end effects.

In the 1997 paper of Tamura et al. [184], it was shown that the LDOS could be tuned continuously by the presence of an apical magnetic flux through the relation

$$N_i^{(n)}(E) = \frac{1}{n}\sum_{j=0}^{n-1} N_i\left(\frac{2\pi}{n}[j+\Phi], E\right) \tag{34}$$

between the expression (29) for the LDOS at the ring defect and the normalized magnetic flux Φ. These ideas were developed further by Lammert and Crespi. According to their continuum model, the normal period of intrinsic magnetoconductance is extinguished for the case of one pentagonal defect. This is referred to as the anomalous Aharanov-Bohm [56, 188–190] effect in [166]. The non-Abelian nature of the gauge fluxes enabled a connection between the relative configuration of two pentagonal defects and the LDOS around the Fermi level, and agreement was achieved between the predictions of Tamura et al. in [182] and this continuum model. Additional subtle magnetic properties persistent in the ray-optics limit were predicted in [170].

3.5.4. The Origin of Petal Superstructures

In the 2000 paper of Kobayashi [173], the effective mass theory, as outlined in Section 3.4 was used to explain the origin of observed [191] petal superstructures in STM images of conical protuberances on graphitic nanoparticles. (The observations, by An et al., were published first in 2001.) A similar analysis, in terms of TB, have been done for tubes [192].

The simulation of the STM images included only the LDOS $\rho(\mathbf{r}, E)$ of the sample cluster, and the tunneling current I at bias voltage V and distance $\mathbf{r}$ from the tip is then expressed as [193]

$$I \propto \int_{E_F}^{E_F+eV} \rho(\mathbf{r}, E)\, dE \tag{35}$$

where E_F is the Fermi energy. Figure 15 shows a simulated STM image for a cone with one pentagon at the tip. The model system for the calculation of Kobayashi was a $C_{500}H_{50}$ cluster with a single pentagon in the center. The geometry was optimized by the MM potential of Burgos, Halac, and Bonadeo [110] discussed in Section 3.1.

The pronounced petal superstructure of Fig. 15 agreed with the previous observations of An et al. Figure 16 shows the close agreement between EMT and DFT simulations of the STM images. The sector corresponds to the map of the factor $f(\mathbf{R})$ (Fig. 11) appearing in the LCAO expression for the EMT wave function. The good agreement between experiment, DFT, and EMT enabled a clear interpretation of the superstructure: Recall from Section 3.4 the EMT expression for the coefficient

$$C(\mathbf{R}) \propto f(\mathbf{R}) z^{(3m+2)/5}$$

of the LCAO expansion of the wave function. It is seen that the solution with $m = -1$ gives the main contribution to the electron transport in STM; the solutions with other integer

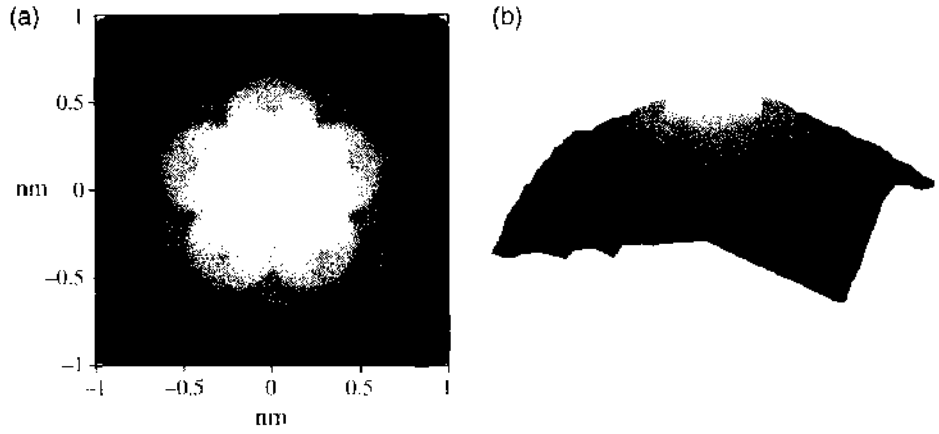


Figure 15. Simulated STM image of a nanocone with one pentagon at the tip; grayscale (a) and three-dimensional view of the constant-current surface (b). Reprinted with permission from [173]. K. Kobayashi, *Phys. Rev. B* 61, 8496 (2000). © 2000, American Physical Society.

m values are more localized. The lower STM image of Fig. 16 was in fact simulated with $m = -1$ only, and the petal superstructures can thus be explained by the phases of $f(\mathbf{R})$ shown previously in Fig. 11.

These superstructure are different those originating from scattering of extended Bloch waves at the K and K' points of the first Brillouin zone [194], commonly observed around defects of flat graphite surfaces. The qualitative difference is emphasized by the $r^{-1/5}$ decay of the wave function of [173] away from the pentagon. A later DFT calculation of charge densities by Yamaguchi [195] were also in agreement with these superstructures. Similarity between the electronic states of cones with two and three pentagons and those of C_{60} fullerenes adsorbed on silicon surfaces were additionally found in that work.

3.5.5. Large Ring Defects and Screw Dislocations

An interesting connection between large ring defects and screw dislocations in graphite, pointed out in the 1996 paper of Ihara et al. [185], was elaborated in the 1997 paper of Tamura et al. [184]. Their TB expression (29) for the LDOS at any site i of a single n-membered ring in an otherwise hexagonal graphene sheet takes continuous values in the limit of infinitely large rings. It was found that the LDOS at the screw dislocation center could be closely approximated by those of an 18-membered ring and that they quickly approached those of perfect graphite with increasing distance from the defect, as for small rings. A model of the screw dislocation is shown in Fig. 17. This connection has renewed relevance due to the recent discoveries of cones with open tip topologies (Terrones et al. [26]) and naturally occurring graphite cones (Jaszczak et al. [2]). As discussed in Section 2.3, the measured apex angle distribution for the latter supports the cone-helix growth conjecture of Double and Hellawell [86], which implies a central screw dislocation. With regard to LDOS, both these

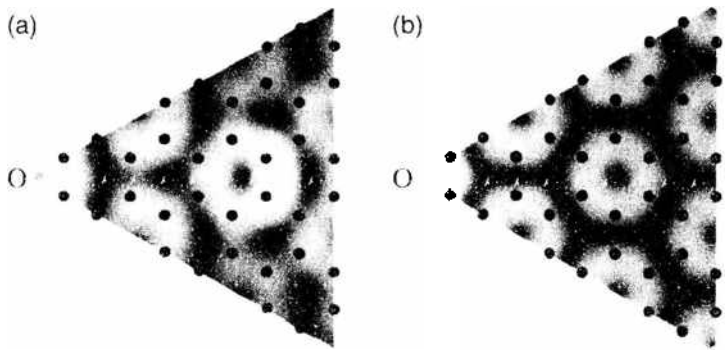


Figure 16. Comparison of simulated STM images of the nanocone around the apex (O) based on DFT (a) and EMT (b). Reprinted with permission from [173]. K. Kobayashi, *Phys. Rev. B* 61, 8496 (2000). © 2000, American Physical Society.

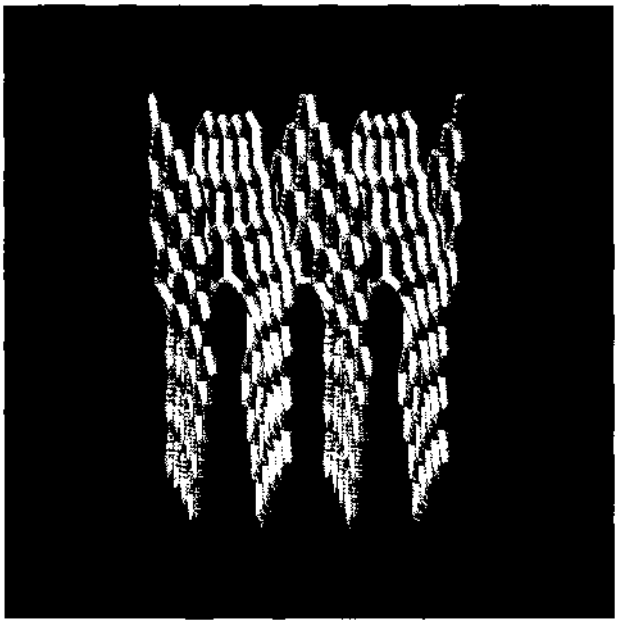


Figure 17. Model of a screw dislocation in graphite. Reprinted with permission from [184], R. Tamura et al., *Phys. Rev. B* 56, 1404 (1997). © 1997, American Physical Society.

classes of cones may therefore, within the TB scheme, be treated on the same footing as the nanocones.

3.5.6. Effect of Substitutional Atoms

Very recently, the effects of introducing boron or nitrogen at the tips of carbon nancones with one or two pentagons have been investigated by Azevedo [196]. Previously, there have been done similar theoretical studies on silicon-doped nanotubes [197] and boron-nitride fullerenes [19]. Figure 18 shows the locations of the substitutional atoms at the cone tips. The open ends where terminated by hydrogens.

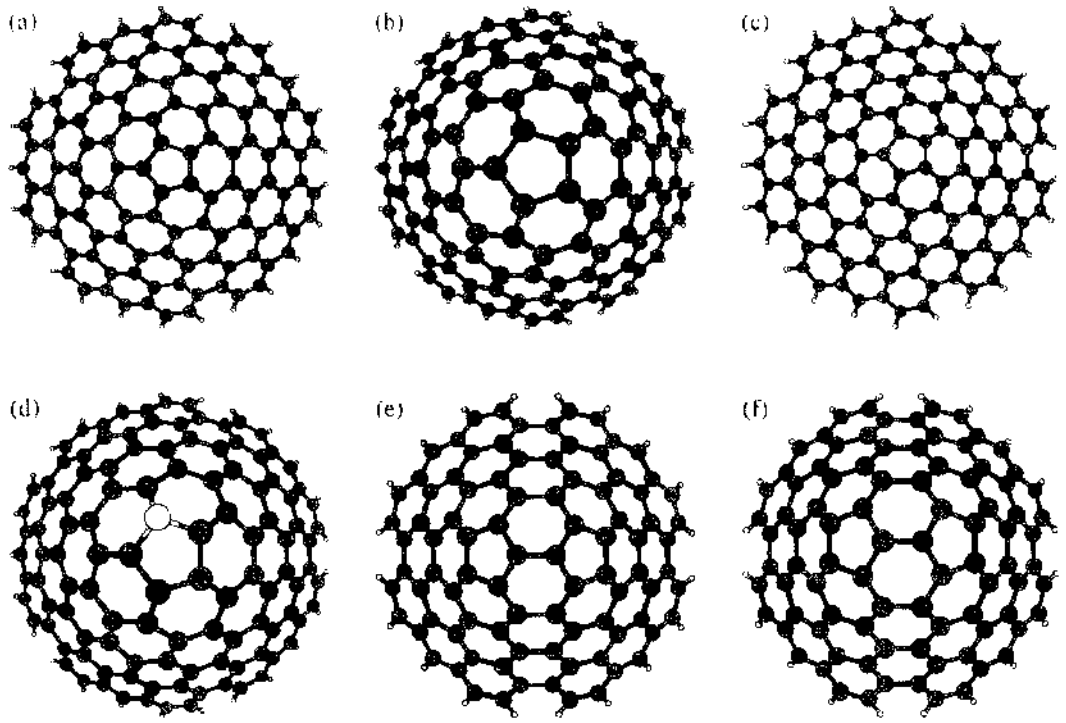


Figure 18. Nanocones with one (a–c) and two (d–f) pentagons and substitutional atoms (b, c, d, f) at various locations around the tip. Reprinted with permission from [196], S. Azevedo, *Phys. Lett. A* 325, 283 (2004). ©2004, Elsevier.

The formation energies of the various cones were calculated as

$$E_{form} = E_{tot} - n_C\mu_C - n_X\mu_X - n_{HC}\mu_{HC} \tag{36}$$

where E_{tot} is the total energy of the cluster, n_C is the number of carbon atoms, n_X is the number of boron or nitrogen atoms, n_{HC} is the number of HC bonds, and the μs are chemical potentials. For the HC bonds, the chemical potentials were extracted from the total energy of a flat graphene sheet;

$$E_f^{sheet} = n_{CC}\mu_{CC}^{bulk} + n_{HC}\mu_{HC} \tag{37}$$

obtained by ascribing a null value to its formation energy. Here, n_{CC} and μ_{CC}^{bulk} are, respectively, the number of CC bonds and the bulk chemical potential.

The energies were determined by quite comprehensive DFT computations, involving the generalized gradient approximation (GGA) [132] for the exchange correlation potential. It was found that the nitrogen substitutions shown in Fig. 18 lowered the formation energy of the cones by reducing the cost of the pentagons. Boron, on the other hand, had the opposite effect. It was further found that both substituents decreased the HOMO-LUMO gap and that nitrogen gave the most enhanced reactivity in this sense.

4. THE NUCLEATION PUZZLE

Several models for the nucleation and subsequent growth of fullerenes and nanotubes under the various environments of synthesis have been proposed [57, 198–215], but complete nucleation theories for curved graphene surfaces are still missing. In this section, we collect the pieces of experimental facts and theoretical predictions that seem relevant for an attempt to formulate a nucleation theory for nanocones.

The method of synthesis invented by Terrones et al. [26] is distinguished by the presence of a heavy catalyst, and its theoretical description requires a nucleation model significantly different from those proposed for cone formation in the hot vapor phase [22] or under pyrolysis of heavy oil [25]. A schematic model of nucleation and growth of cones on the palladium particle is given in [26]. In this section, we will discuss nucleation processes in the two amorphous states [22, 25], where the various nucleation models for fullerenes and tubes developed over the years presumably have larger relevance.

The effective plasma temperature in the experiment of Krishnan et al. [25] were estimated to be at least 2000°C. Because hydrocarbons are known to dehydrogenate at high temperatures [199], the nucleation mechanisms for the cones, although synthesized from heavy oil in this experiment, should have similarities to those determining the distribution of morphologies originating from a pure carbon plasma. The two most well-known models for formation of fullerenes and tubes under these conditions are the "pentagon road" [198, 200, 202] and the "ring-stacking" model [201, 206–208]. The question of nucleation is more pressing than the nature of the subsequent growth. Once a single-walled cone is formed, additional layers, seen in the samples of [25], are likely to grow on its surface in accordance with the growth model for tubes suggested by Iijima, Ajayan, and Ichihashi [204].

The "pentagon road" model assumes the pre-formation of a flat hexagonal sheet, which subsequently folds into itself by incorporating pentagons in order to get rid of the dangling bonds along the edges. The inherent weakness of this model is the origin of the initial sheet. In 2003, a very clarifying series of total energy calculations in terms of DFT were published by Fan et al. [57]. In short, their conclusion was that the bending energies of the various curved graphene surfaces were small compared to the energy of a dangling bond, which suggests that the pentagons, which reduce the number of dangling bonds per atom, are incorporated in a very early stage in the formation of single-walled carbon nanotubes. This is a severe objection to the "pentagon road" model for formation of curved graphene surfaces in general.

To explain the distribution of cone apex angles observed by Krishnan et al., arguments based on the "ring-stacking" model were put forward in [25, 216]. In particular, from simple geometrical stability considerations as those in Section 2.2.1, one should expect

predominance of the largest apex angle, 112.9°, not 60.0° as observed. The ring-stacking approach to this problem was refined in a 2001 paper of Treacy and Killian [217] where the apex angle distribution was explained by differences in the Gibb's free energy

$$\Delta G_{ij} = \Delta H_{ij} - T\Delta S_{ij} \tag{38}$$

where ΔH_{ij} is the enthalpy of formation when a chemical system transforms from state i to state j at temperature T. The entropy change S_{ij} serves in this context primarily as a measure of the number of reaction pathways from state i to j. The proposed mechanism leaned on topologically flexible seeds, primarily, but not necessarily, carbon rings. Large rings were found to favor multiple disclinations (high enthalpies) because of the entropy increase resulting from the calculated large numbers of pathways to the resulting cones. Ring seeds of 18 and 24 carbons favored cones with two and four pentagons, respectively.

The occurrence of only one apex angle, 19°, for the cones synthesized in the hot vapor phase by Ge and Sattler [23] was understood as a result of similar nucleation seeds for the cones and the simultaneously synthesized tubes, namely imperfect or perfect C_{60} hemispheres, respectively. This intuitive picture is in agreement with the above precursor to a complete nucleation theory for cones and tubes.

REFERENCES

1. R. Bacon, *J. Appl. Phys.* 31, 283 (1960).
2. J. A. Jaszczak, G. W. Robinson, S. Dimovski, and Yury Gogotsi, *Carbon* 41, 2085 (2003).
3. S. Iijima, T. Ichihashi, and Y. Ando, *Nature* 356, 776 (1992).
4. S. Iijima, M. Yudasaka, R. Yamada, S. Bandow, K. Suenaga, F. Kokai, and K. Takahashi, *Chem. Phys. Lett.* 309, 165 (1999).
5. S. Berber, Y.-K. Kwon, and D. Tomanek, *Phys. Rev. B* 62, R2291 (2000).
6. R. C. Mani, X. Li, M. K. Sunkara, and K. Rajan, *Nano Lett.* 3, 671 (2003).
7. L. Bourgeois, Y. Bando, W. Q. Han, and T. Sato, *Phys. Rev. B* 61, 7686 (2000).
8. F. Banhart, M. Zwanger, and H.-J. Muhr, *Chem. Phys. Lett.* 231, 98 (1994).
9. F. Banhart, M. Zwanger, and H.-J. Muhr, *Chem. Phys. Lett.* 231, 98 (1994).
10. N. G. Chopra, R. J. Luyken, K. Cherrey, V. H. Crespi, M. Cohen, S. G. Louie, and A. Zettl, *Science* 269, 966 (1995).
11. F. Jensen and H. Toftlund, *Chem. Phys. Lett.* 268, 352 (1997).
12. G. Seifert, P. W. Fowler, D. Mitchell, D. Porezag, and T. Frauenheim, *Chem. Phys. Lett.* 268, 352 (1997).
13. D. Golberg, Y. Bando, K. Kurashima, and T. Sasaki, *Appl. Phys. Lett.* 72, 1966 (1998).
14. W. Han, Y. Bando, K. Kurashima, and T. Sato, *Appl. Phys. Lett.* 73, 3085 (1998).
15. D. Golberg, W. Han, Y. Bando, L. Bourgeois, K. Kurashima, and T. Sato, *J. Appl. Phys.* 86, 2364 (1999).
16. O. Stéphan, Y. Bando, A. Loiseau, F. Willaime, N. Schramenko, T. Tamiya, and T. Sato, *Appl. Phys. A: Mater. Sci. Process.* 67, 107 (1998).
17. D. Golberg, Y. Bando, O. Stéphan, and K. Kurashima, *Appl. Phys. Lett.* 73, 2441 (1998).
18. T. Oku, T. Hirano, M. Kuno, T. Kusunose, K. Niihara, and K. Suganuma, *Mater. Sci. Eng. B* 74, 206 (2000).
19. S. S. Alexandre, H. Chacham, and R. W. Nunes, *Phys. Rev. B* 63, 045402 (2002).
20. A. T. Balaban, D. J. Klein, and X. Liu, *Carbon* 32, 357 (1994).
21. H. Terrones, *J. Math. Chem.* 15, 143 (1994).
22. M. Ge and K. Sattler, *Chem. Phys. Lett.* 220, 192 (1994).
23. K. Sattler, *Carbon* 33, 915 (1995).
24. Kvaerner's patent no. PCT/NO98/00093 for production of micro domain particles by use of a plasma process.
25. A. Krishnan, E. Dujardin, M. M. J. Treacy, J. Hugdahl, S. Lynum, and T. W. Ebbesen, *Nature* 388, 451 (1997).
26. H. Terrones, T. Hayashi, M. Muñoz-Navia, M. Terrones, Y. A. Kim, N. Grobert, R. Kamalakaran, J. Dorantes-Dávila, R. Escudero, M. S. Dresselhaus, and M. Endo, *Chem. Phys. Lett.* 343, 241 (2001).
27. A. C. Dillon, K. M. Jones, T. A. Bekkedahl, C. H. Kiang, D. S. Bethune, and M. J. Heben, *Nature* 386, 377 1997).
28. A. Chambers, C. Park, R. Terry, K. Baker, and N. M. Rodriguez, *J. Phys. Chem. B* 102, 4253 (1998).
29. C. Park, P. E. Anderson, A. Chambers, C. D. Tan, R. Hidalgo, and N. M. Rodriguez, *J. Phys. Chem. B* 103, 10572 (1999).
30. A. C. Dillon and M. J. Heben, *Appl. Phys. A* 72, 133 (2001).
31. S. J. V. Frankland and D. W. Brenner, *Chem. Phys. Lett.* 334, 18 (2001).
32. K. Murata, K. Kaneko, H. Kanoh, D. Kasuya, K. Takahashi, F. Kokai, M. Yudasaka, and S. Iijima, *J. Phys. Chem. B* 106, 11132 (2002).
33. M. Volpe and F. Cleri, *Surf. Sci.* 544, 24 (2003).
34. D. Stojkovic, P. Zhang, P. E. Lammert, and V. H. Crespi, *Phys. Rev. B* 68, 195406 (2003).
35. Yu. S. Nechaev and O. K. Alexeeva, *Int. J. Hydrogen Energy* 28, 1433 (2003)

36. A. T. Skjeltorp and A. J. Macland. U.S. Patent No. 6,290,753 B1. Sept. 18, 2001.
37. J. R. Dias, *Chem. Phys. Lett.* 185, 10 (1991).
38. J. R. Dias. *Chem. Phys. Lett.* 187, 334 (1991).
39. J. R. Dias, *J. Chem. Inf. Comput. Sci.* 32, 2 (1992).
40. J. R. Dias, *J. Chem. Inf. Comput. Sci.* 32, 203 (1992).
41. J. R. Dias, *J. Chem. Inf. Comput. Sci.* 35, 148 (1992).
42. S. J. Cyvin, *J. Molec. Struct.* 262, 219 (1992).
43. S. J. Cyvin, B. N. Cyvin, and J. Brunvoll, *J. Molec. Struct.* 281, 229 (1993).
44. S. J. Cyvin, B. N. Cyvin, J. Brunvoll, Z. Fuji, G. Xiaofeng, and R. Tosic, *J. Molec. Struct.* 285, 179 (1993).
45. J. Brunvoll, B. N. Cyvin, S. J. Cyvin, G. Brinkmann, and J. Bornhoft, *Z. Naturforsch.* 51a, 1073 (1996).
46. J. Brunvoll, B. N. Cyvin, and S. J. Cyvin, *J. Chem. Inf. Comput. Sci.* 36, 91 (1996).
47. G. G. Cash and J. R. Dias, *Z. Naturforsch.* 57a, 650 (2002).
48. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, *Science* 283, 80 (1999).
49. S. Li, C. P. Hill, W. I. Sundquist, and J. T. Finch, *Nature* 407, 409 (2000).
50. U. K. von Schwedler, K. M. Stray, J. E. Garrus, and W. I. Sundquist, *J. Virol.* 77, 5439 (2003).
51. E. Hückel, *Z. Phys.* 60, 423 (1930).
52. D. M. Cvetković, M. Doob, and H. Sachs, "Spectra of Graphs," Academic Press, New York, 1980.
53. R. C. Haddon, *Acc. Chem. Res.* 21, 243 (1988).
54. R. C. Haddon, *Science* 261, 1545 (1993).
55. R. Hoffmann, *J. Chem. Phys.* 39, 1397 (1963).
56. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes," Academic Press, New York, 1995.
57. X. Fan, R. Buczko, A. A. Puretzky, D. D. Geohegan, J. Y. Howe, S. T. Pantelides, and S. J. Pennycook, *Phys. Rev. Lett.* 90, 145501 (2003).
58. O. A. Shenderova, B. L. Lawson, D. Areshkin, and D. W. Brenner, *Nanotechnology* 12, 191 (2001).
59. D. Tománek, W. Zhong, and E. Krastev, *Phys. Rev. B* 48, 15461 (1993).
60. S. Timoshenko and J. N. Goodier, "Theory of Elasticity," McGraw-Hill, New York, 1951.
61. B. L. Zhang, C. Z. Wang, and K. M. Ho, *Chem. Phys. Lett.* 193, 225 (1992).
62. K. Raghavachari and C. M. Rohlfing, *J. Phys. Chem.* 95, 5768 (1991).
63. K. Raghavachari, *Chem. Phys. Lett.* 190, 397 (1992).
64. D. Tománek, G. Overney, H. Miyazaki, S. D. Mahanti, and H.-J. Güntherodt, *Phys. Rev. Lett.* 63, 876 (1989); 63, 1896(E) (1989).
65. J.-C. Charlier and J.-P. Michenaud, *Phys. Rev. Lett.* 70, 1858 (1993).
66. D. Langbein, "Theory of Van der Waals Attraction," Springer Verlag, Berlin, 1974.
67. J. N. Israelachvili and D. Tabor, *Proc. Roy. Soc.* A331, 19 (1972).
68. G. B. Adams, O. F. Sankey, J. B. Page, M. O'Keeffe, and D. A. Drabold, *Science* 256, 1792 (1992).
69. N. Park, K. Lee, S. Han, J. Yu, and J. Ihm, *Phys. Rev. B* 65, 121405(R) (2002).
70. H. W. Kroto, *Nature* 329, 529 (1987).
71. R. Taylor, *Tetrahedron Lett.* 32, 3731 (1991).
72. R. Taylor and R. M. Walton, *Nature* 363, 685 (1993).
73. W. H. Mills and I. G. Nixon, *J. Chem. Soc.* 2510 (1930).
74. R. Taylor "Electrophilic Aromatic Substitution," Wiley, Chichester, 1989.
75. R. C. Haddon and K. Raghavachari, in "Buckminsterfullerenes" (W. E. Billups and M. A. Ciufolini, Eds.), VCH Press, New York, 1993.
76. K. Raghavachari and C. M. Rohlfing, *J. Phys. Chem.* 96, 2463 (1993).
77. D. J. Klein, T. G. Schmaltz, T. G. Hite, and W. A. Seitz, *J. Am. Chem. Soc.* 108, 1301 (1986).
78. T. G. Schmaltz, W. A. Seitz, D. J. Klein and T. G. Hite, *J. Am. Chem. Soc.* 110, 1113 (1988).
79. P. W. Fowler and J. I. Steer, *J. Chem. Soc., Chem. Commun.* 1403 (1987).
80. D. E. Manolopoulus, D. R. Woodall, and P. W. Fowler, *J. Chem. Soc. Faraday Trans.* 88, 2427 (1992).
81. P. W. Fowler, *J. Chem. Soc. Perkin Trans.* 2, 145 (1992).
82. P. W. Fowler, *J. Chem. Soc. Faraday Trans.* 86, 2073 (1990).
83. H. Heiberg-Andersen and A. T. Skjeltorp, *J. Math. Chem.* 38, 589 (2005).
84. D. D. Double and A. Hellawell, *Acta Metallurgica* 17, 1071 (1969).
85. D. D. Double and A. Hellawell, *Acta Metallurgica* 19, 1303 (1971).
86. D. D. Double and A. Hellawell, *Acta Metallurgica* 22, 481 (1974).
87. M. G. Day, *B.C.I.R.A. Jl.* 18, 17 (1970).
88. B. Lux, M. Grages and D. Sapey, *Prakt. Metallogr.* 5, 587 (1968).
89. B. Lux, W. Bollmann, and M. Grages, *Prakt. Metallogr.* 6, 530 (1969).
90. H. Nieswaag and S. J. Zuithoff, 36th International Foundry Congress, Belgrade, 16 (1969).
91. M. B. Haanstra, W. F. Knippenberg, and G. Verspui, "Proceeding 5th European Congress on Electron Microscopy," p. 214, Institute of Physics, Manchester, 1972.
92. M. B. Haanstra, W. F. Knippenberg, and G. Verspui, *J. Crystal Growth* 16, 71 (1972).
93. S. Amelinckx, W. Luyten, T. Krekels, G. van Tendeloo, and J. van Landuyt, *J. Crystal Growth* 121, 543 (1992).
94. B. A. Van der Pluijm, and K. A. Carlson, *Geology* 17, 161 (1989).
95. P. Ball, *Nanozone News, Nature* (2003).
96. S. Dimovski, J. A. Libera, and Gogotsi, *Mater. Res. Soc., Symp. Proc.* 706, Z6.27.1 (2002).
97. J. Libera and Y. Gogotsi, *Carbon* 39, 1307 (2001).

98. V. I. Merkulov, M. A. Guillorn, D. H. Lowndes, and M. L. Simpson, *Appl. Phys. Lett.* 79, 1178 (2001).
99. V. I. Merkulov, A. I. Melechko, M. A. Guillorn, D. H. Lowndes, and M. L. Simpson, *Chem. Phys. Lett.* 350, 381 (2001).
100. L. R. Baylor, V. I. Merkulov, D. E. Ellis, M. A. Guillorn, D. H. Lowndes, A. V. Melechko, M. L. Simpson, and J. H. Whealton, *J. Appl. Phys.* 91, 4602 (2002).
101. G. Zhang, X. Jiang, and E. Wang, *Science* 300, 472 (2003).
102. J. C. Tully, in "Modern Methods for Multidimensional Dynamics Computations in Chemistry" (D. L. Thompson, Ed.). World Scientific, Singapore, 1998.
103. J. C. Tully, in "Classical and Quantum Dynamics in Condensed Phase Simulations" (B. J. Berne, G. Ciccotti, and D. F. Coker, Eds.). World Scientific, Singapore, 1998.
104. D. Marx and J. Hutter, "Ab Initio Molecular Dynamics: Theory and Implementation" (J. Grotendorst, Ed.). Vol. 1. John von Neumann Institute for Computing, NIC Series, 2000.
105. D. W. Brenner, *Phys. Rev. B* 42, 9458 (1990).
106. J. Tersoff, *Phys. Rev. Lett.* 56, 632 (1986).
107. G. C. Abell, *Phys. Rev. B* 31, 6184 (1985).
108. J. Ferrante, J. R. Smith, and J. H. Rose, *Phys. Rev. Lett.* 50, 1385 (1983).
109. J. H. Rose, J. R. Smith, and J. Ferrante, *Phys. Rev. B* 28, 1835 (1983).
110. E. Burgos, E. Halac, and H. Bonadeo, *Chem. Phys. Lett.* 298, 273 (1998).
111. D. Frenkel and B. Smit, "Understanding Molecular Simulation." Academic Press, New York, 1996.
112. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* 220, 671 (1983).
113. R. G. Parr and W. Yang, "Density-Functional Theory of Atoms and Molecules." Oxford University Press, Oxford, 1989.
114. R. M. Dreizler and E. K. U. Gross, "Density-Functional Theory." Springer, Berlin, 1990.
115. F. Jensen, "Introduction to Computational Chemistry." Wiley, New York, 1999.
116. ADF: Developed and distributed by Scientific Computing & Modelling NV, Vrije Universiteit, Theoretical Chemistry, Amsterdam.
117. *ADF User's Guide*. SCM/Vrije Universiteit, Theoretical Chemistry, Amsterdam.
118. CPMD: Developed by J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, Max-Planck Institute für Festköperforschung and IBM Zurich Research Laboratory.
119. Gaussian: See http://www.gaussian.com.
120. NWChem: Developed and distributed by Pacific Northwest National Laboratory.
121. CASTEP: Molecular Simulations Inc.
122. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* 64, 1045 (1992).
123. G. Kresse and J. Furthmüller, *Phys. Rev. B* 54, 11169 (1996).
124. B. L. Hammond, W. A. Lester, and P. J. Reynolds, "Monte Carlo Methods in Ab Initio Quantum Chemistry." World Scientific, Singapore, 1994.
125. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, 864 (1964).
126. W. Kohn and L. J. Sham, *Phys. Rev.* 140, 1133 (1965).
127. J. P. Perdew, *Phys. Rev. B*, 33, 8822 (1986); Erratum: *Phys. Rev. B* 34, 7406 (1986).
128. R. O. Jones and O. Gunnarson, *Rev. Mod. Phys.* 61, 689 (1989).
129. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolais, *Phys. Rev. B* 46, 6671 (1992).; Erratum: *Phys. Rev. B* 48, 4978 (1993).
130. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
131. C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* 37, 785 (1988).
132. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996); Erratum: *Phys. Rev. Lett.* 78, 1396 (1997).
133. R. M. Nieminen, M. J. Puska, and M. J. Manninen, "Many-Atom Interactions in Solids." Springer, Berlin, 1989.
134. J. A. Majewski and P. Vogl, "The Structure of Binary Compounds." North-Holland, Amsterdam, 1989.
135. L. Goodwin, L. Skinner, and A. J. Pettifor, *Europhys. Lett.* 9, 701 (1989).
136. R. Virkkunen, K. Laasonen, and R. M. Nieminen, *J. Condens. Matter* 3, 7455 (1991).
137. C. Z. Wang, C. T. Chan, and K. M. Ho, *Phys. Rev. B* 45, 12227 (1992).
138. C. Molteni, L. Colombo, and L. Miglio, *Europhys. Lett.* 24, 659 (1993).
139. H. Hellmann, *Z. Phys.* 85, 180 (1933).
140. R. P. Feynman, *Phys. Rev.* 56, 340 (1939).
141. G. Galli and M. Parrinello, *Phys. Rev. Lett.* 69, 3547 (1992).
142. F. Mauri, G. Galli, and R. Car *Phys. Rev. B* 47, 10891 (1993).
143. P. Ordejon, D. Drabold, M. Grunbach, and R. Martin, *Phys. Rev. B* 48, 14646 (1993).
144. W. Kohn, *Chem. Phys. Lett.* 208, 167 (1993).
145. X.-P. Li, W. Nunes, and D. Vanderbilt *Phys. Rev. B* 47, 10891 (1993).
146. M. S. Daw, *Phys. Rev. B* 47, 10895 (1993).
147. A. Alavi and D. Frenkel, *J. Chem. Phys.* 97, 9249 (1992).
148. W. Yang, *Phys. Rev. Lett.* 66, 1438 (1991).
149. M. Aoki, *Phys. Rev. Lett.* 71, 1438 (1993).
150. S. Goedecker and L. Colombo, *Phys. Rev. Lett.* 73, 122 (1994).
151. M. S. Tang, C. Z. Wang, C. T. Chan, and K. M. Ho, *Phys. Rev. B* 53, 979 (1996).
152. J. C. Slater and G. F. Koster, *Phys. Rev.* 94, 1498 (1954).

153. B. N. Davidson and W. E. Pickett, *Phys. Rev. B* 49, 11253 (1994).
154. R. Peierls, *Z. Physik* 80, 763 (1933).
155. R. Peierls, *Z. Physik* 81, 186 (1933).
156. G. H. Wannier, *Phys. Rev.* 52, 191 (1937).
157. J. C. Slater, *Phys. Rev.* 76, 1992 (1949).
158. J. M. Luttinger, *Phys. Rev.* 84, 814 (1951).
159. E. N. Adams II, *Phys. Rev.* 85, 41 (1952).
160. P. Feuer, *Phys. Rev.* 88, 92 (1952).
161. E. N. Adams II, *J. Chem. Phys.* 21, 2013 (1953).
162. J. M. Luttinger and W. Kohn, *Phys. Rev.* 97, 869 (1955).
163. D. P. DiVincenzo and E. J. Mele, *Phys. Rev. B* 29, 1685 (1984).
164. E. A. Kochetov and V. A. Osipov, *J. Phys. A* 32, 1961 (1999).
165. P. E. Lammert and V. H. Crespi, *Phys. Rev. B* 61, 7308 (2000).
166. P. E. Lammert and V. H. Crespi, *Phys. Rev. Lett.* 85, 5190 (2000).
167. V. A. Osipov and E. A. Kochetov, *JETP Lett.* 73, 562 (2001).
168. V. A. Osipov, E. A. Kochetov, and M. Pudlak, *JETP* 96, 140 (2003).
169. R. Pincak and V. A. Osipov, *Phys. Lett. A* 314, 315 (2003).
170. P. E. Lammert and V. H. Crespi, *Phys. Rev. B* 69, 0354606 (2004).
171. G. S. Painter and D. E. Ellis, *Phys. Rev. B* 1, 4747 (1970).
172. H. Matsumura and T. Ando, *J. Phys. Soc. Jpn.* 67, 3542 (1998).
173. K. Kobayashi, *Phys. Rev. B* 61, 8496 (2000).
174. M. Nakahara, "Geometry, Topology and Physics," A. Hilger, Bristol, UK, 1990.
175. J. Han and R. Jaffe, *J. Chem. Phys.* 108, 2817 (1998).
176. L. Salem, "The Molecular Orbital Theory of Conjugated Systems," W. A. Benjamin, Inc., 1966.
177. J.-C. Charlier and G.-M. Rignanese, *Phys. Rev. Lett.* 86, 5970 (2001).
178. S. Matsumoto and Y. Matsui, *J. Mater. Sci.* 18, 1785 (1983).
179. D. Shechtman, J. L. Hutchinson, J. L. Robins, E. N. Farabaugh, and A. Feldman, *J. Mater. Res.* 8, 473 (1993).
180. O. A. Shenderova and D. W. Brenner, *Phys. Rev. B* 60, 7053 (1999).
181. W. A. de Heer, A. Châtelain, and D. Ugarte, *Science* 270, 1179 (1995).
182. R. Tamura and M. Tsukada, *Phys. Rev. B* 49, 7697 (1994).
183. R. Tamura and M. Tsukada, *Phys. Rev. B* 52, 6015 (1995).
184. R. Tamura, K. Akagi, M. Tsukada, S. Itoh, and S. Ihara, *Phys. Rev. B* 56, 1404 (1997).
185. S. Ihara, S. Itoh, K. Akagi, R. Tamura, and M. Tsukada, *Phys. Rev. B* 54, 14713 (1996).
186. A. De Vita, J. C. Charlier, X. Blase, and R. Car, *Appl. Phys. A* 68, 283 (1999).
187. R. Haydock, V. Heine, and M. J. Kelly, *J. Phys. C* 5, 2845 (1972).
188. A. G. Aronov and Yu. V. Sharvin, *Rev. Mod. Phys.* 59, 755 (1987).
189. R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, *Phys. Rev. Lett.* 54, 2696 (1985).
190. H. Ajiki and T. Ando, *J. Phys. Soc. Jpn.* 62, 2470 (1993); Erratum, *J. Phys. Soc. Jpn.* 62, 4267 (1993).
191. B. An, S. Fukuyama, K. Yokogawa, M. Yoshimura, M. Egashira, Y. Korai, and I. Mochida, *Appl. Phys Lett.* 78, 3696 (2001).
192. V. Meunier and Ph. Lambin, *Phys. Rev. Lett.* 81, 5588 (1998).
193. J. Tersoff and D. R. Hamann, *Phys. Rev. B* 31, 805 (1985).
194. H. A. Mizes and J. S. Foster, *Science* 244, 559 (1989).
195. T. Yamaguchi, *Surf. Sci.* 493, 597 (2001).
196. S. Azevedo, *Phys. Lett. A* 325, 283 (2004).
197. R. J. Baierle, S. Fagan, R. Mota, A. da Silva, and A. Fazzio, *Phys. Rev. B* 64, 085413 (2001).
198. Q. L. Zhang, S. C. O'Brien, J. R. Heath, Y. Liu, R. F. Curl, H. W. Kroto, and R. E. Smalley, *J. Phys. Chem.* 90, 525 (1986).
199. H. W. Kroto and K. McKay, *Nature* 331, 328 (1988).
200. R. E. Smalley, *Acc. Chem. Res.* 25, 98 (1992).
201. T. Wakabayashi and Y. Ashiba, *Chem. Phys. Lett.* 190, 456 (1992).
202. D. H. Robertson, D. W. Brenner, and C. T. White, *J. Phys. Chem.* 96, 6133 (1992).
203. M. Endo and H. W. Kroto, *J. Phys. Chem.* 96, 6941 (1992).
204. S. Iijima, P. M. Ajayan, and T. Ichihashi, *Phys. Rev. Lett.* 69, 3100 (1992).
205. Y. Saito, T. Yoshikawa, M. Inagaki, M. Tomita, and T. Hayashi, *Chem. Phys. Lett.* 204, 277 (1993).
206. S. W. McElvany, M. M. Ross, N. S. Goroff, and F. Diederich, *Science* 259, 1594 (1993).
207. G. von Helden, N. G. Gotts, and M. T. Bowers, *Nature* 363, 60 (1993).
208. J. Hunter, J. Fye, and M. F. Jarrold, *J. Phys. Chem.* 97, 3460 (1993).
209. P. J. F. Harris, S. C. Tsang, J. B. Claridge, and M. L. H. Green, *J. Chem. Soc. Faraday Trans.* 90, 2799 (1994).
210. F. G. Gamaly and T. W. Ebbesen, *Phys. Rev. B* 52, 2803 (1995).
211. A. Maiti, C. Brabec, C. Roland, and J. Bernholc *Phys. Rev. B* 52, 14850 (1995).
212. J.-C. Charlier, A. DeVita, X. Blase, and R. Car, *Science* 275, 646 (1997).
213. H. Kanzow and A. Ding, *Phys. Rev. B* 60, 11180 (1997).
214. P. Zhang and V. Crespi, *Phys. Rev. Lett.* 83, 1791 (1999).
215. H. Kanzow, C. Lenski, and A. Ding, *Phys. Rev. B* 63, 125402 (2001).
216. T. W. Ebbesen, *Acc. Chem. Res.* 31, 558 (1998).
217. M. M. J. Treacy and J. Killian, *Mater. Res. Soc. Symp. Proc.* 675, 1 (2001).

CHAPTER 11

Simulation and Optimization of Composite Doped Metamaterials

Christian Hafner, Jasmin Smajic, Daniel Erni

Laboratory for Electromagnetic Fields and Microwave Electronics, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland

CONTENTS

1. Introduction 538
2. Metamaterials 540
   2.1. Frequency Selective Surfaces: Gratings 540
   2.2. Artificial Anisotropic Media 544
   2.3. Artificial Chiral Media 546
   2.4. Negative Index Materials: Left-Handed Media 548
3. Photonic Crystals 553
   3.1. Perfect Photonic Crystals 555
   3.2. Waveguides in Photonic Crystals 560
   3.3. Waveguide Discontinuities in Photonic Crystals 564
4. Simulation of Composite Doped Metamaterials 577
   4.1. Selection of Numerical Methods 578
   4.2. Band Diagram Computation 582
   4.3. Guided Waves 585
   4.4. Waveguide Discontinuities 590
5. Optimization of Composite Doped Metamaterials 595
   5.1. Probabilistic Optimizers 599
   5.2. Deterministic Optimizers 603
   5.3. Sensitivity Analysis 604
   5.4. Towards Phenomenological Two-Dimensional Modeling of Planar Photonic Crystal Devices 605
6. Conclusions 611
   References 612

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (537–613)

1. INTRODUCTION

Currently, there is much research on artificial media that do not exist in nature. Such media typically have a structure of the order the wavelength or smaller than the wavelength they are designed for. Because the optical wavelength is in the nanometer regime, nanotechnology is required for fabricating artificial media. Although nanotechnology is a rather new branch of science, artificial media were already fabricated a long time ago.

Obviously, no nanotechnology is required for the fabrication of artificial media in the radio frequency (RF) regime, where the wavelengths are in the millimeter regime or even longer. Thus, it is not surprising that artificial RF media were already invented a short time after the invention of RF signal transmission by RF pioneers. In 1919, Marconi [1] had the first patent on a structure that would now be considered as a frequency selective surface (FSS) [2]. A FSS is a periodic structure on a dielectric plate that only transmits a part of the frequency spectrum through the plate (i.e., it can be considered as a filter for electromagnetic waves). Although the fabrication of FSS devices was possible at Marconi's time, these structures gained not much interest until printed circuit technology made it really easy to fabricate such structures and computer codes provided useful means for the analysis and design. Therefore, FSS devices were intensively studied after 1960.

Despite the much smaller size of artificial media structures for optics, such structures were fabricated and used even a long time before Marconi. Already in 1785, David Rittenhouse reported the fabrication of a diffraction grating consisting of 53 apertures [3, 4]. Unaware of this work, Fraunhofer began in 1821 to use gratings for spectroscopy, namely for the analysis of the spectrum of solar light. After this work, diffraction gratings were continuously improved.

In radar technology, chiral media were considered to be very promising for building new absorbers and reflectors. It is interesting to note that such structures were studied by Bose at the early time of radio engineering in 1898 [5]. Because the chiral effect of natural materials was too weak, artificial chiral media were fabricated by inserting small helical antenna into dielectric emulsions [6]. This could be considered as an extension of the FSS technology because this artificial material was no longer planar. It became obvious that artificial materials could be designed in such a way that material properties that do not exist in nature may be obtained.

In 1967, Veselago [7] proposed that artificial materials might have both negative permittivity and negative permeability and therefore can be described by a negative refraction index. Several terms were used for such materials, namely "negative index materials" and "left-handed media." Already simple theoretical considerations of negative index materials led to attractive effects, namely lenses with a higher resolution that traditional optical lenses built of natural dielectrics [8].

The exaggerations of the artificial inelligence (AI) community finally discredited the term artificial. This is probably the reason why the most natural term *artificial material* was replaced by *metamaterial*. In the following, we consider metamaterials as artificially fabricated structures consisting of many similar substructures that are of the order the wavelength or smaller and establish the desired macroscopic material properties. These macroscopic material properties may be similar to the properties of natural materials (anisotropic, chiral, biisotropic, etc.) but much more pronounced, or they may be new in the sense that no natural materials with similar properties are known—which does not mean that such natural materials do not exist. Note that the term *metamaterial* is sometimes defined in a more restrictive way. Namely, the size of the substructures is sometimes assumed to be much smaller than the wavelength. This considerably simplifies the numerical and analytical treatment. At the same time, the strength of the electromagnetic effects obtained under this restriction becomes rather weak, whereas strong effects are obtained as soon as one is in the resonant regime (i.e., when the structure size is in the order of the wavelength).

When dealing with natural materials, one often focuses on isotropic media because their mathematical description is relatively simple. When one intends to fabricate isotropic metamaterials, one must compose them of isotropic natural materials. Furthermore, one must make sure that the cells of the metamaterial are randomly oriented. For example, when small

helical antennas are used for obtaining a chiral metamaterial, these antennas should be randomly oriented. For the fabrication process, this is more difficult than systematic orientation. For this reason, metamaterials are often anisotropic and exhibit some periodic symmetry like natural crystals. This makes the mathematical description of metamaterials more demanding, but it also produces many interesting effects. In 1987, Yablonovitch [9] proposed to fabricate photonic crystals (PhCs) [10] as photonic counterparts of electronic semiconductor crystals. The main feature that makes semiconductors attractive is the existence of electronic band gaps (i.e., energy bands where no electrons can be). Similarly, a photonic crystal is a periodic metamaterial that exhibits photonic band gaps, that is, an energy band or a frequency band where no photon (i.e., no electromagnetic wave) can propagate through the PhC. Because semiconductors are the most important material for electronics, the PhC concept is extremely promising and has gained much attention. We therefore will consider PhCs in more detail in the following.

Although interesting objects may consist of a single material, most of the attractive devices are composed of at least two different materials. In fact, all metamaterials are composed of at least two different natural materials. The design of new devices is therefore mainly a composition of different materials. When working with metamaterials, one can not only compose natural materials for obtaining a metamaterial, but one can also compose different metamaterials in order to obtain devices with higher functionality (i.e., *composite* metamaterials are attractive from the engineering point of view).

In semiconductor technology, most of the functionality is obtained from composing semiconductors with differently *doped* areas. Doping means that impurities are introduced in the semiconductor crystals. This may heavily affect the band structure, and it causes interesting effects that are essential for all semiconductor devices. Impurities are atoms that do not belong to the pure crystal. They are more or less randomly distributed because there currently is no way to place a doping atom at a precise location. Therefore, doped semiconductors must consist of rather many atoms. Because photonic crystal cells consist of cells containing many atoms (typically of the order 100 nm), doping of PhCs seems to be not very attractive because one expects doped PhCs to become big. Fortunately, the rather big PhC cells offer the opportunity to dope precisely a perfect PhC in various ways. Here, one is not restricted to a small set of atoms that may be used for doping. Instead of inserting doping materials into a PhC, one can also modify the geometry of a cell to be doped. For example, when a perfect PhC consists of a set of dielectric spheres located on a regular lattice, one can not only replace a sphere by another sphere made of a different material, but one can also omit it entirely, one can replace it by a smaller one, by a dislocated one, by a nonspherical object, or even by a composite object such as a dielectric ellipsoid that is coated with some metal.

From this it becomes evident that composite doped metamaterials (CDMs) provide an extremely high degree of freedom for the design of new photonic structures and devices. Although this is highly desirable and promising, it also causes strong difficulties for the design process. Currently, no design rules are known for the design of CDM devices with a desired functionality. It is not even known what functionality may be achieved in the near future. For example, it is not known if any useful, competitive device can be obtained from negative index materials. When PhCs are considered, one has a long list of devices that might become feasible in the near future: waveguides that permit sharp bending without radiation loss and with almost zero reflection over a wide frequency range, couplers, resonators with very high Q, filters, splitters, diplexers, multiplexers, modulators, sensors, and probably also active devices such as amplifiers and switches. It is therefore imaginable that PhCs might be used for fabricating highly complex optical systems such as entire computers including sensors for data input, signal processing, storage devices, and optical displays. Although this is imaginable, we do not yet know whether a simple optical filter with predefined characteristic can be fabricated with a reasonable effort and structure size. We know that PhC filters can be obtained, but we have no useful design rules (as for electronic filters) for the filter synthesis. Currently, the only way is to analyze promising structures using numerical codes and to try optimizing such structures numerically. For this reason, numerical simulation and optimization are crucial for the development of PhC and more general CDM structures.

Despite the rapid growth of computer power, current computers do not allow one to optimize efficiently and accurately even relatively simple CDM structures because of their high complexity, big size (compared with the wavelength), and often "counter-intuitive or unsystematic" behavior that will be demonstrated in this chapter. In order to keep the computational effort small enough, we will focus on 2D models (i.e., on cylindrical geometry). Such models are certainly less realistic than 3D models but they often exhibit the fundamental effects and can be analyzed in reasonable time, which allows one not only to analyze structures obtained from intuition but also to use optimizers for finding much better structures.

2. METAMATERIALS

2.1. Frequency Selective Surfaces: Gratings

Printed circuit boards (PCBs) consist of metal structures on a dielectric substrate as shown in Fig. 1. The first patent on such structures was submitted by A. P. Hanson in 1903. PCBs are widely used in RF engineering and can easily be fabricated. Note that a single metallic layer is shown in Fig. 1 for reasons of simplicity, but already Hanson's board had two layers of wires, and multilayer boards are very common. When a printed circuit board is illuminated by an electromagnetic wave, the wires act as antennas and cause a complicated response. Depending on the size and arrangement of these antennas, one can obtain a strong frequency dependence of the total power that is reflected at the board or transmitted through the board, that is, the structure acts as a filter for the incoming wave and is therefore called frequency selective surface (FSS) [2]. In general, the response of the FSS depends not only on the frequency but also on the angle of incidence and on the polarization of the incoming wave. One now can try to synthesize a FSS with special properties, for example, one that acts like a bandpass filter and only transmits power within a specified frequency band. Because one wants to obtain one and the same property along the entire surface, one usually designs a FSS as a periodic array of antennas as shown in Fig. 2. The periodicity considerably simplifies the design and the numerical analysis of such structures.

We already have mentioned that optical gratings were fabricated before FSSs were invented. From the theoretical point of view, there is no essential difference between a FSS and a diffraction grating [4]: Both structures exhibit periodic symmetry in one or two directions. For reasons of simplicity, we only consider structures that are periodic in one direction as shown in Fig. 3. Assume that a is the period in x direction. This means that the geometry of the entire structure at $x + nd$ is the same as the geometry at x, where n is an integer number. The entire structure is therefore completely described, when a single period, for example, in the interval from $x = 0$ to $x = a$ is defined.

The periodic geometry does not imply a periodic symmetry of the electromagnetic field, especially when the incident wave that illuminates the structure is not periodic. The first step

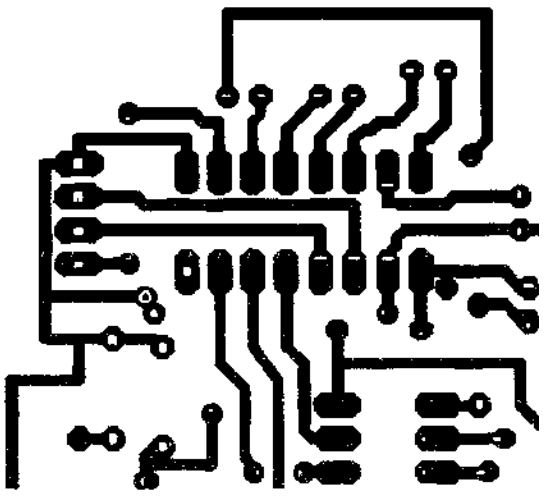


Figure 1. Printed circuit board. Black: copper layer.

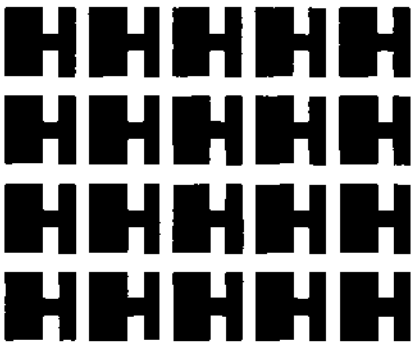


Figure 2. Frequency selective surface (FSS).

is to decompose the incident wave into plane waves because plane waves exhibit periodic symmetry in all directions of the 3D space, namely in x direction. However, depending on the frequency and on the angle of incidence, the geometric period a and the period p of the incident plane wave are usually different.

$$Field(x, y, z, t) = \mathrm{Re}\left(A \times e^{i(k_x^{inc} x + k_y^{inc} y + k_z^{inc} z - \omega t)}\right) \tag{1}$$

describes the electromagnetic field of a plane wave in Cartesian coordinates x, y, z with a complex field amplitude A and a given angular frequency ω. $(k_x^{inc}, k_y^{inc}, k_z^{inc})$ describes the wave vector in Cartesian coordinates. From this, one can see that the incident field is periodic in x direction and that the period p is equal to $2\pi/k_x$. When any component of the incident field has the complex amplitude A at $x = 0$, it has the complex amplitude $A \cdot C_x = A \cdot e^{ik_x^{inc} a}$ at $x = a$. From this, we obtain the periodic condition

$$F(x + d, y, z) = C_x \times F(x, y, z); C_x = e^{ik_x^{inc} a} \tag{2}$$

where F is the complex amplitude of any field component of the incident field as well as of the scattered field and of the total field. Obviously, all waves with x dependence of the form

$$e^{i(k_x^{inc} + 2m\pi/a)x} \tag{3}$$

fulfill the condition (2), when m is any (positive or negative) integer number. Thus, all plane waves with $k_x = k_x^{inc} + 2m\pi/a$ fulfill Eq. (2). These waves are called Rayleigh or Floquet modes of order m. When we assume cylindrical symmetry of the structure along the y direction, the y components of all wave vectors must be equal to k_y^{inc}. Furthermore, the lengths k of the wave vectors are given by the frequency and the material properties. Thus, the z components are obtained from

$$k_z = \pm\sqrt{k^2 - k_x^2 - k_y^2} = \pm\sqrt{k^2 - (k_x^{inc} + 2m\pi/a)^2 - (k_y^{inc})^2} \tag{4}$$

Note that k_z becomes imaginary for loss-free materials with real wavenumber k when the absolute value of the order m is high enough. The corresponding plane waves are then

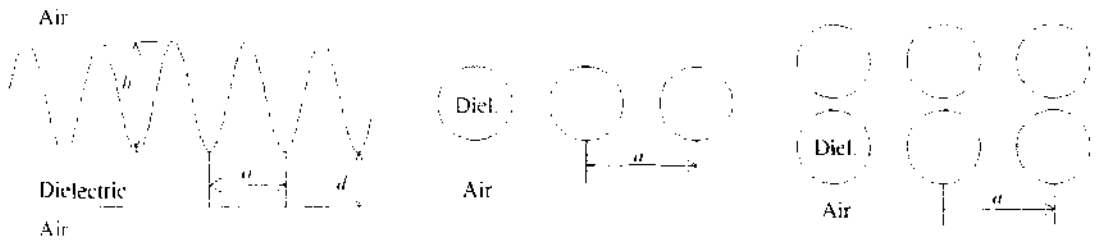


Figure 3. Structures with periodic symmetry in a single direction. Left: sinusoidal grating; middle: grating with a single layer of dielectric rods; right: grating with two layers of dielectric rods.

evanescent, that is, they do not propagate in z direction and are exponentially dampened with the distance from the grating or FSS. Because the wavenumber k is proportional to the frequency, propagating higher order Floquet modes are only obtained for sufficiently high frequencies or for sufficiently long period a. The highest orders of propagating Floquet modes are obtained for $k_y = 0$. For this reason, one often focuses on the special case $k_y^{inc} = 0$. For the even more special case of vertical incidence, one also has $k_x^{inc} = 0$. One then easily can see that first-order modes propagate when

$$k = \omega/c > 2m\pi/a \rightarrow f > f_m = mc/a \rightarrow \lambda < \lambda_m = a/m \tag{5}$$

holds, where c is the speed of light, λ is the wavelength, and f is the frequency. Below the cutoff frequency f_m, the orders $+m$ and $-m$ are evanescent. Note that the Floquet modes of order $m = 0$ are always obtained, even for arbitrarily low frequencies and short periods a. For the special case $a = 0$, the FSS or grating degenerates to a slab. Then all cutoff frequencies for $m \neq 0$ become infinite, and the zero order modes are the well-known reflected and transmitted plane waves.

Although FSSs and gratings can be used at low frequencies (i.e., when the period a is large compared with the wavelength), when only zero-order Floquet modes propagate, the effects of these structures become much more colorful and pronounced when a is of the order the wavelength (i.e., when also higher order modes can propagate). This is illustrated in Figs. 4–7. Figure 4 shows the wavelength dependence of the zero- and first-order reflected and transmitted efficience for a vertically incident plane wave. Obviously, higher order Floquet modes only propagate when condition (5) is met (i.e., for wavelengths below 1 micrometer in the case considered here). Note that the $+1$ and -1 order modes are identical because of the vertical incidence. The transmitted efficiencies depend on the polarization, which indicates anisotropy, but this effect is rather weak at long wavelength. The effects become much more pronounced but also more complicated with decreasing wavelength, especially when higher order Floquet modes propagate. Then, the slab acts not only as a filter (i.e., as a FSS) but also as a mode converter. Here, mode conversion means that a part of the transmitted power propagates in a direction that is different from the direction of the incident wave. We will encounter similar effects for negative index materials but also for planar structures with $a = 0$.

When we work in the area where no mode conversion takes place, we still observe very strong for sufficiently small wavelength. Here, sharp resonance peaks are obtained. Because the resonance peaks are at different locations for s and p polarization, the grating becomes strongly anisotropic.

The power transmission of a grating depends not only on the wavelength, but also on the angle of incidence, as one can see from Fig. 5. Here, a grating consisting of circular dielectric rods is simulated. Because such a grating can be considered as a thin slice of a photonic

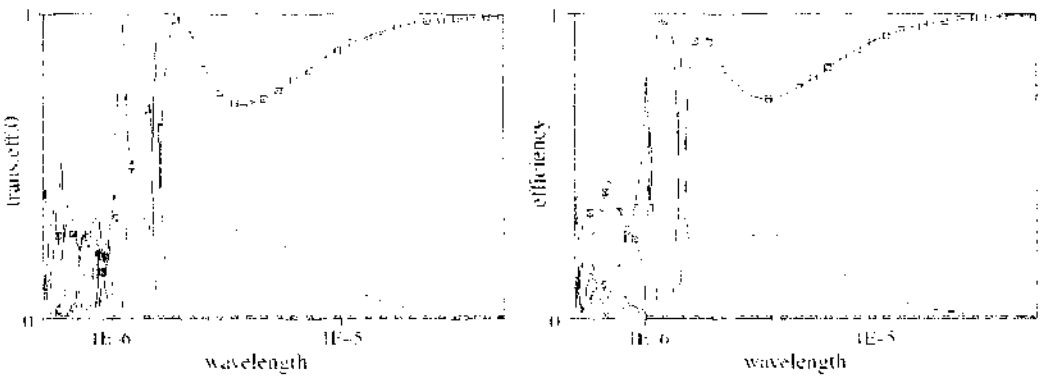


Figure 4. Sinusoidal grating (see Fig. 3): wavelength dependence of the reflected and transmitted efficiencies of orders 0, -1, and $+1$ for a vertically incident plain. Relative permittivity of the grating 4, period $a = 1$ μm, $h = 0.4$ μm, $d = 0.2$ μm. Left: s polarization ($H_z = 0$); right: p polarization ($E_z = 0$).

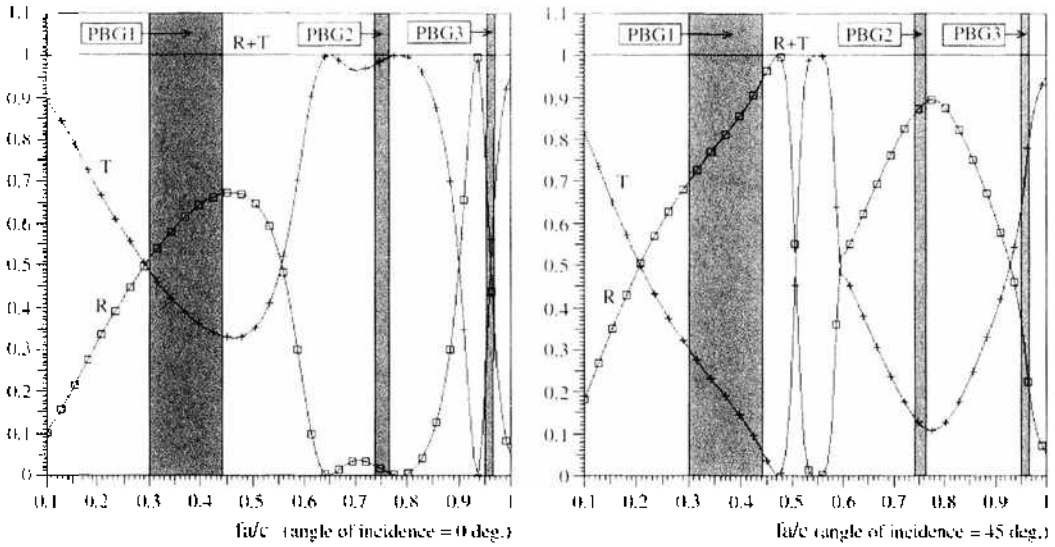


Figure 5. Grating consisting of a single layer of dielectric rods (see Fig. 3), frequency dependence of the reflected and transmitted efficiencies of orders 0 for s polarization (i.e., $H_z = 0$). Left: vertically incident plane wave, right: angle of incidence 45 degrees. The gray areas indicate the photonic band gaps (PBG) of the corresponding perfect photonic crystal.

crystal, we replace the wavelength dependence by the normed frequency dependence that is usually used for photonic crystals:

$$f\frac{a}{c} = \frac{a}{\lambda} \tag{6}$$

where c is the speed of light and a the lattice constant (i.e., the period of the grating). Figure 6 shows that all higher order Floquet modes are evanescent when fa/c is slightly below 1 in the case of vertical incidence. As one can see, there is a nonuniform pattern of the Poynting vector field below the grating. Such a pattern is typical when higher order modes are excited. Because it disappears with the distance from the grating, the higher order

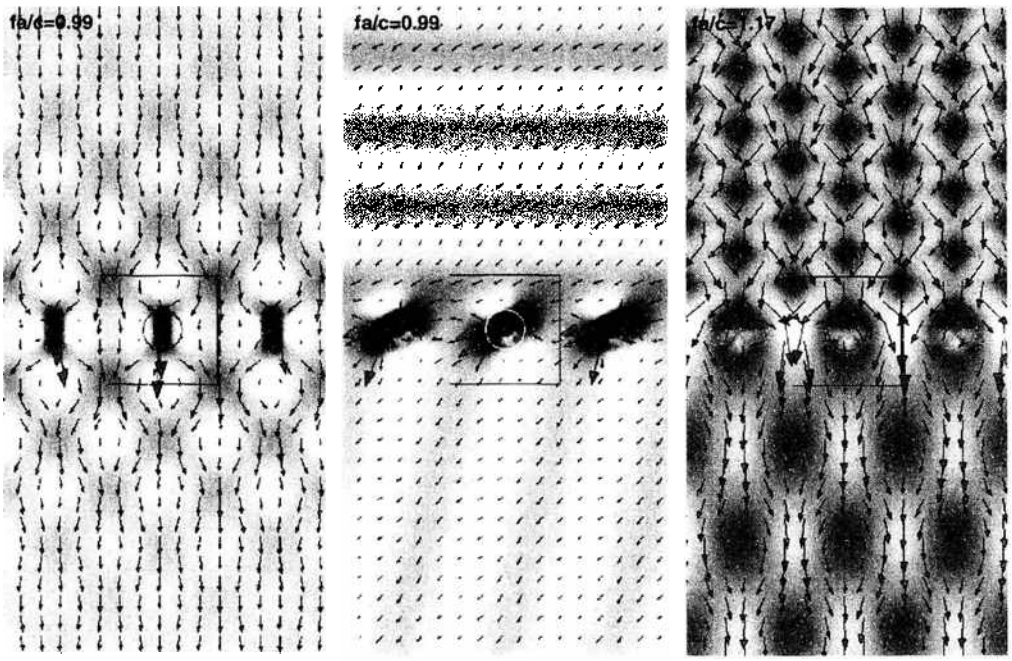


Figure 6. Grating consisting of a single layer of dielectric rods (see Fig. 3), time average of the Poynting vector field, s polarization. Dark areas indicate a strong field. Left: vertically incident plane wave at $fa/c = 0.99$; middle: angle of incidence 45 degrees at $fa/c = 0.99$; right: vertically incident plane wave at $fa/c = 1.17$.

modes do not propagate (i.e. are evanescent). For a nonvertical incidence of the plane wave, a nonuniform pattern is observed. This pattern does not disappear with the distance from the grating (i.e., higher order propagating Floquet modes are excited). The same holds for vertical incidence as soon as fa/c is bigger than 1.

The photonic band gaps (PBG; see section on photonic crystals for more details) of the corresponding perfect photonic crystal are obviously only weakly correlated with the areas of low transmission. When we add a second layer of dielectric rods, the correlation becomes clearer as one can see from Fig. 7. Obviously, one has a low transmission within the first PBG. Note that one can also have low transmission coefficients at frequencies outside the PBG, but these exhibit a strong dependence on the angle of incidence. For a sufficiently high number of layers, the transmission coefficients become almost zero for all angles of incidence.

The Floquet theory outlined above can easily be extended for structures that are periodic in two different directions (e.g., x and y). The Floquet modes are then described by two different integer numbers m, n, and one obtains

$$k_z = \pm\sqrt{k^2 - k_x^2 - k_y^2} = \pm\sqrt{k^2 - (k_x^{inc} + 2m\pi/a_x)^2 - (k_y^{inc} + 2n\pi/a_y)^2} \tag{7}$$

Note that the periods a_x and a_y in x and y directions can be different. We will consider structures with such symmetries in the following section. Note that the formalism becomes slightly more complicated when the directions of the two periods are not perpendicular. For more information, see Ref. [11].

2.2. Artificial Anisotropic Media

Natural anisotropic media are obtained when the atoms or molecules favor a certain direction of the polarization and are arranged in a systematic way, for example, on a regular grid as in crystals. We therefore expect to obtain artificial anisotropic media when we arrange small nonspherical particles, for example, on a cubic grid. From the numerical point of view, the cubic symmetry (i.e., periodic symmetry in three perpendicular directions) is very simple, and it allows us to explicitly model only a single cell as shown in Fig. 8. In order to observe the polarization effects of an anisotropic material block, we want to consider an arrangement of a finite number of cells, illuminated by a plane wave. This completely breaks the cubic symmetry and leads to high numerical costs even when the number of cells is small. We therefore consider an artificial anisotropic slab that still exhibits periodic symmetry in

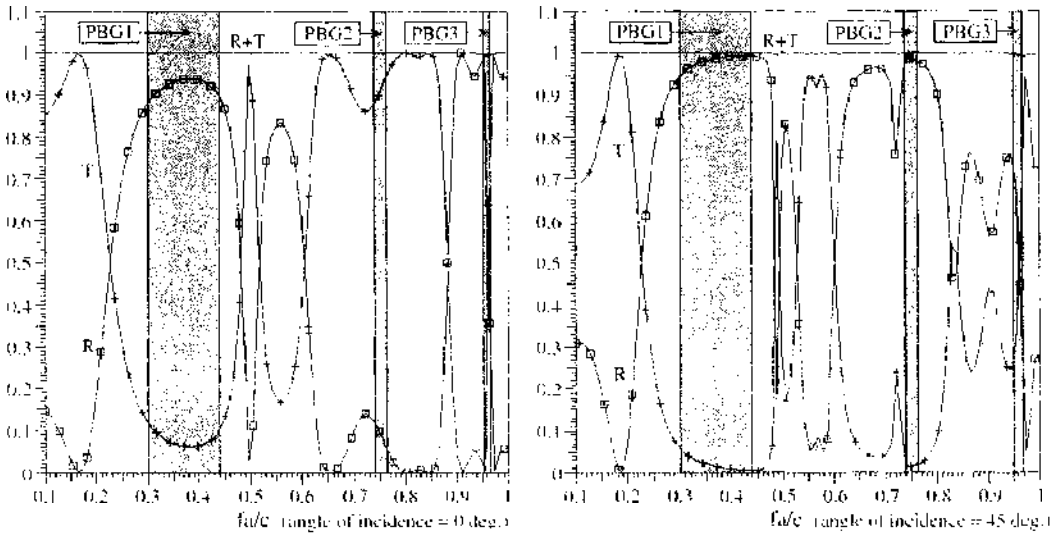


Figure 7. Grating consisting of two layers of dielectric rods (see Fig. 5). Frequency dependence of the reflected and transmitted efficiencies of orders 0 for s polarization (i.e., $H_z = 0$). Left: vertically incident plane wave, right: angle of incidence 45 degrees. The gray areas indicate the photonic band gaps (PBG) of the corresponding perfect photonic crystal.

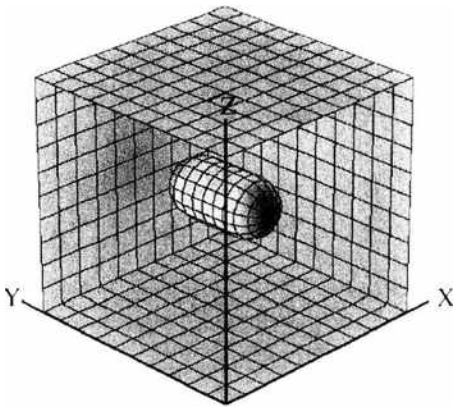


Figure 8. Cell containing a nonspherical object for obtaining an anisotropic metamaterial.

x and y directions and consists of a single layer of cells in z direction as shown in Fig. 8. This structure is essentially the same as a FSS or a grating with two perpendicular periods.

For reasons of simplicity, we consider a plane wave incident perpendicularly on the slab. We expect that the transmission coefficients for the polarization parallel and perpendicular to the axis of the "molecules" are different. As one can see from Fig. 9, the transited zero-order efficiencies exhibit a strong dependence on the wavelength for wavelength slightly above 1 μm. Because the slab is built of an infinite set of bean-shaped "molecules" arranged on a square lattice with lattice constant $a_x = a_y = 1$ μm, no higher order Floquet modes are excited, and the zero-order efficiencies represent the total transmitted power. For wavelength above 1.25 μm, the curves become rather flat and tend toward 1 (i.e., almost 100% of the power is transmitted). Despite this, there is a considerable difference between the transmission for the polarization parallel and perpendicular to the axes of the molecules, which clearly indicates the anisotropy. For wavelength below 1.2 μm, one observes strong changes in the curves due to resonance effects. Note that the relative permittivity of the "molecules" is 11.56, that is, the wavelength inside the molecules is 3.4 times shorter than in free space (i.e., outside the "molecules"). Therefore, resonances are observed also when no higher order Floquet modes propagate. Because the resonances depend on the polarization of the incident wave, one obtains areas where one has almost zero transmission for the p polarization and almost 100% transmission for the s polarization and *vice versa*.

Another artifical anisotroptic slab is obtained when we use metals for fabricating the "molecules" instead of dielectrics. It is well-known that metals loose their conductivity

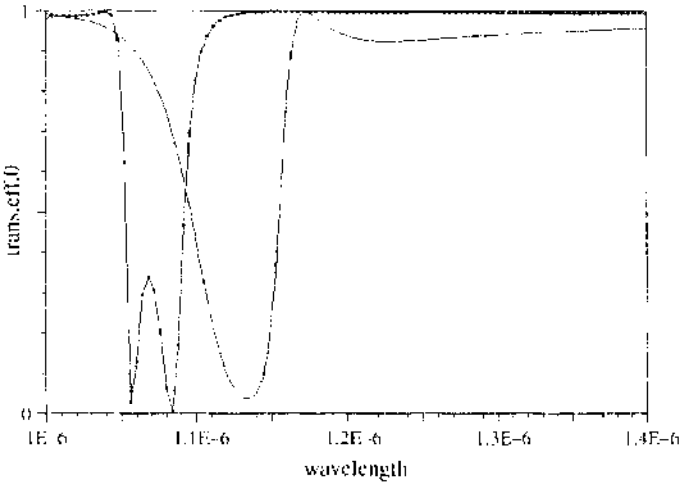


Figure 9. Wavelength dependence of the transmission coefficients of an anisotropic metamaterial slab consisting of a single layer of cells (shown in Fig. 8, for cell size 1 μm in x, y, and z directions) for vertical incidence of a plane wave. Relative permittivity of the dielectric "molecules" is 11.56. Curve with marker: polarization of the incident wave parallel to the axis of the "molecules"; without marker: orthogonal polarization.

at optical frequencies and become very lossy. For some frequency range, metals may be described by a complex permittivity with negative real part. When this happens, plasmon effects [12, 13, 14] are observed: When an arbitrarily small particle with negative permittivity is illuminated by a plane wave, it becomes resonant for certain values of the permittivity. The plasmon resonances depend not only on the material property and geometric shape of the particle but also on particle interactions. Because we are interested in a slab consisting of many particles, particle interactions play an important role and cannot be neglected. Figure 10 shows the dependence of the transmission coefficient on the relative permittivity for the same artificial slab as in Figs. 8 and 9. Obviously, two strong plasmon resonances are observed near the relative permittivities -1.3 and -1.75 when the electric field is polarized parallel to the axes of the molecules. It is important to note that these resonances exist at long wavelength (i.e., the "molecules" can be short compared with the wavelength). Despite this, one can obtain an almost perfect filter that transmits only one polarization. Because the location of the plasmon resonances do not depend much on the wavelength, such filters would have broadband characteristics despite the sharp plasmon resonances.

The model considered above is not very realistic because the permittivity is assumed to be real (i.e., losses are ignored). All known metals exhibit rather strong losses (i.e., the permittivity has an imaginary part that cannot be neglected). In order to see the consequences of the losses in the "molecules," we assume that the relative permittivity has an imaginary part of 0.1, which is very small for realistic metals. As one can see from Fig. 11, the resonance peaks become much broader and less pronounced. Despite this, metals obviously help to strengthen the anisotropic effect of a metamaterial even at low frequencies (i.e., when the period of the artificial crystal is long compared with the wavelength).

2.3. Artificial Chiral Media

Chiral media are obtained from molecules like sugar that have some handedness, that is, these molecules are either left- or right-handed (Fig. 12). Note that this has nothing to do with the so-called left-handed media or negative index materials that will be considered in the following subsection. Chiral molecules act on the electromagnetic field like helical antenna. As a consequence, the chiral media provide interactions of the electric and magnetic field that can be described by macroscopic material properties of the form

$$\vec{D} = \varepsilon\vec{E} - \chi\frac{\partial\vec{H}}{\partial t}; \quad \vec{B} = \mu\vec{H} + \chi\frac{\partial\vec{E}}{\partial t} \tag{8}$$

where χ is the chiral parameter [15]. Chiral slabs are attractive because they rotate the direction of the polarization. Because the chirality of natural media is usually very weak, this effect can only be exploited by sufficiently thick slabs or when metamaterials with a strong chirality can be fabricated. At radar frequencies, one cannot afford thick slabs. At the same time, it is relatively easy to fabricate small helical antennas for such frequencies. When one

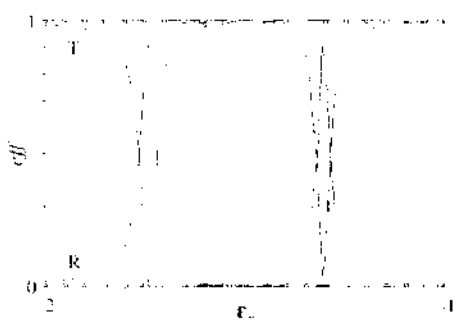


Figure 10. Permittivity dependence of the reflection (R) and transmission (T) coefficients of an anisotropic metamaterial slab consisting of a single layer of cells shown in Fig. 8 for vertical incidence of a plane wave. The wavelength is $a/10$, where $a = 1$ μm is the period of the structure in x and y directions. Curves without markers: polarization of the incident wave parallel to the axis of the "molecules" curves with markers: orthogonal polarization

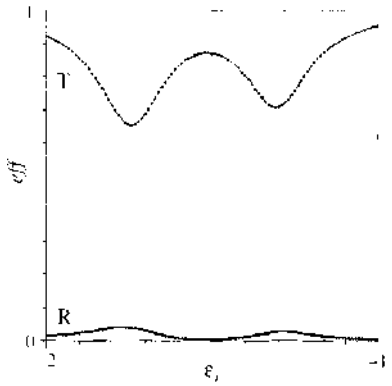


Figure 11. Same as in Fig. 10, when the relative permittivity of the "molecules" is assumed to be complex, with an imaginary part equal to 0.1. Only the polarization of the incident wave parallel to the axis of the "molecules" is considered.

distributes them in a dielectric, one obtains a metamaterial with strong chirality. For this reason, artificial chiral materials first attracted interest within radar applications [16, 17].

When the chiral elements, for example the helical antennas, are arranged on a regular grid in such a way that the axes of all elements point in the same direction, the artificial material becomes not only chiral, but also anisotropic. The analytic treatment of such materials is very demanding (i.e., even simple cases such as the transmission of a plane wave through a slab becomes tedious). Because no commercial field solvers can handle such materials, it is best to compute an artificial anisotropic chiral slab by directly computing the array of cells that are used to fabricate it. This can be done with any field solver that can handle periodic symmetries. In Ref. [11], it has been demonstrated that a periodic structure consisting of helical antennas rotates the polarization of the incident wave by some angle α as it is expected for a chiral slab. The angle α depends on the orientation of the antennas, which indicates that the slab is anisotropic, but for a short helical antenna, this effect is weak. Beside this, a strong frequency dependence is observed (i.e., a strong rotation of the polarization is only obtained when the helical antennas become resonant). One then can obtain a rotation of several degrees with a slab consisting of a single layer of antennas. Because the length of the helical antenna can be considerably longer that the periods in x and y directions of the slab, the first resonance of the antennas can be reached before higher order Floquet modes are excited (i.e., one can operate the chiral slab at a frequency where only a single, transmitted plane wave is observed). It is interesting to note that almost 100% transmission is obtained when one is sufficiently well below the first resonance frequency, whereas the transmission coefficient drops down very much near the resonance frequency.

Because metals lose their conductivity at optical frequencies, the model of perfectly conducting helical antennas in Ref. [11] is not realistic. In the previous subsection, we have seen that metals can be interesting also in the optical regime when one wants to obtain

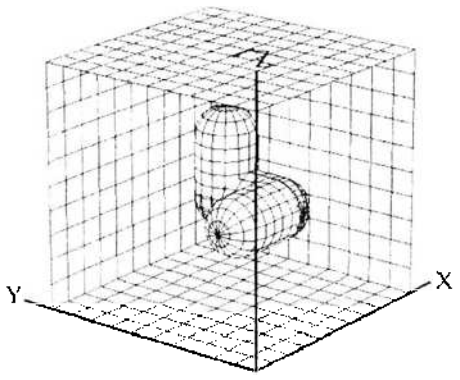


Figure 12. Cell with a right handed chiral "molecule."

metamaterials with pronounced effects. Therefore, we assume that metals will play an important role for the fabrication of chiral metamaterials.

2.4. Negative Index Materials: Left-Handed Media

In 1968, Veselago [7] predicted negative refraction for media with both negative permittivity and negative permeability (i.e., with a negative refraction index). For example, a plane wave incident from free space (with refraction index $n = 1$) on a negative index material with refraction index $n = -1$ should be refracted as shown in Fig. 13. In fact, simulations with well-known electromagnetic field solvers support this idea. Although this seems to provide no problems when an optical ray model is used, difficulties are observed in more realistic wave models. According to Valanju [18], negative refraction would violate the fundamental limit of speed of light, and according to Garcia and Nieto-Vesperinas [19], a perfect lens made of negative index material—as proposed by Pendry [20]—would require an infinite amount of energy to operate. Furthermore, numerical simulations in time-domain, as performed by Correia and Jin [21], become unstable when one simple assumes that the permittivity and permeability are equal to -1. Because no negative index materials have ever been observed in nature, the discussion of such materials was obsolete until Pendry suggested fabrication of such metamaterials based on small antenna structures such as wires and metallic split-ring resonators.

The main effect of negative index material is that a plane wave incident from free space on a negative index material is refracted in the "wrong" direction as shown in Fig. 13 (i.e., the component of the wave vector tangential to the surface becomes negative). This is also called negative refraction. Note that also gratings can exhibit negative refraction, namely when a negative-order Floquet mode is dominant. Because the amplidudes of Floquet modes strongly depend on the angle of incidence, this effect is usually only observed for certain angles of incidence, where it is present for all angles in the case of a negative index material. For this reason, one also uses the term *all-angle negative refraction* (AANR). In Ref. [22], it has been shown that AANR can also be obtained for metamaterials without negative effective index, namely with photonic crystals. We will consider this case in the section on photonic crystals.

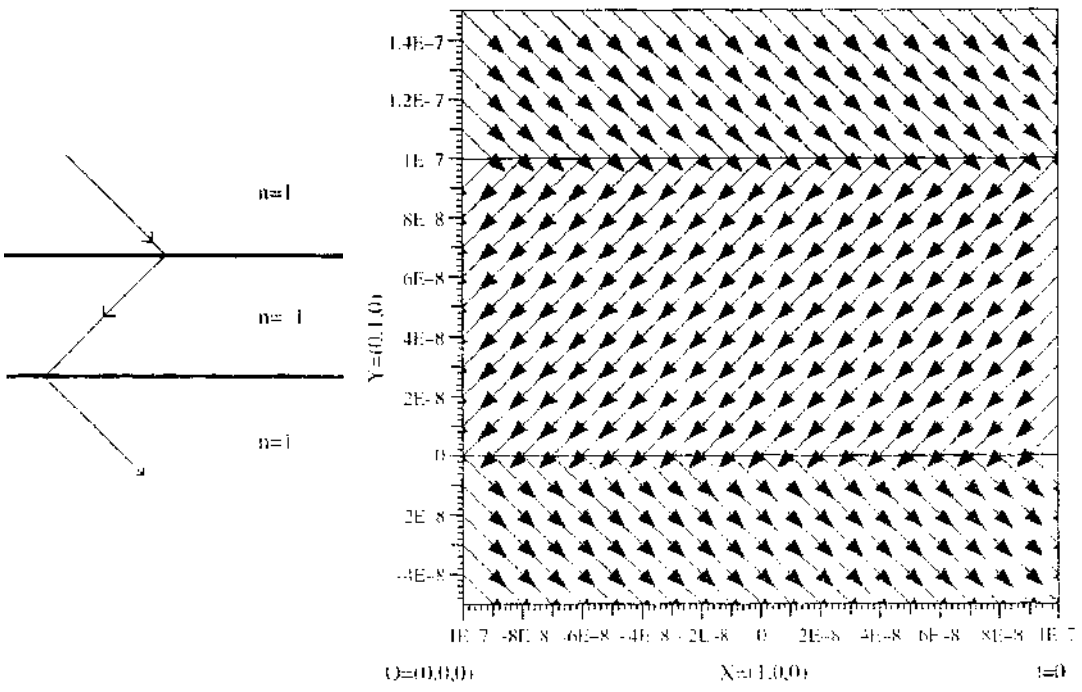


Figure 13. Plane wave refraction at a negative index material slab. Left: ray model; right: MaX-1 simulation, time average of the Poynting vector field.

Negative index materials are also called *left-handed materials*, although they do not consist of left-handed molecules like chiral media. To understand the term *left-handed materials*, we best consider the energy conservation law for electromagnetic fields. Maxwell theory assumes that the electric field is a vector field, whereas the magnetic field is a pseudo-vector or axial vectors. Pseudo-vectors are obtained from vector products of vectors. Vectors describe line-like objects (i.e., 1D objects), whereas pseudo-vectors describe 2D objects in 3D space [23]. The Pointing vector is obtained from a vector product of the electric and magnetic field (i.e., of a vector and a pseudo-vector). As a consequence, the Poynting vector is a true vector that is perpendicular to the electric and magnetic field vectors in such a way that these three vectors are right handed (i.e., when the electric field points in the direction of the thumb and the magnetic field points in the direction of the index finger, the Poynting vector points in the direction of the middle finger of our right hand). This only holds in an ordinary, "right-handed" material as we can see when we now consider the fundamental energy conservation law; that is, the divergence of the Poynting vector field should be the negative time derivative of the energy density:

$$\mathrm{div}\vec{S} = \mathrm{div}\left(\vec{E} \times \vec{H}\right) = \vec{H} \times \mathrm{curl}\vec{E} - \vec{E} \times \mathrm{curl}\vec{H} = -\left(\mu\vec{H}\frac{\partial\vec{H}}{\partial t} + \varepsilon\vec{E}\frac{\partial\vec{E}}{\partial t}\right) \tag{9}$$

In Eq. (9), the first equation is obtained from the definition of the Poynting vector, the second one is a mathematical identity, and the third is obtained from Maxwell's equations for a linear, homogeneous, isotropic material that may be described by the scalars μ and ϵ (i.e., the permittivity and permeability). The term on the right-hand side is essentially the negative time derivative of the energy density of the electromagnetic field. From the negative sign of this term, it follows that the Poynting vector points away from an area where the energy density is decreased—provided that μ and ϵ are positive. Because μ and ϵ are negative in a negative index material, Eq. (9) would violate the energy conservation law in such a material. To avoid this, the definition of the Poynting vector field $\vec{S} = \vec{E} \times \vec{H}$ must be replaced by $\vec{S} = -\vec{E} \times \vec{H} = \vec{H} \times \vec{E}$ in such a medium. This means that the three vectors $\vec{E}$, $\vec{H}$, $\vec{S}$ only obey the right-hand rule in an ordinary medium, whereas a left-hand rule applies for these vectors in a left-handed medium. However, in order to avoid confusion with left-handed molecules, we do not use the term *left-handed medium* in the following.

The extraordinary effect of AANR at the surface of negative index materials is certainly not sufficient for explaining the tremendous interest in such materials that was initiated by Pendry's paper [24]. The main reason for this is the promise of a "perfect" lens that would image a point source on a sharp focus point that is not limited by the wavelength as when an ordinary lens is used. Such a lens could be used for obtaining optical microscopes with higher resolution than conventional ones. Because higher optical resolution is already obtained from scanning near-field optical microscopes (SNOMs) [12], a "perfect" lens or a more modes "super-resolution" lens should have advantages compared with SNOM; that is, it should provide higher resolution or it should allow one to remove the main drawbacks of SNOMs, namely, the fact that a SNOM acts in the near field of the object to be observed. Although microscopes are useful instruments, super-resolution becomes much more attractive from the engineering point of view because it would allow one to overcome the limit of optical storage devices such as CD ROMs, and this would be attractive for the mass market of computer industry. The demands for supper-resolution CD ROMs are much higher than for super-resolution microscopes (speed, stability, durability, fabrication costs). Therefore, the promise that SNOM technology might provide such CD ROMs is still neglected, although SNOMs are widely used in microscopy.

Pendry proposed that a very simple "perfect" lens could be obtained using a simple slab of negative index material as illustrated in Fig. 14. As one can see from the ray model, such a lens would not behave like an ordinary lens: When the source point moves to the left, also the image point moves to the left, and when the source point moves closer to the lens, the image point moves away. Unfortunately, but not surprisingly, calling such a slab a "perfect" lens is in fact an extreme exaggeration as we can see from accurate numerical simulations based on Maxwell theory. Before we do this, we would like to point out that all structures that were

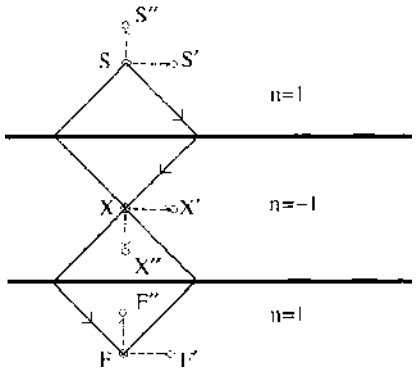


Figure 14. Ray model of a "perfect" lens consisting of a slab with refraction index $n = -1$. The rays originated at the source point S cross each other inside the slab at the point X and again at the focal point F. The distances of S and X from the upper boundary of the slab are equal, and the distances of X and F from the lower boundary are equal as well. When the source point S is moved in any direction, the focal point is moved in the same direction. This only holds as long as the focal point is below the lower boundary of the slab.

proposed for constructing negative index materials are operated near a resonance. As we have seen before, resonance can be important for obtaining metamaterials with sufficiently strong effects. As a consequence, such materials exhibit the desired effect only in a very narrow frequency band. To overcome this difficulty, Tretyakov [25] proposed the use of dipole and loop antenna that are loaded with electronic circuits. This is a strong extension of the metamaterial concept that might be feasible for sufficiently low frequencies. At the same time, it makes it questionable that a "perfect" lens with focus size considerably smaller than the wavelength might be obtained using such an advanced metamaterial at least as long as the cells, that is, the loaded antennas, that constitute the metamaterial are not small compared with the wavelength. However, in the following we assume that all fabrication problems can be solved and that a material with negative permeability and negative permittivity exists. In order to study the effects of wave propagation through a simple negative index slab, we apply the MMP solver of MaX-1 [26]. Because MMP is a semianalytical method, it allows one to obtain highly accurate frequency domain results, which is important because there is currently no way to compare such results with measurements.

First, we consider a simple 2D model of a negative index slab illuminated by a monopole source. In order to avoid numerical problems when the relative permittivity and permeability are equal to -1, we insert small losses (i.e., we assume that the relative permittivity and permeability are complex and have a small imaginary part). As one can see from Fig. 15, the field becomes rather complicated. In addition to the field radiating away from the monopole

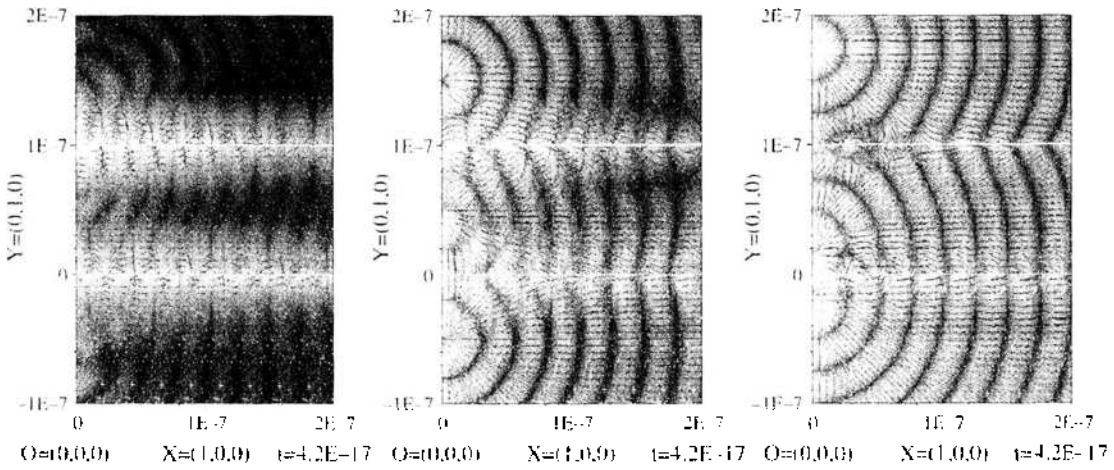


Figure 15. Poynting vector field for a 2D model of a negative index slab illuminated by a monopole source at distance d above the slab. The width of the slab is 2λ. Left: $d = \lambda/2$; middle: $d = \lambda$; right: $d = 3\lambda/2$. The relative permittivity and permeability of the slab are $-1 - 0.001i$.

source, penetrating the slab and radiating out of the slab on the opposite side as suggested by the ray model, an additional field near the surfaces of the slab is observed. This field becomes stronger when the monopole moves closer to the surface and when the losses in the slab are reduced. This clearly supports the ideas of Garcia and Nieto-Vesperinas [19]. The strong field near the surfaces of the slab makes it difficult to see a strong focus as suggested by the ray model. As one can see from Fig. 16, the field near the surface can be suppressed well by introducing higher losses in the slab, but then, no sharp focus, that is, not even a "super-resolution lens" is obtained.

It is well-known that the field of a 3D dipole source decays much more rapidly with the distance that the field of a 2D monopole source. Therefore, one can assume that the undesired effect observed above is less strong for a more realistic 3D model. At the same time, such a model becomes more complicated because of polarization effects and it is much more difficult to visualize and inspect the electromagnetic field in 3D space. Fig. 17 shows a quarter of a circular negative index disk illuminated by a dipole that is oriented perpendicular to the disk. Although this structure is axisymmetric, the electromagnetic field is not axisymmetric because the dipole is not oriented along the axis. Because a dipole along the axis would not radiate toward the disk, this case is simpler, but not interesting. However, when we place the dipole at a distance $d = w/2$ from the upper surface of the disk, the ray model suggests a focus at the distance d below, where w is the width of the disk. As one can see from Fig. 17, the field is concentrated very much near the dipole source. Therefore, the disk is sufficiently large for avoiding undesired effects caused by its finite size. Despite this, the field is extremely complicated. Because the main purpose of the negative index lens is to focus the field radiated by the dipole on a spot that is expected to be small compared with the wavelength, we consider the power flux through a plane that contains the hypothetical focus point, perpendicular to the axis of the disk. Figures 18 and 19 show the intensity pattern on this plane for different wavelengths. As long as the width of the disk is small compared with the wavelength, a strong intensity is only obtained inside an area of radius r around the focal point, and r is of the order $d = w/2$. Close to the focus point one has at least one strong spot with negative intensity, that is, here the focal plane is not illuminated from the back side. This is possible because one is in the near field of the disk. Energy radiated from the disk at some distance from the axis turns back and illuminates the focal plane from the "wrong" side. This has an important consequence for the measurement. Any equipment used for measuring the field near the focus will strongly interact with the field of the disk. Thus, it will be difficult to observe the "super-resolution." The fact that several spots are present in the focal plane causes another difficulty and makes the negative index disk unattracive as an imaging device. Finally, when we increase the frequency—as shown in Fig. 19—the areas of negative illumination (i.e., illumination from the back side), disappear. But then, one still observes several spots in the focal plane, and the total spot area is no longer small compared with the wavelength. Obviously, a conventional dielectric

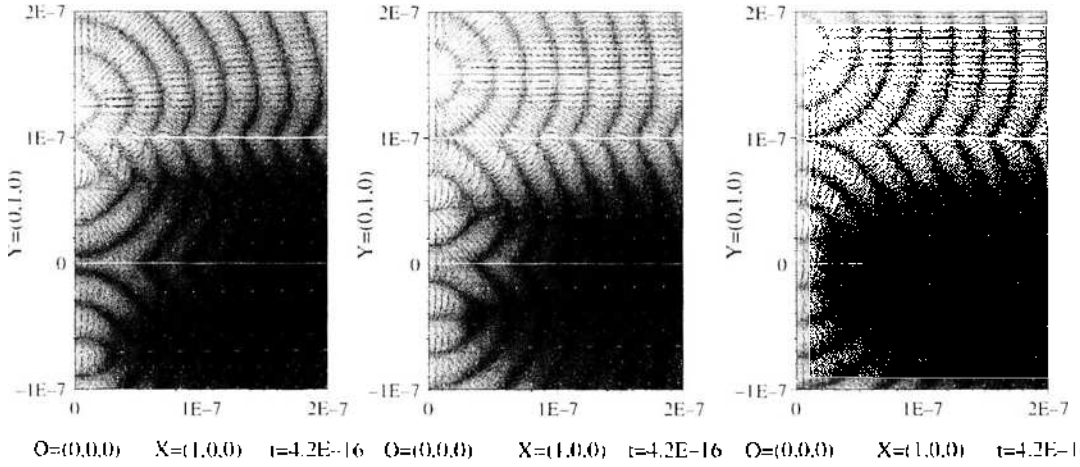


Figure 16. Same as in Fig. 15 but with relative permittivity and permeability of the slab equal to $-1 + 0.1i$.

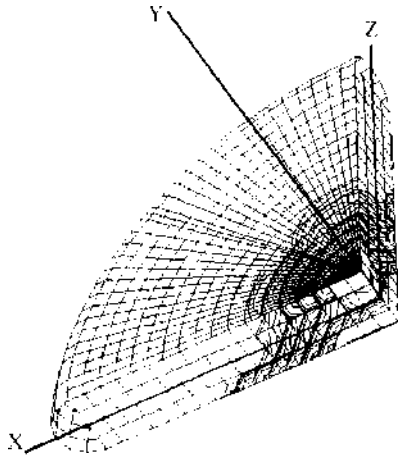


Figure 17. Negative index material disk illuminated by a dipole on the axis of the disk, polarized perpendicular to the axis. Distance d of the dipole from the disk $w/2$, where w is the width of the disk. According to the ray diagram, the focal plane is also at the distance d from the disk but on the opposite side. The radius of the disk is $5w$. The relative permittivity of the disk is $-1 + 0.0001i$; the relative permeability is -1. Iso-lines of the time average of the Poynting vector field are shown with a logarithmic spacing.

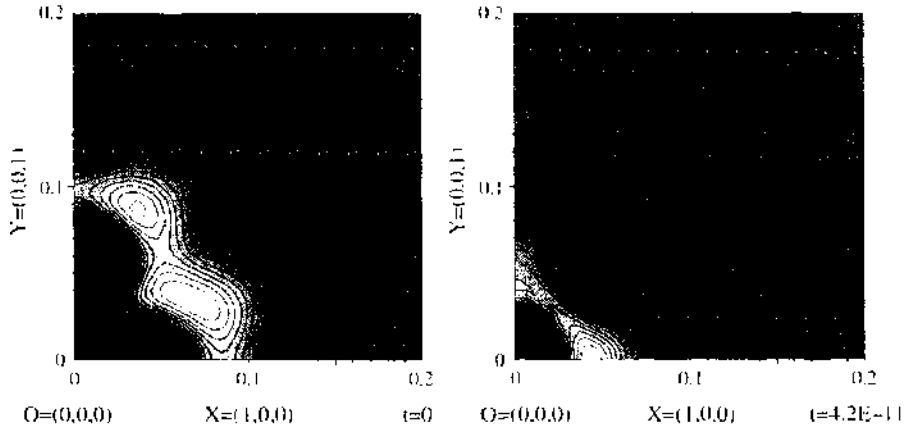


Figure 18. Intensity of the time average of the Poynting vector component perpendicular to the focal plane for the negative index disk shown in Fig. 17. White indicates strong positive, black strong negative values. Left-hand side: wavelength $\lambda = 10w$, where w is the width of the disk. Right-hand side: $\lambda = 2w$.

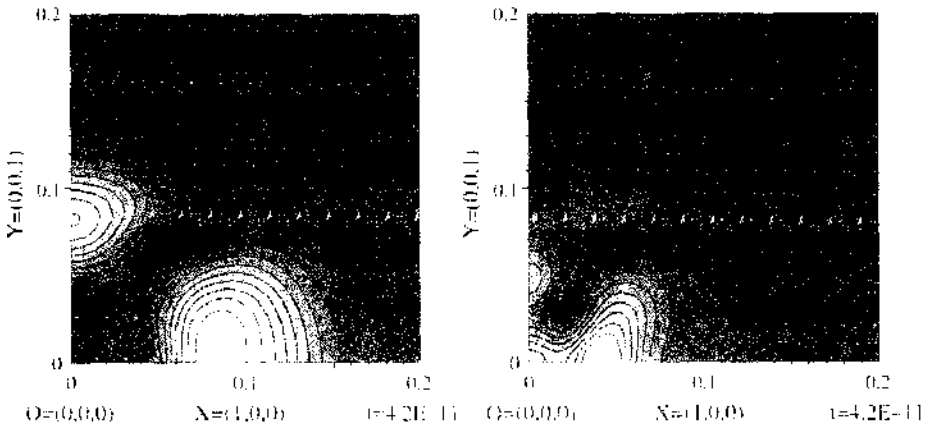


Figure 19. Same as Fig. 18. Left-hand side: wavelength $\lambda = 1w$, where w is the width of the disk. Right-hand side: $\lambda = 0.5w$.

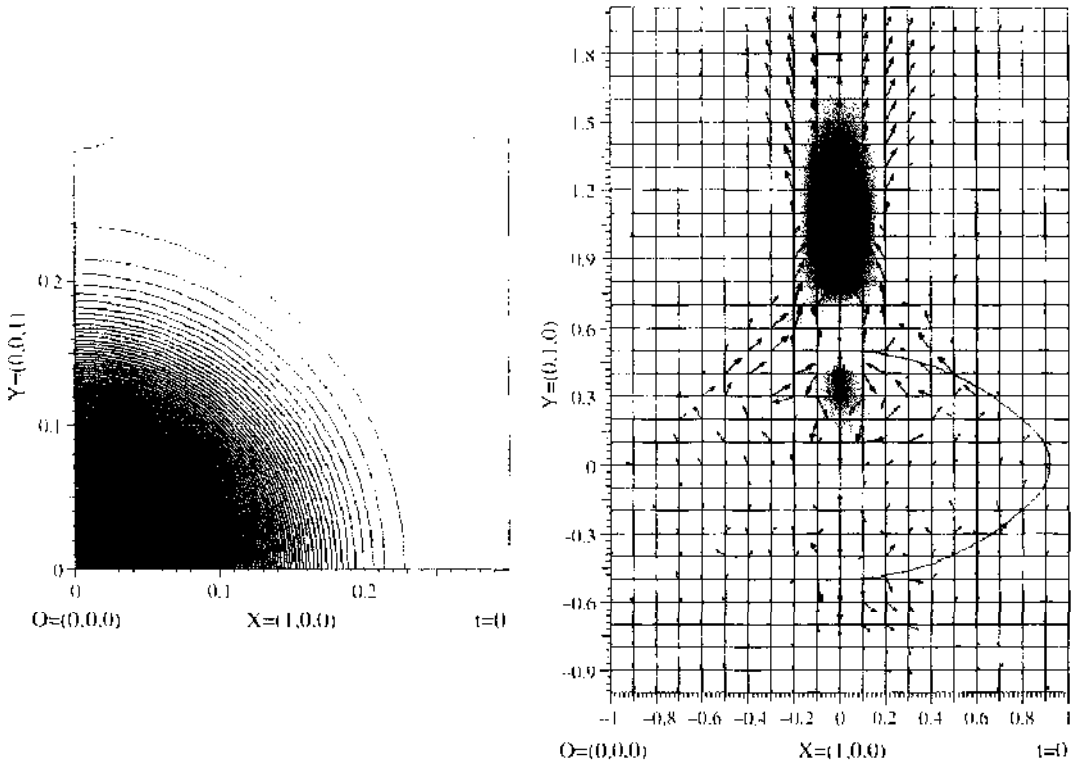


Figure 20. Small conventional dielectric lens with relative permittivity 2.25, diameter 6 wavelengths, illuminated by a plane wave propagating along the axis of the lens. Left-hand side: intensity plot in the focal plane. right-hand side: time average of the Poynting vector field in the plane parallel to the electric field (left half of the plot) and in the plane perpendicular to the electric field (right half of the plot).

lens as illustrated in Fig. 20 has much nicer characteristics than a negative index lens that is far from being perfect. As one can see, the conventional lens has a single, bright spot around the focus that is almost circular and has a diameter of the order half the wavelength.

From these simulations, we assume that the promises of super-resolution lenses based on negative materials are wishful thinking that has gained a position in modern optics similar to the philosopher's stone in alchemy. Although our results concerning the "perfect lens" are rather discouraging, the study of negative index materials is fascinating because it provides unexpected and highly complicated effects even when most simple configurations are considered. These effects are inspiring and motivating for the research on metamaterials. Maybe, this will push material research forward just as the vain search for the philosopher's stone finally initiated fruitful chemical research. In the following section, we therefore focus on photonic crystals as a more promising category of metamaterials that also exhibits many surprising effects and provides a variety of new structures for ultradense integrated optics with attractive engineering applications. As we will see, photonic crystals also provide all-angle negative refraction and could therefore be candidates for super-resolution lenses [22].

3. PHOTONIC CRYSTALS

In the previous section we have considered a simple grating consisting circular dielectric rods. The analysis of the results presented in Fig. 20 and more extended studies suggest that light incident from any direction on such a grating with sufficiently many layers of rods will be totally reflected within some frequency ranges. As mentioned in the introduction to this chapter, Yablonovitch [27] proposed to consider such structures, called photonic crystals (PhCs), as photonic counterparts of electronic semiconductor crystals (i.e., the photon in a photonic crystal propagates in a photonic crystal in a similar way as an electron in a semiconductor). The most interesting physical effect is that there are energy bands where no electron can propagate in a semiconductor and where no photon can propagate in a photonic crystal. Such energy bands in a PhC are called photonic band gaps (PBGs). Because the

energy of a photon is proportional to its frequency, the energy bands can also be considered as frequency bands. When a photonic crystal slab is illuminated by a plane wave, a stop band is observed (i.e., in terms of filter technology, each PBG corresponds to a stop band). Thus, it is evident that PhCs are interesting for the design of optical filters.

Semiconductors as well as PhCs have interesting physical properties, but what makes them really attractive from the engineering point of view is the possibility to modify their characteristics by doping. Semiconductors are usually doped by a more or less statistical infiltration of dopants (i.e., atoms that do not belong to the pure crystal). Therefore, doped semiconductor areas typically consist of many atoms. Because the wavelength of photons is much longer then the wavelength of electrons, the cells in a photonic crystal—that correspond to the atoms in a semiconductor—are big compared with the atoms, typically of the order 100 nm or more. Therefore, one might expect that doped photonic crystal become big compared with the wavelength and this would make them unattractive for integrated optics, especially when one intends to design ultrasmall configurations that should not be much bigger than the wavelength. Here, it is important to note that the size of the cells of a PhC—that correspond to the atoms—is big enough for permitting a precise doping. Furthermore, because these cells consist of many atoms, one has a much higher degree of freedom in the doping process, that is, instead of introducing an atom (selected from a small list of acceptable atoms) as dopant, one may modify the location, the orientation, and the geometric shape of the "photonic atom" as well as the material used for fabricating it. When we consider, for example, a PhC built of circular dielectric rods, we can have a variety of "dopants" as shown in Fig. 21. Although this high degree of freedom is attractive, it also poses a big problem: A good design of a doped PhC structure requires an optimal or at least a good selection of the different parameters. Because the parameter space is huge, this is an extremely demanding task. Furthermore, we will demonstrate that doped PhC structures have a high complexity and often behave in a counterintuitive way (i.e., intuition of experienced engineers can be insufficient or even misleading during the design process). Furthermore, simple and useful design rules that are usually applied by engineers are currently missing for the design of doped PhC structures. In order to overcome these difficulties, we will take advantage of extensive numerical simulations and optimizations.

Because perfect PhCs are periodic structures, the doping process always breaks at least one of the periodic symmetries. This leads to a simple classification of doped PhC structures: In 3D crystals, one can have one, two, or three broken symmetries. PhC structures for integrated optics are usually obtained from drilling holes in dielectric plates. Such structures can be approximated by 2D PhCs that exhibit only two periodic symmetries. For reasons of simplicity, we focus on such structures in the following. Obviously, one only can break one or two of the periodic symmetries of a 2D PhC. When only one symmetry is broken, a waveguide structure is obtained. Although PhC waveguides are important, the more general case where both symmetries are broken is much more attractive and fundamental for the design of various structures such as resonators, filters, modulators, and so on. From the engineering point of view, waveguides are used to connect different devices. For example, a resonator that is not linked to some other part of a circuit by some waveguide makes

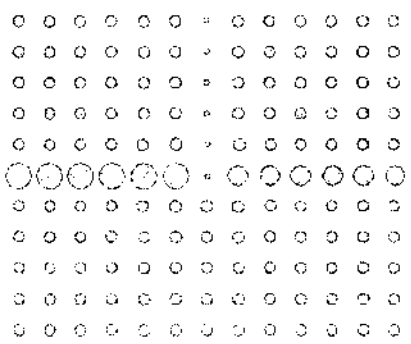


Figure 21. Photonic crystal with different "dopants" (i.e., cells containing rods of bigger or smaller size than the rods of the original crystal).

not much sense. A resonator that is attached to a waveguide may also be considered as a discontinuity of the waveguide. Therefore, we will consider the following classes of problems in the following: (1) perfect PhCs (without any broken symmetry), (2) PhC waveguides (with one broken symmetry), and (3) waveguide discontinuities (with two broken symmetries). Note that a waveguide in a 3D PhC breaks two symmetries, and a waveguide discontinuity in a 3D PhC breaks all three symmetries. A single broken symmetry in a 3D PhC is of not much practical interest. This case corresponds to parallel plate waveguides.

In the following, we consider the analytical background and the numerical treatment of the three classes (1) perfect crystals, (2) PhC waveguides, and (3) PhC waveguide discontinuities. From the engineering point of view, the third class is most interesting because all attractive PhC devices belong to this category, but as these devices are accessed mainly through waveguides, one should study category 2 beforehand. This study also requires the knowledge of the locations of the PBGs because PhC waveguides are operated within a PBG of the corresponding perfect crystal.

3.1. Perfect Photonic Crystals

Although a perfect PhC that occupies the entire 3D space is not realistic and not very attractive from the engineering point of view, the idealization behind a perfect PhC is required for analytical studies and it is essential for understanding the fundamental properties of PhCs. Essentially, this idealization is the same as when one considers plane wave propagation in a natural material that occupies the entire 3D space (i.e., we want to find out how electromagnetic waves propagate within infinite photonic crystals). Because PhCs are structured, a strong dependence on the direction of the wave propagation is expected. This makes the analysis rather demanding.

Perfect PhCs are closely related to perfect gratings. In both cases, periodic symmetry is involved (i.e., the structures consist of identical cells that are repeated at least in one direction). The main difference is that a grating consists of at least two different materials, usually free space and a periodic metamaterial or a slab of metamaterial with free space or a natural material on both sides of the slab. For this reason, one can use the same techniques for handling the periodicity of gratings and PhCs. Above all, it is sufficient to know the electromagnetic field of the grating or PhC within a single periodic cell together with a symmetry description.

Because a grating consists of sections with different natural materials or metamaterials, it is also different from a perfect PhC. Usually, one assumes that a grating is illuminated by a plane wave that is defined within a natural material, such as free space. This known plane wave excites the field in the entire structure (i.e., one essentially has a scattering problem with some periodic symmetry). For the simple case where the grating is periodic in x direction (see Figs. 3–7) we have seen that the angle of incidence of the plane wave and the x component of its wave number play an essential role. In fact, the incident plane wave "impresses" the symmetry condition in x direction

$$Field(x + na, y, z) = Field(x, y, z) \times e^{jnk_x^{inc}a} = Field(x, y, z) \times (C_x)^n \tag{10}$$

where n is an arbitrary integer number. When the field is known, for example, in the interval from $x = 0$ to $x = a$, one can use Eq. (10) for computing the field everywhere outside this interval.

Equation (10) can easily be extended for the case where the structure is periodic in two or even three directions. Therefore, we can write for a 3D PhC with a periodic symmetry in x, y, z directions with the lattice constants a_x, a_y, a_z:

$$Field(x + na_x, y + ma_y, z + la_z) = Field(x, y, z) \times (C_x)^n \times (C_y)^m \times (C_z)^l \tag{11}$$

Now, the essential difference between the grating and a PhC must be considered: Because the PhC is assumed to extend to infinity, there is no space left where we can define the incident plane wave. This, we do not have an incident wave vector with known components in the x, y, z directions. Therefore, we cannot compute the periodicity constants C_x, C_y, C_z for

the electromagnetic field. In fact, as no excitation is defined, we no longer have a scattering problem. We now have an eigenvalue problem with the characteristic constants C_x, C_y, C_z. We therefore must find nontrivial solutions or eigenmodes that fulfill Eq. (11) as well as the Maxwell equations within a single cell of finite volume, for example, within $0 \leq x < a_x$, $0 \leq y < a_y$, $0 \leq z < a_z$. As in resonator problems, nontrivial solutions are only found for certain frequencies that therefore can be considered as eigenvalues. These eigenfrequencies depend on the characteristic constants C_x, C_y, C_z. For an arbitrary, given set of constants C_x, C_y, C_z, we can find infinitely many eigenfrequencies. For practical reasons, we can restrict ourselves to the lowest eigenfrequencies. Finding them for the space spanned by C_x, C_y, C_z is still quite demanding.

Because such problems are very common in crystallography, we take advantage of the knowledge and formalisms known from crystallography. Here, one usually considers the energy as eigenvalue. Because the frequency of a photon is proportional to its frequency, it is easy to convert the energy into the frequency and vice versa.

Despite of this, the analysis of perfect PhCs is still complicated. Therefore, we use two additional assumptions for reasons of simplicity: (1) we assume that the PhC is two-dimensional (i.e., cylindrical in the z direction perpendicular to the two periodic directions) and (2) we assume that all materials are loss-free. The first assumption is reasonable because it allows us to work with 2D illustrations and 2D symmetry considerations. Furthermore, many realistic PhC structures can be approximated well with such 2D models and the extension to 3D models is straightforward. The second assumption is reasonable because one usually tries to use materials with low losses in optics. Note that this assumption is rather crucial for all kinds of eigenvalue problems for the following reason: The proper definition of an eigenvalue problem requires the assumption that the system is energetically closed. In the case of a simple resonator, this means that no energy is transferred from outside into the resonator and no energy leaves it. Each cell of a PhC should also be energetically closed in the sense that the total energy exchanged with the neighbor cells is zero. When losses are present, some part of the field energy is converted to thermal energy. To take this energy into account, one would need to consider Maxwell's equations together with thermodynamic equations, which is extremely demanding. In order to avoid this, one usually works with simplified, complex models that assume that the eigenvalue becomes complex due to the losses and that the imaginary part of the eigenvalue can represent the losses. In fact, a complex frequency describes a dampened oscillation that is observed when a resonator is lossy. Note that a perfect resonator cannot be observed at all because it is perfectly closed. In order to observe it, one always must open it. Usually, one then has an input port that feeds the resonator and an output port for the measurement. Such a structure is no longer described by an eigenvalue problem. It can be well described as a waveguide discontinuity problem that has a known incident field as an ordinary scattering problem. One then observes a signal in the output port that is frequency dependent and exhibits peaks near the resonance frequencies. As in the case of the plasmon resonances shown in Figs. 10 and 11, one observes a broadening of the reasonance peaks with increasing losses. The width of these peaks is a good indicator for the quality of the resonator (i.e., for the losses). Unfortunately, opening a resonator for observing it also introduces losses, namely radiation losses that also broaden the resonance peaks. This makes it difficult to precisely measure the losses of high-quality resonators.

Another problem of complex eigenvalue problems is the following: In the loss-free case, the modes are characterized by a real eigenvalue that can usually be restricted to a finite interval as we will see in the following. The number of solutions in this interval is always finite. Beside this finite number of modes, one can also have an infinite number of evanescent modes that are characterized by imaginary eigenvalues. In most cases, the evanescent modes are of not much interest and need not be computed. When a complex eigenvalue problem is defined because of the losses, both types of modes are characterized by complex eigenvalues. This makes it not only difficult to separate the evanescent modes, it also leads to an eigenvalue search over the entire complex plane that is much more difficult than the eigenvalue search over a finite real interval. For this reason, one avoids complex eigenvalue problem formulations whenever possible. We will do this also in the following

subsection, but we will consider a complex eigenvalue formulation when we consider PhC waveguides.

3.1.1. Symmetry Considerations

A general 3D photonic crystal is periodic in three directions, the so-called lattice vectors $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$ as shown in Fig. 22. We therefore obtain the following equation for the material properties of the PhC:

$$Material(\vec{r} + n_1\vec{r}_1 + n_2\vec{r}_2 + n_3\vec{r}_3) = Material(\vec{r} + \vec{R}) = Material(\vec{r}) \tag{12}$$

where n_1, n_2, n_3 are arbitrary integer numbers. Note that the vectors $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$ are the three shortest vectors that transform a cell to its neighbor cells. For a rectangular lattice, these vectors are perpendicular, but in general, they are arbitrary. As for the grating, we now have the symmetry condition for the field:

$$Field(\vec{r} + n_1\vec{r} + n_2\vec{r}_2 + n_3\vec{r}_3) = Field(\vec{r} + \vec{R}) = Field(\vec{r}) \times e^{i\vec{K}\vec{R}} \tag{13}$$

We now can define a reciprocal lattice space for the wave vector $\vec{K}$ that is spanned by three reciprocal lattice vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$ as shown in Fig. 23:

$$\vec{K} = \kappa_1\vec{k}_1 + \kappa_2\vec{k}_2 + \kappa_3\vec{k}_3 \tag{14}$$

Note that the reciprocal lattice vectors are not orthogonal in general. Because we now have an eigenvalue problem, where no excitation is given, none of the components of the wave vector $\vec{K}$ is known, but for all nontrivial solutions we must have

$$e^{i\vec{K}\vec{R}} = 1 \Rightarrow \vec{K}\vec{R} = 2N\pi \tag{15}$$

where N is an integer number. This allows us to construct a set of reciprocal lattice vectors as follows:

$$\vec{k}_1 = 2\pi\frac{\vec{r}_2 \times \vec{r}_3}{\vec{r}_1 \times (\vec{r}_2 \times \vec{r}_3)}, \quad \vec{k}_2 = 2\pi\frac{\vec{r}_3 \times \vec{r}_1}{\vec{r}_1 \times (\vec{r}_2 \times \vec{r}_3)}, \quad \vec{k}_3 = 2\pi\frac{\vec{r}_1 \times \vec{r}_2}{\vec{r}_1 \times (\vec{r}_2 \times \vec{r}_3)} \tag{16}$$

Note that this construction is not unique, but it obviously fulfills the condition (15). As a consequence of the periodic symmetry of the original lattice space, also the reciprocal lattice space has a periodic symmetry. Therefore, it is sufficient to know the field of a nontrivial solution only in finite, limited cell, spanned by the lattice vectors $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$ and to consider only a finite, limited zone in the reciprocal lattice space spanned by the reciprocal lattice vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$. When additional symmetries are present, this zone can be further reduced. The smallest, irreducible zone near the origin of the reciprocal lattice space is called first irreducible Brillouin zone (IBZ).

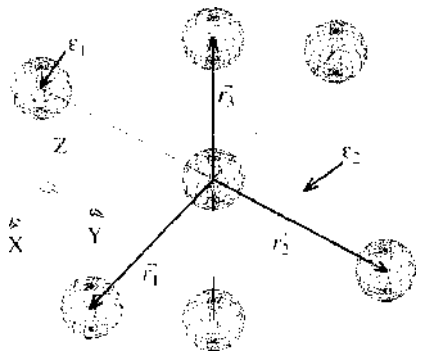


Figure 22. Three-Dimensional photonic crystal consisting of dielectric spheres periodically distributed in 3D space. Lattice vectors and permittivity constants are depicted.

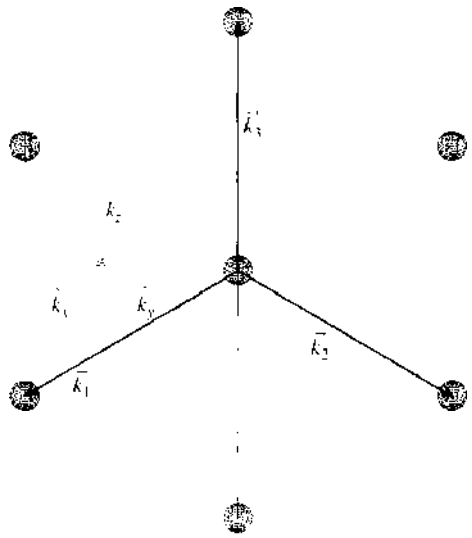


Figure 23. Reciprocal lattice space of the 3D photonic crystal presented in Fig. 22.

We mentioned that we will focus on 2D PhCs for reasons of simplicity. In fact, the IBZ of a 3D PhC is limited but three-dimensional and this makes the visualization of the non-trivial solutions over the IBZ not simple. A 2D PhC is periodic in two directions $\vec{r}_1$, $\vec{r}_2$ and cylindrical in the third direction $\vec{r}_3 = \vec{e}_z$, where z is the coordinate along the cylinder axis. The reciprocal lattice space now becomes two-dimensional and we obtain from Eq. (16)

$$\vec{k}_1 = 2\pi \frac{\vec{r}_2 \times \vec{e}_z}{\vec{r}_1 \times (\vec{r}_2 \times \vec{e}_z)}, \quad \vec{k}_2 = 2\pi \frac{\vec{e}_z \times \vec{r}_1}{\vec{r}_1 \times (\vec{r}_2 \times \vec{e}_z)} \tag{17}$$

We now consider the two most frequently used 2D cases of a square and a hexagonal (or triangular) lattice. In both cases, we assume that the two lattice vectors have the same length a. For the square lattice, the two vectors are perpendicular and the hexagonal lattice has an angle of 60 degrees between the two lattice vectors as shown in Fig. 24. In both cases, the

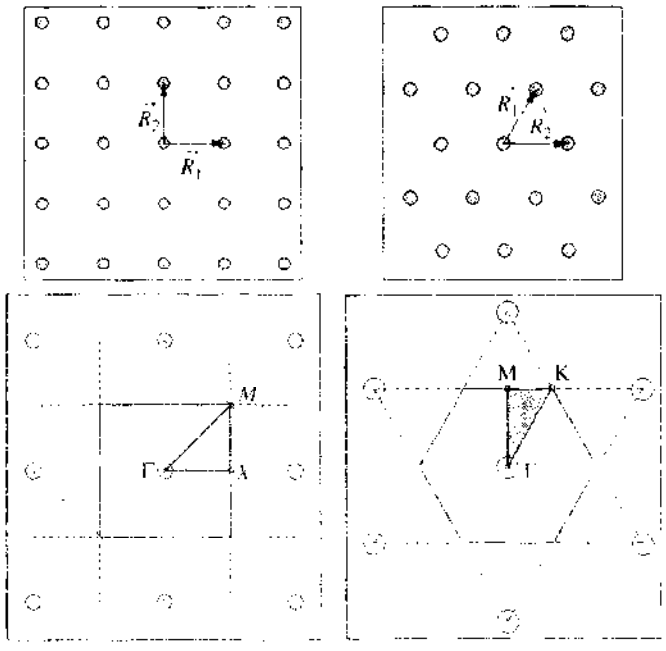


Figure 24. Original lattice space of the perfect photonic crystal with a square lattice (top left) and with a hexagonal lattice (top right); reciprocal lattice and irreducible Brillouin zone for the square (bottom left) and hexagonal (bottom right) lattice

first IBZ becomes triangular and is described by the three corners Γ, M, X and Γ, M, K, respectively. Note that the Γ point is always in the origin (i.e., at $\vec{K} = 0$), but the two other corners are not unique.

3.1.2. Band Diagrams

Each nontrivial solution in a photonic crystal is characterized by a (resonance) frequency that is a function of the $\vec{K}$ vector. As mentioned before, it is sufficient to describe this function only over the first IBZ because of the symmetry properties. Furthermore, extreme values of this function are always on the contour of the first IBZ. Therefore, it is sufficient to plot the resonance frequencies only along the sides of the triangles with the corners Γ, M, X and Γ, M, K for the 2D PhC with square and hexagonal lattice as shown in Fig. 25. These band diagrams allow one to obtain a quick overview over all solutions. Usually, one can even guess the values of the resonance frequencies *within* the first IBZ.

Because identical band diagrams are obtained when the geometry is multiplied by any factor and the frequency is divided by the same factor, one usually plots the normed frequency fa/c, where a is the lattice constant (for the square and hexagonal lattice) and c is the speed of light in free space.

What makes photonic crystals most attractive is the occurrence of band gaps, that is, frequency ranges where no eigenmode exists (for any value of $\vec{K}$). This means that light cannot propagate at these frequencies in any direction. When a finite—but sufficiently large—PhC is illuminated by a plane wave incident from any direction, it will totally reflect the wave when its frequency is within such a PBG. Note that such a complete PBG is usually only observed for one polarization for 2D PhCs.

In some applications, the direction of the incident plane wave is restricted to some sector of angles. Then, it is sufficient that the photonic crystal allows no penetration within this sector only. This means that an incomplete band gap may be sufficient. Furthermore, there are applications where one wants light to propagate through a PhC. In this case, there is no need for a PBG at all. A typical example is a PhC waveguide structure: such a structure is essentially composed of at least two different PhCs or of a PhC and an ordinary material. Within the waveguide channel one then wants to have wave propagation, whereas no propagation is required through the walls of the channel. We will consider such waveguides in the following.

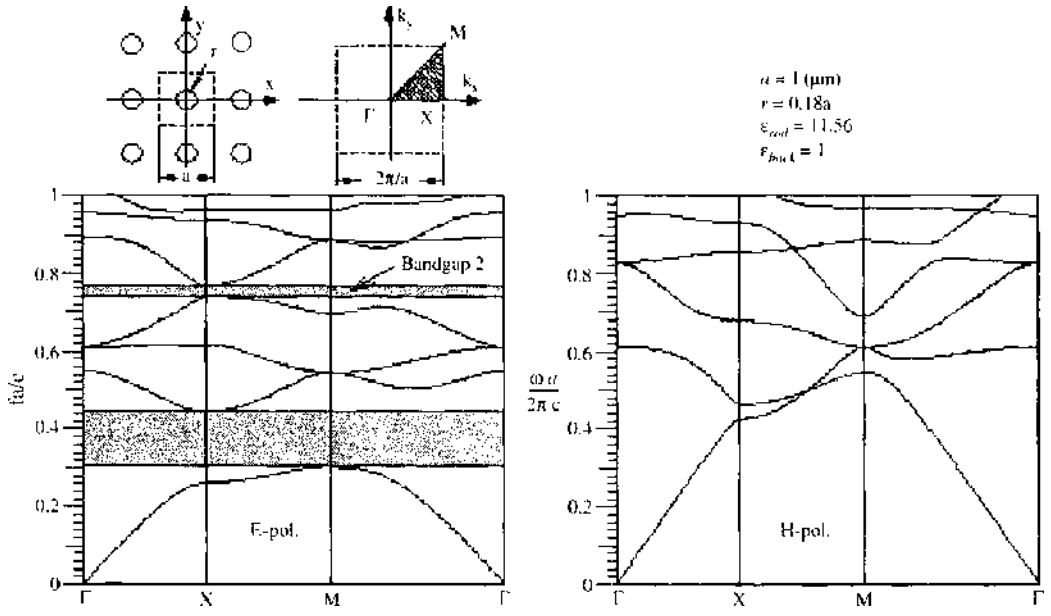


Figure 25. Band structure (dispersion relation) for E-polarization (bottom left) and H-polarization (bottom right) of a perfect 2D photonic crystal consisting of circular dielectric rods on a square lattice: original lattice (top left), reciprocal lattice and first irreducible Brillouin zone (center), geometrical and material properties (top right). As one can see, the bandgap exists only in the case of E-polarization.

It is important to know that a complete PBG is only obtained for certain geometries and certain material properties. Then a PhC is composed of two dielectrics; the dielectric contrast must be high enough for obtaining a complete PBG. This considerably limits the selection of materials that may be used for the fabrication of PhCs. Most frequently, materials with a relative permittivity around 10 are used.

From the fabrication point of view, it is most simple to obtain a 2D PhC structure by drilling holes in a dielectric as shown in Fig. 26. Because of the finite size of the dielectric slabs used for such crystals, the 2D model is somehow inaccurate. It is obvious that the finite size will cause some radiation effect. This radiation can be accounted for by inserting lossy materials in the 2D model. As in the case of lossy resonators, the consideration of lossy PhCs is not trivial. Because the perfect PhC model analysis is mainly used for obtaining a first overview and for finding the band gaps, it makes not much sense to carry out sophisticated complex eigenvalue problems that take losses into account.

3.2. Waveguides in Photonic Crystals

When we embed a slab of an arbitrary material (natural material or metamaterial) within a photonic crystal, we obtain a PhC waveguide as shown in Fig. 27. When we operate at a frequency within the band gap of the PhC on the sides of the waveguide channel, all waves within the channel will be totally reflected at the walls (i.e., at the PhC). The channel itself may be constructed of a natural material or of another PhC or of any metamaterial. The only requirement is that electromagnetic waves may propagate through the channel material at least in some directions. This gives one a huge degree of freedom in the design of PhC waveguides. One therefore can obtain single-mode and multimode waveguides with very different characteristics. Because it requires extensive much experience to synthesize such waveguides with desired characteristics, one usually takes advantage of numerical simulations and optimizations. This is of great importance because one currently does not know much about the limitations of PhC waveguides. The following, very simple waveguide structures demonstrate that it is possible to fabricate waveguides within a PhC that have very different properties. For reasons of simplicity, we will consider the same rod-type PhC we already have considered in the grating section of this chapter.

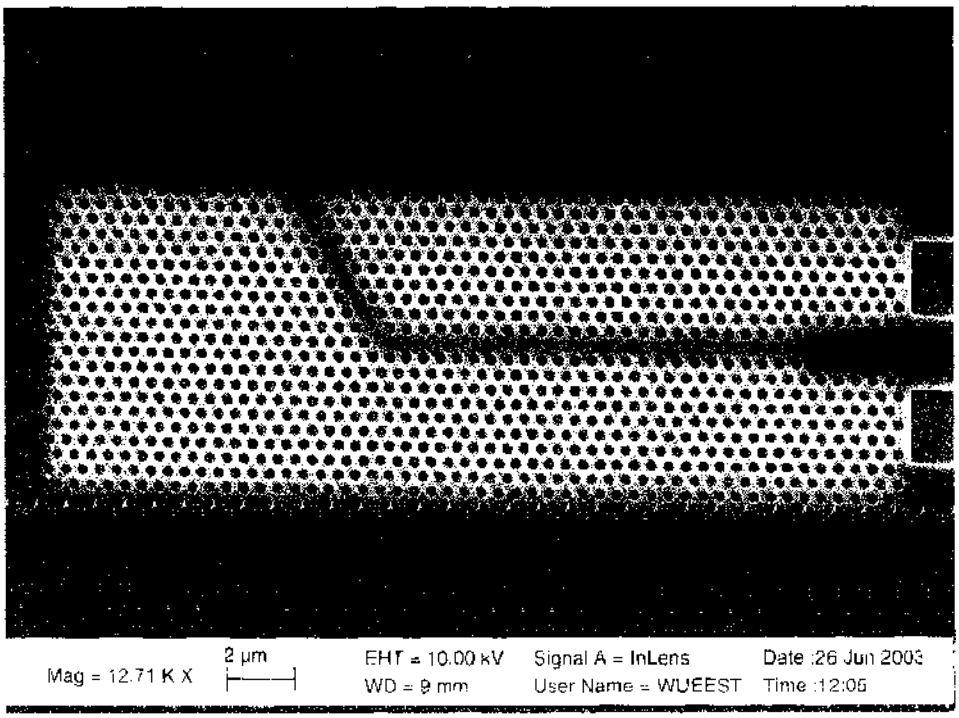


Figure 26. SEM image of a holey photonic crystal structure with holes arranged on a hexagonal lattice and a defect waveguide with a 60-degree bend.

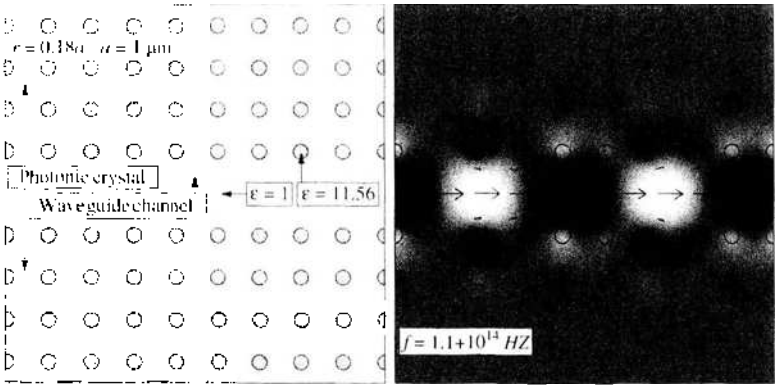


Figure 27. Photonic crystal waveguide consisting of a line defect (vacancies) in the perfect crystal shown in Fig. 25. Left: geometry and material properties; right: z-component of electric field (grayscale) and time-average Poynting vector (arrows).

From the analytical point of view, the waveguide channel destroys the periodic symmetry of the original PhC. Only the periodic symmetry along the channel remains. Thus, the corresponding IBZ becomes one-dimensional, which seems to be much easier than the 2D or 3D IBZ of perfect 2D or 3D crystals. Although this is correct, one should not forget that the cell in the original lattice space now extends to infinity, whereas it was finite for the perfect PhC. This excludes an analytical treatment of PhC waveguides and makes the numerical computation much more demanding. The most commonly used way out of this is the so-called supercell approach. In this approach, one considers a periodic set of parallel waveguides. This replaces the previously broken symmetry by a new symmetry with a larger lattice constant as shown in Fig. 28. As a consequence, one can use the same procedures for perfect PhCs also for PhC waveguides—under the assumption that the interactions of neighbor waveguides may be neglected. As one can see from Fig. 29, the lines in the band diagram split in several lines when a supercell is analyzed. The number of additional lines is increased with the size of the supercell (i.e., with the distance between neighbor waveguides). Although this is numerically annoying, it is of not much interest for the waveguide design. What is interesting, however, is the occurrence of new lines (i.e., modes within the band gap of the original crystals). These lines characterize the modes that propagate within the waveguide. As one can see from Fig. 27, the field of such modes within the channel is

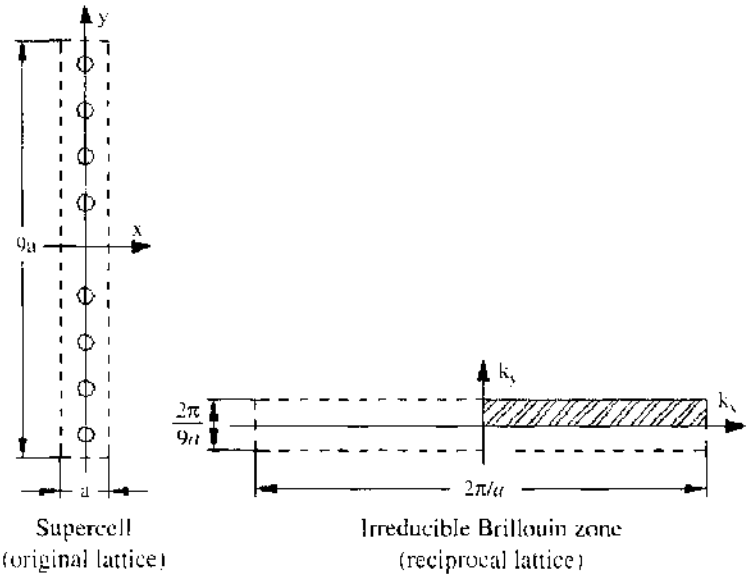


Figure 28. Supercell approach for photonic crystal waveguides. As one can see, fictitious periodicity in y-direction is introduced. This allows us to define periodic boundary conditions in x and y directions and to perform eigenvalue analysis of the waveguide as in the case of the perfect crystal.

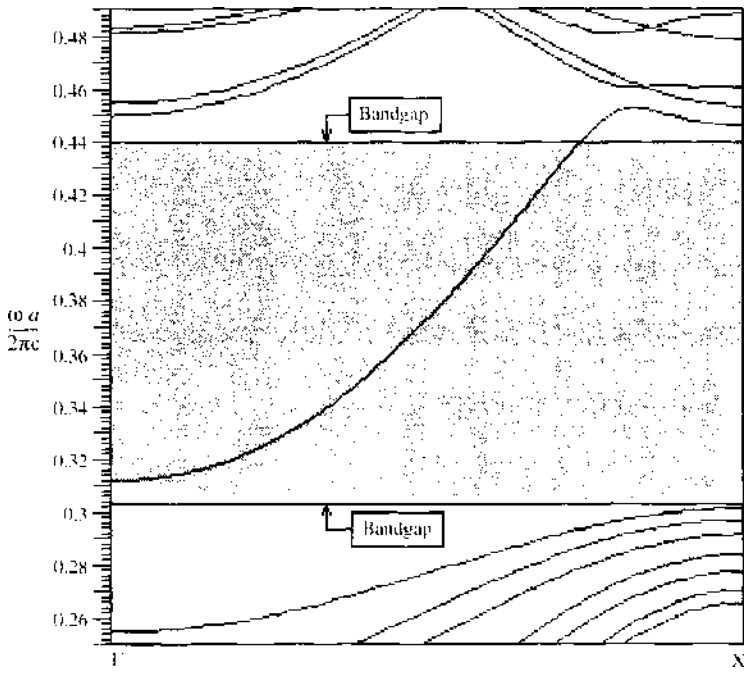


Figure 29. Dispersion relation (band structure) of the photonic crystal waveguide depicted in Fig. 28. A single defect mode covers almost the entire bandgap of the original perfect crystal shown in Fig. 25. Note that the modes of the original crystal outside the bandgap split in several modes. The number of additional modes outside the bandgap depends on the size of the supercell.

dominant and decays rapidly with the distance from the channel. When the field between two channels is weak enough, one can assume that the interaction between the neighbor channels is weak enough and can be neglected. This finally justifies the supercell model.

An alternative to the supercell solution is obtained from an extension of the general procedure for the computation of waves in cylindrical waveguides. In cylindrical waveguides, the longitudinal field dependence is obtained from the separation of the z dependence, where z is the cylinder axis, in the frequency-domain analysis. One then simply obtains

$$Field(\vec{r}_T, z) = Field(\vec{r}_T, 0) \times e^{i\gamma z} \tag{18}$$

where γ is the propagation constant and $\vec{r}_T$ is a two-dimensional vector in the transverse plane, perpendicular to the z direction. This means that the field is known everywhere when the propagation constant is known and the field in any transverse plane, for example, at $z = 0$, is known. This reduces the dimension of the problem by one. In the transverse plane, an eigenvalue problem is to be solved. As for photonic crystals, one could now search for the resonance frequencies as functions of γ and draw the corresponding 1D band diagrams. Because cylindrical symmetry is the limit of periodic symmetry with lattice constant $a = 0$, the corresponding 1BZ is one-dimensional, but infinite.

In practice, a waveguide mode is excited by a generator at the end of the waveguide. This generator determines the frequency. Thus, it is more reasonable to fix the frequency and to search for the corresponding propagation constant as a function of the frequency. When one normalized the propagation constant with the free-space wavenumber, the normalized propagation constant becomes finite for all frequencies and for all modes.

Damping is one of the most important issues in the design of waveguides. Therefore, one usually is interested in the attenuation due to material losses. The attenuation is described by the imaginary part of the propagation constant. Thus, one has much experience with complex eigenvalue problems in the description of waveguides. When we consider a PhC waveguide as a waveguide with periodic rather than cylindrical symmetry, we can profit from this experience. For this purpose, we need to consider (1) the replacement of the cylindrical symmetry by a periodic and (2) the replacement of attenuation due to material loss by attenuation due to radiation loss. The latter has the following reason: When we do not

establish periodic symmetry as in the supercell approach, we must truncate the model (i.e., we must consider a finite width of the PhC walls on both sides of the waveguide channels). As we have seen in the grating section, some energy may propagate through a PhC slab of finite width. As a consequence, some energy will be transmitted through the walls, and this energy will radiate into the space outside the PhC walls. This radiation loss will also cause an attenuation of the waveguide mode (i.e., it will contribute to the imaginary part of the propagation constant). Because any realistic PhC waveguide can only have walls of finite widths, such a model is more realistic than the supercell model and it provides useful information on the attenuation and radiation loss.

We now replace the cylindrical symmetry in z direction by a periodic one in x direction with period a. Instead of computing the field in the transverse plane at $z = 0$ we then must evaluate the field in a section, for example from $x = 0$ to $x = a$. Out side this section, we then obtain the field everywhere else from the periodic condition

$$Field(\vec{r} + na\vec{e}_x) = Field(\vec{r}) \times e^{jk_x na} \tag{19}$$

From this, it is obvious that the constant k_x now plays the role of the propagation constant. As mentioned before, the propagation constant for lossy waveguides becomes a complex function of the frequency, and its imaginary part describes the attenuation. Now, k_x becomes a complex function of the frequency, and its imaginary part describes the attenuation. To compute k_x, we can apply the same eigenvalue solvers as for cylindrical waveguides. More details will be given in the next section.

3.2.1. Defect Waveguides

The most simple waveguide is obtained from a perfect photonic crystal when one or several rows of cells or "molecules" are omitted, as shown in Fig. 27. Depending on the width of such a defect, we can obtain single-mode or multimode waveguides. As for ordinary, cylindrical waveguides, one observes the more modes the broader the waveguide channel is.

Defect waveguides can be obtained for rod-type PhCs as well as for hole-type PhCs. When we consider such structures with an effective index model, the hole-type PhC waveguide has a higher effective index in the channel because the holes are missing there, whereas the rod type has a lower index in the channel. In both cases, we have something like a symmetric slab waveguide, and it is well-known that such a structure can guide electromagnetic waves only when the channel has a higher index. Thus, it is somehow surprising that also the rod-type PhC defect waveguide can guide electromagnetic waves.

3.2.2. Other Waveguides

The degree of freedom in PhC defect waveguides is rather low because one only can omit a finite number of rows of cells of the original PhC. When one wants to obtain single-mode operation, one can usually only omit a single row. Sometimes even a single-row defect waveguide supports more than one mode. The design and composition of waveguides with different characteristics is interesting for the design of filters and filtering junctions. Therefore, there is a strong need for a more general waveguide design that provides more flexibility. One might be tempted to simply generalize the defect waveguide concept by admitting an arbitrary, real channel width w instead of the width na for standard defect waveguides with n missing rows of width a. Although this would be feasible, this has a strong drawback when one wants to have a bend in such a waveguide. In fact, sharp waveguide bends are one of the most attractive features of PhC waveguides as we will see later. For this reason, one should usually not consider PhC waveguides of a channel width different from na.

As mentioned before, the channel of a PhC waveguide can consist of any natural material. Thus, we also could modify the permittivity within the channel instead of the width. This would not cause any essential problems for waveguide bends, but it would cause strong fabrication problems. For this reason, it is certainly best to use a photonic crystal structure also within the waveguide channel, but one with a different geometry, namely one that does

not have a band gap at the same location as the original PhC or one that has no band gap at all.

When we fabricate a PhC waveguide by a PhC channel embedded in another PhC, we can obviously have two different types. The more simple one has the same period a along the channel for both crystals (channel and outside), whereas the second one has different periods a_{in} and a_{out}. Obviously, the former is also much simpler from the analytical and numerical point of view because it does not break the periodic symmetry. For this reason, it is best to focus on this case whenever possible (i.e., as long as the desired waveguide properties may be obtained with such a design). Incidentally, the second case can be handled with a supercell approach in longitudinal direction when a_{in}/a_{out} is rational. Otherwise, its handling becomes very demanding. However, in the following, we only consider the simple case that is periodic along the channel.

When a photonic crystal consisting of circular rods or circular holes is given, the simplest version of a channel waveguide is obtained when we either increase or decrease the radii of the rods or holes. Figure 30 shows that the radius as a single tuning parameter already allows us to shift the propagation constant of a mode rather freely within the PBG of the original crystal. As one can see, one can design waveguides that carry modes over almost the entire band gap or only over some fraction of the band gap.

When we connect PhC waveguides with different characteristics, we obtain frequency-dependent transmission and reflection coefficients exactly as when we connect conventional, cylindrical waveguides. Such waveguide junctions can therefore be used for the design of filters. Waveguide junctions break the symmetry along the waveguide as to the more general class of waveguide discontinuities. This causes additional analytical problems that are outlined in the following.

3.3. Waveguide Discontinuities in Photonic Crystals

In practice, the main purpose of a waveguide is to transfer energy or information from one device to another one, for example from a sender to a receiver, from a signal generator to a filter, from a filter to a power divider, and so on. Although the analytical and numerical models of waveguides usually assume infinitely long structures—for obtaining cylindrical or at least periodic symmetry—all realistic waveguides have a finite length and at their end points one must have a discontinuity that breaks the symmetry. Because devices like filters, modulators, and so forth, are connected through waveguides, one can consider them also as a waveguide discontinuity. In general, each device has at least one waveguide port. When a

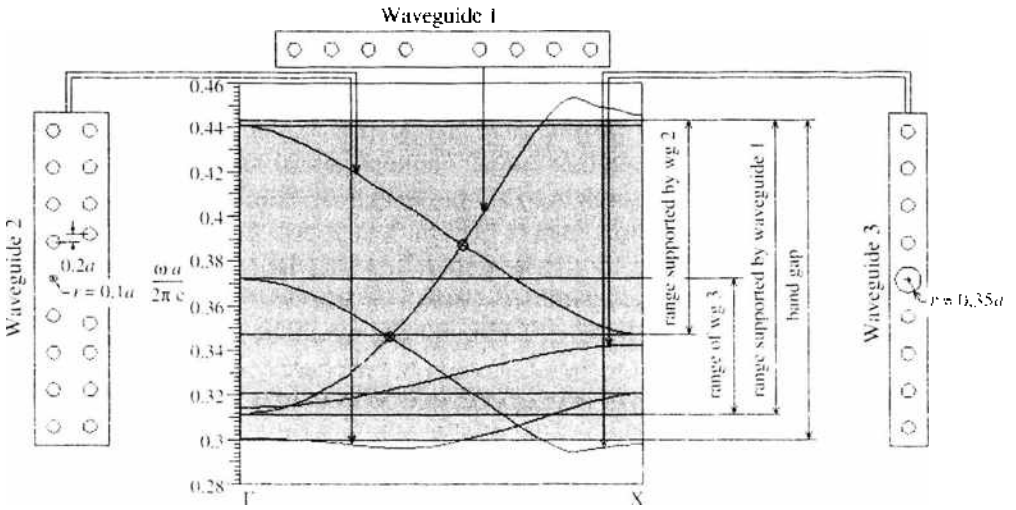


Figure 30. Design of different types of photonic crystal waveguides. Top, vacancy line defect; left: substitutional line defect with smaller rods; right: substitutional line defect with larger rods. Combining appropriate waveguides with optimally tuned characteristics allows us to design photonic crystal filters, filtering T-junctions, and other interesting structures.

device has N ports (Fig. 31), it should obtain energy at least from one port, the input port. The input port then plays the same role as the excitation of a standard scattering problem as, for example, the scattering of a plane wave at an optical lens. Therefore, the mathematical description of waveguide discontinuities is closely related to the the mathematical description of scattering problems.

Because one has much experience in the description of waveguide discontinuities at radio frequencies in electrical engineering, it is reasonable to take advantage of this knowledge. First, we consider an arbitrary field that is impressed in the cross section of a loss-free cylindrical wave guide. According to mode matching theory [28], this field can be expanded as follows:

$$Field(\vec{r}_T) = \sum_{m=1}^{M} A_m Field_m^{guided}(\vec{r}_T) + \sum_{m=M+1}^{\infty} A_m Field_m^{evanescent}(\vec{r}_T) + \int_s a(s) Field^{radiating}(s, \vec{r}_T)\, ds \tag{20}$$

that is, one has a finite set of guided modes, an infinite set of evanescent modes, and a continuous spectrum of radiating modes. Because each evanescent mode becomes guided at sufficiently high frequencies, it is reasonable to number both types of modes with the same index. From this, it is also evident that mode matching theory works in the frequency domain [i.e., Eq. (20) only holds at a certain frequency]. Now, a scalar product can be defined as an integral over some field components over the cross section of the waveguides, and it can be shown that all modes are orthogonal. Here, it should be mentioned that the definition of the scalar product is not unique, and that different definitions are preferred depending on the type of the waveguide. However, the scalar product can be used to normalize the fields of all modes. Because of the orthogonality, it is possible to compute the amplitudes A_m of the guided and evanescent modes without an explicit computation of the radiating modes. Because the evanescent modes are characterized by an imaginary propagation constant, they do not contribute to the energy transfer along the waveguide. Furthermore, waveguides and waveguide discontinuities are often designed in such a way that radiation is minimized [i.e., the most cumbersome third term in Eq. (20) can be neglected]. However, the modal expansion Eq. (20) allows us to handle all waveguide discontinuities.

It has been mentioned that the distinction between guided and evanescent modes becomes difficult when losses are present. In this case, one can account for both types of modes in a single, infinite sum in the modal expansion [i.e., Eq. (20) becomes formally even simpler].

In practice, the infinite sum must be truncated. Because the modes are usually ordered in such a way that higher order modes have bigger attenuation constants, this poses no severe problems, even in the lossy case. However, a previous computation of the propagation constant and field of all modes of order $m < m_{big}$ is required for the proper truncation, where m_{big} is so high that the corresponding mode and all higher order modes are dampened so rapidly that they can be neglected. Note that this condition depends on the desired accuracy and on the length of the considered waveguide. The shorter the waveguide, is the bigger m_{big} must be. Because of the frequency dependence of the modes, m_{big} must also be increased with increasing frequency.

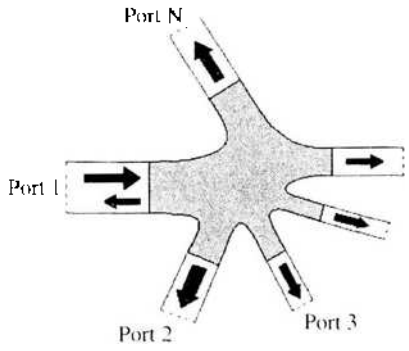


Figure 31. Analysis of general waveguide discontinuities. A structure with N waveguide ports is depicted (Port 1 is the input port and all other ports are pure outputs).

Obviously, the integral over the radiating modes is the most cumbersome term in the modal expansion. Usually, it is not possible to evaluate the field of all radiating modes. Therefore, mode matching often is restricted to nonradiating cases. Although waveguide discontinuities are often designed in such a way that radiation is negligible, radiation is essential for two special cases: (1) when energy of a waveguide shall be radiated and (2) when radiation shall excite a waveguide mode. The first case is a radiating, active antenna, and the second case is a receiving, passive antenna. Because of the reciprocity theorem, both cases are closely related. Therefore, it is sufficient to consider explicitly only one case, for example, the radiating antenna. Here, the waveguide carries usually a single, guided mode that feeds the antenna structure that is designed in such a way that most of the energy of the mode is radiated. The radiated waves can then be expanded, for example, by a multipole series [i.e., the third term in Eq. (20) can be replaced by a more simple multipole approximation].

For a junction of two different waveguides, one can now take advantage of the modal expansion. When both waveguides are linear, one can assume that only one of the guided modes is incident on the junction. This mode plays the role of the incident wave in a scattering problem. It excites a set of guided, evanescent, and radiating waves for both waveguides. When the discontinuity is abrupt, one only has a single discontinuity plane, and one then can expand the known field of the incident mode in this plane using a superposition of the modal expansions of both waveguides. From the scalar products, one then directly obtains the amplitudes of all modes of interest. In most cases, only guided modes are of interest because only these modes transmit energy along the waveguides. Furthermore, one often considers single-mode propagation. In this case, the situation becomes most simple (i.e., a part of the incoming energy is reflected and a part is transmitted). The first part is then described by the reflection coefficient R and the second part by the transmission coefficient T. When no radiation is present, the sum $R + T$ should be equal to one because of the energy conservation law. Otherwise, $1 - R - T$ is equal to the total radiated power. When M_1 propagating modes exist on the waveguide with the incident mode and M_2 propagating modes exist on the second waveguide, one obtains M_1 reflection coefficients R_m and M_2 transmission coefficients T_m. Besides this, the formalism remains essentially the same. This is the fundamental procedure of the mode matching technique.

In practice, one often uses tapered waveguide transitions instead of abrupt ones in order to minimize undesired reflections. Mode matching technique then replaces the tapered section in a sequence of sufficiently short cylindrical sections with abrupt transitions that may be handled as before (Fig. 32). One then obtains a system of N waveguide junctions. Now, the evanescent and radiating modes can no longer be neglected because the waveguide sections are short. Obviously, the radiating terms are much more cumbersome here. For this reason, mode matching technique is usually restricted to nonradiating junctions. Then, one also defines a set of reflection and transmission coefficients for the lowest order evanescent modes that cannot be neglected. As a result, one obtains M_n important modes for each junction. Each of these modes acts on the neighbor junction. Thus, one needs to know for each mode on a waveguide what reflected and transmitted modes it excites at the discontinuity. This is usually described by the so-called scattering matrix, or S matrix, that has M_n rows and columns. The element s_{ik} of S indicates how big the amplitude of the mode number k of the junction is when the mode number i is incident with an amplitude 1. When all S matrices of all junctions are known, the entire system may be computed (i.e., the S matrix of the entire tapering section may be obtained).

Figure 32. Waveguide transition obtained by tapering. As one can see, the mode matching approach replaces the continuous tapering by a finite set of abrupt transitions.

The mode matching technique can also be applied to more general waveguide discontinuities with more than two ports under certain restrictions that are caused by the modal expansion that is performed only in a cross-section of an abrupt waveguide transition. One then obtains a big S matrix that describes the transmission of energy from any mode in any port to any mode in any port. Because this procedure is not only very general but also tedious and approximative (higher order modes and radiating modes are neglected), one can replace it by a numerical technique that only matches the important modes at the N ports as illustrated in Fig. 33 and approximates the field in the discontinuity by any numerical field approximation exactly as an arbitrary scattering problem. One then can use any Maxwell solver to compute directly the S matrix of the entire N port waveguide discontinuity. Mode matching is then only required at each waveguide port. When the length of each waveguide (i.e., the distance from the neighbor discontinuity) is big enough, one can define each port (i.e., each reference plane at such a distance from the discontinuity that all evanescent modes can be neglected), which further simplifies the mode matching and reduces the size of the S matrix.

The application of mode matching and S matrix formalisms to PhC waveguides is straightforward. The only speciality to be considered is the fact that the PhC waveguides are periodic rather than cylindrical. As we have seen, the field of a PhC waveguide is there entirely described when it is given within a section of length a, where a is the period along the waveguide. Thus, mode matching can no longer be defined in a transverse reference plane. It should be defined in a reference section of length a instead. This makes the definition and numerical evaluation of the scalar products for the modal expansion more complicated. Fortunately, the field pattern of the important guided modes in a PhC waveguide varies often not much along the axis of the waveguide. Therefore, one can also obtain a reasonable approximate solution when one restricts the mode matching on a reference plane as in the case of cylindrical waveguides.

In the following, we outline several simple cases of PhC waveguide discontinuity in order to illustrate the procedure and to show the variety of structures that may be considered as waveguide discontinuities.

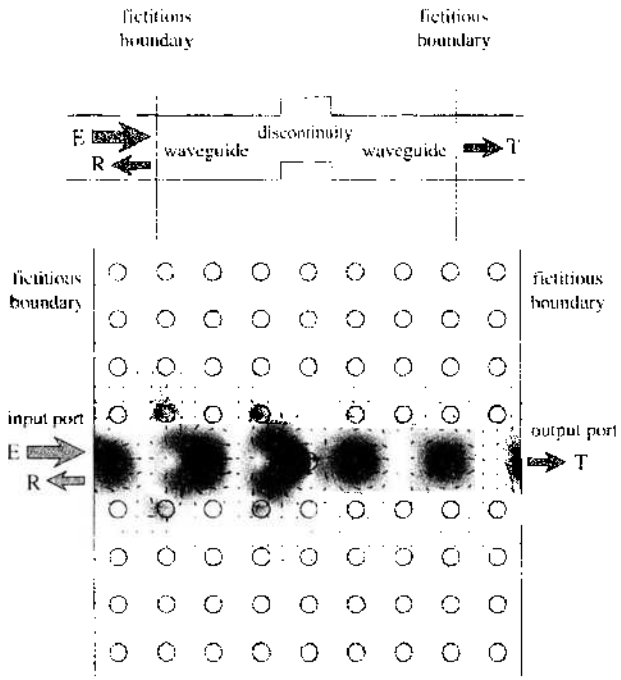


Figure 33. Numerical treatment of a simple PhC waveguide discontinuity. Fictitious boundaries are introduced in order to be able to match the eigenfield of all important modes in the two waveguide ports with the field in the discontinuity area. Top: schematic representation; bottom: plot of the Poynting vector field.

3.3.1. Coupling Energy into Photonic Crystals

The first practical problem of PhC waveguides is the need for structures that allow one to efficiently excite any PhC waveguide mode (i.e., to couple energy into a PhC waveguide). This is rather demanding from the practical point of view because the PhC waveguides are typically very thin (i.e., of the order less than a wavelength).

There are two essentially different cases for doing this: (1) One can design some receiving antenna structure that collects energy from a not-guided electromagnetic energy and transfers it to the PhC waveguide. (2) One can design a junction of a traditional waveguide with a PhC waveguide.

In the first case, one has a waveguide discontinuity with a single waveguide port, whereas one has two waveguide ports in the second case. Obviously, radiating waves are important in the first case, whereas one tries to suppress any radiation in the second case. Despite these fundamental differences, we can use one and the same formalism (outlined above) for handling both cases.

Figure 34 shows the most simple case of an abruptly terminated PhC waveguide illuminated by a plane wave. As one can see, even this simple structure acts as an antenna that transfers energy into the PhC waveguide.

We can easily modify the properties of a receiving PhC antenna with the goal to couple energy more efficiently into the PhC waveguide. For doing this, one has a huge degree of freedom because one can use arbitrary elements with arbitrary material properties, shape, and locations near the ending of the PhC waveguide. Another simple example is shown in Fig. 35.

In order to illustrate the coupling of a conventional waveguide with a PhC waveguide, Fig. 36 shows another simple 2D configuration of a slab waveguide in front of the same PhC waveguide as in Fig. 27. As one can see, the energy is transferred rather well, but some radiation energy is lost at this junction. In order to reduce the radiation loss, some tapering section might be used for the transition of the two waveguides. This would be a natural procedure, but the high degree of freedom in the design of PhC junctions allows one to improve considerably the performance by optimizing the locations and radii of the PhC rods near the end of the photonic crystal. As one can see from Fig. 37, even the optimization of a single rod on each side of the PhC channel allows us to reduce drastically the radiation loss and reflection. Similar optimizations may also be used for the design of PhC antennas and all other PhC waveguide discontinuities. We will consider such optimizations in more detail at the end of this chapter.

3.3.2. Coupling Energy Out of Photonic Crystals

Because of the reciprocity theorem, coupling energy out of a PhC waveguide is not much different from coupling energy into a PhC waveguide. We also have the two categories of a (radiating) PhC antenna and of a transition to a conventional waveguide.

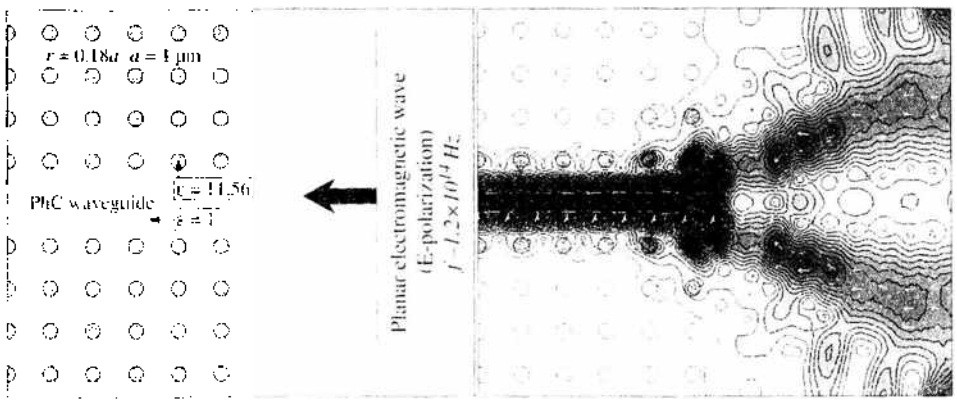


Figure 34. Photonic crystal waveguide termination illuminated by a planar electromagnetic wave incident from the right-hand side. Waveguide geometry, material properties (left) and time average of the Poynting vector field (right). As one can see, this simple structure behaves like a receiving antenna.

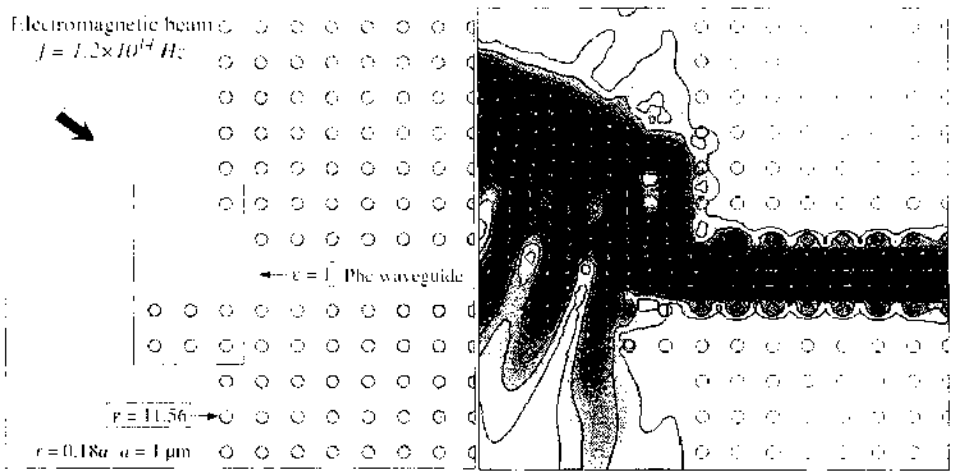


Figure 35. Modification of the photonic crystal waveguide termination shown in Fig. 34 for obtaining a different antenna pattern (directivity). Geometry and material properties (left) and time average of the Poynting vector field (right).

An initial design of a photonic crystal waveguide termination (antenna) is presented in Fig. 38. Obviously, this waveguide termination is not so bad when we look at the frequency characteristics and returned losses (reflection) are acceptable. However, this design can be optimized. Initially, we have decided that the optimization variables are a position and radii of the last rod of waveguide, as it can be seen in Fig. 38. The fitness function is defined as an integral of the reflection over the bandgap frequency range, and a Newton-like (gradient) optimizer was applied. The optimization result is presented in Fig. 39. As one can see, our optimization was actually very efficient and we obtained a high-quality photonic crystal antenna (reflection is negligible over an entire bandgap).

Light coupling out of a photonic crystal waveguide is also an interesting problem. An initial design of this coupler is presented in Fig. 40 together with the frequency characteristics. Obviously, this design suffers significant radiation (50%) and poor transmission (<50%) and it has to be optimized. Applying the same algorithm as in the case of the previously

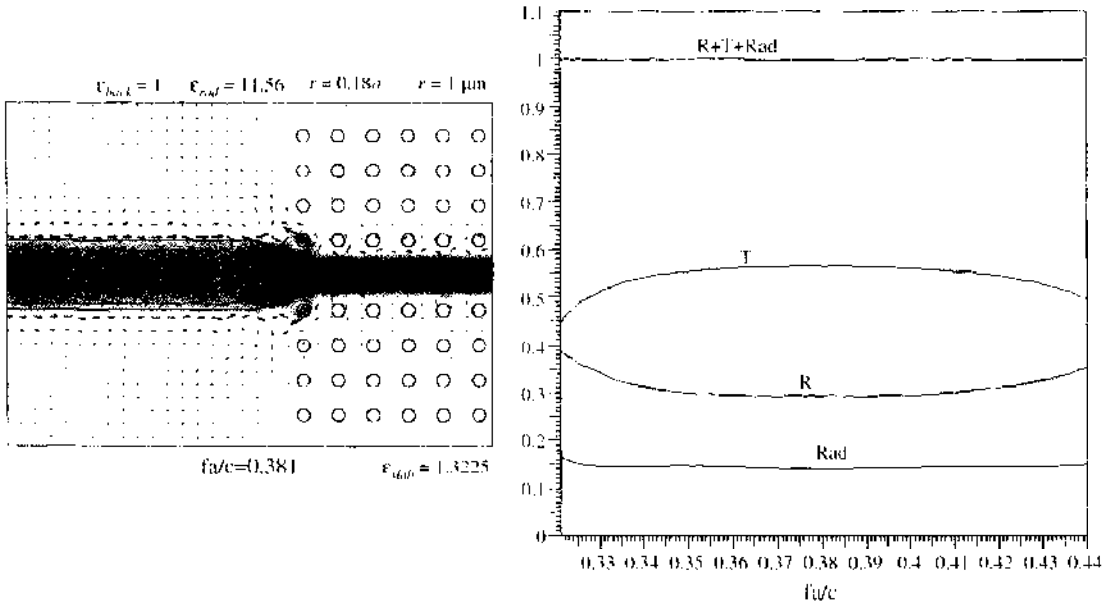


Figure 36. Coupling of a dielectric slab waveguide and a defect waveguide in a 2D PhC. Left: geometry, material properties, and time average of the Poynting vector field; right: frequency characteristics (reflection coefficient R, transmission coefficient T, and total radiation Rad) of the coupler. As one can see, this design exhibits considerably strong radiation and reflection.

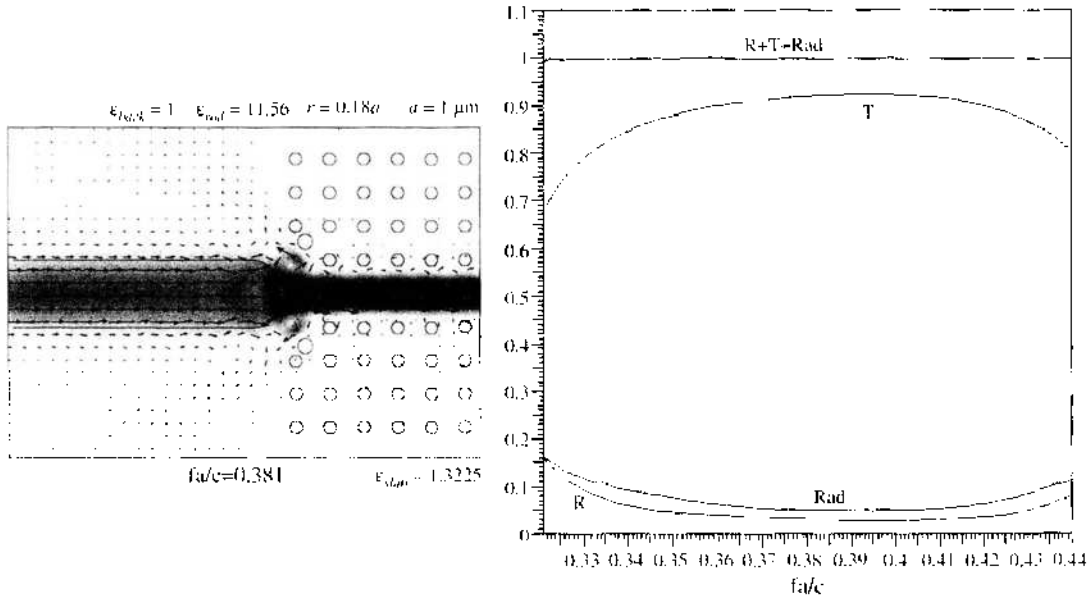


Figure 37. Improved optical coupler. Positions and radii of the two rods of the photonic crystal closest to the slab waveguide were optimized. As one can see, the reflection and the radiation losses were significantly reduced over the entire frequency range (i.e., the bandgap of the original PhC).

mentioned photonic crystal antenna, we have obtained significant improvement, as presented in Fig. 41. After these two examples, it seems to be very useful to apply the gradient-based optimizer in the case of photonic crystal structures. It is also important to see that gradient-based optimizer gives us a good result only in the case when we have smooth character of fitness function, and obviously the photonic crystals are such a case.

3.3.3. Waveguide Bending

We have already mentioned that sharp bends in PhC waveguides may be obtained without any radiation loss—as long as one operates within the band gap and provided that the walls of the PhC waveguide extend to infinity. Theoretically, a waveguide bend is nothing else than a waveguide discontinuity with two ports of identical waveguides. When radiation is present,

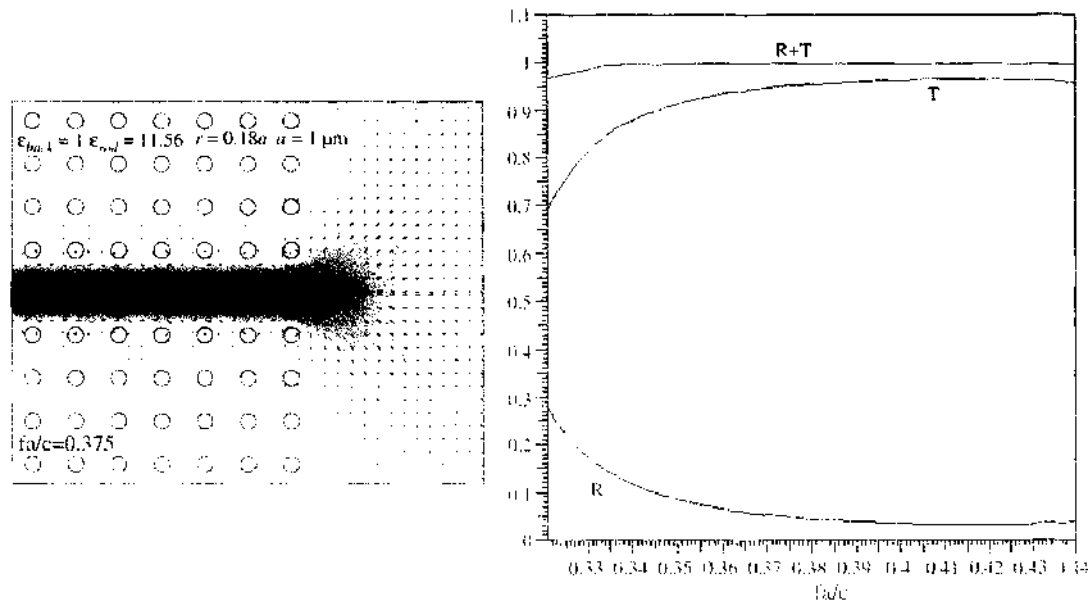


Figure 38. Simple open photonic crystal waveguide (photonic crystal antenna). Left: geometry, material properties, and time average of the Poynting vector field; right: frequency characteristics.

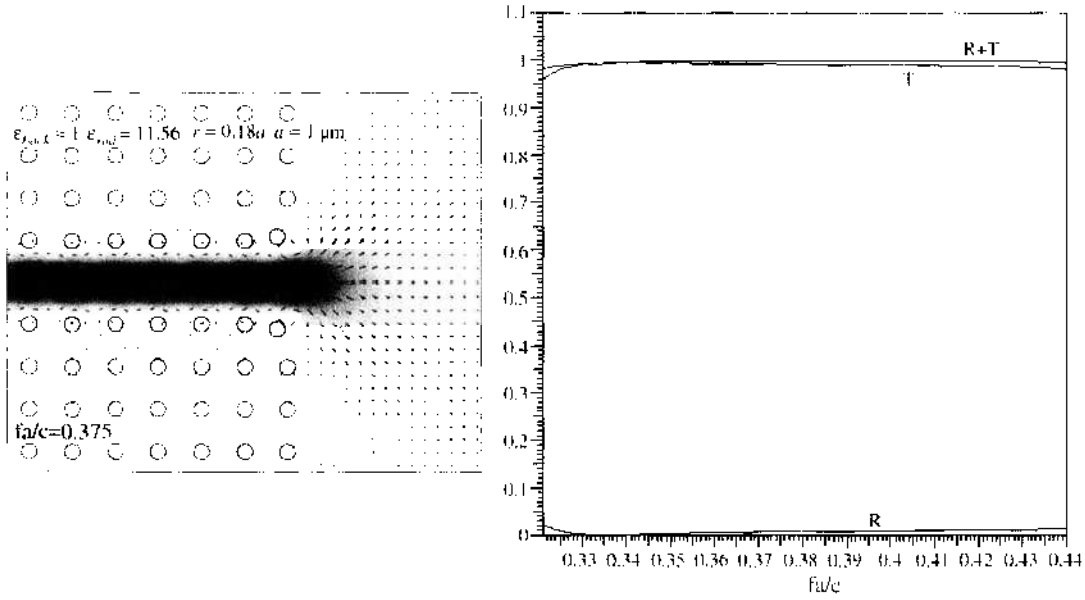


Figure 39. Optimized design of the open photonic crystal waveguide. The initial design from Fig. 38 was optimized using the gradient approach. Only two rods were optimized. Left: geometry, material properties, and time average of the Poynting vector field; right: frequency characteristics.

the description is more demanding than when radiation may be neglected. Therefore, the treatment of a PhC waveguide bend is simpler than the treatment of a conventional waveguide bend. A special problem is caused in both cases because the planes of the two ports intersect each other at some distance d from the waveguides. This causes a field region where modes of both waveguide ports are assumed to be present. Such regions are not assumed to be present in the mode matching theory. When d is big enough, the field of all modes in this region is small enough that it can be neglected. Fortunately, d can be increased to any desired value by moving the two ports away from the bend area. Note that this problem is more pronounced when strong radiation is present or when some waveguide modes extend far into the PhC walls, which mainly is the case when the PhC is operated near the end of the band gap.

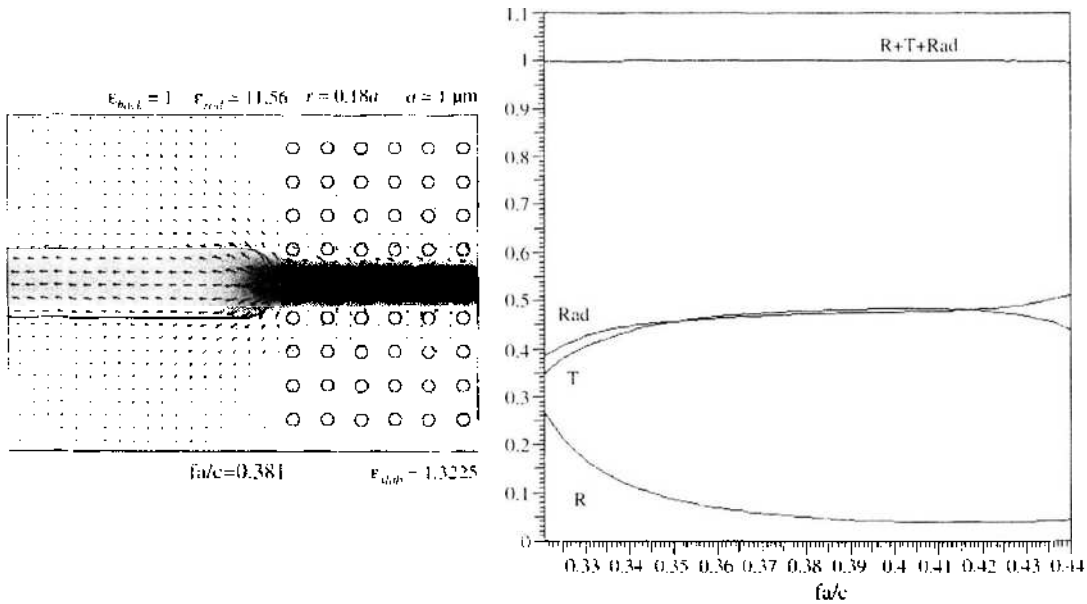


Figure 40. Light coupling out of a photonic crystal waveguide into a dielectric slab waveguide (same structure as in Fig. 36, but opposite coupling direction). Obviously, this initial design exhibits large radiation and low transmission.

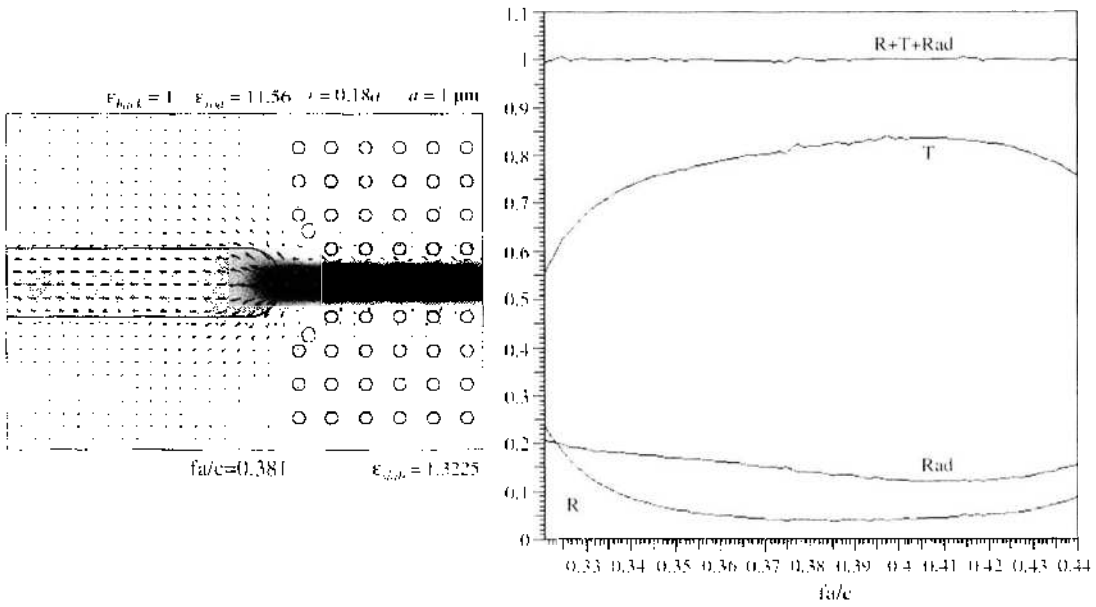


Figure 41. Optimized design of the coupler presented in Fig. 40. Optimization variables are the position of the planar dielectric waveguide and the position and radii of the two rods of photonic crystal waveguide near the slab. A gradient-based optimizer was applied once more. As one can see, the transmission is significantly improved, and the radiation losses are reduced. For better characteristics, additional rods or the shape of the slab termination should be optimized.

As one can see from Fig. 42, strong radiation is observed near the bend of a conventional slab waveguide, whereas a sharp PhC waveguide bend exhibits no radiation when its walls are thick enough. As one can see from Fig. 43, even a rudimentary PhC waveguide with a single layer of rods on each side of the channel can be bent in such a way that much less radiation loss is obtained as for the slab waveguide.

Because the transition from a conventional waveguide to a PhC waveguide is not trivial, one can also suppress radiation at a conventional waveguide bend by inserting a small PhC piece as shown in Fig. 44.

Although a sharp PhC waveguide bend exhibits no radiation loss, it is not perfect in general because a part of the energy is reflected. It has been shown in Ref. [29] that the reflection coefficient can be zero for certain frequencies that depend on the geometry of the PhC near the bend. As in the design of PhC waveguide couplers, one can optimize the bend

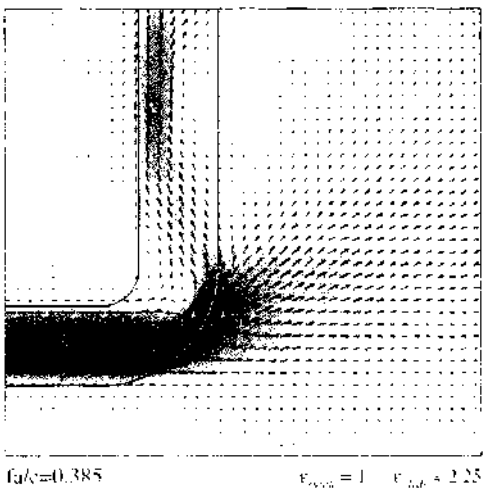


Figure 42. Classical 90-degree bend in a dielectric slab waveguide with two guided modes. The time average of the Poynting vector field is shown. This bend exhibits a very strong radiation loss (Rad = 69% of the input power). The reflection (R = 1% of input power) is not significant.

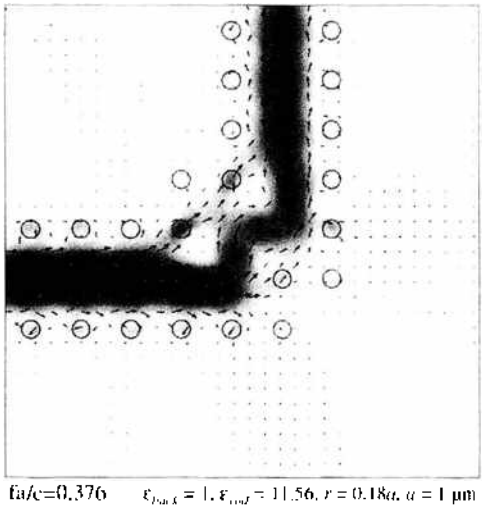


Figure 43. Time average of the Poynting vector field for a "single layer photonic crystal waveguide." At frequencies close to the upper limit of the bandgap, even this primitive PhC waveguide is better than the slab waveguide presented in Fig. 42. The reflection coefficient (R ≈ 7%) is bigger, but the radiation loss is considerably smaller (Rad ≈ 33%).

area in such a way that even almost zero reflection is obtained over the entire band gap, that is, that an achromatic waveguide bend is obtained [30]. This is illustrated in Fig. 45.

3.3.4. Filters

We have already mentioned that the combination of waveguides with different characteristics of the guided modes may be used for constructing filters. Such a filter consists of at least one section of a different waveguide type inserted in a given PhC waveguide. One therefore has at least two waveguide discontinuities at the two ends of the inserted section. What is important for such a design is that one should match the propagation constant and its derivative with respect to the frequency—this derivative corresponds to the group velocity—as well as possible at the frequencies where one wants to have low reflection at the junction. We will consider such filtering sections when we consider more advanced structures.

Alternative filters along PhC waveguides are obtained when defects are either introduced in the waveguide channel or in the walls of the channel. As one can see from Figs. 46 and Fig. 47, this allows one to obtain simple band-pass and band-stop filters. High-pass and

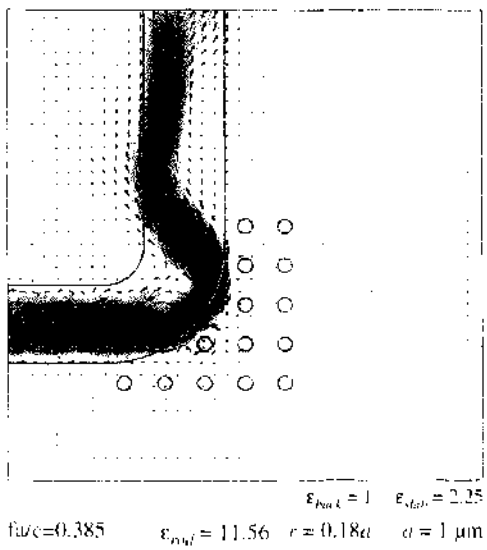


Figure 44. Slab waveguide bend improved by small PhC patch near the bend. Much lower radiation (Rad ≈ 2% of input power) and acceptable reflection (R ≈ 6% of input power) is obtained.

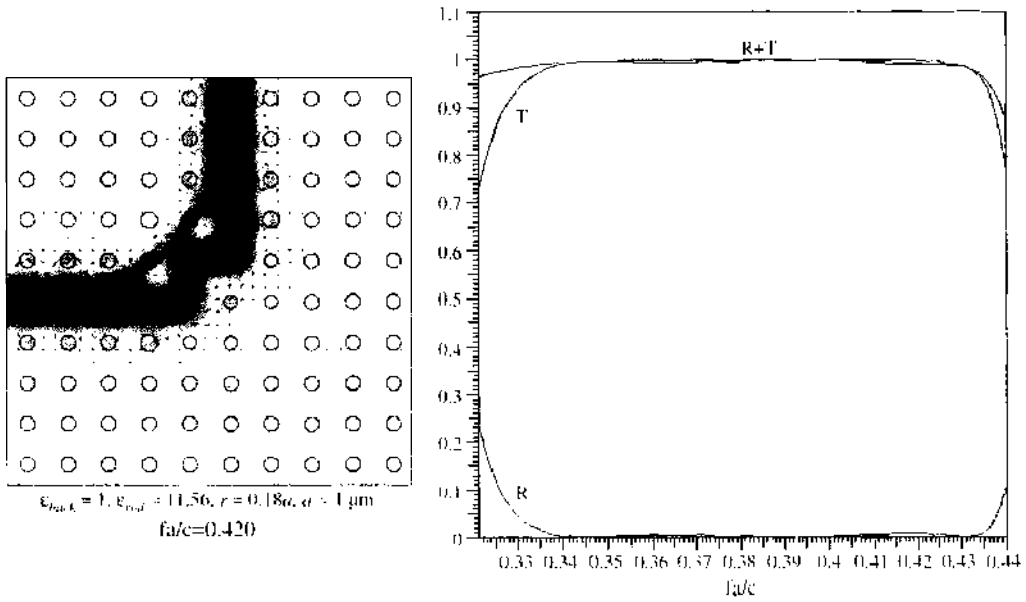


Figure 45. Achromatic 90-degree bend in a PhC defect waveguide. This design is a result of a gradient optimization of the rod radii in the bend area. Details can be found in Ref. [30].

low-pass filters make not too much sense because the band gap limits the area where a PhC waveguide is operated. The main concept here is that the defects create PhC resonators that are coupled to the waveguide. When more than one resonator is coupled, as in Fig. 48, the different resonators also couple with each other because the PhC walls that separate the resonators are finite. When this filter design only uses simple defects, one only can obtain resonators with specific frequencies and a finite number of coupling coefficients between the resonators and between each resonator and the waveguide. For a reasonable filter design (i.e., for filter synthesis), a much higher degree of freedom is required. For this reason, it is necessary to consider also more sophisticated modifications of the photonic crystal, namely modifications of the position, size, and shape of the cells in the filter area. This allows one to tune the resonance frequencies as well as the coupling coefficients.

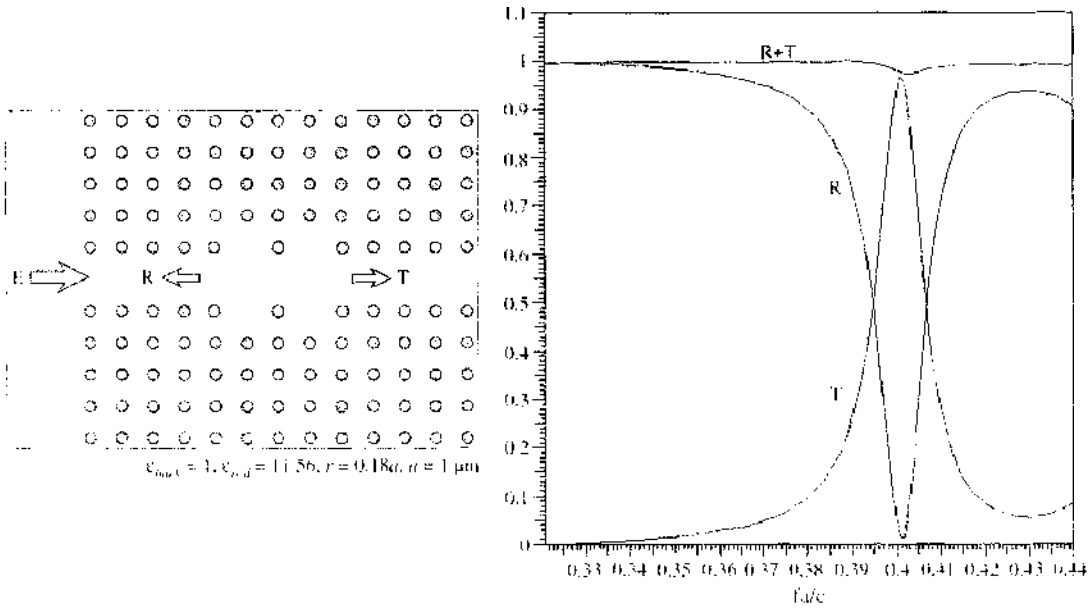


Figure 46. Simple bandpass photonic crystal filter consisting of two strongly coupled cavities inserted in the waveguide channel. Left: geometry and material properties; right: frequency characteristics (reflection and transmission coefficients R and T respectively).

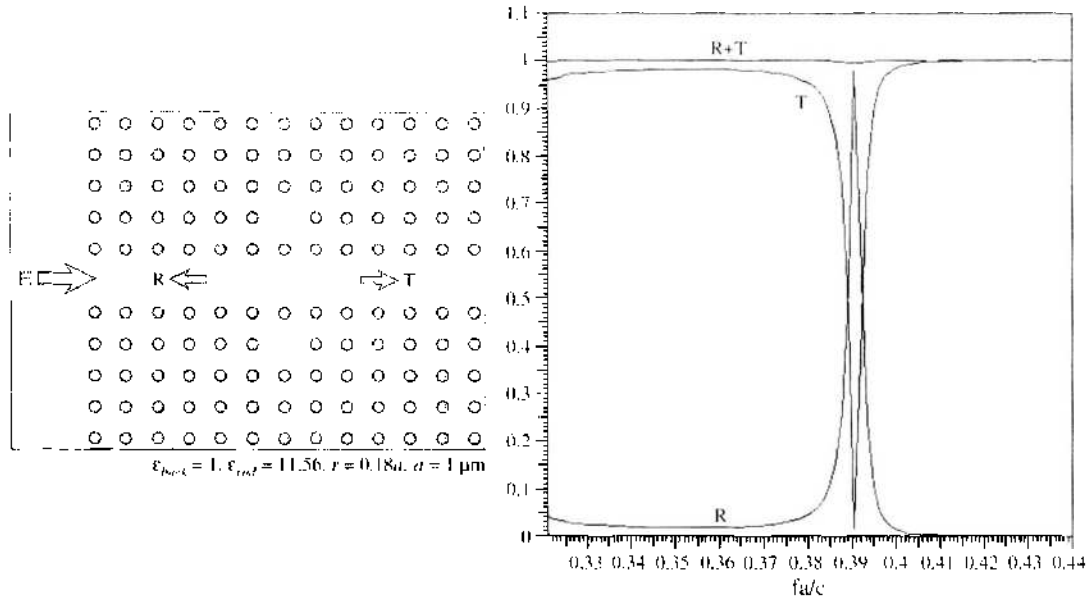


Figure 47. Simple band-stop photonic crystal filter consisting of two symmetric cavities coupled to a defect waveguide. Left: geometry and material properties; right: frequency characteristics.

For the synthesis of RLC filters in electrical engineering, one has many simple design rules. For this reason, more sophisticated microwave filters are also usually designed with equivalent RLC networks. In the case of photonic crystal filters, the situation is much more complicated because of the coupling of all parts involved. Currently, one therefore has no simple design rules. Therefore, the only way of efficient PhC filter synthesis is numerical optimization, which will be discussed later.

3.3.5. Power Dividers

A power divider is a structure that distributes the power of a mode in an input channel over two or more output ports. Thus, the total number of ports of such structures is higher than for the structures considered above. This does not cause essential new problems.

As an illustration, we consider a simple power divider with two output ports. When we want that equal amounts of energy are transmitted to both output ports, it is reasonable to design a symmetric structure. The most primitive design is a T-shaped structure as shown in Fig. 49. Although the field at a frequency near the center of the band gap looks nice, the reflection coefficient at the input port is not small enough from the engineering point of view. As one can see from Fig. 49, the reflection coefficient is even higher near the ends of the band gap.

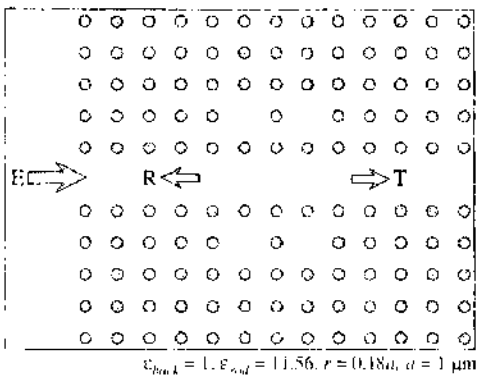


Figure 48. Photonic crystal filter consisting of four coupled cavity resonators.

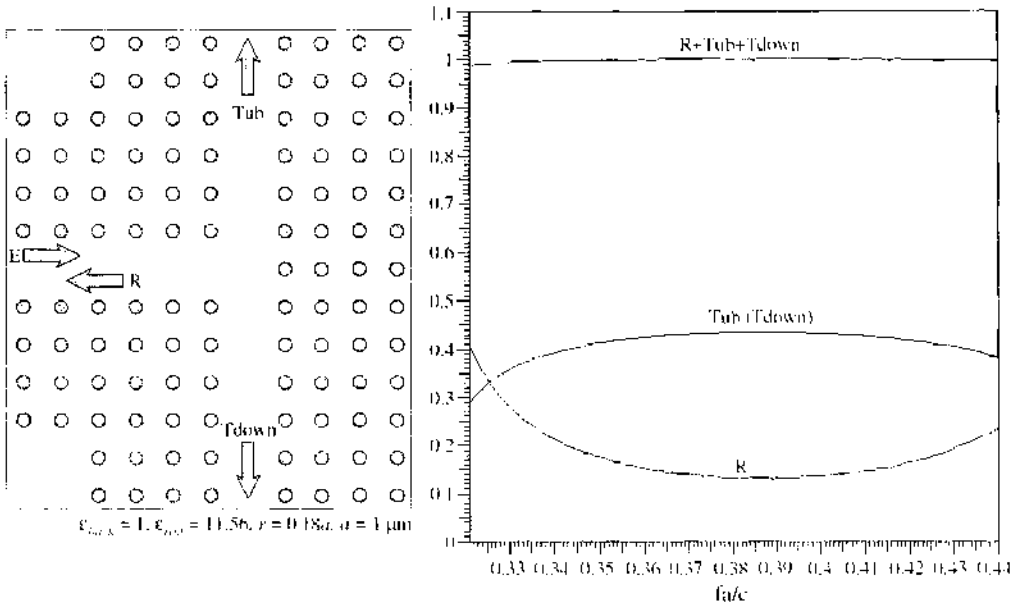


Figure 49. Initial, "intuitive" design of a photonic crystal power divider. Left: geometry and material properties; right: frequency characteristics. For an optimized design, see Figs. 61–63.

An intuitive idea—motivated by the design of microwave power dividers—is to insert a little "mirror" or to insert a small Y-shaped part as in Fig. 50. As one can see from Fig. 50, this even increases the reflection coefficient everywhere in the band gap frequency range. Anyhow, this demonstrates that intuition and experience from non-PhC design may be misleading. The fact that intuitive design often leads to suboptimal design is another reason for numerical optimization. In the section on optimization, we will consider the power divider in more detail, and we will find much better solutions than the "intuitive" ones considered here.

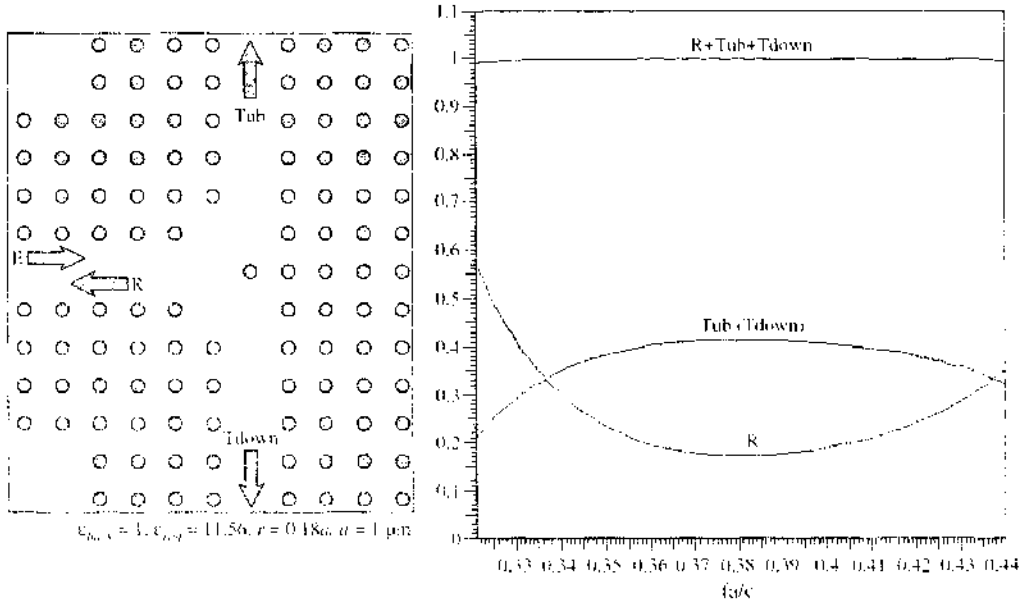


Figure 50. Alternative "intuitive" design of a photonic crystal power divider with a 45-degree "mirror" in the junction area. Left: geometry and material properties; right: frequency characteristics. For an optimized design, see Figs. 61–63.

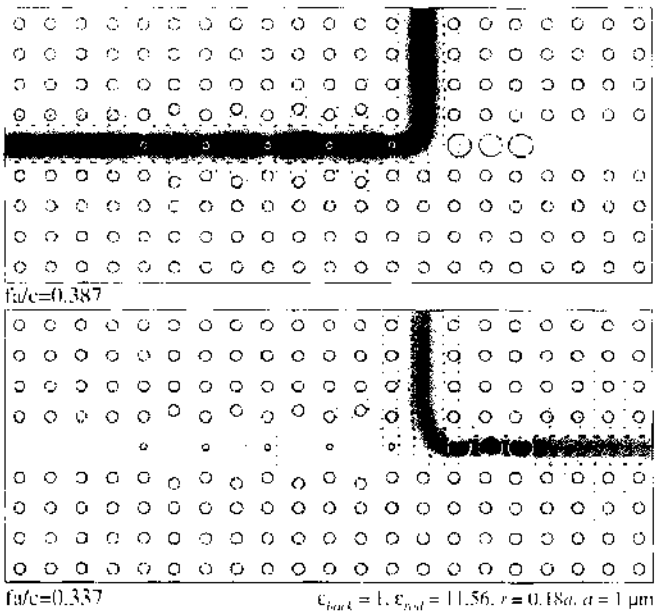


Figure 51. Photonic crystal filtering T-junction. For a certain frequency range, we obtain left propagation (top), and for another frequency range, we obtain right propagation (bottom).

3.3.6. More Advanced Structures

More advanced PhC discontinuities may be obtained from combinations of the structures considered up to now. A simple example is the filtering T junction shown in Fig. 51. This structure is a combination of a power divider with two different filter sections in the two output ports. Because power dividers and filters require optimization for obtaining a good design, we will also consider the filtering T junction in the optimization section.

Beside more advanced structures obtained from combinations of simple waveguide discontinuities, we can also obtain waveguide discontinuities with higher functionality by inserting other materials in a photonic crystal. Of course, nonlinear and active materials are very attractive in this context because such materials allow us to fabricate modulators, lasers, amplifiers, and switches within a photonic crystals (i.e., devices that might play an essential role in future PhC design). Because we currently have not enough experience in the treatment of such structures, we cannot provide any details here.

4. SIMULATION OF COMPOSITE DOPED METAMATERIALS

Composite doped metamaterials (CDMs) are a natural generalization of photonic crystals. They exhibit very rich and often complicated, sometimes even counterintuitive effects. Therefore, both analytical solutions and the intuitive design of such structures are very much limited as we have seen in the previous section. For obtaining experience and for finding devices that meet the strong requirements of engineers, one therefore needs extended numerical simulations and optimizations.

Currently, many numerical methods for the analysis of PhC structures are available and frequently used. Many of these methods provide a good insight to the mechanisms of photonic crystals and can also be extended for the simulation of more CDM structures.

We have already seen that the intuitive design of PhC structures often leads to subobtimal solutions and that numerical optimization is therefore highly desirable and promising. This is even more important for the more general CDM case. When CDM structures shall be optimized, one theoretically can use any field solver that can evaluate all models that will be generated by the numerical optimizer. Thus, all available CDM simulation tools might be linked with an optimizer. Although this is true in principle, it may lead to cumbersome, tedious, and inefficient procedures. First of all, an optimizer might generate rather crazy designs. When such designs let the field solver crash, the optimization process will be stopped, which requires undesired interventions of the user. In order to avoid this, the field

solver should be extremely robust (i.e., it should never crash). A second problem is that an optimizer may easily generate thousands or millions of different structures. In order to keep the computation time reasonably short, the field solver should therefore also be fast and efficient. Fast results are usually obtained at the cost of lower accuracy. As a result, the results become "noisy" and this can considerably disturb the optimizer. Therefore, it is not good advice to use the fastest field solver. In fact, many structures of composite doped metamaterials exhibit extremely sensitive parts that require a rather high accuracy. For this reason, the optimizer might miss the best solutions when the field solver is not accurate enough. All in all, the field solver and the optimizer should be adapted to each other and also to the problem to be solved. In the following, we will compare the most important field solvers. After this, we outline the main methods and procedures for handling the three main tasks in the design of CDM devices: band diagram computations, waveguides, and waveguide discontinuities.

4.1. Selection of Numerical Methods

In principle, two different classes of field solvers are available. The first, more widely used class contains domain methods that require the discretization of the entire domain where the electromagnetic field exists. This includes the most prominent finite differences (FD) [31, 32], finite integrals (FI) [33], finite volume (FV) [34, 35], and finite elements (FE) [36, 37] techniques. Boundary methods such as boundary elements (BE) [38], and generalized multipole techniques (GMT) [39, 40] are very promising for PhC and CDM structures but less widely used and therefore also not very well-known. In the following, we outline the main features of these methods with a special focus on their applicability to PhC and CDM analysis.

4.1.1. Domain Methods

Domain methods require the discretization of the entire space occupied by a PhC or a CDM. Because this space is infinite in most of the idealized models, the discretization must be truncated to a finite size. For doing this, special absorbing boundary conditions (ABCs) are required. Although special problems with ABCs may be encountered when infinite waveguides must be truncated, and although the implementation of good ABCs is often more difficult than the implementation of the core of domain methods, powerful ABCs are available. Despite this, the truncation of the model may be crucial. For example, when we consider a PhC waveguide discontinuity, we would like to truncate the model as close to the discontinuity as possible. In order to avoid strong inaccuracies near the truncation of the waveguide ports, some appropriate mode matching should be introduced. This is not easy for domain methods, and many of the available codes do not provide features for matching modes of PhC waveguides. Usually, these codes do not even provide any feature for solving the eigenvalue problem associated with the waveguide modes. In order to circumvent mode matching at the waveguide ports, one usually introduces sufficiently long waveguide sections (i.e., one moves the truncation lines at rather big distances from the discontinuity). One then can set a fictitious point source at the end of the input channel. This source will not only excite the desired input mode but also undesired modes. In the simple case of single-mode waveguides, the undesired modes are evanescent and will be exponentially dampened with the distance from the source. Thus, these modes can be neglected at the discontinuity – provided that the distance from the truncation line is big enough. Obviously, moving the truncation line away from the discontinuity is not sufficient when multimode waveguides are considered. For this case, additional techniques must be developed and implemented. Furthermore, moving the truncation line away makes the discretization large and this drastically increases the memory requirement and the computation time. As a result, the simulation becomes inefficient. This can only be tolerated for the analysis of a few special structures, but it should be strictly avoided when one intends to optimize the structure.

Domain methods always lead to large, sparse matrices. In FD and FI codes, one usually uses a uniform grid for the domain discretization that allows one to work without the explicit storage of the system matrix. Although this leads to simple codes, it also leads to

unacceptable inaccuracies when small features should be taken into account. According to our experience, this is often the case within numerical optimization. FD and FI codes usually provide additional procedures such as subgridding techniques to refine locally the discretization, but these techniques are still subject to research and improvement. FD and FI are known to be very closely related and lead to practically identical codes. The main promise of FI was that it is easier to develop algorithms for irregular grids with the FI formalism. Although this is correct, no efficient FI solver for really irregular discretizations is available. The main reason for this is the fact that second-order schemes are obtained almost for free on regular grids. As soon as irregular grids are introduced, the corresponding algorithms obtain first-order accuracy only.

In contrast to FD and FI, FV and FE use irregular meshes that usually are based on geometrical simplexes (i.e., triangles in 2D and tetrahedral in 3D). As a consequence, small details may be approximated rather well. At the same time, often first-order schemes are used for reasons of simplicity. For this reason, it is often difficult to know what methods provide the most accurate solution for a given problem on a computer with limited memory.

The large, sparse matrices of domain methods are most efficiently solved iteratively. For this reason, it is natural to develop time-domain versions in such a way that one iteration per time-step is used. In fact, all prominent FD, FI, FV implementations work in the time domain, whereas prominent FE codes are available both in time and frequency domains. We have seen that several useful techniques for the analysis of PhCs and CDMs are formulated only in the frequency domain. This holds for the treatment of periodic symmetries when the field has not the same periodicity as the geometry as for gratings and PhCs, for the computation of eigenfrequencies, for the mode matching technique, and finally for material properties with strong frequency dependence. Time-domain code users often circumvent these problems by brute-force models. For example, to obtain a PhC waveguide mode, one simulates a "sufficiently long" section of the waveguide instead of the single cell section required in frequency-domain codes. Similarly, one considers the scattering at a "sufficiently big" PhC block instead of modeling a single PhC cell required in frequency-domain codes. As a result, such brute-force models are neither efficient nor accurate enough for numerical optimization. In consideration of the fact that CDM structures exhibit many features that are only well studied in frequency domain, the dominance of time-domain codes is rather surprising. The main reasons for this dominance are the relatively simple implementation and the success of these codes in radar and radio frequency applications. When one considers the research activities in experimental verification of negative index materials and effects, one can see that many of the researchers in this area come from radar and radio frequency rather than from optics and therefore are used to time-domain considerations. Despite this, we focus on frequency-domain methods in this chapter because we are convinced that these methods are better suited for the analysis and especially for the optimization of CDMs.

Up to now, we did not mention a prominent domain method, the method of moments (MoM). The MoM is different from the other domain methods because it only discretized the domain of the field sources. When losses are present, the discretization domain is a natural domain, but without losses, only the boundaries of the natural domains are discretized as in boundary methods. For this reason, MoM stands somehow between domain and boundary methods. It should be mentioned that typical MoM matrices are considerably smaller and less sparse than typical matrices of the other domain methods. As a consequence, iterative time domain implementations of MoM exhibit severe stability problems. For this reason, all prominent MoM implementations work in the frequency domain. Although this makes MoM attractive for PhC and CDM simulations, it seems that FE and standard boundary domain methods are more powerful.

Before we consider the boundary methods, we apply the frequency domain solver of FEMLAB [41]—a well-known commercial FE solver—to a simple PhC structure, that is, a rectangular 2D PhC block illuminated by a point source as shown in Fig. 52. This configuration has been presented in Ref. [20] as metamaterial without negative effective index that provides "all angle negative refraction" as required for fabricating a perfect lens. Note that the right part of Fig. 52 is in good agreement with the time-domain FD solution presented in Ref. [20]. Obviously, there is some stronger field around a point near the PhC block on

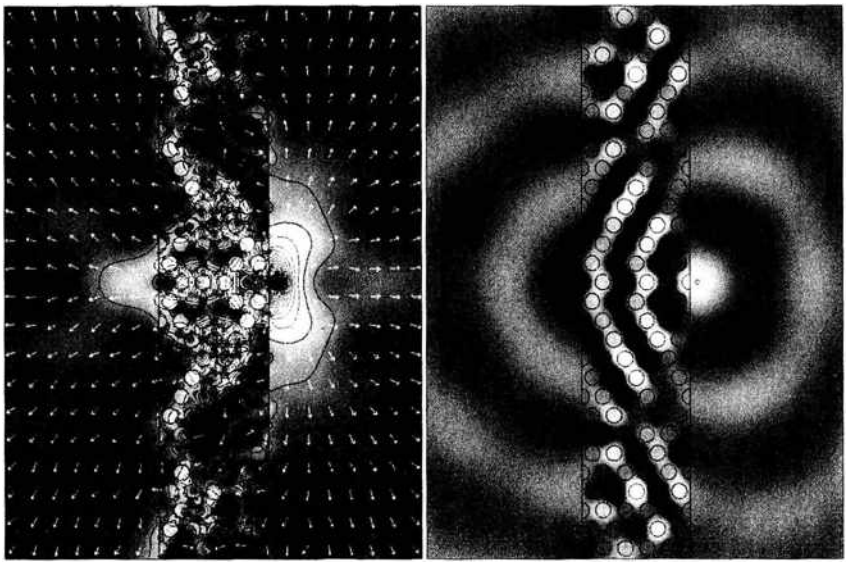


Figure 52. Metamaterial slab with "all angle negative refraction" illuminated by a monopole source on the right-hand side near the slab. A photonic crystal slab with holes arranged on a hexagonal lattice is used. The time average of the Pointing vector field (left) and the z-component of the magnetic field (right) are presented at the relative frequency $fa/c = 0.191$.

the side to the source point. One can also see that one has no sharp focus at all. In fact, the field plot for a certain time point provides a special "tuning parameter" (i.e., one can choose the time in such a way that the field in the focus point seems to be more concentrated than in other time points). A much better, that is, more honest representation is therefore obtained from time-average plots as shown in the left part of Fig. 52. From this figure, it becomes obvious that no sharp focus is obtained at all (i.e., this structure is far away from a "perfect lens"). It cannot even compete with a small conventional lens shown in Fig. 20. A closer look at the source point in Fig. 52 shows that this point is extremely close to the metamaterial block. Furthermore, the metamaterial is truncated in such a way that the circular holes are split along the borders, that is, the metamaterial has some sort of "coating" consisting of half circles and the source point is almost within such a half circle. In order to call this structure an optical lens, we should therefore also verify that the movement of the source point causes the expected movement of the focus point as indicated in Fig. 14. The result shown in Fig. 53 is crushing. As soon as the source point is moved, no focusing is observed at all. Despite this, the structure is interesting because it essentially consists of a photonic crystal that is not operated within a band gap. This is necessary because one intends to obtain energy transmission through the crystal. Furthermore, Fig. 52 shows that one really can observe negative refraction at such a metamaterial and this demonstrates that even geometrically simple metamaterials exhibit rich and attractive physical effects.

4.1.2. Boundary Methods

Boundary methods only discretize the boundaries of the natural domains. As a consequence, the corresponding system matrices become relatively dense and small. Thus, iterative matrix solvers are not very important and time-domain implementations are expected to be as problematic as for the MoM, that is, all prominent boundary method codes work in the frequency domain and are therefore well suited for the computation of periodic structures, resonance frequencies, waveguide modes, and so forth.

Because only the boundaries are discretized, the field equations inside the natural domains must be fulfilled by the series expansions that approximate the solution. This means that boundary methods use a set of basis functions that must fulfill the field equations, that is, Maxwell equations in each domain analytically. Such basis functions may be found with different techniques, namely the separation of variables and Green's techniques. Although many analytical solutions have been found for domains with relatively simple material properties, the derivation and implementation of a useful basis is the most difficult part of the

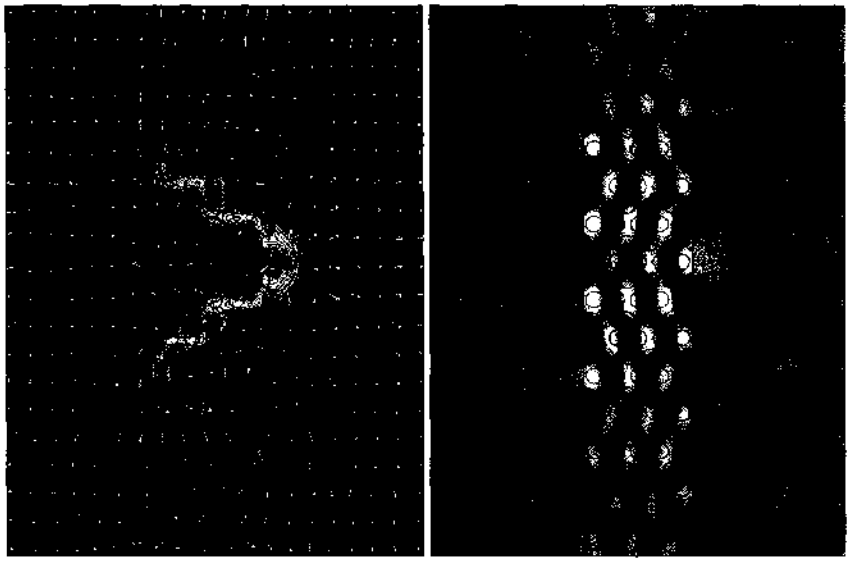


Figure 53. Same as in Fig. 52 with a different position of source (at the same distance from the slab as in Fig. 52!). As one can see, the sharp focus does not exist any more (i.e., a weak "super-resolution" effect is only observed for a very special location of the source).

boundary method. For this reason only, analytically experienced professionals are able to develop boundary methods. Furthermore, this considerably limits the application of boundary methods. Above all, such methods may not be used when nonlinear materials are present.

The simplest class of basis functions for boundary methods are the plane waves that are obtained from the separation of variables of the wave equations in Cartesian coordinates. Plane wave solutions are known for rather complicated materials including chiral, anisotropic, and bi-isotropic materials. Plane waves are closely related to spatial Fourier transforms. The main drawback of plane waves basis functions is the fact that these functions have the same amplitude everywhere in space when no losses are present and that on infinite spectrum of plane waves is available. The former can cause numerical cancellation problems in shadow areas. The latter causes difficulties in the proper selection of an appropriate set of finitely many plane waves for a good field approximation, especially when small features of a structure play an important role.

A numerically more attractive set of basis functions is obtained when the wave equation is separated in polar coordinates (2D) or in spherical coordinates (3D). This leads to well-known multipole expansions that exhibit a singularity in the origin. This leads to a very desirable local behavior of these basis functions that allows one an accurate modeling of small details. Multipole expansions can easily be derived only for simple materials. For complicated materials, one can obtain multipole approximations from plane wave superpositions [42]. For these reasons, multipole expansions have the same limitations as plane wave expansions. Multipole expansions are the basis of generalized multipole techniques (GMT) [39, 40]. The most advanced GMT implementation currently is the multiple multipole program (MMP) [23] of the MaX-1 software [26] that was used for most of the examples shown in this chapter. MMP includes not only multipoles and plane wave expansions but also more sophisticated expansions and linear superpositions of expansions such as waveguide modes of cylindrical and periodic structures. For this reason, MMP is very well suited for PhC and CDM simulations.

There are numerous special cases of GMT, namely codes based on lowest order multipoles (i.e., monopoles in 2D and dipoles in 3D). Several methods focus on these special cases, namely the method of auxiliary sources (MAS). Furthermore, these special cases are also the basis of Greeen's techniques that essentially integrate over monopole or dipole fields. In this sense, MoM and BE could be considered as special cases of GMT.

Because boundary methods discretize the boundaries, the transition from 2D to 3D models means that one has a transition from 1D to 2D boundaries, which is a much bigger step for boundary methods than for domain methods. For this reason, boundary methods

are very attractive, efficient, and accurate when 2D models are considered but they lose much attractivity in 3D. Here they are still interesting for band diagram and waveguide computations, but for 3D waveguide discontinuity computations, current implementations of boundary methods are very time-consuming.

4.2. Band Diagram Computation

As we have seen, the computation of band diagrams of photonic crystals is based on the eigenvalue solution of a periodic resonator problem. This also holds for more general periodic metamaterial structures. Before the eigenvalue problem should be tackled, the symmetry should be analyzed, and the corresponding first IBZ should be constructed. Once this has been done, any code that can handle periodic symmetries and eigenvalue problems may be used for the band diagram computation. Both the symmetries and the eigenvalue problems can be handled in different ways that also depend on the field solver to be applied.

First of all, time-domain codes cannot handle periodic symmetries properly. Therefore, they require brute-force solutions that model sufficiently big blocks of metamaterial. When such a block is illuminated by a time-pulsed plane wave, one can observe the time dependence of the field in the center of the block. Fourier transform then allows one to find the locations of the resonance frequencies for the given angle of incidence. When this is repeated for all angles corresponding to the positions on the contour of the first IBZ, the band diagram is obtained. This brute-force method is not efficient because it does not take advantage of the symmetry condition, but it includes an eigenvalue handling technique that is closely related to how resonances are observed in practice (i.e., some excitation is introduced and the response of the structure is measured at some point inside the resonator). This technique can also be applied to frequency-domain methods with either domain or boundary discretization.

In the frequency domain, one can easily take advantage of the periodic symmetry condition (13) for the field. One then must only discretize one single cell of the periodic structure, which makes the discretization area finite. We call this "original cell" in the following. Furthermore, the symmetry conditions link the field values on a boundary of such a cell with the field values on the opposite boundary. As a consequence, a hypercube topology is obtained. Although this topology is somehow unconventional, its implementation is trivial. Therefore, this technique can easily be applied to both domain and boundary methods. For domain methods, the procedure is trivial.

In the case of boundary methods, the periodic symmetry condition (13) for the field can be formulated on the boundary of the original cell only when a natural boundary is there. Because the natural boundaries usually do not coincide with the boundaries of the original cell, one must introduce fictitious boundaries. The domains on the two sides of such a boundary have identical material properties but are considered as different domains. Fictitious boundaries need no special implementation and can be always introduced. Because these boundaries are discretized as natural ones, the system matrix size is increased when fictitious boundaries are introduced. When the structure inside the original cell is simple, the discretization of the fictitious boundaries may become dominant. It is important to note that the definition of the original cell is pretty much arbitrary. For example, when a crystal with a square lattice is considered, it seems to be reasonable to define a square origina cell with four straight fictitious boundaries. First of all, the location of this square is arbitrary. Second, the boundaries can be curved lines, the only requirement is that the boundaries on the opposite sides of the original cells are indentical (i.e., are obtained from parallel translation). In order to minimize the discretization effort for the fictitious boundaries one should place them within those areas where the field is expected to be simple because this allows one to discretize the boundaries more roughly than when the field varies rapidly along them. Because the field is usually most complicated near the natural boundaries and becomes smooth with the distance from these boundaries, the original cell should be defined in such a way that all points of its boundaries are as far away from the natural boundaries as possible. This will reduce the numerical effort more than the minimization of the length of the fictitious boundaries (i.e., the introduction of straight lines).

In fact, boundary methods can also handle periodic problems without fictitious boundaries that increase the resulting matrix size. When no fictitious boundaries shall be introduced, one must approximate the field in the original cell by a series expansion with basis functions that fulfill the periodic symmetry condition (13) for the field. For plane wave basis functions, this is trivial. When more sophisticated basis functions, namely multipoles, are used, their symmetry decomposition is trickier. To illustrate this, we consider a simple PhC consisting of circular rods on a square lattice as shown in Fig. 54. The field outside a circular rod can easily be approximated with a multipole expansion located in the center of the rod. Obviously, the corresponding basis function does not fulfill the periodic symmetry condition (13). Because the original cell is only a section of the entire crystal, this is no surprise, and it is clear that an infinite array of multipoles placed in all centers of the rods are required. The condition (13) is fulfilled when the amplitudes of all multipole expansions outside the original cell are computed from corresponding amplitude of a multipole expansion inside the original cell according to

$$A(n_1\vec{r} + n_2\vec{r}_2 + n_3\vec{r}_3) = A(\vec{R}) = A_{original} \times e^{i\vec{k}\vec{R}} \tag{21}$$

When one now sums up over the infinite array of multipoles, one obtains the desired periodic multipole expansion. This sum must be truncated at some distance from the original cell in order to make it finite. Unfortunately, no rapid convergence is obtained for increasing distance, which causes a very high numerical effort for the evaluation of the basis function. For monopole arrays, special techniques to accelerate the convergence have been developed, and it is likely that such acceleration procedures may also be derived for arbitrary multipoles. However, the derivation of efficient algorithms and the corresponding implementation is tedious and always increases the matrix setup time compared with a standard multipole expansion that may be used when fictitious boundaries are introduced. Therefore, it is questionable whether this approach can outperform the "fictitious boundary," approach although the resulting system matrix is smaller.

We have outlined two different numerical techniques for handling periodic symmetries in boundary methods. We now will outline how three prominent boundary methods compute band diagrams.

4.2.1. Plane Wave Approximation

Currently, the plane wave approximation (PWA) code MPB developed at MIT [43] is the most frequently used code for the band diagram computation of photonic crystals. This code takes advantage of fast Fourier transforms and is highly efficient and user-friendly, namely because it is a specialized code for band diagram computations rather than a general field solver.

Figure 54 shows a typical output of the MPB code for a simple 2D PhC consisting of circular dielectric rods on a square lattice. Plane wave code MPG is based on the Fourier expansion of the dielectric constant of photonic crystal cell and it has consequently the same

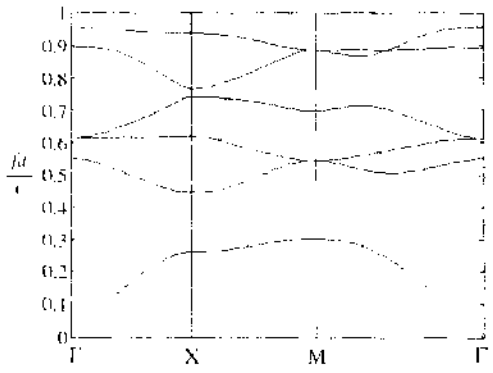


Figure 54. Perfect photonic crystal band structure calculation with the MPG and MMP codes. The disagreement is less than 1% everywhere (i.e., below the graphical resolution).

accuracy and convergence properties as Fourier. For this typical example (perfect photonic crystal with dielectric rods arranged in a square lattice), MPG code has generated the matrix of size 150 × 150. We have obtained the same accuracy (disagreement was bellow 1%) with the MMP code that generates the matrix 324 × 82 (overdetermined system). Unfortunately, the CPU time was not possible to compare because a free, available MPG code works under UNIX and our MaX-1 code runs on a PC. However, it is possible to say that the computation time is of the same order of magnitude.

4.2.2. MAS Computation

The method of auxiliary sources (MAS) [44] can be considered as a generalized multipole technique that focuses on the lowest order multipoles (i.e., monopoles in the 2D case). The MAS is practically the same as the method of fictitious sources (MFS) [45]. Because acceleration techniques for the numerical evaluation of monopole arrays are known, these codes need no fictitious boundaries for periodic problems as the PWA codes. Beside this, the MAS uses a different procedure for the computation of the resonances (i.e., another eigenvalue solver). This solver is motivated by the practical measurement of resonances like the time-domain procedure that was outlined before. Because only the original cell is modeled, one cannot specify an "exterior" excitation like in the brute-force model that considers a large but finite PhC block. For this reason, the MAS introduces a ficititious excitation somewhere inside the original cell and a fictitious measurement point P at some other place. Because the MAS uses monopoles to approximate the field, it is natural to also use a monopole excitation at some point S. One then can compute the strength of any field component in P as a function of the frequency and will observe peaks of this function at the resonances. For detecting these peaks, numerical search routines are required. These routines are heavily disturbed by a strange "double peak" phenomenon [46] (i.e., an accurate analysis shows that one may obtain two peaks that are very close to each other instead of a single one). Furthermore, the strengths and widths of the peaks depend very much on the location of the source point S and of the field point P. When one of these points is set at a position where the field of a mode is zero, the peak for the corresponding mode is suppressed. This is especially the case when points are set on special locations, namely on symmetry axes. For randomly set points P and S, it is very unlikely that a peak indicating one of the eigenvalues is suppressed, but it may happen that some of the peaks are much weaker than other ones. This problem may be reduced when more than one observation point P is introduced and the sum of the field in all observation points is considered. Similarly, one can introduce several source points and introduce excitations with equal amplitudes in all source points. This causes higher numerical effort but it also makes the procedure more robust. Unfortunately, the double peak phenomenon is not removed. For this reason, the MAS solver is currently not competitive with PWA solvers for the band diagram computation, although MAS is more attractive than PWA for scattering problems.

4.2.3. MMP Computation

The multiple multipole program (MMP) is the most sophisticated implementation of generalized multipole techniques GMT. In addition to multipole expansions, it also contains plane wave expansions and more sophisticated expansions. The implementation of the periodic counterparts of all expansions contained in the MMP library would be tedious. Because the band diagram computation is only a very special task for MMP, such periodic expansions are not available. Therefore, the MMP band diagram computation is based on the fictitious boundary technique mentioned above.

In addition to a technique for handling the symmetries, an eigenvalue solver is required. MMP allows one to use the same eigenvalue solver as for the MAS with one or several fictitious excitations.

Because eigenvalue problems are frequently solved by MMP users, an alternative eigenvalue solver without fictitious excitations is also available. When no excitation is available, the resulting matrix equation becomes homogeneous. Nontrivial solutions are then obtained from the zeros of the determinant of the system matrix. This common technique cannot

be applied in most of the MMP models because overdetermined systems with rectangular matrices are preferred. Instead of this, the MMP eigenvalue solver searches for the minims of a special eigenvalue search function that is obtained from the residual vector of the matrix equation and from the field amplitudes in a set of test points P similar to the MAS. MMP also contains a sophisticated eigenvalue search procedure that considers both the MAS search function and the residual-based search function at the same time in order to overcome the numerical problems caused by the doube peak phenomenon.

Numerically, the nonlinear search for maxims and minims of search functions is rather demanding and therefore not very efficient even when good search strategies are applied. To compensate this drawback in the band diagram computation, MMP takes advantage of the fact that the locations of the eigenvalues vary slowly when one traces them along the border of the first IBZ. As soon as the resonance frequency (i.e., the eigenvalue of a mode) is found in any point of the diagram, one therefore can estimate the eigenvalue in a point nearby. A good estimate of the location of an eigenvalue drastically reduces the iterations required for detecting the maxims or minims of the eigenvalue search function. When several points in a band diagram are known, one can take advantage of higher order extrapolation techniques for obtaining very accurate estimates. This eigenvalue estimation technique (EET) [23, 26] drastically reduces the computation time of band diagrams. Unfortunately, it sometimes exhibits numerical problems, especially near points in the band diagram where degenerated modes are present, that is, where two (or more) different modes have identical eigenvalues. For this reason, the MMP band diagram computation is either less robust (with EE) or less efficient (without EET) than the MPP computation.

Figure 54 shows the MMP result of the same PhC as computed with the MPG. In fact, an excellent agreement is obtained.

4.3. Guided Waves

As we have seen before, guided waves on cylindrical structures and on periodic structures such as photonic crystals are closely related. Because cylindrical waveguides are more common and analytically simpler than periodic ones, many methods for cylindrical waveguide simulations are available. As in the case of band diagram computations, time-domain techniques use brute-force models that are neither efficient nor convincing from the theoretical point of view. Therefore, we focus on the frequency-domain approach that is much more natural for eigenvalue problems. We have mentioned that only a cross section of a cylindrical waveguide structure needs to be analyzed and discretized. This simplifies the numerical model considerably. In optics, most of the waveguide structures are assumed to have modes with electromagnetic fields that tend to infinity. Usually, one has some core region where a strong field is present. Outside this core region, the field decays with the distance. Thus, it is not too difficult to truncate the model at a sufficiently big distance from the core region, except when a mode is close to its cutoff frequency, because the decay of the field with the distance from the core is very slow near cutoff. This mainly affects domain methods that always require a model truncation, but it also affects boundary methods when waveguide structures with infinite boundaries are considered.

It is important to know that the cutoff frequency of a mode is only defined for loss-free models. Above cutoff, the mode is then guided and characterized by a real propagation constant, where it is evanescent and characterized by an imaginary propagation constant below cutoff. When losses are present, the propagation constant is always complex, the transition from "guided" to "evanescent" becomes smooth, and the cutoff frequency is not sharply defined. Interestingly, even the most simple case of the wave propagation along a lossy circular wire becomes rather complicated, as shown in Ref. [12]. For this reason, many prominent codes for waveguides are restricted to loss-free cases or take the losses approximately into account using techniques for small disturbances (i.e., for small losses). Because the MMP eigenvalue solver, mentioned above, is very sophisticated and can also handle complex eigenvalue problems, this solver can also directly solve lossy waveguide problems.

For periodic waveguide structures, one must not only provide an eigenvalue solver but also a method for handling the periodic symmetry. Because codes for the computation of band diagrams must include both features, there seems to be no problem to use such

solvers for periodic waveguides. In fact, there are two important differences between perfect crystals and periodic waveguides. First of all, the periodic waveguides are described by an original cell that extends to infinity and this causes truncation problems for both domain methods and boundary methods because both domains and boundaries extend to infinity. Thus, an appropriate model truncation technique is important. The second problem is that engineering models of waveguides consider the propagation constant as an eigenvalue that is frequency dependent because the frequency is usually defined by the excitation (i.e., the generator at the end of the waveguide). In the band diagram formulation, the frequency is the eigenvalue that depends on the location in the reciprocal lattice space. Because a certain value of the propagation constant corresponds also to a point in the reciprocal lattice space, this is a minor problem as long as only the loss-free case is considered. When losses are present, it makes a big difference when the frequency is considered to be complex (in band diagram models) or when the propagation constant is considered to be complex (in conventional waveguide models). In the following, we therefore outline two different approaches, one that essentially uses the band diagram model and one that extends the conventional waveguide model.

4.3.1. Supercell Approach

The main problem of the computation of periodic waveguides with specialized band diagram computation codes is the broken symmetry (i.e., that the structure is only periodic along the waveguide). Because band diagram codes can also handle periodic structures with a single symmetry, this seems to be no big problem. In fact, codes like MPP assume that such structures consist of planar slabs, which is not the case for periodic waveguide structures at all. Therefore, an alternative technique must be found. The most elegant technique is to replace the given waveguide by an infinite set of parallel waveguides that are periodically repeated. This establishes periodic symmetry in a second direction for 2D models and in a second and third direction for 3D models. Consequently, the same procedure as for a perfect PhC can then be used for arbitrary (PhC and non-PhC) periodic waveguide structures. In the case of a PhC waveguide, the resulting original cell—that needs to be discretized—contains the waveguide channel and a bunch of cells of the original photonic crystal. For this reason, this cell is usually called *supercell*. Obviously, the matrix size of the supercell model becomes the larger the bigger the distance between the parallel waveguides is. Furthermore, the band diagram also becomes bigger with increasing distances of the parallel waveguides (i.e., with the size of the supercell). The reason for this is that each mode splits in more and more submodes with increasing supercell size. As one can see from Fig. 29, the band diagram becomes rather complicated even when the size of the supercell is not very big. As a consequence, the numerical effort grows drastically with the distance of neighbor waveguides. Because the field in these waveguides is coupled, the coupling of the waveguide modes becomes strong when the distance becomes short. The coupling causes undesired effects (i.e., it shifts the modes and therefore causes inaccuracies). The mode coupling depends not only on the distance of neighbor waveguides, but also the frequency. Near cutoff, the field does not decay rapidly with the distance from the waveguide channel, and this causes strong coupling even for big distances. As a consequence, the supercell computation becomes numerically expensive or inaccurate near the cutoff frequencies.

Because PhC waveguides are usually operated only within a PBG of the original perfect crystal, it is not necessary to evaluate the entire band diagram (i.e., one can restrict the computation to the PBG area of interest). This drastically reduces the number of modes to be evaluated. Note that only modes that occur within a PBG of the original PhC may represent guided waves. Note also that this technique can easily be applied to arbitrary periodic waveguides that are not obtained from photonic crystals, but it cannot easily be extended to lossy periodic waveguides.

4.3.2. Direct Approach

When we introduce a procedure for handling periodic symmetries in one direction (along a waveguide) to any frequency-domain code for the computation of cylindrical waveguides, we easily obtain a solver for periodic waveguide problems. Usually, this causes no

problems for boundary methods as long as no boundaries extend to infinity within the original cell that must be discretized. For domain methods, the original cell must be truncated at some distance from the waveguide core. If this distance is so big that the resulting field is almost zero on the truncation boundary, one can assume that it is zero there. Such a zero field boundary condition causes a perfect mirroring of the structure (i.e., the model becomes very similar to the periodic model of the supercell approach). Because domain methods provide ABCs, one can also move the truncation boundary closer to the waveguide channel—which reduceds the model size and computation costs—and apply ABCs on these boundaries. Here, it is important to note that the field of a guided mode outside the core does not propagate like a plane wave, and plane wave propagation is what usually is assumed for the development of ABCs. For this reason, one should carefully analyze the ABCs before one tries to take advantage of them for periodic waveguides.

When the walls of a periodic waveguide are blocks of PhCs or of arbitrary metamaterials that extend to infinity, model truncation becomes also an issue for boundary methods. When this happens, one can introduce special boundary conditions on the truncation boundary. A special case is the application of periodic truncation boundary conditions. This directly leads to the supercell approach mentioned before. The same holds for zero field or perfect mirror boundary conditions.

In fact, a periodic waveguide with infinite walls is not realistic at all. In practice, the width of the walls must be finite. The finite width has an important consequence for the solution. Even when total reflection is assumed at the waveguide wall, for example, because a PhC waveguide operates within the PBG, some energy may tunnel through the walls even when no losses are present. One then observes essentially an exponential decay of the field in the walls, but the field is not zero at the outer boundary of the walls, and therefore a radiating wave may be obtained (i.e., the mode is no longer guided and it becomes difficult to separate "guided" and radiating modes on such structures). Despite this, it is clear that assuming zero field on the outer boundaries of the walls is reasonable for sufficiently thick walls and therefore leads to reasonable (i.e., accurate) solutions. Because this argumentation is rather intuitive, we now establish and study a model that does not neglect the radiation outside the walls.

For reasons of simplicity, we consider a rod-type photonic crystal with a defect waveguide as in Fig. 28. We assume that the walls consist of finitely many PhC cells. Thus, the model of the original cell shown in Fig. 55 becomes very similar to the supercell model with one important difference. The wall is not terminated by a periodic boundary. Instead, we have a

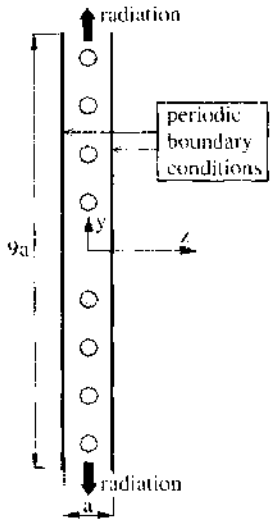


Figure 55. Direct approach for the photonic crystal waveguide analysis. The periodic boundaries of the supercell approach in x-direction are replaced by fictitious boundaries. Outside these boundaries, free space is assumed instead of the periodic repetition of the supercell approach. In free space, the radiating field is accurately modeled. The corresponding eigenvalue problem and the periodic constant in x-direction (propagation constant) become complex.

fictitious boundary that separates the PhC walls from free space. Incidentally, for a hole-type PhC, this would be a natural boundary. Now, we have a situation that is very similar to a grating, where we often also have a periodic structure embedded in free space, for example, a periodic metamaterial slab. When the free-space sections of a grating are truncated by straight (fictitious or natural) boundaries, one can expand the field in free space by Rayleigh expansions or Floquet modes (i.e., propagating and evanescent plane waves). From the propagating Rayleigh expansions, we obtain the radiation loss of the structure. Because of the energy conservation law, the total power that enters the original cell through the boundary on the left side of Fig. 55 should be equal to the sum of (1) the total power that leaves the original cell through the boundary on the right side of Fig. 55, (2) the total radiated power, and (3) the power loss within the original cell due to material loss. Note that no power reflection is present because the original cell is a part of a waveguide structure. Because the periodic conditions (10) and (19) hold for the left and right boundaries of the original cell with $n = 1$, the field on the right boundary is simply C_x times the field on the left boundary. From this, one easily can see that the time average of the total power flux through the right boundary is given by

$$P_{right} = P_{left} \times C_x \times C_x^* = P_{left} \times e^{jk_x a} \times e^{-jk_x^* a} = P_{left} \times e^{-2\mathrm{Im}(k_x)a} \tag{22}$$

where * indicates the conjugate complex. Obviously, the imaginary part of the periodic propagation constant k_x indicates the attenuation and must be positive when radiation losses or material losses are present.

When no material losses are present, the radiated power loss is obtained from Eq. (22) when the imaginary part of the propagation constant is known. To obtain the complex k_x, we can run a complex eigenvalue solver such as the MMP eigenvalue solver. In order to make sure that big radiation loss is observed, we first consider a PhC defect waveguide with only a single PhC cell on each side of the waveguide channel as shown in Fig. 56. Figure 57 illustrates the MMP eigenvalue search function over the first quadrant of the complex k_x plane for a given frequency. All modes of the structure are characterized by minims of this function. Although the search space is infinite and infinitely many modes may be found, one may restrict the eigenvalue search to a region near the real axis, because modes far away exhibit strong dampening and therefore decay rapidly. Such modes are usually of no practical interest. When the frequency is modified, the landscape in the complex eigenvalue plane changes and all minims move around. The EET of MMP allows one to follow easily the traces of the minims as indicated for the fundamental mode with relatively low loss in Figure 57.

When the width of the PhC walls is increased, the attenuation constants (i.e., the imaginary parts of the fundamental waveguide modes) are reduced as expected (Fig. 58). Consequently, the minims of the complex eigenvalue search function move closer to the real axis. As soon as the distance from the imaginary axis (i.e., the radiation losses) are small enough, one can obtain a good approximation of the real part of k_x by a real eigenvalue search along the real axis, which drastically reduces the computation time. One can use such a real eigenvalue search for obtaining a good start point of the complex search, but one can also use the eigenfield obtained from the real eigenvalue search for obtaining an approximation of the radiated power. One can then take advantage of Eq. (22) for obtaining an approximation of the imaginary part of k_x (i.e., of the attenuation constant). Figure 59 shows that such approximations are rather accurate for sufficiently thick waveguide walls. Incidentally, no good supercell approximations can be obtained when the radiated power of the corresponding direct model is so high that the approximation of the imaginary part of k_x—that uses the results of the real eigenvalue search—is not accurate enough. The direct solution can therefore always rely on the simple real eigenvalue search when the corresponding supercell model is accurate and *vice versa*. Thus, the direct model provides not only information on the radiation that is of practical interest, but also information that is useful for the validation of the supercell approach.

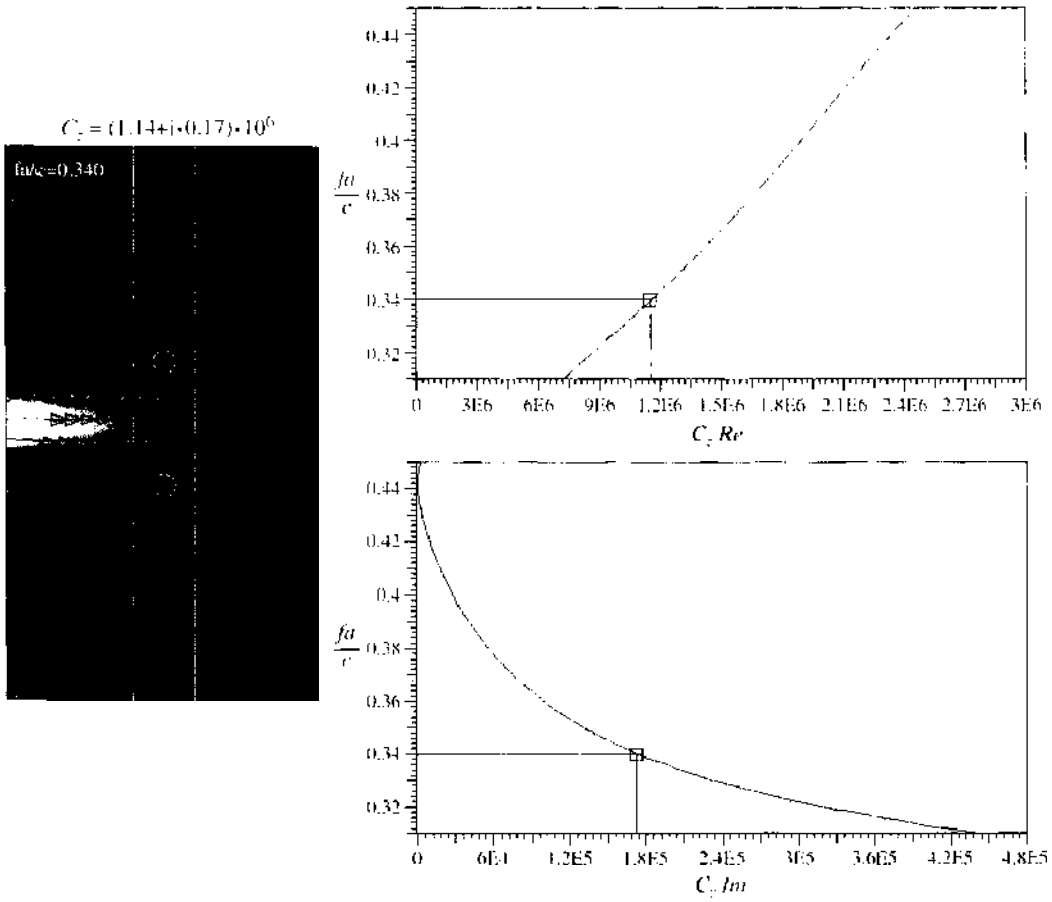


Figure 56. Single-layer photonic crystal waveguide. Time average of the Poynting vector field (left), real part (right top), and imaginary part (right bottom) of the complex periodic constant in x-direction (propagation constant) over the frequency range of the bandgap of the original perfect PhC. Results obtained from a complex eigenvalue search procedure.

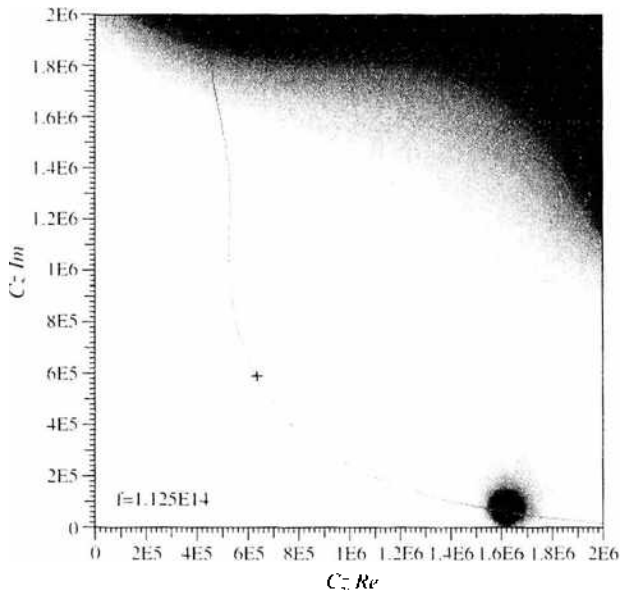


Figure 57. Eigenvalue search function in the complex plane of the periodic constant (propagation constant) for a radiating PhC waveguide as shown in Fig. 56. Eigenvalues are characterized by local minima of this function (dark spots). The solid line shows the trace of the fundamental mode for varying frequency.

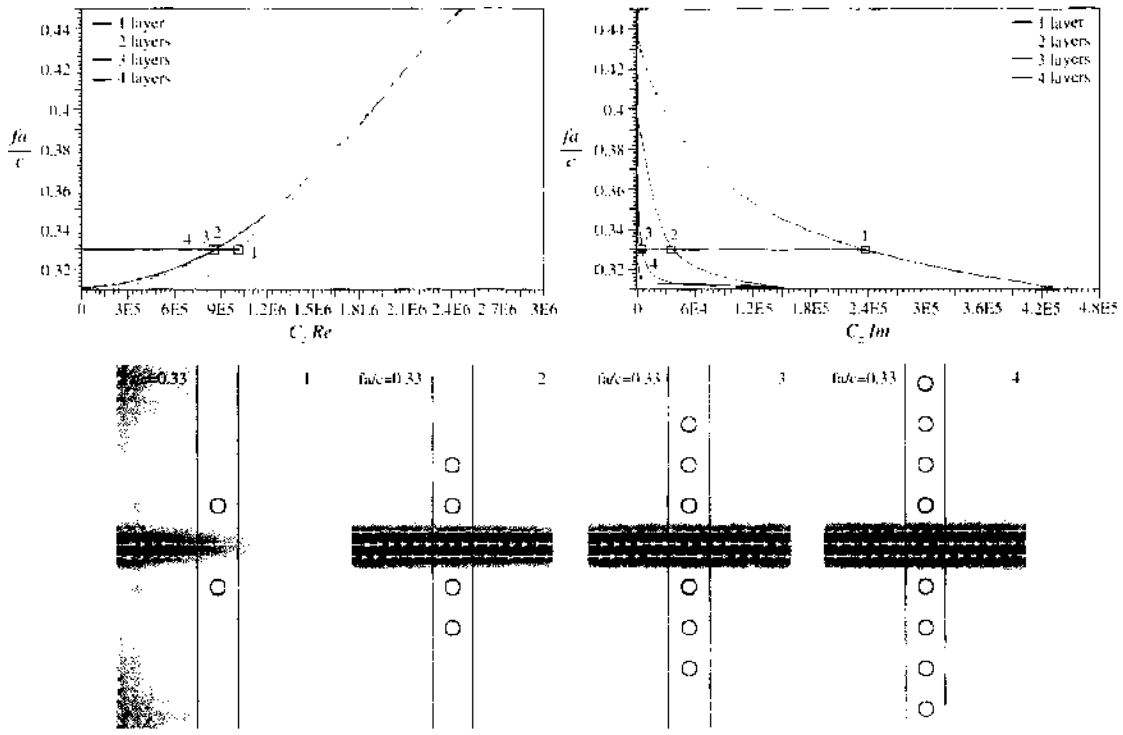


Figure 58. Complex periodic constant (propagation constant) analysis with respect to the number of PhC layers on both sides of the waveguide channel. Real part (top left) and imaginary part (top right) of the periodic constant for one up to four crystal layers and plots of the time average of the Poynting vector field for the four different waveguides (bottom).

4.4. Waveguide Discontinuities

We have seen that time-domain methods use some brute-force approach for problems with periodic symmetry. In this brute-force approach, sufficiently large blocks of CDM structures are modeled explicitly, which makes the procedure rather inefficient. When waveguide discontinuity problems are considered, all periodic symmetries are broken, and the missing handling of symmetries in time-domain codes is no longer any handicap. For this reason, users of time-domain codes usually directly tackle waveguide discontinuity problems without any previous analysis of the waveguide modes that are important for the structure under consideration. Such "quick and dirty" solutions provide a high risk of obtaining inaccurate or even useless results especially because CDMs often exhibit very complex effects that are not easily understood. Because the complex effects make the CDM concept attractive, careful studies—starting with the band diagram analysis and continuing with an analysis of all possible waveguide modes—are strongly recommended.

When efficient and accurate solutions are desired, one should first understand the physical effects that may be caused by the most primitive waveguide discontinuities; namely, those outlined in the previous section. All such models assume that one or several waveguides are connected to the discontinuity area and that these waveguides extend to infinity. Because only finite models may be discretized, model truncation is essential for all types of numerical modes. As we will see below, there are many different techniques for model truncation that have different advantages and drawbacks. Most of the attractive truncation techniques are limited to frequency-domain formulations. Therefore, typical time-domain solutions keep their quick and dirty image also here.

It is important to keep in mind that each discontinuity can have a strong frequency dependence and usually does not only reflect incoming modes but also converts them into different modes, including evanescent modes. Because the evanescent terms decay exponentially with the distance from the discontinuity, one usually assumes that the evanescent modes can be neglected when the model truncation is sufficiently far away from the discontinuity. The problem of this concept is that one does not know how long this distance should be as long

as one has not computed at least the lowest order evanescent modes. For "safety" reasons, quick and dirty modelers often define rather large models with long distances from the truncation lines to the discontinuity area. Such models obviously become rather inefficient and are certainly not "quick" in terms of computation time. Even relatively small waveguide discontinuity models are often big compared with the wavelength and require an accurate full-wave analysis based on Maxwell's equations. As a consequence, the resulting system matrices often reaching the memory limit of modern computers. Thus, more sophisticated model truncation techniques that allow one to reduce the required model size are highly desirable. We present some promising concepts in the following.

4.4.1. Model Truncation

The first step of a reliable model truncation is the analysis of the lowest order evanescent modes (i.e., the modes with the smallest attenuation constant). This can be done with the same methods as for the analysis of guided modes—provided that imaginary propagation constants or the more general complex propagation constants can be analyzed with the corresponding codes. Although this provides no essential difficulty from the theoretical point of view, special codes for the waveguide computation may be restricted to real propagation constants. From the attenuation constant of the first evanescent mode and from the desired accuracy, one can easily estimate the required distance from the discontinuity to the truncation line.

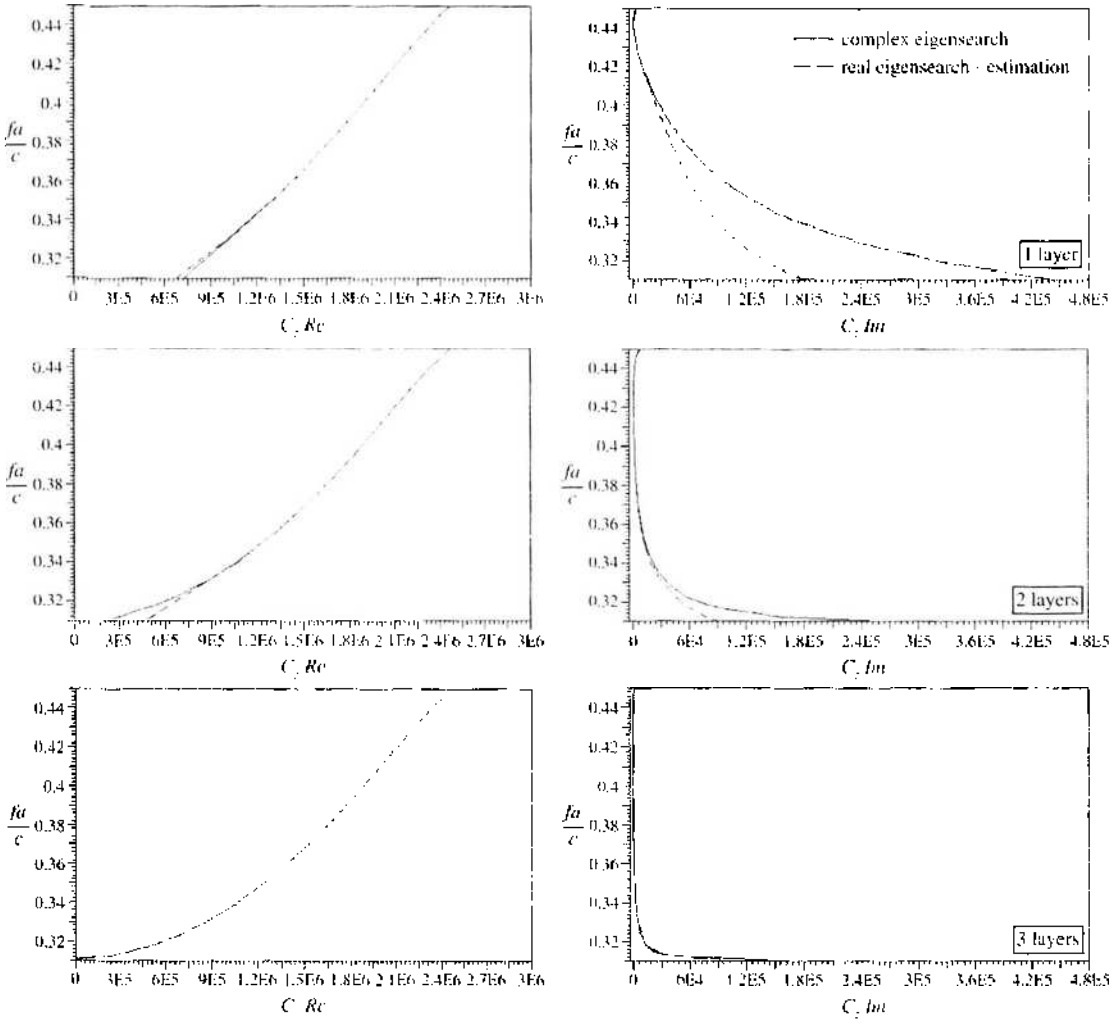


Figure 59. Comparison of the frequency dependence of the complex eigenvalues obtained by a complex eigenvalue search and by a real eigenvalue search with an estimation of the imaginary part of periodic constant using the energy conservation law. As one can see, the imaginary part can accurately be estimated when sufficiently many crystal layers are present on both sides of the waveguide channel.

Specialists of absorbing boundary conditions (ABCs) tend to shrink the model size as much as possible for demonstrating the value of their ABCs. ABCs can also be used for absorbing waveguide modes. When an excellent ABC is available, one is tempted to move the truncation lines closer to the discontinuity region than mentioned above (i.e., to set the ABCs in the area where the evanescent modes are not small). When one does this, one should first make sure that the ABCs also absorb evanescent modes. If this is not the case—as usual—unexpected inaccuracies can occur. For testing the ABCs, it is reasonable but not sufficient to model a short waveguide section terminated with ABCs on both sides. This is not sufficient because the discontinuities at the ABCs do not necessarily excite the disturbing evanescent modes as strongly as a real waveguide discontinuity.

Note that the attenuation constant of an evanescent mode is strongly frequency dependent. Near the cutoff frequency, it tends to zero. Therefore, it may be impossible for some frequencies to move the truncation lines sufficiently far away. In frequency-domain codes, one can then apply mode matching techniques that also include low-order evanescent terms. In time-domain codes, one can only hope that only a small fraction of the spectrum is near the cutoff frequency of an evanescent mode and therefore no troubles are caused.

4.4.2. Excitation of Waveguide Modes

For all output ports of a waveguide discontinuity, the model truncation outlined above is sufficient, but an additional technique is required in the input port, where at least one incident mode is assumed. Obviously, a proper definition of the incoming mode is only possible, when it has been computed before with an appropriate eigenvalue solver. Theoretically, different numerical solvers may be applied for computing the waveguide modes and waveguide discontinuities, but it is often cumbersome or even impossible to insert results obtained from analytical solutions or from another code into a code for waveguide discontinuities. This especially holds when the former code works in the frequency domain and the latter in the time domain. Even when it is impossible to feed a known mode into a code, the knowledge of the important mode patterns is very helpful to excite efficiently the desired waveguide modes with other means. This again supports the recommendation to study carefully all important waveguide modes before modeling any waveguide discontinuity.

The most simple waveguide mode excitation is obtained from a fictitious point source placed somewhere in the input waveguide. When single-mode waveguides are considered, this causes no big problems. Usually, the source will excite not only the propagating mode but also evanescent modes. As in the case of the model truncation, these evanescent modes become negligible at a sufficient distance from the source point. Because the field of the fundamental waveguide mode is usually strong in the center of the waveguide, it is reasonable to put the source point somewhere in the center. For the excitation of higher order modes, this technique is useless. First of all, some of the higher order modes will have a zero field in the center of the waveguide. These modes cannot be excited by a source point in the center. Furthermore, an arbitrarily placed source point will usually excite all guided and evanescent modes because its field is singular. From mode matching, it then becomes evident that all modes must be excited.

There are different methods to overcome the problem of mode excitations in multimode waveguides. First of all, one can apply the rigorous mode matching technique—provided that one works in the frequency domain. This technique has already been outlined and will be considered later again. An alternative is to extend the source point technique as follows: When N guided modes are present, one can insert N different source points at different locations. Each source will excite all modes. When one analyzes the field at a sufficient distance from the source, one will therefore observe a superposition of the N modes for each of the N models. From this, one can finally find linear superpositions of the N models with appropriate amplitudes of the sources in such a way that each superposition only excites one single mode. Once more, for this procedure, a previous analysis of the waveguide modes is helpful. Note that the superpositions mentioned above are frequency dependent in general. Therefore, it is not easy to apply this technique to time-domain methods.

The multisource methods outlined above represent a relatively simple alternative to the sophisticated mode matching technique. Its main drawback is the fact that a point source also

strongly excites evanescent modes. In order to overcome this difficulty, the introduction of a "complex origin" source has been proposed in the MAS [47, 48]. This technique essentially moves the origin of the point source, that is, of the monopole expansion in the source point into the complex plane. As a result, a beam-like source field is obtained. The width and direction of the beam can easily be steered with the imaginary part of the complex origin. This technique allows one to excite the fundamental waveguide mode very efficiently. By introducing appropriate complex origin multipole, one can also efficiently excite higher order modes, but it is not possible to excite only one single mode at once within a multimode waveguide. Thus, a multi-complex-origin-source method is required for multimode cases.

The model truncation technique and the multisource method outlined above are sufficient in principle for handling waveguide discontinuity problems. In the following, we present some more advanced techniques for the same purpose.

4.4.3. S Matrix Computation

When we need to compute the S matrix of a waveguide discontinuity, we must repeat the procedures mentioned above with each guided mode in each of the ports as an excitation in one of the models. When the total number of modes in all ports is N, the S matrix is an N by N matrix that is obtained from solving N models. Each of these models is represented by a matrix equation of the form $\mathrm{M}\mathrm{X}_n = \mathrm{A}_n$, $n = 1 \ldots N$ where the excitation is defined in the right-hand side vector A_n. Because the geometry and material properties are identical for all models, the system matrix M is one and the same for all n. Many direct matrix solvers allow one to solve such systems with multiple right-hand sides at once. This allows one to reduce drastically the computation time by a factor of almost N. Of course, this can only be done when the system matrix M is explicitly stored.

As soon as the vectors X_n are known, one can compute the field of all outgoing modes in all ports—provided that no reflections are caused by the truncation of the ports. Usually, this is no problem when the mode matching technique is applied. The elements s_{ik} of S are then simply given by $\mathrm{P}_{kn}/\mathrm{P}_n$, where P_{kn} is the outgoing power of the mode k in the nth model that is excited with a total power P_n in the input port. Note that these powers must be computed at well-defined reference planes in each of the ports. When mode matching is used, the reference planes usually coincide with the truncation planes, but for other methods it might be reasonable to move the reference planes away from the truncation planes.

4.4.4. Mode Matching Techniques

For simple cylindrical waveguide discontinuities, one can apply mode matching for obtaining the S matrix by subdividing the entire discontinuity in sufficiently many slices of cylindrical cross sections as illustrated in Fig. 32. One then has $k = 1 \ldots K$ abrupt transition planes. The transmission and reflection of the modes at each transition plane are then described by the S matrices S_k. From these matrices, one finally obtains the S matrix of the entire structure. For complicated geometries and for noncylindrical waveguides, this procedure is not efficient or even applicable. Furthermore, all relevant waveguide modes must be known in all waveguide sections. The numerical evaluation of these modes is extremely tedious for waveguides with complicated geometry. For this reason, the standard mode matching technique is only applied to simple waveguides where analytical solutions of the modes are known. Thus, it is obvious that we need alternatives for CDM waveguide discontinuities.

The most simple way to take advantage of mode matching without undesirable restrictions is to apply mode matching only at the model truncation lines (or planes in 3D) and to approximate the electromagnetic field in the entire discontinuity area exactly as a scattering problem by any numerical method as illustrated in Fig. 33. This partial mode matching technique requires only the computation of the relevant modes in the waveguide ports. Furthermore, the scattering problem solver must include a feature for importing the waveguide modes. Unfortunately, several commercial scattering problem solvers do not provide such features.

Because the knowledge of the waveguide modes is also useful for the synthesis of a waveguide discontinuity with desired properties, partial mode matching is currently our preferred technique although the numerical evaluation of all waveguide modes can be tedious.

4.4.5. Compensation Scheme for Reflected Waves

When a scattering problem solver does not provide any feature for taking advantage of the partial mode matching technique outlined above, one usually computes the S matrix of the discontinuity as mentioned before. Here, it is important to note that a perfect or at least a good suppression of modes reflected at the output ports is required. Some codes do not provide sufficiently good absorbing boundary conditions for the waveguide ports. In this case, one can take advantage of a scheme that compensates the reflected waves instead of absorbing them [48].

For reasons of simplicity, we now assume that only single-mode waveguides are present. Then N is not only the total number of modes but also the number of ports. In order to compute the S matrix, N systems of equations of the form $MX_n = A_n$, $n = 1 \ldots N$ are set up and solved all together using a matrix solver for multiple right-hand sides. Because of the reflections at all ports, each model has an incoming and an outgoing wave in each port. Thus, X_n represents the solution for a set of incoming waves from all ports rather then a solution with a single incoming wave. Assume that the reflection coefficients R_{mn} represent the reflection at port m for the model n. It is well-known that reflection coefficients can be measured in reality by observing the standing wave patterns along a waveguide, that is, the reflection coefficient is obtained from the standing wave ratio (SWR) analysis. The SWR is obtained in a numerical simulation when the field is evaluated along a sufficiently long waveguide section that is free of evanescent waves. As soon as the reflection coefficients are known, one also knows the strengths of the incoming and outgoing waves in all ports for each of the N models. One now can compute a superposition of these N models in such a way that all incoming waves cancel each other in all but one port, for example, port i. From this, one obtains a solution where no energy is incident from any port, except port i. This is the solution that also would be obtained when perfectly absorbing terminations of the ports would be available. When one repeats this for $i = 1 \ldots N$, one has all information for computing the S matrix. Note that the numerical effort for this reflection compensation is small. In addition to the computation of the reflection coefficients from the SWR, one only needs to solve a linear system of N equations with N unknowns, where N is the number of ports that is usually very small. For more details, see Ref. [48].

4.4.6. Optimized Open Port Approach

Up to now, we have three techniques for handling the output ports of a waveguide discontinuity: (1) sufficiently good absorption of the outgoing waves (i.e., suppression of reflected waves by ABCs), (2) mode matching at the ports, and (3) compensation of the reflections from a SWR analysis evaluation of the reflection coefficients. Because excellent ABCs are known for free space but not for complicated waveguide structures, a fourth alternative is to add a transition of the waveguide to free space at all ports as illustrated in Fig. 39. Such a transition is essentially a radiating antenna fed by the waveguide. As one can see from Fig. 39, even a simple, abrupt termination of a PhC waveguide can radiate rather well into free space. In general, such abrupt terminations cause rather high reflections that are not desirable. Of course, we could now use the reflection compensation scheme outlined above, but if so, we need no ABC as well. An alternative is to minimize the reflection by optimizing the termination of the waveguide. For doing this, we can take advantage of the methods presented in the following section. As soon as such an optimal termination is known, it may be applied to all ports of a waveguide discontinuity. Because the numerical optimization can be very time-consuming, this procedure is only useful when one frequently considers problems where always one and the same waveguide structure is involved.

It should be mentioned that it is often surprisingly simple to optimize an open waveguide port with zero reflection at a single frequency. For example, Fig. 39 shows such a port where only one rod of the PhC on both sides of the waveguide channel was optimized in a symmetric way. Finding broadband solutions, for example, solutions with almost zero reflection over the entire PBG of a photonic crystal, is much more difficult and requires much more extensive optimizations. We now will consider such optimizations in more detail.

5. OPTIMIZATION OF COMPOSITE DOPED METAMATERIALS

During the past few years many promising CDM devices, above all PhC structures, were proposed and simulated by many researchers. Most of these structures were found by experience obtained from conventional devices based on waveguide structures and from intuition. Although many of these structures demonstrate that CDM devices might become very attractive also outside high-tech labs in the future, the quality of these devices is usually not competitive with traditional structures that exhibit the same functionality. To illustrate this, we consider the simple PhC power divider shown in Fig. 49. In conventional waveguide techniques, one often uses T- or Y-shaped junctions. Such junctions can easily be obtained also within a photonic crystal as shown in Fig. 49. The reflection constants of both structures turn out to be above 10% at a frequency near the center of the first PBG of the crystal. This is nice in the sense that it proves that the concept works in principle, but it is by far not enough for making such structures competitive or for meeting the requirements of industrial applications. The fact that the reflection coefficients of the T and Y structures are rather similar lets one assume that modifications (i.e., additional "doping" of these structures) would not strongly affect the performance of the junction. In fact, this assumption is completely wrong. When we introduce, for example, a single additional defect in the T structure, the reflection coefficient becomes drastically higher, as illustrated in Fig. 60. When one tries to reduce the reflection coefficient of either the T or the Y structure by small modifications, one will see that intuition is often misleading because our brains have not much experience with CDM structures. After a while of trying, one might assume that the T junction is optimal in this case. This would have dramatic consequences for the industrial applicability of PhCs. Fortunately, this assumption also is wrong. In fact, even a simple PhC junction provides extremely many parameters that can be optimized. This leads to huge, high-dimensional optimization problems that only can be solved numerically, but it also provides enough freedom in the CDM design for obtaining really powerful structures, for example, power dividers with almost zero reflection.

Numerical optimization essentially searches for minima of a cost function or maxima of a fitness function. Whether the former or the latter is preferred is a question of taste. Here, we consider the search for optimal fitness. The first step is therefore the definition of the goal (i.e., of the fitness function). For the power divider mentioned above, we want to minimize the reflection coefficient R. Because R must be in the range from 0 to 1, we can define the fitness as $1 - R$, but this definition is not unique at all. Simple alternatives might be $1/R$ and $1 - R^2$. When we want to obtain a broadband solution, we need to minimize the reflection coefficient over some frequency range. This might be done by a weighted integral of one of the definitions above over the given frequency range. More complicated fitness definitions are required when several different goals shall be met at the same time, for example, when one wants to design a T junction that transmits the energy to one output

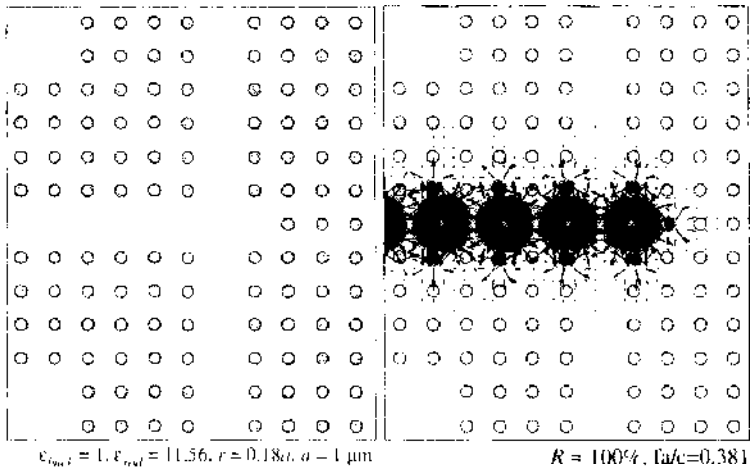


Figure 60. Photonic crystal power divider. Compared with the structure presented in Fig. 49, only a single rod is removed. As a result, the reflection coefficient increases from 13% to almost 100%.

port for some frequency and to the other output port for another frequency in such a way that (1) the reflection coefficient at the input port is small for both frequencies and (2) that the transmission coefficients to the wrong output port is also small for both frequencies. The number of conditions that have an influence on the fitness may be high. When the fitness function is not well defined, it may happen that the solution perfectly meets some of the conditions, whereas others are not fulfilled at all. A good fitness definition is a very important task that is often underestimated. It plays an extremely important role and requires much experience from the user because one has a high degree of freedom, almost no rules how to obtain good fitness definition, and because the results of an optimization depend strongly on the fitness definition.

The fitness is always defined over some search space. For example, when we want to optimize the location of a certain rod in a PhC, its position is defined by two real values, and the search space is the 2D real plane spanned by the coordinates of the rod. The size and dimension of the search space have a strong impact on the speed of the numeric optimization but also on the quality of the solutions that can be found. For example, when we reduce the 2D search space above on a small area (i.e., when we allow the optimizer to modify the original position of the rod only slightly), we cannot expect to obtain a much better solution. In most cases, moving a single rod in a PhC will not be sufficient for finding sufficiently good solutions. Because even a simple PhC structure consists of hundreds of cells, the search space obtains a very high dimension when we want to optimize the parameters of all cells.

As a consequence, the optimizer then can explore only an extremely small fraction of the entire search space, and it becomes not very likely that a good solution can be found. For this reason, the second important task for the user is to specify a reasonable big, reasonably limited search space. This requires as much experience as the definition of the fitness function. Because CDM structures exhibit very rich and complicated effects, enough experience can only be obtained from many numerical simulations. Chess players know that they learn most when they fight against stronger players. Similarly, it is best to study excellent CDM designs for getting a good feeling for CDM structures. We consider the quality of intuitive designs with quick and dirty field simulations to be insufficient for this purpose, that is, the student who studies such structures is trained by solutions far from optimal ones, and this does not allow him to realize how good solutions work. This is the main reason why we recommend intensive numerical optimizations. The numerical optimizer can design thousands of structures and filter out only the most promising ones. After this, the user can focus his study on the best solutions and learn from these like a chess player who learns only from the analysis of games that were played by chess masters.

Numerical optimization can be based on two different classes of algorithms: probabilistic or stochastic and deterministic optimizers. In early years of computing, only deterministic optimizers were considered because these optimizers usually require only a few fitness evaluations for finding a local optimum near a start point defined by the user. Each fitness evaluation in the CDM optimization requires the numerical field analysis of the corresponding structure. This field analysis can be very time-consuming even on fast computers. Therefore, the promise of deterministic optimizers to work with a few fitness evaluations is very attractive. At the same time, one should see that this requires a good, user-defined start point. As long as a user has not much experience with CDM structures, it is likely that the start point he provides is not good enough for allowing the deterministic optimizer to find a sufficiently good solution. This becomes obvious when we consider the simple power divider mentioned above.

We have noticed that most of the modifications of the initial T structure increase the reflection coefficient and therefore reduce the fitness. When we select a small search area of five times four cells near the junction—as illustrated in Fig. 61—and consider all possible effects in this area, we obtain $2^{5\times4} = 1,048,576$ possible models. Because we want that equal amounts of energy are transferred to both output ports, it is sufficient to consider only the $2^{5\times4} = 4096$ different symmetric models. This number is small enough for a brute-force evaluation of all possible solutions. We then can see that most of the models provide very high reflection coefficients far above 10%. It is therefore very unlikely to find good solutions by randomly selecting the defects. From this point of view, one can say that our intuitive

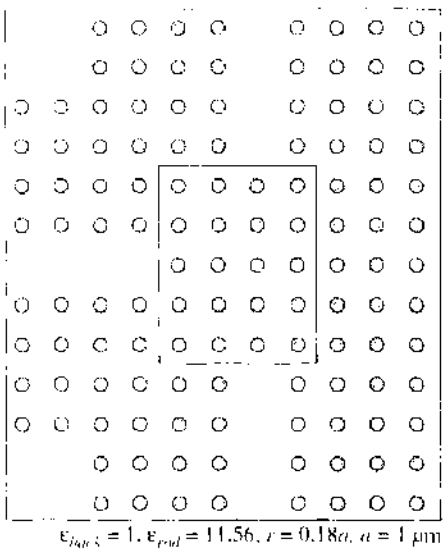


Figure 61. Photonic crystal power divider search area. As one can see, we have defined a relatively small area of five by four rods around the junction of the three waveguides. In this area, defects shall be introduced in such a way that the reflection coefficient at the input port (left-hand side) is minimized and that equal amounts of energy are transmitted to the output ports (top and bottom).

T- and Y-shaped designs are not bad at all. But the inspection of all 4096 solutions also shows that there are some solutions with reflection coefficients below 1% (i.e., solutions that are much better than our intuitive ones). The best two solutions are shown in Fig. 62. As one can see, these structures are rather different. What is common to both structures is that one can observe two strongly coupled cavity areas. When we have a look at the frequency response (Fig. 62) of both structures, this becomes clearer. One cavity has a resonance above the center frequency where the fitness function was defined previously, and the other cavity has a resonance above this frequency. As one can also see, the distance of the two resonances of the best solution is bigger than the distance of the second-best solution, but the resonance peaks for both solutions are within the band gap. Although we obtain a low reflection coefficient over a rather broad frequency band for the best solution, low reflection over the entire frequency band is not obtained.

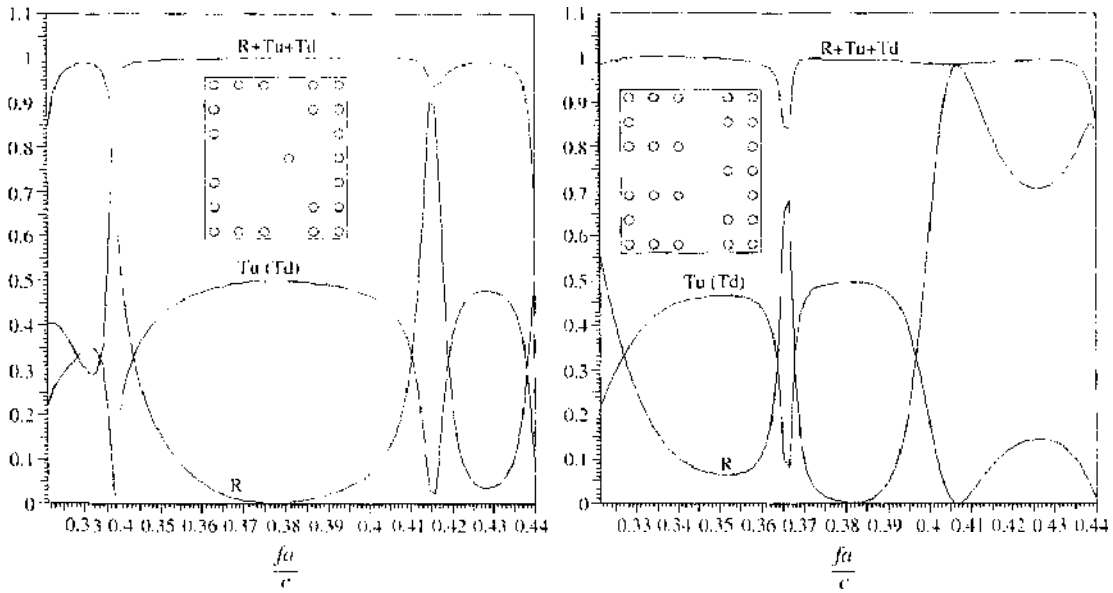


Figure 62. Frequency characteristics and geometry of the best (left) and second-best (right) solutions of a photonic crystal powr divider—obtained from the definition of defects in the search area shown in Fig. 61. These two solutions cannot easily be found from intuition or human experience.

We now can run a deterministic parameter optimizer that tries to improve the solutions by modifications of the locations and radii of all rods in our small search area. This leads to a real search space with five dimensions because only three rods are present in the search area of the best model and two of them are symmetric. For a pair of symmetric rods, we can optimize three parameters (x and y coordinates and the radius r). For a rod on the symmetry axis, we can optimize x and r. Numerical optimization now allows us to find a slightly better solution, as one can see from Fig. 63. Maybe this is not surprising because the six-dimensional search space is rather small, but it is unlikely that we could find a better solution when we would use a deterministic optimizer that starts with the intuitively designed T-shaped structure. This structure has 13 rods in the search area, six of them are symmetric. Thus, we obtain a real parameter search space of 20 dimensions. In order to estimate numerically a single gradient in this search space, we must compute at least 20 slightly different numerical models, which is rather time-consuming and does not allow us to perform many iterations during the optimization. Such optimizations can easily reach the limitations of modern computers, even when simplified 2D models and fast Maxwell solvers are applied with sophisticated port truncation schemes for obtaining a minimal model size.

When we want to find a better solution than the one presented above, we must increase the area where we select the defects. When we consider a single additional layer around the optimization area shown in Fig. 61, we already obtain $2^{4+6} = 16,777,216$ possible symmetric models. This is far beyond the reach of brute-force search. Therefore, we need a strategy to find efficiently good configurations. This placement of the defects can also be considered as an optimization problem but in a binary search space. When we have a search area of N cells, we can characterize each cell by a bit that indicates whether the cell contains a defect or not. The entire model is then characterized by a bit string of N bits (i.e., we have an N-dimensional binary search space). In such search spaces, no gradient may be defined and also the definition of distances in such spaces is of questionable value. For this reason, traditional optimizers cannot be applied. In the following, we outline the most important probabilistic optimizers for this purpose.

When high-dimensional spaces shall be explored, systematic procedures always become inefficient. This already holds for simple tasks such as numerical integration. Therefore, it is no surprise that deterministic routines become inefficient for high-dimensional spaces.

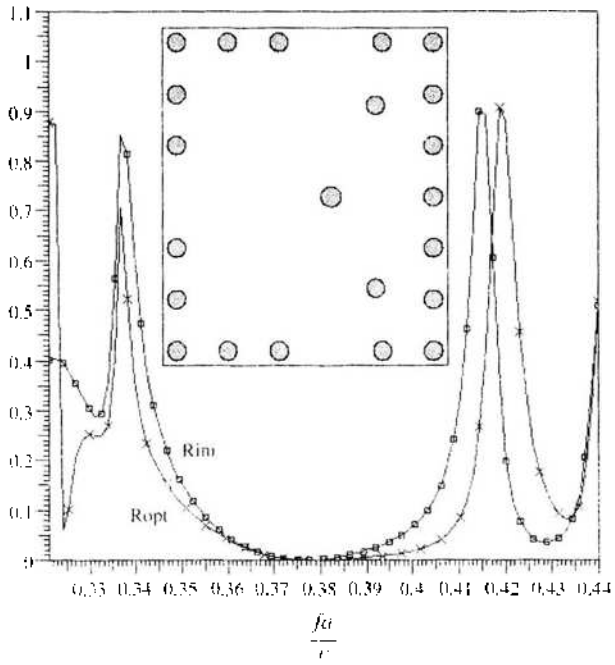


Figure 63. Further optimization of the best solution shown in Fig. 61. The three most crucial rods (positions and radii) of this power divider were optimized by a gradient optimizer in such a way that a broader band with low reflection is obtained.

At sufficiently high dimensions, Monte Carlo integration outperforms systematic integration routines. Because Monte Carlo routines rely on random numbers, it also becomes obvious that random search algorithms become more efficient than deterministic ones when the dimension of the search space is high enough. Unfortunately, we cannot specify a certain dimension limit for deterministic algorithms because this depends extremely much on the problem type, i.e., on the properties of the fitness function. Because CDM structure optimizations are very complex, deterministic optimizations can only be applied for very low-dimensional spaces with less than 10 dimensions.

In optimization problems, one often has fitness functions with many local maxima. Deterministic optimizers usually converge toward one of these local maxima. Whether an acceptable solution is found or not depends very much on the start point of the search that is usually specified by the user. In CDM applications, the fitness function often has very many local optima. For this reason, one only should start deterministic optimization when one is sure that one has an initial design that provides a sufficiently good start point. Because random search explores the search space in a uniform way in the sense that it does not focus on some special area, it always can find the global optimum or at least a solution close to the global optimum even when the fitness function is not a continuous function. Furthermore, random search is not limited to continuous search spaces like real number spaces. It also can be applied to discrete spaces including binary spaces.

Although random search is the simplest probabilistic optimizer that is even the best one for extreme cases with fitness functions that are discontinuous everywhere, this strategy is not efficient enough for most of the practical applications, even in difficult situations of CDM optimizations. For this reason, we search for better probabilistic optimizers that clearly outperform random search for typical CDM structures.

5.1. Probabilistic Optimizers

A probabilistic optimizer that outperforms random search must be somewhere between random search and deterministic search. The simplest way to obtain a search strategy between random search and deterministic search is to combine any deterministic algorithm with random search. For example, one can randomly set the start point of a deterministic search routine and repeat this for several start points. Because deterministic search is only efficient when the dimension of the search space is low enough and when the search space is at least locally continuous, such algorithms are not very general and not good enough for CDM structure optimization. We therefore start with random search and try making it smarter by adding some reasonable search strategies.

Nature is sometimes considered as a huge lab where extremely complicated systems are optimized in a very efficient way, which then explains why organisms like human beings can be obtained in nature, although the probability of obtaining such organisms by a random assembly of atoms is almost zero. This means that nature takes advantage of optimization strategies that outperform random optimization by far. Thus, one can try to analyze "natural optimization algorithms" and try to translate them in computer codes. This has really been done when several modern optimizers were designed, namely simulated annealing and algorithms that mimic the evolution. Because simulated annealing can also be considered as a most simple case of an evolutionary strategy, we focus on the latter.

The characteristics of evolution are the population-based procedures (i.e., one always has a population of similar individuals). An individual alone can learn, but it cannot evolve. A group of individuals, for example, cats, dogs, and so forth, is always characterized by genotypes that are modified from one generation to the next one by genetic operations, namely mutation and crossover (i.e., mixing of the chromosomes of two parents). This concept is implemented in genetic algorithms (GAs). On the level of species, evolution works in a similar way: One always has a population of different species like one has a population of individuals. Individuals may die and also species may die and new species occur like new individuals are born. Finally, new species are also created by mutations, but it seems that crossover is not used for obtaining new species (i.e., a new species is usually not a combination of two previous species). For this reason, GAs and evolutionary strategies (ESs) that

mimic the evolution of species exhibit some similarities as well as a fundamental difference. Because GAs [49] and ESs [50] are currently widely used for numerical optimization, we outline these procedures in the following.

5.1.1. Genetic Algorithms

When we translate the chromosomes into computer language, it is very natural to represent chromosomes, that is, the genotype of an individual, by a bit string, although natural DNA uses three rather than two pairs of molecules for storing the genetic information. Therefore, considering a model of our power divider as an individual characterized by a bit string is trivial. The bit string then is the genotype of the individual, and the waveguide discontinuity defects are determined by the different bits is the phenotype. The fitness definition then associates a fitness value to each individual. The first step of defining the genotype (i.e., the bit string) is trivial here; also, it is not unique at all because we can select the PhC cells to be optimized differently. For nonbinary optimizations, for example, when we want to use a GA to optimize a real-valued parameter like the rod radius, some coding is required (i.e., we must discretize the real value and associate a sequence of bits). The length of this sequence then depends on the accuracy of the discretization. When the accuracy is too low, it may happen that a good solution cannot be found, and when it is too high, the bit strings become too long. Therefore, a proper discretization and coding of real model parameters into bit strings requires some experience. Beside this, the procedure is very general.

Now, we consider a population of individuals rather than a single one. Because we usually know nothing about good solutions at the beginning, standard GAs randomly create a population of N_{pop} individuals. Then, the genotype of each individual decodes, and the corresponding numerical model may be solved by an appropriate field solver. From the results of the field solver, one then can evaluate the fitness value for each individual.

Once the fitness values for all individuals are known, one can analyze these values and try to take advantage of them in such a way that good solutions are more quickly found than with random search. This is usually done by selection mechanisms that favor the individuals in a population with highest fitness values as parents for obtaining the next generation. Here, one is very free in the definition of the selection mechanism. When one uses very strict mechanisms that keep only a few of the best individuals and discard all others, one usually obtains a rather rapid convergence toward a local optimum, but after a while, the progress of the algorithm becomes very slow. For this reason, one often uses not too strict selection methods that give also poor individuals a chance to become parents.

Once the parents for the next generation are selected, one can start generating the children using a crossover algorithm. Here, one is again rather free in the definition of the procedure. A simple technique is single-point crossover, where a position in the bit string is selected randomly. One then obtains two children from the two parents by exchanging the parts of the bit string before the randomly selected position as illustrated in Fig. 64.

After crossover, one may apply some mutation to the genotype of each of the children. Also here one is rather free. Usually, one applies mutation with a rather small rate (i.e., low probability to a randomly selected bit in the bit string). Higher mutation rates let the algorithm work more like random search.

Because we already have different choices like the population size, the selection procedure, the crossover definition, the crossover rate, the mutation definition, and the mutation rate, we already have rather many tuning parameters. More sophisticated GAs use additional techniques, namely elitism, dynamic mutation rates, and dynamic population sizes, which

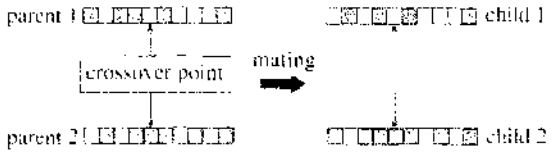


Figure 64. Basic principle of genetic programming—crossover and mating. First, a crossover point is randomly selected. After that, all bits of the two parents before the crossover point are exchanged.

leads to additional tuning parameters. The optimal selection of these parameters requires much experience. For obtaining experience, one often considers extremely simple mathematical fitness definitions that permit an extremely quick fitness evaluation. It is important to note that such fitness definitions for getting experience are similar to the procedure where a chess player analyzes a bunch of extremely stupid games played by primitive chess algorithms. In order to avoid this, we have studied RLC filter optimizations and similar examples that are much easier than CDM structure optimization but much closer to CDM structures because of several similarities. Because the fitness evaluation of RLC filters is quick, we can obtain enough experience from this for tuning the GA parameters mentioned above.

When we applied a standard GA with reasonable parameters to our simple PhC power divider, the result was extremely disappointing because the global optimum was often not found within 4096 fitness evaluations, whereas the random search finds it within 2048 fitness evaluation in the average. The main reason for the bad performance of the GA is that it frequently recomputes individuals near a local optimum several times. Because the fitness evaluation is extremely time-consuming in the case of CDMs, we therefore should avoid duplicate fitness evaluations. This can easily be done by maintaining a table of all known fitness evaluations. A GA with such a table turned out to be still not better than random search when we insist on finding the global optimum. The reason is that the algorithm tends to focus on a local optimum found in an early stage. From this, it becomes clear that even a simple CDM structure can be very hard for numerical optimizers.

Because standard GAs require too many fitness computations for most of the engineering applications, MicroGAs were developed. MicroGAs outperform the standard GAs in most cases. They are based on extremely small populations of typically 5 five individuals only, which is clearly not a realistic imitation of nature. Furthermore, the convergence is increased by entirely omitting mutation, which makes the algorithm simpler. After a few generations, all individuals are identical and the algorithm stops because no new individuals can be found from crossover of identical parents. One then reinitializes the entire population except one of the individuals that is copied from the previous generation. Thus, one essentially moves much from the stochastic part to the reinitialization. The probability of reevaluations of individuals in microGAs is even much higher than in standard GAs. Therefore, it is extremely important to maintain tables of known fitness evaluations. Surprisingly, such tables are not available in well-known microGAs. However, with our table-based microGA, we were able to outperform the random search in the average (i.e., we found the global optimum within around 1200 fitness evaluations in the average). Although this is encouraging, we consider it to be not good enough for efficient CDM optimization. Therefore, we look for alternatives in the following.

5.1.2. Evolutionary Optimizations

There are many different types of evolutionary optimizations. GAs are probably the most frequently used category. We have already mentioned that the coding of real parameters into bit strings for GAs provides some difficulties. Because one can consider other objects than bit strings as genotypes, a GA can also be extended to nonbinary forms. Namely, one can use integer arrays or even real arrays instead of bit strings. When this is done, appropriate genetic operators must be defined. The mutation provides no essential problems, but when we define a crossover like in the binary case, we see that it cannot modify any of the real parameters. For this reason, mutation becomes much more important.

Evolutionary strategies (ESs) mimic the evolution of species, where crossover is assumed to pay no role. Therefore, an individual in an ES corresponds to a species. Usually, each individual is characterized by a vector in the real search space (i.e., ES focuses on the optimization in an N-dimensional real space). Like in GAs, an initial population is obtained by random creation, and like in MicroGAs, one often works with small populations. Once all individuals of a generation have been evaluated (i.e., the fitness values have been computed), one can proceed with a selection procedure as before. Usually, a rather strict selection is done (i.e., only a small fraction of the individuals survive). According to the literature and our experience, it seems to be best to keep approximately 1/7 of the previous population and to create six new individuals by mutation from each surviving individual.

One of the important steps is now to define mutation. Adding a random vector $\vec{d}$ to a given vector $\vec{r}$(describing an individual) would not focus the search on the neighborhood of a good individual. Therefore, one uses standard deviation with a given variance s for generating the mutation vector $\vec{d}$. Because it can be favorable to have different variances in the different directions of the search space, one may also use a vectorial variance $\vec{s}$. When one comes closer to a local optimum, the variance should be reduced for speeding up the convergence. Therefore, both $\vec{r}$ and $\vec{s}$ should be optimized simultaneously (i.e., each individual is characterized by the location $\vec{r}$ and the variance). One therefore needs an additional algorithm for the evolution of the variance. In addition to these algorithms, also crossover operators and more sophisticated algorithms were invented. According to our experience, ES always outperformed GA for realistic parameter optimization in real-valued N-dimensional search spaces. For this reason, it is certainly attractive to develop ES-like algorithms also for binary search spaces. We will do this below.

5.1.3. Other Methods

In addition to GA and ES, many alternatives were proposed and studied. A prominent one is simulated annealing (SA), which mimics crystallization mechanisms known from physics. In fact, SA can be considered as the simplest case of an ES with a single individual. Selection necessarily selects and mutates this using standard deviation. Then, one decides whether the parent or the child shall survive. In SA, the temperature controls the variance. Unlike in ES, the "evolution" of the temperature is usually defined by the user, which provides an additional burden for the user. As a consequence, only experienced SA users are able to outperform the simple one-individual ES.

Alternatives to GA and ES are also obtained when one mimics the search strategies of groups of "smart animals." This leads to concepts like *ant colonies* and *particle swarm optimization* (PSO) that mimic the search procedures of swarms of birds and fishes, where each individual is described by a position vector, but also by velocity and acceleration vectors. Furthermore, it is assumed that there is some communication between the individuals (i.e., an individual can know when another one is at a better position). Although PSO is more complicated and seems to be smarter than ES, we never found a PSO implementation that outperformed ES when realistic problems were considered. For this reason, GA and ES are our favorites among the well-known probabilistic optimizers.

5.1.4. Development of New Methods

When we observe how a MicroGA optimizes our simple PhC power divider, we see that it converges toward a local optimum. Then, random initialization (i.e., a completely "blind" technique) is applied. This technique does not take advantage of any knowledge that might be obtained by an intelligent observer. Thus, it may happen that the algorithm restarts in an area that is already known to be not promising. The standard MicroGA does not "know" this, because it always completely deletes all information of previous generations. Because our table-based MicroGA does not discard this information, it is much faster although it does not really take advantage of the contents of the table in a very smart way (i.e., it simply checks in the table if the fitness value of the current individual is already known).

How can a "smart" observer take advantage of the fitness table? Because this table tends to become huge, it is certainly reasonable to do some statistical analysis. Furthermore the bits of each bit string (that corresponds to a fitness value in the table) are not meaningless because they correspond to the defects in the crystal. Therefore, each bit has some meaning. There are certain cells in the structures where it is very likely that a defect should be there, cells where this is very unlikely, and cells where one does not know. One therefore can try to find out for each bit how likely it should be either 0 or 1. As soon as the table contains sufficiently many entries, one can analyze it as follows. For each bit, we compute the average fitness a_1 of all individuals with bit equal to 1 and the average fitness a_0 of all individuals with bit equal to 0. When the ratio $r = a_1/a_0$ is big, it is very likely that the bit should be 1. When it is small, the bit should probably be 0, and when it is near one, we do not know. From this, we can define a bit fitness for each bit, for example, as $b = 2 * \mathrm{Arctg}(\ )/\pi$.

Note that this definition is not unique at all. As soon as we have allocated a reasonable bit fitness value b_i to each of the bits in the string, we can take advantage of this information when we manipulate any bit. For example, instead of randomly creating a new individual by the instruction "create a random number R uniformly distributed in the interval 0...1; set the bit equal to 1 when R > 0.5 and 0 otherwise," we can simply use the instruction "create a random number R uniformly distributed in the interval 0...1; set the bit equal to 1 when $R * b_i > 0.5$ and 0 otherwise." This small modification may considerably speed up the table-based MicroGA.

We know from our experience that ES usually outperforms GA in real parameter optimization, and we also obtained the impression that crossover is not very helpful for our CDM optimizations. For this reason, we developed several table-based algorithms without crossover that mimic ES. The simplest one is the following: (1) Create a random population. Here, seven individuals were best for our power divider and similar applications. (2) Evaluate the fitness values, store them in the table, and perform the statistical analysis for obtaining the bit fitness values. (3) Keep the best individual only and fill the populations with single-bit mutations of it. Take advantage of the bit fitness values for this mutation (i.e., mutate the bit that is most likely wrong with a higher probability). Here, one should make sure that all new mutations are different from each other. (4) Repeat step 3. Usually it is reasonable to display continuously the best solution that was found and to let the user stop the procedure when he is satisfied with it. Otherwise, the procedure is stopped when all possible configurations were evaluated. For our simple power divider test case, this strategy outperformed the table-based Micro GA by far. In the average, it found the global optimum within less than 300 fitness evaluations. It is interesting to note that it also outperformed a table-based MicroGA with bit fitness evaluation, which supports our impression that crossover is not helpful here. Despite this, we would like to take advantage of even faster optimization schemes, and we hope that appropriate solutions for the efficient CDM structure design will soon become available.

5.2. Deterministic Optimizers

When we have found a promising structure from experience, from intuitive design, or from a probabilistic optimizer, we can apply some deterministic optimizer (i.e., the goal of such an optimizer will be fine-tuning rather than design). When the initial design (i.e., the start point for the deterministic search) is not close enough to an optimum, there is little hope that the deterministic optimizer can find it within reasonable time (i.e., with not too many of the time-consuming fitness evaluations). Because deterministic optimizers are well-known, we only focus on the aspects that are important for the CDM structure optimization.

First of all, deterministic optimizers become numerically more efficient when the first derivatives (gradient vector) of the fitness function or even the second derivatives (Hessian matrix) are known. In an N-dimensional search space, the gradient has N and the Hessian matrix N^2 elements. CDM structures are always so complex that neither the gradient nor the Hessian matrix is known analytically. The numerical approximation of any derivative requires at least one additional fitness evaluation (in addition to the evaluation in the point where the derivative is required). For the second-order derivatives in the Hessian matrix, at least $2N^2$ additional fitness evaluations are required. This causes a huge numerical effort. Therefore, one should work with algorithms without an explicit computation of the Hessian matrix.

A second problem associated with numerical derivatives is that small steps are required for obtaining a good accuracy of the gradient (i.e., one must compute the fitness values of $N + 1$ models that are only slightly different). When the numerical accuracy of the fitness evaluation is not high enough, the gradient approximation becomes inaccurate, and this will disturb the optimization procedure very much. For this reason, we favor accurate field solvers for numerical optimizations.

The third problem is that one should keep the dimension of the search space as low as possible. Experienced users might know what cells in a CDM structure have a strong influence on the overall fitness. For waveguide discontinuities, it is rather obvious that only

the cells near the discontinuity should be optimized, but even this recommendation usually leads to too high values of N. For this reason, one best computes the gradient for the initial model for all parameters that might have an important influence. This can be numerically demanding, but it is helpful not only for the numerical optimization but also for the fabrication as we will see below. Once the gradient for the initial model is known, one can focus on the most important parameters (i.e., those directions in the original search space with the biggest components of the gradient). In many cases, one then can only optimize a few parameters as we will see below.

5.3. Sensitivity Analysis

Assume that we have any initial model obtained from a previous probabilistic optimization (Fig. 62) or from intuitive design (Fig. 45 [30]) or from some detailed waveguide analysis (Fig. 51 [51]). We now want to improve the design using deterministic optimization. In the simple examples mentioned above, we only have rod-type PhCs with circular rods. In order to keep the optimization space small, we only modify the coordinates or radii of the rods near the discontinuity. We now can compute the gradient for the most promising positions and display the results as in Fig. 65. This allows us to reduce further the search space (i.e., to focus our optimization on the biggest components of the gradient). For more information, see Ref. [51].

A relatively big component of the gradient indicates that the fitness will vary strongly when we modify the corresponding parameter in our model. For example, the derivative of the fitness function with respect to the radius of the rod at the outer corner of the 90-degree bend is relatively strong, as one can see in Fig. 66. This indicates that the optimization of this radius is promising. At the same time, this also means that an inaccurate fabrication of the corresponding rod might make the entire structure useless when the derivative is big enough (i.e., the strengths of the components of the gradient indicate how sensitive the corresponding model parameter will be with respect to fabrication tolerances).

Sensitive model parameters are crucial for the fabrication at least as long as CDM structures cannot be fabricated with low tolerances. Parameter optimizations do not only improve the fitness of a structure. They usually also tend toward solutions that exhibit extremely sensitive parts. When the fabrication tolerances are not good enough, this will spoil the optimized solution. For this reason, one should not only perform a sensitivity analysis of the initial design but also a sensitivity analysis of the optimized solution (Fig. 45 [30]). Depending on the quality of the fabrication, it may even be reasonable to focus on suboptimal solutions with lower sensitivity.

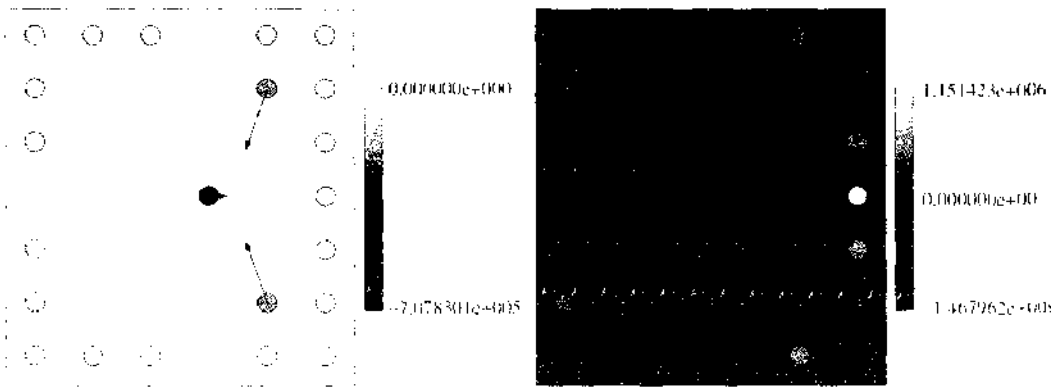


Figure 65. Photonic crystal power divider sensitivity analysis. Sensitivity analysis of the positions of the three most important rods of power divider (left) and of the radii of all surrounding rods (right). The reflection coefficient R at the input port is used to define the fitness function: Fitness = 1 − R. Long arrows (left) indicate that the rods must be moved far away from the initial position for improving the fitness. Dark and bright rods (right) are sensitive with respect to the corresponding radii. The fitness is increased when the radius of a dark rod is increased or when the radius of a bright rod is decreased.

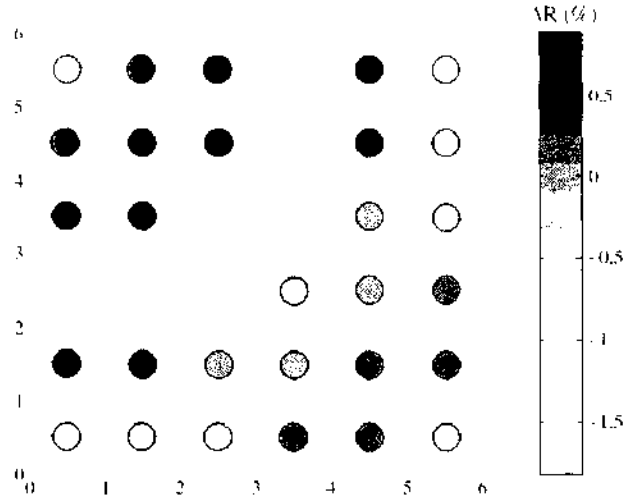


Figure 66. Sensitivity analysis of the 90-degree photonic crystal bend. The radii of all rods are changed by 5%. The resulting change of the reflection coefficient is shown by the grayscale.

Although crucial (i.e., sensitive) parts of a structure cause fabrication problems, they also provide attractive aspects. First of all, there are techniques for performing some modifications of given CDM structures, for example, focused ion beam (FIB) technology. This allows one to trim a structure that was fabricated. When this shall be done efficiently, one should obviously try to modify the sensitive parameters of the structure. Furthermore, it is possible to modify, for example, the permittivity of a rod in a PhC structure or of an area in a CDM cell by an external electric field, heating, mechanical forces, and so forth. This allows one to tune a CDM structure. Because the effect of changing the material property of a rod is very similar to the effect of changing its size, the sensitivity analysis also shows the way to the most efficient tuning. Finally, highly sensitive areas also open the door to attractive CDM-based sensors.

5.4. Towards Phenomenological Two-Dimensional Modeling of Planar Photonic Crystal Devices

Present planar PhC realizations are typically based on a semiconductor substrate material technology (AlGaAs/GaAs or InGaAsP/InP) that allows the integration into passive and active optoelectronics devices for optical telecommunications. Due to the small index contrast as provided by these semiconductor material systems, such substrate-type realizations of planar PhCs usually suffer from weak vertical light confinement leading to considerable out-of-plane scattering. Alternatively, strongly confining membrane-type PhC devices using either silicon on insulator (SOI) layer systems, or Si, InP (see Fig. 67), or GaAs in air as membrane structure, have already shown record low out-of-plane scattering loss [52]. Due to technological reasons, hole-type PhCs are much easier to fabricate than the rod-type counterpart that is mostly used within the conceptual studies as elucidated along our previous examples. Regarding the 2D lattice symmetry, triangular lattices are favored because they offer larger photonic band gaps (PBGs) at the same air-hole filling factor f compared to, for example, a corresponding square lattice, where f denotes the air hole's volume fraction (i.e., the cross section of the air hole normalized by the area of the PhC 2D lattice's primitive cell).

Even if our conceptual studies using generic examples are among the very first numerical structural optimization of PhC devices, a question still remains open, namely whether a 2D device design is reliable enough to cope with 3D planar PhC realizations. A contemporary "real-world" design based on planar PhCs has therefore to address the following aspects: First, device concepts may be explored using efficient 2D computational optics tools such as, for example, MMP, where promising PhC device topologies may emerge from 2D structural optimizations as proposed earlier. Here, corresponding phenomenological models [53] have

Figure 67. 3D-FDTD simulation of a W1-waveguide (one line of vacancies) that is implemented in an InGaAsP membrane-type PhC. The TE-polarized mode is excited in front of the vertical plane. The residual substrate radiation is an undesired feature of the modal mismatch at the waveguide junction (left). SEM image of a fabricated membrane-type PhC W1-waveguide. The membrane consists of a 434-nm-thick InGaAsP core layer surrounded by air (right).

already proven to be best suited to bridge the gap between a realistic (planar) PhC structure and its proper 2D representation. Second, the simulation of realistic PhC devices requires true 3D modeling capabilities that are numerically much more demanding. Emergent planar PhC device topologies are thus evaluated using, for example, 3D-FDTD or more efficient and more accurate 3D PhC simulation tools that have still to be developed. Despite the severe challenges posed by such realistic device design, we are prone to claim 2D structural optimization as highly appropriate to exploit the peculiar nature of PhCs toward most compact device topologies.

5.4.1. The Phenomenological 2D Model

One major advance of the phenomenological model lies in its proper terminology: it can be grounded on a real-world PhC characterization, such as, for example, the internal light source (ILS) technique. The ILS technique was introduced at the Ecole Polytechnique Palaiseau in 1996 to study successfully substrate-type, planar PhC structures that are deeply etched in a GaAs layered material system. It has been further developed at the Swiss Federal Institute in Lausanne (EPFL) and is now among the most reliable means to characterize accurately planar PhCs [53]. The ILS technique is based on the photoluminescence (PL) of quantum dots or quantum wells that have to be present in the material system and that are excited by a focused pump laser beam. The excitation spot is positioned in front of the PhC, enabling light propagation through the PhC bulk section, whereas the transmitted light is collected by a reflecting achromatic objective at the rear face of the chip. Facet reflections, multiple output signals (such as, e.g., the substrate radiation) and the proper renormalization of potential propagation paths defines the major challenges of the ILS setup. To cope with the large PBG lithographic tuning (i.e., simultaneously up- or down-scaling of the lattice constant a and the hole radius r while keeping the filling factor f constant) has been applied because of the limited bandwidth provided by PL spectra ($\delta\lambda \approx 150$ nm for GaAs and $\delta\lambda \approx 100$ nm for InP, respectively).

In order to dispose of a comprehensive data set, the sample is usually characterized along both the Γ-M and Γ-K crystal orientation. It is worth mentioning that the ILS is the only characterization setup that is capable of providing absolute data at very high accuracy. The measured transmission spectra is then compared to the spectra provided by a corresponding 2D simulation model (FDTD [53] or MMP), where the out-of-plane scattering loss is represented by an imaginary contribution ϵ_{im} to the dielectric constant of the air holes. In general, the air band (upper band edge of the PBG) is more sensitive to losses than the dielectric band (lower band edge), as the light is mainly concentrated in the air holes. The fitting procedure actually accounts for the total intrinsic loss where a separation of all constituents such as material absorption and the various out-of-plane scattering contributions

has to involve an additional model. The interrelation between the out-of-plane scattering loss and the hole quality is thus developed using an advanced perturbation model [53], taking into account both the hole depth and the proper cylindroconical hole shape. The background material in the 2D model is assumed to coincide with the effective refractive index of the planar multiplayer slab, where the latter provides the vertical confinement in the planar PhC structure.

A numerical 3D analysis of the ILS measurement scenario is depicted in Fig. 68. When comparing 2D-MMP and 3D-FDTD [54] field simulations, we observed that (a) in-the-gap propagation turned out to be negligible, whereas (b) the hole diameter in the phenomenological model has to be increased by 3% in order to reproduce the PBG of either the measured data or the 3D simulations. While relating the band gap signatures of measured transmission spectra to statement (a), one has to keep apart the underlying mechanisms that contribute to similar band gap features, namely, in-the-gap propagation from the PhC sample's improper in-coupling or poor facet quality.

5.4.2. Modeling of Planar Photonic Crystal Tapers

The incoupling of light into a PhC is still a task that has to be mastered before the predicted advantages of PhCs in ultradense guiding of light can be exploited. Tapers making the transition from a W3-waveguide (a defect channel in the Γ-K-direction having a width of three vacancies) to a W1-waveguide (a single line defect acting as a light channel in the Γ-K-direction) have turned out to provide a feasible interface between the PhC device and its accessing waveguides. The width of a conventional deeply etched ridge waveguide is around a few micrometers, while a W1-waveguide has a width of only a few hundred nanometers. Therefore, the conventional ridge waveguide has to be tapered down to the width of a W3-waveguide, whereas a subsequent confinement is conducted within the PhC providing a transition from a W3- to the W1-waveguide while gradually altering the sizes of the corresponding holes. The material system under consideration is a InGaAsP/InP slab structure. The filling factor of the implemented 2D PhC lattice has a value of 39%, which opens up a bandgap between $c/a = 0.20$ and 0.35 for TE polarization.

The conceptual taper studies are performed in the frequency domain using the 2D-MMP code MaX-1 in conjunction with the phenomenological model for the bulk PhC with a loss parameter of $\epsilon_{im} = 0.14$. For the interface, a linear tapering over five stages is assumed, starting with a filling factor of 20% corresponding to the smallest hole diameter (160 nm) that is producible without any considerable lag effect with respect to the hole depth. Without taking out-of-plane loss into account, the 2D simulation of a 5-ary linear taper structure yields a maximal power transmission of 94% and power reflection of 2.2%, whereas the residual 3.8% is attributed to in-plane power leakage (Fig. 69). The subsequent 3D-FDTD simulation of the corresponding realistic taper version exhibits a significantly lower power transmission of 72% comprising the considerable amount of out-of-plane scattering.

Figure 68. 3D-FDTD simulation of an ILS experiment. The excitation is modeled by a TE-polarized waveguide mode coming from the left. Here the transmission of light along the Γ-K direction of a PhC bulk section with a thickness of 20 rows is used to retrieve the effective hole size for the phenomenological 2D model. The out-of-plane scattering loss is translated into a loss parameter ϵ_{im} that is derived from a corresponding fitting procedure.

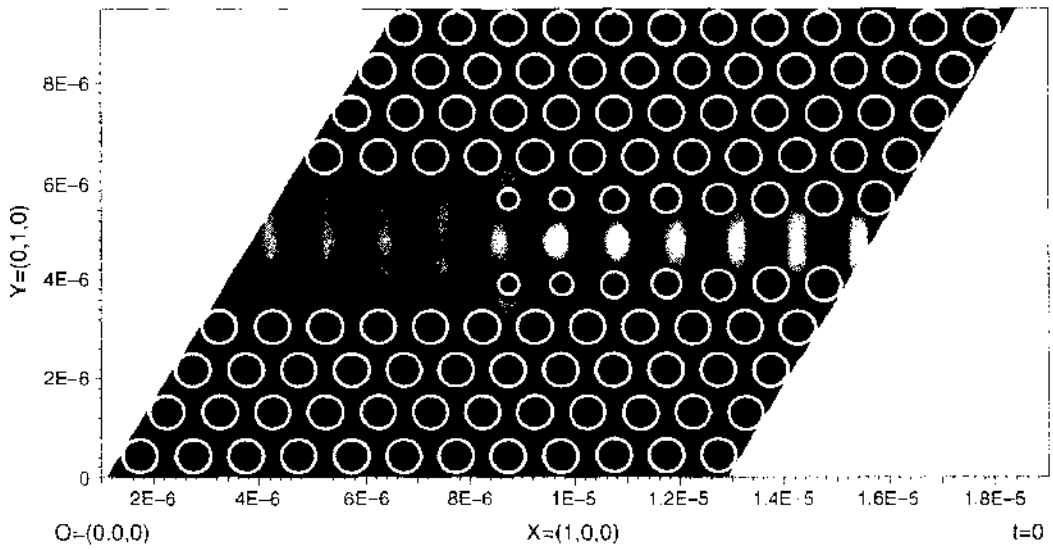


Figure 69. Improving the in-coupling of light into a PhC: The light is excited with the TE wave from the left. The picture shows the Hz-field, and for the lossless case the taper provides a power transmission of 94%.

Smoother transitions are obtained when starting the 5-ary linear tapering from even smaller hole diameters (with correspondingly reduced hole depths). To account for the lag effect, the refractive index of the affected air holes are slightly increased [55] following a perturbation argument. The 3D-FDTD analysis shows a distinct increase in power transmission up to 84.4%, whereas the lossless 2D model delivers a value of 90.5%, which is less than the uncorrected one. This somehow counterintuitive behavior is attributed to the presence of a so-called mini-stopband that divides the PBG into two parts. It has been observed that the periodic nature of the PhC waveguide boundaries can give rise to contradirectional coupling between modes of the same orders at the Brillouin zone edge (i.e., Bragg scattering) [56]. Because of the presence of higher order modes in hole-type W1-waveguides, contradirectional coupling also occurs between the fundamental and higher order modes [56, 57]. The resulting sharp transmission dips are termed *mode gaps* or *mini-stopbands* (MSBs). To exploit properly the PhC taper's efficiency, one has to direct the design to an appropriate operation wavelength on the opposite side of the MSB (i.e., at the low-energy side). The resulting taper provides a power transmission spectra peaking at 100% with a nearly flat transmission band centered at $c/a = 0.252$ having a 90%-bandwidth of around 10%.

The remaining discrepancies between the 2D and the 3D models are now clearly assignable to the omission of the radiation loss contribution. In order to set up a reliable 2D device model for the taper structure, the phenomenological loss parameter $\epsilon_{im} = 0.14$ has to be assigned to the underlying PhC lattice. The resulting 2D simulation still does not coincide with the transmission spectrum of the 3D analysis, because the proper tapering operation actually represents a symmetry breaking with respect to the propagation direction, and hence increasing the number of potential out-of-plane scattering channels. Accurate modeling should therefore include some additional loss contributions that are localized in the proper tapering area. To obtain a refined 2D taper model, the output of the 2D simulation was fitted to the transmission spectra of the 3D reference model while assigning additional loss contributions to the air holes involved in the tapering process (Fig. 70). Perfect agreement has been achieved when the loss parameter ϵ_{im} in the first air hole of the 5-ary linear taper is increased by 79% and the subsequent two by 39%.

Additional ideas [58] to provide simultaneously tapering and smooth S-bending are depicted in Fig. 71. Even if this very special concept still needs further refinement and most of the parts are experiencing highly multimodal fields, one may intuitively deduce that the small holes within the local taper sections are exposed to larger light intensities than the regular ones underpinning our approach to assign locally different values for the loss parameter in the phenomenological 2D model.

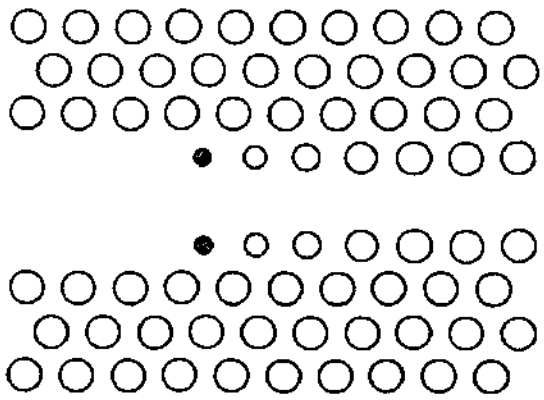


Figure 70. Modified phenomenological 2D taper model. A loss parameter of $\epsilon_{im} = 0.14$ is assigned to all holes of the underlying PhC (holes with white filling). Modifications comprise an increase of the loss parameter to $\epsilon_{im} = 0.25$ (dark filling) and $\epsilon_{im} = 0.18$ (light filling), respectively.

5.4.3. Planar Photonic Crystal Bends and Power Splitters

As a further example we have investigated and optimized 60-degree PhC waveguide bends that are implemented in a planar photonic crystal (PhC) with triangular lattice symmetry [59]. The in-plane guiding within the planar PhC stucture is based on a W1 defect waveguide, whereas for the vertical light confinement we rely on a slab waveguide formed by the low index contrast material system InGaAsP/InP. To achieve a reasonable bandgap around 1.55 nm, the planar PhC consists of a lattice of holes with a filling factor of 39%. Our conceptual device analysis is again carried out along the phenomenological 2D model in the frequency domain using the 2D-MMP method.

In contrast to the 90-degree bend of the generic square lattice case, the initial 60-degree PhC waveguide bend shows a very narrow transmission bandwidth of only 0.0065 around a normalized frequency of $c/a = 0.245$ laying at the low-energy side of the MSB (Fig. 72). The power transmission peaks at a value of around 94% for the lossless 2D case. A sensitivity analysis with respect to small displacements of lattice sites was performed for the most crucial holes around the proper bending region in order to increase the transmission efficiency and to enhance its bandwidth. Here the high-energy side proximate to the MSB has turned

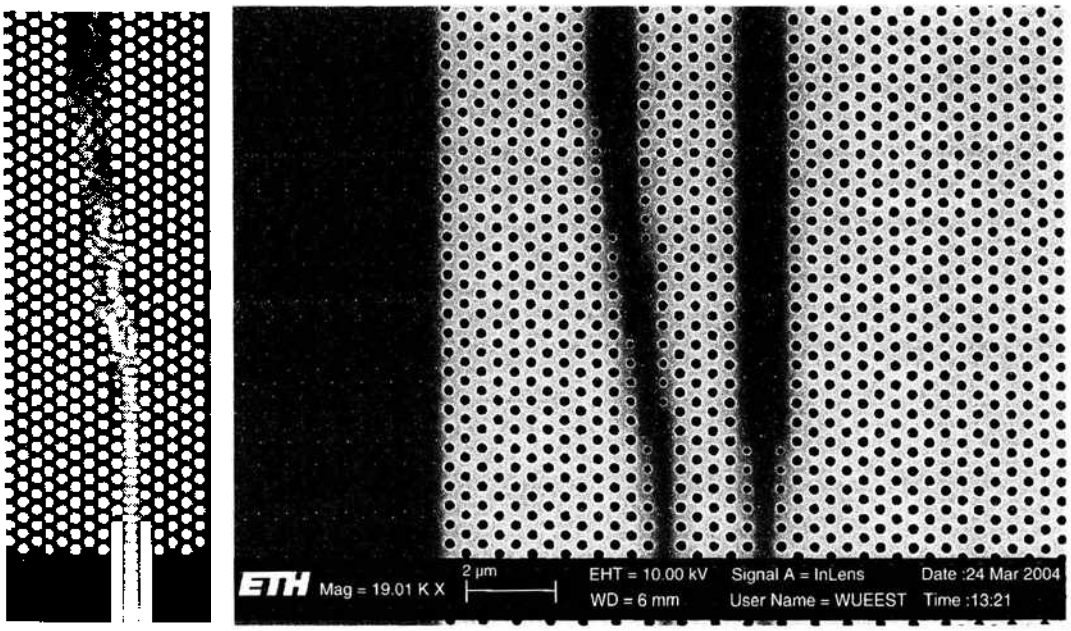


Figure 71. Top view on staggered PhC taper structures. Left: The 3D-FDTD simulation shows a power transmission around 70%. Right: SEM picture of the patterned device containing both the staggered taper and a regular taper structure.

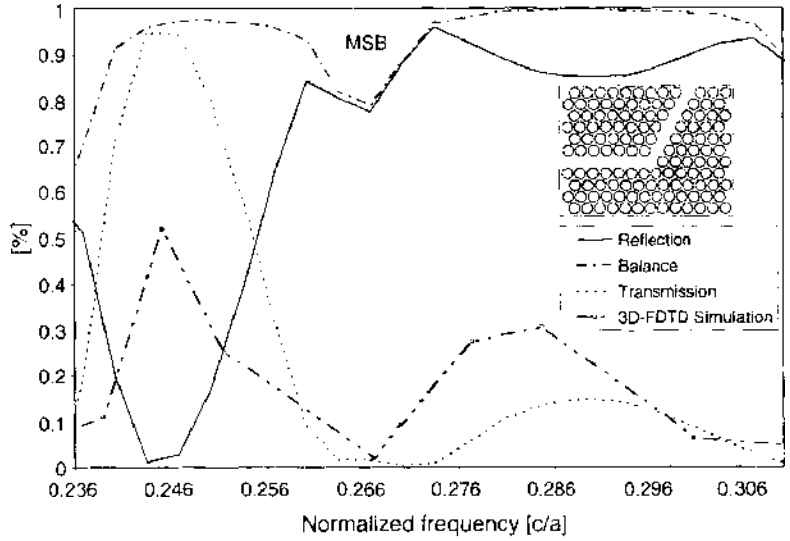


Figure 72. Frequency response of the initial 60-degree PhC waveguide bend. As a characteristic feature of the implemented 2D MMP model, the MSB is indicated by a dip in the spectral response of the power conservation. Regarding an efficient bend design, the high-energy side proximate to the MSB ($c/a = 0.27 \ldots 0.30$) turns out to be more promising with respect to proper optimization.

out to be more promising with respect to our optimization task. After several optimization steps, we finally succeeded to find a bend topology (see Fig. 74a) that improves the values for the power transmission around $c/a = 0.287$ from 14.48% (power reflection 85.35%) up to 96.8% (power reflection 2.84%) providing an overall transmission of more than 86% for a normalized frequency range of 0.02, which corresponds to a wavelength range of 290 nm. While adopting the (lossy) phenomenological 2D model with $\epsilon_{im} = 0.14$, the overall transmission spectrum drops nearly homogeneously according to the peak power transmission decrease from 96.8% to 85.7%. Nevertheless, 3D simulations predict a power transmission of below 66%. The discrepancy between the latter two values is again assigned to additional local loss contributions in the bending region of the 3D bending structure. This is also indicated by the anomalous field distribution in the proper bending region (see Fig. 74b). Nearly perfect agreement between the transmission spectra of the 60-degree bend's 3D model and the modified phenomenological 2D model has been achieved within the frequency band of interest when charging the two opposite holes at the outer and inner side of the proper waveguide "knee" with a loss parameter of $\epsilon_{im} = 0.5$ and some of their neighboring holes with a lower value of $\epsilon_{im} = 0.3$ (for the proper assignment, see Fig. 73). From this lesson, one may also learn that further improvements in power transmission are only achievable if the inherent resonant feature of the bending area is inhibited. Preliminary attempts toward non-resonant 60-degree bends are presented using a so-called steering effect [60]. A convincing

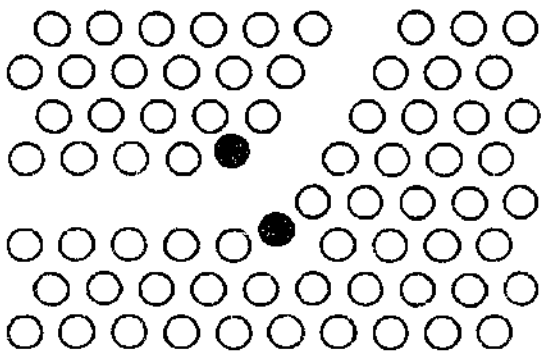


Figure 73. Modified phenomenological 2D model of the 60-degree waveguide bend. A loss parameter of $\epsilon_{im} = 0.14$ is assigned to all holes of the underlying PhC (holes with white filling). Modifications comprise an increase of the loss parameter to $\epsilon_{im} = 0.5$ (dark filling) and $\epsilon_{im} = 0.3$ (light filling), respectively.

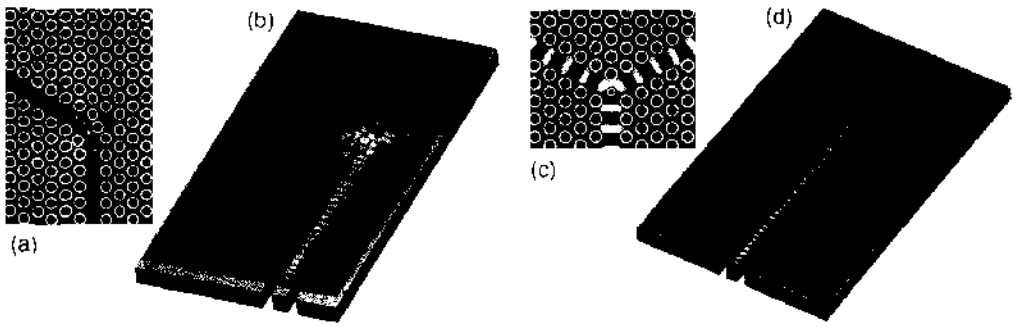


Figure 74. InP PhC device simulation of the transversal H-field for TE excitation. Optimized 60-degree bend with (a) 2D-MMP and (b) 3D-FDTD; optimized power splitter with (c) 2D-MMP and (d) 3D-FDTD.

demonstration of the aforementioned design principle has been recently provided in the framework of a 120-degree S-bend [61].

The reasoning for additional localized loss contributions due to local symmetry breaking may gain additional support from another device topology, namely the symmetrical PhC power splitter. Here, the strength of asymmetry at the proper branching point seems less severe because mirror symmetry with respect to the optical axis of the input waveguide is still preserved. This argument becomes intuitively plausible when comparing the localization of the field distributions in the asymmetry points of both topologies, that is, the 60-degree bend (Fig. 74b) and the symmetric power splitter (Fig. 74d), respectively. It is still the subject of our further research whether other characteristic features are assignable to simple symmetry considerations with respect to the topology of the involved waveguide channels. For instance, a preliminary study has demonstrated the PhC power splitter's inherent potential for large operation bandwidths including resonant and nonresonant realizations [62, 63]. Our optimized power splitter (Figs. 74c and 74d) relies on an increased-index type of defect in the splitting region, giving rise to a relative 90% bandwidth of 29.2% (i.e., 70% of the PBG), which is actually split in two bands by the presence of the MSB. The resulting splitter design yields nearly ideal performance when showing power transmission efficiencies for both arms of 44% (2D model) or 42% (3D model), respectively.

6. CONCLUSIONS

Metamaterials (i.e., artificially manufactured structures that may be considered as new materials that do not exist in nature) have attracted increasing interest by the scientific community during the past decade. Thanks to the latest achievements of nanotechnology, metamaterials also became very promising in the optical regime. Here, periodic metamaterials such as photonic crystals might open the door to new, ultradense integrated optical devices with a wide range of potential devices.

In order to make metamaterials attractive from the engineering point of view, additional techniques, namely composing and doping, are required. This leads to the rich, new research area of composite doped metamaterials (CDMs). We have demonstrated that CDMs often exhibit counterintuitive behavior and that the design of such structures based on experience and intuition usually leads to suboptimal structures that are not good enough from the engineering point of view, that is, such structures are attractive in physics because they show that CDMs may be used for obtaining optical waveguides, couplers, filters, and so forth, but these devices are not competitive with existing electronic counterparts.

Because of the lack of experience, intuition, and appropriate design rules, and because of the highly complicated physical mechanisms inside CDMs, extensive numerical simulation is currently the only feasible method for CDM design. From numerical field solvers linked with appropriate numerical optimizers, we have found structures that are much better than those obtained from intuitive design. These optimizations exceed our initial expectations and make it very likely that CDMs will be a key technology for integrated optics as soon as appropriate fabrication processes are established.

In this chapter, we have presented a computer-based design strategy for CDM structures that is necessary because of the complexity of such structures. Because this complexity is caused by strong interactions that also result from the densification of nonoptical structures, we are convinced that similar design strategies will replace traditional ones in all areas of engineering where densification is an important issue and especially in all areas of nanotechnology.

REFERENCES

1. G. Marconi and C. S. Franklin, U.S. Patent No. 1, 301, 473 (1919).
2. C. Mias, C. Tsakonas, and C. Oswald, "An Investigation into the Feasibility of Designing Frequency Selective Windows Employing Periodic Structures (Ref. AY3922)," The Nottingham Trent University, Nottingham, 2001. Available at http://www.ofcom.org.uk/static archive/ra/topics/research/topics/emc/ay3922/ay3922.pdf.
3. D. Rittenhouse, "American National Biography Online," Oxford University Press, Oxford, UK. Available at http://www.anb.org/articles/13/13-01396.html (2003).
4. C. Palmer, "Diffraction Grating Handbook," Richardson Grating Laboratory, Rochester, NY. Available at http://www.optics.arizona.edu/Palmer/OPTI380B/PDFdocs/RGL_Diff_Grat_Hdbk.pdf (1997).
5. J. C. Bose, *Proc. Roy. Soc.* 63, 146 (1898).
6. I. Tinoco and M. P. Freeman, *J. Phys. Chem.* 61, 1196 (1957).
7. V. G. Veselago, *Sov. Phys. Usp.* 10, 509 (1968).
8. J. B. Pendry, *Phys. Rev. Lett.* 85, 3966 (2000).
9. E. Yablonovitch, *Phys. Rev. Lett.* 58, 2059 (1987).
10. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, "Photonic Crystals–Molding the Flow of Light," Princeton University Press, Princeton, NJ, 1995.
11. Ch. Hafner, *J. Opt. Soc. Am. A* 12, 1057 (1995).
12. L. Novotny, Light propagation and light confinement in near-field optics, Thesis ETH 11420, Zurich (1996).
13. E. Moreno, MMP modeling and optimization of photonic crystals, optical devices, and nanostructures, Thesis ETH 14553, Zurich (2002).
14. J. P. Kottmann, Numerical study of plasmon resonant nanowires, Diss. ETH 14247, Zurich (2001).
15. J. A. Kong, "Electromagnetic Wave Theory," EMW Publishing, Cambridge, MA, 2000.
16. V. Losada, R. R. Boix, and F. Medina, *IEEE Trans. Antennas Propagat.* 49, 1603 (2001).
17. D. L. Jaggard and Ph. Pelet, *IEEE Trans. Antennas Propagat.* 40, 1273 (1992).
18. P. Valanju, *Phys. Rev. Lett.* 88, 187401 (2002).
19. N. Garcia and M. Nieto-Vesperinas, *Phys. Rev. Lett.* 88, 207403 (2002).
20. C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, *Phys. Rev. B* 65, 201104(R) (2002).
21. D. Correia and J.-M. Jin, *Microwave Opt. Technol. Lett.* 40, 201 (2004).
22. C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, *Opt. Express* 11, 746 (2003).
23. Ch. Hafner: "Post-Modern Electromagnetics: Using Intelligent Maxwell Solvers," John Wiley & Sons, Chichester, 1999.
24. J. B. Pendry, *Phys. Rev. Lett.* 84, 4184 (2000).
25. S. A. Tretyakov, *Microwave Opt. Technol. Lett.* 31, 163 (2001).
26. Ch. Hafner, "MaX-1: A Visual Electromagnetics Platform," John Wiley & Sons, Chichester, 1998.
27. E. Yablonovich, *Phys. Rev. Lett.* 58, 2059 (1987).
28. B. Z. Katsenelenbaum, "Theory of Non-Regular Waveguides with Slowly Varying Parameters," (Russian), Moscow, U.S.S.R.: Academy Sci. USSR, Moscow, 1961 (in Russian).
29. A. Mekis, J. C. Chen, I. Kurland, S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, *Phys. Rev. Lett.* 77, 3787 (1996).
30. I. Smajic, Ch. Hafner, and D. Erni, *Opt. Express* 11, 1378 (2003).
31. K. S. Yee, *IEEE Trans. Antennas Propagat.* 14, 302 (1966).
32. A. Taflove, "The Finite-Difference Time-Domain Method," Artech House, Norwood, MA, 1995.
33. T. Weiland, *Electronics and Communications AEUE* 31, 116 (1977).
34. J. S. Chen and K. S. Yee, *IEEE Trans. Antennas Propagat.* 45, 354 (1997).
35. S. M. Rao, "Time Domain Electromagnetics," Academic Press, New York, 1999.
36. O. C. Zienkiewicz, "The Finite Element Method in Engineering Science," McGraw-Hill, London, 1971.
37. P. P. Silvester, R. F. Ferrari, and R. L. Ferrari, "Finite Elements for Electrical Engineers," 3rd Edn, Cambridge University Press, Cambridge, UK, 1996.
38. P. K. Banerjee, "The Boundary Element Methods in Engineering," McGraw-Hill, London, 1994.
39. A. C. Ludwig, *IEEE Trans. Antennas Propagat.* 34, 40 (1986).
40. Ch. Hafner, "The Generalized Multipole Technique for Computational Electromagnetics," Artech House, Boston, 1990.
41. Comsol FEMLAB 3.0, The Finite Element Based Simulation Software, Available at http://www.comsol.com
42. N. B. Piller, Coupled-Dipole Approximation and Generalized Multipole Technique in Electromagnetic Scattering, Thesis ETH 12727, Zurich (1998).
43. MIT Photonic-Bands (MPB) software package for computing the band structures (dispersion relations) and electromagnetic modes of periodic dielectric structures, Available http://ab-initio.mit.edu/mpb/.

44. E. G. Bogdanov, D. D. Karkashadze, and R. S. Zaridze, in "Generalized Multipole Techniques for Electromagnetic and Light Scattering" (T. Wriedt, Ed.), p. 143. Elsevier, Amsterdam, 1999.
45. D. Maystre, M. Saillard, and G. Tayeb, "Special Methods of Wave Diffraction" in "Scattering" (P. Sabatier and E. R. Pike, Eds.). Academic Press, London, 2001.
46. J. Smajic, Ch. Hafner, and D. Erni, Applied Computational Electromagnetics Society Journal (*ACES*) 18, 172 (2003).
47. A. Boag and R. Mittra, *J. Opt. Soc. Am. A* 11, 1505 (1994).
48. D. Karkashadze, R. Zaridze, A. Bijamov, Ch. Hafner, J. Smajic, and D. Erni, *Applied Computational Electromagnetics Society Journal (ACES)* 19, 10 (2004).
49. J. H. Holland, "Adaption in Natural and Artificial Systems." MIT Press, Boston, 1975.
50. D. B. Fogel, "Evolutionary Computation." IEEE Press, New York, 1995.
51. J. Smajic, Ch. Hafner, and D. Erni, *Opt. Express* 11, 566 (2003).
52. Y. Sugimoto, Y. Tanaka, N. Ikeda, Y. Nakamura, and K. Asakawa, *Opt. Express* 12, 1090 (2004).
53. R. Ferrini, R. Houdré, H. Benisty, M. Qiu, and J. Moosburger, *J Opt. Soc. Am. B* 20, 469 (2003).
54. The 3D-FDTD simulation platform SEMCAD is provided by SPEAG. Further information is available at www.semcad.com.
55. K. Rauscher, D. Erni, J. Smajic, and Ch. Hafner, EOS Topical Meeting on Optics in Computing 2004, pp. 17-18. Engelberg, Switzerland, 2004.
56. M. Qiu, K. Aziz, A. Karlsson, M. Swillo, and B. Jaskorzynska, *Phys. Rev. B* 64, 155113 (2001).
57. M. Agio and C. Soukoulis, *Phys. Rev. E* 64, 055603 (2001).
58. This tentative concept is based on a preliminary idea provided by R. Wüest and further investigated by K. Rauscher. For further information refer to www.photonics.ee.ethz.ch.
59. K. Rauscher, D. Erni, J. Smajic, and Ch. Hafner, "Progress in Electromagnetics Research Symposium PIERS 2004," p. 25. Pisa, Italy, Session 36, 2004.
60. D. F. G. Gallagher, *Europhotonics* February/March, p. 20 (2004).
61. P. I. Borel, A. Harpoth, L. H. Frandsen, and M. Kristensen, *Opt. Express* 12, 1996 (2004).
62. Ch. Hafner, J. Smajic, and D. Erni, EOS Topical Meeting on Optics in Computing 2004, pp. 15-16. Engelberg, Switzerland, 2004.
63. J. Smajic, Ch. Hafner, and D. Erni, *J. Soc. Am. A* 21, 2223 (2004).

CHAPTER 12

Theoretical and Computational Atomic-Scale Studies of Complex Catalytic Materials

Karl Sohlberg
Department of Chemistry, Drexel University, Philadelphia, Pennsylvania, USA

Sergey N. Rashkeev
Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee, USA

CONTENTS

1. Introduction 615
2. Theoretical Methodologies 616
   2.1. Cluster Models 616
   2.2. Models Employing Periodic Boundary Conditions 617
   2.3. Analysis Techniques 620
   2.4. Leveraging Experiments 620
3. Typical Studies and Results 620
   3.1. Nanocatalysts 620
   3.2. Metal Oxide Materials 625
   3.3. Catalytic Systems Based on Supported Metal Particles 630
   3.4. Stability of the Support Material 634
   3.5. Photocatalysis 638
4. Future Developments 640
   4.1. Some Selected Outstanding Problems 640
   4.2. Leveraging Infrared Spectroscopy 641
   References 642

1. INTRODUCTION

By some accounts, catalysis affects 30% or more of GDP in developed countries [1]. Catalysis is the enabling technology for petroleum production, for control of gaseous emissions from petroleum combustion, and for the production of industrial and consumer chemicals. Future

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (615–645)

applications of catalysis are potentially even more far-reaching. There is an ever-growing need to move the economy from a fossil-fuel energy base to cleaner alternatives. Hydrogen-based combustion systems and fuel cells could play a dominant role, given a plentiful and inexpensive source of hydrogen. Photocatalysis is the most promising clean technology for hydrogen production because it relies solely on water and sunlight, but performance enhancements in photocatalysis are needed to make this technology economically competitive.

Given the enormously widespread utilization of catalysts, even incremental performance enhancements have far-reaching benefits for multiple end-use sectors. In the area of fuel and chemical production, such improvements translate into vast reductions in energy consumption. At the consumption end, improvements in the catalysts involved yield tremendous reductions in pollution. In the area of photocatalysis, such efficiency improvements could finally render hydrogen an economically viable fuel.

The promise of theory and computation to facilitate advances in materials chemistry has been widely promoted; witness the numerous glossy covers of trade journals with computer-generated renderings of materials structures. Large-scale computations are even heralded by some enthusiastic futurists as a replacement for traditional experimental research. Indeed, theory and computation *are* making important contributions to materials research. Herein we review the state of the art in theoretical and computational modeling of complex catalytic materials and nanocatalysts. Related experimental work will occasionally be discussed, especially where theoretical work has leveraged experiments, but the focus of this review is on selected theoretical and computational studies of the materials themselves, as opposed to catalysis in general.

2. THEORETICAL METHODOLOGIES

The techniques for computational modeling of complex catalytic materials may generally be divided into two broad categories, the essential difference being in the choice of boundary conditions. The first category of techniques may be called the cluster models. In a cluster model, an isolated (typically small) cluster of atoms is assumed to be representative of the material. The second category of techniques employs periodic boundary conditions. In this second category, a small unit cell is modeled, and the boundary conditions at the edges of the unit cell are set so that translationally symmetric virtual images of the unit cell continue *ad infinitum* in all spatial directions. In effect, one is modeling a perfect periodic system of infinite extent. A detailed review of software for quantum-mechanical modeling of crystalline materials has been presented by Pisani [2].

2.1. Cluster Models

Generally speaking, cluster models have evolved in the computational chemistry community. For decades, computational chemists have been developing tools to model isolated gas phase molecules. Over time, algorithmic improvements and increases in computer hardware technology have enabled the application of these techniques to ever larger systems. Such techniques may be applied to the study of materials by taking a nanoscopic fraction of the material and modeling that in isolation as representative of its macroscopic whole. Notwithstanding practical limitations of computer power, any of the techniques from the computational chemists' toolbox may be applied to the modeling of clusters.

The techniques of computational chemistry may be divided into three categories: molecular mechanics methods, semiempirical methods, and *ab initio* methods. Molecular mechanics (MM) methods seek a description of the geometric arrangement of the constituent atoms in a system by minimizing an empirical potential energy function. The terms in the function are determined by empirical fitting to known molecular structures. MM techniques are fast and readily allow structural optimizations of systems of thousands, tens of thousands, or even millions of atoms. Because MM methods incorporate electronic structure effects only *implicitly* through empirical parameters, they are not appropriate for investigating electronic structure properties. The theoretically most rigorous techniques are the *ab initio* methods. These incorporate electronic structure effects explicitly. In an *ab initio* method, one attempts to solve the

time-independent Schrödinger equation (TISE) by using only mathematical approximations, without the introduction of empirical fitting factors. Such techniques are computationally very demanding, and typically one is limited to clusters of several dozen atoms or fewer with current computational hardware. Sometimes, only a few atoms (or even just one) are used to represent a surface with which a reactant interacts, although at this extreme, the relationship to any actual heterogeneous catalytic reaction is dubious. Semiempirical (SE) electronic structure methods attempt to capitalize on the strengths of the methods at the two extremes. SE methods explicitly treat the electronic structure but only for the valence electrons. Certain difficult-to-evaluate integrals required for the solution to the TISE are empirically parameterized for greatly increased computational efficiency. Modeling clusters of several hundred atoms is currently practical with SE methods. An in-depth review of the specific techniques from each of these general categories is outside the scope of this review. The text by Jensen [3] is an excellent initial point of reference.

2.2. Models Employing Periodic Boundary Conditions

Historically, the solid-state physics community has favored methods based on modeling a unit cell of a periodic solid and implementing periodic boundary conditions. Notwithstanding practical limitations of computer power, any of these techniques may be applied to the modeling of complex catalytic materials.

The tools of the solid state physics community divide into three categories almost exactly analogous to those of computational chemistry. First are the empirical potential models. Just like the chemists' MM techniques, the geometric arrangement of the constituent atoms in a system is sought by minimizing an empirical potential energy function, but here, instead of a cluster of atoms, a single unit cell is modeled with periodic boundary conditions. Unit cells of 10^3–10^6 atoms are currently within the reach of the empirical potential methods, depending on the complexity of the functions and the computational resource. Analogous to the chemists' *ab initio* methods, there are also first-principles methods for modeling materials, wherein one explicitly models the electronic structure of the system by direct solution of the TISE. Unit cells of 100 or so atoms are currently the practical limit for first-principles methods. Standing between the empirical potential and first-principles methods are tight-binding (TB) methods. These are analogous to the chemists' SE techniques. Because TB methods are computationally more efficient than first-principles methods, they may be routinely applied to larger unit cells. One useful application is to first use TB methods to investigate unit cell size effects [4]. This scheme allows straightforward determination of the smallest appropriate supercell for accurate description of the properties of interest. One then follows with first-principles calculations for this smallest appropriate unit cell.

Although computational chemists have historically favored expanding the electronic wavefunction in a basis set of atom-centered functions (termed an atomic-orbital or AO basis), the natural choice in a periodic system is a Fourier basis, that is, an expansion in sine and cosine functions. These are particle-in-a-box (PIB) wavefunctions. For practical calculations, the basis is truncated by selecting a cutoff energy; all terms are excluded that correspond to PIB eigenfunctions with corresponding eigenvalues in excess of the selected cutoff. A Fourier basis is particularly appealing for convergence studies because the cutoff energy may be systematically increased until convergence is achieved. There are several codes for carrying out plane-wave/pseudopotential/DFT calculations on periodic systems. Three of the most widely used codes that are used also by the authors are CASTEP [5], (Cambridge Sequential Total Energy Package), marketed commercially by Accelrys, a subsidiary of Pharmacopeia Inc; VASP [6, 7] (Vienna *Ab initio* Simulation Package); and DACAPO (from the Center for Atomic-scale Materials Physics: http://www.fysik.dtu.dk/CAMP/).

A complication that arises with the use of a Fourier basis is that it is currently impractical to include enough plane waves to accurately model the wavefunction cusp behavior near singularities in the potential at the positions of the atomic nuclei. Typically, this problem is handled by modeling only the valence electrons and describing the electron–ion interactions by employing pseudopotentials that are parameterized to incorporate the effect of the core electrons. A less common technique is to use a mixed basis, whereby plane waves are used

to describe the valence electrons and AO functions are used to model the core electrons. Mixed basis techniques are computationally very demanding, in part because solution of the TISE in this case involves multidimensional integrals that mix both types of basis functions. Challenging as they are, the all-electron methods are essential for modeling core electron phenomena.

There is some ambiguity in the specification of pseudopotentials that renders their selection somewhat system dependent. The degree to which a given pseudopotential may be used across a variety of chemical environments is termed its *transferability*. We will now present an example of the limits of transferability.

Lin et al. [8] have presented a procedure for developing smooth and transferable pseudopotentials. (The widely used CASTEP software [9] contains a database of pseudopotentials from an implementation of this procedure by Lee [10, 11].) Winkler et al. [12] have used CASTEP to apply density functional theory (DFT) with a plane-wave basis and these pseudopotentials to three hydroxide minerals: brucite [$Mg(OH)_2$], diaspore (AlOOH), and hypothetical anhydrous wadsleyite [$Mg_7Si_4O_{14}(OH)_2$] and reported generally excellent accuracy. This set of pseudopotentials would then appear to be a promising choice for use with related materials. One such class of materials relevant to heterogeneous catalysis is the transition aluminas of the form $H_{3m}Al_{2-m}O_3$ [13] that arise from dehydration of the parent precursor hydroxide boehmite. Remarkably, such calculations can result in a systematic error of as much as 6% in the lattice constant. This discrepancy may be traced to flexibility in the specification of pseudopotentials.

To demonstrate this effect, we have carried out unit cell optimizations for seven oxide and hydroxide systems, including transition aluminas of the form $H_{3m}Al_{2-m}O_3$. These are standard first-principles calculations employing (super)cells of an infinitely periodic bulk and were based on density functional theory (DFT) [14]. The generalized gradient approximation (GGA) to the exchange-correlation energy (V_{xc}) [15] was used predominantly, although some calculations were performed with the local density approximation (LDA) as well, for comparison. Two choices of pseudopotentials were used to describe the electron–ion interactions: (1) the Kleinman-Bylander form [16] of nonlocal pseudopotentials based on Lee's implementations of the scheme of Lin et al. [8, 10, 11], hereafter referred to as LPP, and (2) the "ultrasoft" pseudopotentials of Vanderbilt [17, 18], hereafter referred to as USP. The cutoff energy (E_c) for the plane wave basis was increased until convergence of the material density was achieved. (Typical values of E_c were $E_c = 1500$ eV for LPP calculations and $E_c = 700$ eV for USP calculations.) Integrations over the Brillouin zone employed a grid of k-points with a spacing of 0.1 $Å^{-1}$ chosen according to the Monkhorst-Pack scheme [19]. Unit cell optimizations were taken to be converged when the change between iterations was less than the following values: total energy 2.0×10^{-5} eV/atom, RMS displacement of atoms 1.0×10^{-3} Å, RMS force on atoms 5.0×10^{-2} eV $Å^{-1}$, and the RMS of the stress tensor 1.0×10^{-1} GPa. Results for seven systems are reported in Table 1.

The first system we have considered is brucite. As shown in Table 1, the computed density is in very good agreement with the experimental value and reproduces the value reported by Winkler et al. [12]. The second system, MgO, also has a computed material density in very good agreement with the experimental value. These results are representative of the general reliability of pseudopotential calculations.

Next we turn to the transitional aluminas. The third system in Table 1, HAl_5O_8, is the fully hydrogenated form of γ-alumina [13]. This structure is based on that of Mg-spinel, which is also one of the systems for which calculations are reported. The fully hydrogenated or "hydrogen spinel" form of γ-alumina, HAl_5O_8, is nominally generated by starting with Mg spinel and replacing half of the Mg with Al and the other half with H.

The fourth system is γ-alumina in its fully dehydrated form [13], for which the empirical formula is $\square Al_5O_{12}$, where $\square$ denotes cation vacancy. The unit cell is based on a triple block of the basic spinel unit cells. The unit cell contents are 2$\square$, 16Al, and 24O, which corresponds to three units of HAl_5O_8, with the three H replaced by an Al and two vacancies.

The fifth system, corundum, is the well-known hexagonal Al_2O_3 structure, α-alumina. The sixth system is boehmite. Boehmite is the precursor of γ-alumina in the thermal dehydration

Table 1. Computed material densities for various transition aluminas of the form $H_{2m}Al_{2-m}O_3$ as well as MgO, brucite, and Mg-spinel as predicted with two choices of pseudopotentials, hard (LPP) and ultrasoft (USP). Experimentally determined values taken from the literature as well as those computed with all-electron (LAPW) calculations are reported for comparison.

Material	V_{xc}	Pseudopotential	Density g/cc
Brucite expt. [152]	na	na	2.368
Brucite theo. [12]	GGA	LPP	2.423
Brucite	GGA	LPP	2.41
MgO expt. [152]	na	na	3.583
MgO	GGA	LPP	3.58
HAl_5O_8 expt. [62]	na	na	3.5
HAl_5O_8 lapw [23, 24]	na	na	3.53
HAl_5O_8	GGA	LPP	4.20
HAl_5O_8	GGA	LPP'	3.82
HAl_5O_8	LDA	LPP	4.23
HAl_5O_8	GGA	USP	3.72
HAl_5O_8	LDA	USP	3.76
HAl_5O_8	GGA	LPP'	3.78
$Al_{16}O_{24}$ expt. [153]	na	na	3.65
$Al_{16}O_{24}$	GGA	LPP	4.27
Corundum expt. [22]	na	na	4.00
Corundum lapw [22]	na	na	3.988
Corundum	GGA	LPP	4.80
Corundum	GGA	LPP'	4.28
Corundum	LDA	LPP	4.78
Corundum	GGA	USP	4.14
Corundum	LDA	USP	4.17
Boehmite expt. [154]	na	na	3.08
Boehmite	GGA	LPP	3.65
Boehmite	LDA	LPP	3.85
Boehmite	GGA	USP	3.16
Boehmite	LDA	USP	3.36
Mg-spinel expt. [52]	na	na	3.5827
Mg-spinel lapw. [25]	na	na	3.56
Mg-spinel	GGA	LPP	4.09
Mg-spinel	GGA	USP	3.63

sequence [20] and has the empirical formula $HAlO_2$. The seventh and last system listed in Table 1 is the closely related Mg-spinel.

From Table 1, we see that the first-principles results of calculations with GGA and the LPP choice of pseudopotentials for the aluminas and related Mg-spinel are systematically in error: $HAl_5O_8 = 20\%$, $Al_{16}O_{24} = 17\%$, corundum = 20%, boehmite = 18%, and Mg-spinel = 14%. On average, this error corresponds to an error of about 6% in the predicted lattice constant. Such deviation is too large to account for by physical causes, such as thermal expansion [21]. Table 1 affords a comparison of two different choices of V_{xc}, LDA and GGA. Note that in all cases there is no significant difference in the computed material density, indicating that the choice of V_{xc} is not an issue. One can also dismiss the possibility that the discrepancy arises from errors in the experimental density. Although large error bars may be anticipated for γ-alumina because its typically high porosity may complicate density measurements, the same is not true for boehmite, Mg-spinel, and certainly corundum, where the lattice constant is quite precisely established. Clearly, the source of the systematic error in the predicted material density lies in the calculations. This conclusion is confirmed by the all-electron calculations reported in Table 1. An all-electron calculation of the material density is available for corundum [22]. In addition, we have carried out all-electron calculations of the material density for HAl_5O_8 [23, 24] and Mg-spinel [25]. All-electron calculations on the other systems proved too demanding for our computational resources, but where available, the material density predicted with all-electron calculations is in superb agreement with experimental results. Given that removing the pseudopotential approximation removes the systematic error in the predicted material density, we conclude that flexibility in the construction of pseudopotentials influences the predictions.

These materials are all based on close-packed oxygen, with Al and H occupying interstitial sites among the close-packed oxygens. In these structures, Al-O distances are as short as 1.8 Å. The recommended LPP pseudopotentials in the CASTEP database have core radii of Al = 2.4 au and O = 1.8 au, which sum to about 2.2 Å. This means that in a treatment of these aluminas, the core radii may overlap by as much as 20%, which can result in a degradation of the pseudopotential approximation [26]. Clearly, the short interatomic distances in these transition aluminas render the Al pseudopotential with 2.4 au core radius less than ideal in this case. Employing an oxygen LPP for which the core radius is 1.4 au, combined with the Al core radius of 2.4 au, gives a sum of about 2.0 Å, reducing the overlap to only 10%. Calculations with this choice of pseudopotentials are shown in Table 1 denoted LPP'. The predicted material densities are markedly improved. Finally, employing the oxygen LPP with r_c = 1.4 au and an aluminum pseudopotential (LPP") for which r_c = 2.0 au results in sum of radii of 1.8 Å, which guarantees essentially no overlap. As shown in Table 1, the predicted lattice constant is further improved.

Another approach is to employ the USP pseudopotentials. Results are displayed in Table 1. These are consistent with the results for fully dehydrated cubic aluminas reported by Wolverton and Hass [27]. Overall, these results serve to delineate the fact that the flexibility in construction of pseudopotentials makes the choice of best pseudopotential somewhat system dependent.

2.3. Analysis Techniques

A widely used semiquantitative method of probing electron distributions in gas phase molecules is the Mulliken population analysis. This technique assigns electrons to individual nuclei based on the importance of basis functions centered on each atom to each occupied molecular orbital. This analysis technique is directly applicable to cluster models. As noted previously, in solid-state calculations, the valence electronic wavefunctions are typically expanded in a plane wave basis, not an atom-centered basis. Segall has developed a scheme for performing Mulliken population analysis on a plane-wave-basis expanded wavefunction by first projecting the plane-wave basis onto an atom-centered basis [28, 29]. This technique capitalizes on the computational benefits of the plane-wave basis, while still providing the interpretive benefit of the Mulliken population analysis. The projection is made onto a consistent linear combination of atomic orbitals (LCAO) basis set to help control the typical issues of basis-set dependence of the population analysis.

2.4. Leveraging Experiments

Given the large number of atoms required to reliably model complex catalytic materials regardless of whether cluster or unit-cell techniques are used, comprehensive exploration of configuration space is intractable because of the huge number of internal degrees of freedom in these systems. This is particularly true when manifestations of quantum behavior demand that a true quantum electronic structure treatment be used. No existing computational resource would resolve this problem. The most practical current approach to this problem is to leverage experimental results. High-resolution electron microscopy (HREM) images facilitate intelligently targeted calculations. For example, when modeling supported metal clusters, instead of searching (the countably infinite set of) all possible clusters, calculations can be targeted to those that are actually observed to occur on catalytic surfaces. Several specific examples are highlighted in this review. Other modern atomic-resolution microscopies and X-ray structure analysis are also providing useful data to target computational studies. Other experimental techniques ripe for leveraging will be discussed in Section 4.

3. TYPICAL STUDIES AND RESULTS

3.1. Nanocatalysts

Many materials that are otherwise chemically unremarkable develop catalytic properties when their particle size has physical dimensions on the nanometer length scale. For example, Valden, Lai, and Goodman [30] have been able to correlate the onset of catalytic activity

with appearance of nonmetallic properties as cluster size is diminished in titania-supported Au clusters. Large Au particles are metallic and have no special catalytic properties, but the small nano-scale Au particles exhibit nonmetallic properties and catalytic behavior not known in bulk Au. Tiny particles such as these are termed "quantized colloids" or, more commonly, "nanocrystals" and "nanoparticles." In this section, we discuss the general theory of catalytic nanoparticles and review some computational studies.

As particle size is diminished, the quantum nature of matter typically starts to manifest itself at the nanometer length scale. Finite size effects become apparent in the spectroscopy, and the addition (or subtraction) of a single atom to (from) the structure has a marked effect on its chemistry. At this length scale, a significant fraction of the total number of atoms of a particle resides on its surface. Consequently, surface effects profoundly affect the overall structure and chemical activity. In brief, the tiny size of nanoparticles influences their chemistry in two important ways: (1) by influencing the electronic structure and (2) by giving rise to novel surface structures.

Our intuition from the macroscopic world may start to fail us at the nanoscale, but it is anticipated that with new understanding of this strange world, its atomic precision may be exploited to tailor catalysts with unprecedented specificity. In this nanoregime, there are an insufficient number of atoms to produce true bulklike properties. Electronic structures therefore result from a compromise between bulk and surface effects and produce novel chemistry. Theoretical calculations should prove to be particularly valuable in the design and optimization of nanocatalysts because of the capability to model atomic-scale details of both geometric and electronic structure. We now discuss electronic structure effects, which may be divided into two categories, global and local.

Global electronic structure effects depend on the size, shape, and symmetry of the entire particle [31]. Such effects are particularly evident in clusters smaller than the size threshold for the onset of bulk behavior. The spatial confinement experienced by the electrons in the nanoparticle structure gives rise to quantization effects in the electronic structure. The electronic energy levels of the system are determined not only by the constituent atomic species and their bonding arrangements but also by the physical size of the system. This effect is shown schematically in Fig. 1. Here we show the maximum adjacent energy level spacing $[E(n = 2) - E(n = 1)]$ for an electron confined to an infinite square well potential as a function of the width of the potential (n is the quantum number). This is the famous particle-in-a-box (PIB) problem. The figure shows that spatial confinement starts to split the energy levels with chemically significant energies (>1 kcal/mole) when the particle size decreases below about 1 nm in length. This effect may well be a technological benefit because it suggests that the electronic structure may be, in part, tailored by careful control of the system size.

When the particle size is below the threshold for the onset of bulk behavior, the electronic structure therefore consists of discrete orbitals as in a molecule, instead of bands as in a bulk solid. The degree to which these molecular orbitals are filled with electrons depends

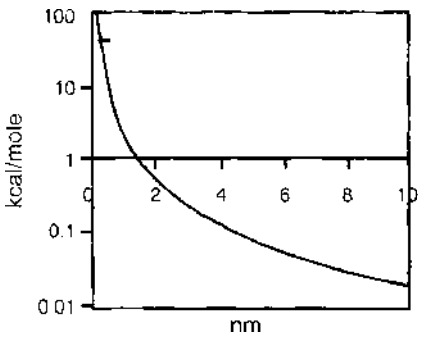


Figure 1 Maximum energy level spacing for an electron confined to an infinite square well potential as a function of the width of the potential. Note that confinement-induced splittings become chemically significant in energy below the 1-nm length scale.

on the exact number and type of atoms in the cluster. Each different cluster therefore has a unique orbital filling and hence a unique electronic structure and unique chemical activity.

That the quantization effects of nanocrystals may in fact directly influence their catalytic properties was noted by Nedeljkovic et al. [32], who reported optical effects due to size quantization in nanoscale colloids, and catalytic properties that similarly correlate with the dimensions. This result suggests that the quantum effects on the electronic structure in these nanocrystals may in fact determine their catalytic chemistry. Particles of HgSe and PtSe exceeding 500 Å in size are black, but below 500 Å the color is dependent on the particle size. With decreasing size, the optical absorption spectrum is shifted to higher energies. This observation is consistent with a simple particle-in-a-box interpretation of the effect of particle size on the electronic energy levels. With decreasing particle size, one anticipates that the electronic energy levels will be pushed up and spread, thereby increasing transition energies. It is thought that as a consequence of this spreading, small particles possess energy levels that can accommodate high-energy electrons with greater reducing capability, and indeed, photoreduction is more efficiently performed by the smaller particles [32].

Supporting this view that the influence of particle size on the electronic structure is a major determining factor in catalytic activity is the work of Pan, Sohlberg, and Ridge [33] In this work, changes in the energy of the lowest unoccupied molecular orbital (LUMO) were shown to be determined by the size and symmetry of the cluster and were seen to correlate precisely with the cluster reactivity. Clusters of Co and Ir were reacted with cyclohexane. The rate of reaction (by C—H bond activation) was seen to correlate to the ability of the cluster to accept electrons from the C—H bond that is being activated. This cluster acceptor molecular orbital was found to result from the combination of the metal 4s atomic orbitals. Clusters where the bonding and symmetry resulted in an empty, metal 4s-derived, molecular orbital at low energy (that is, a good acceptor) were reactive to cyclohexane by C—H bond activation. For clusters where the lowest unoccupied molecular orbital derived from the metal 4s atomic orbitals is at high energy, reactivity is seen to be low or negligible. Here we see that every atom counts. The addition of a single atom changes the symmetry of a cluster. When the symmetry changes, the splitting of the molecular orbitals that arise from the degenerate atomic orbitals also changes. These changes in splittings move the energy level of the LUMO up and down with the addition of each successive new atom to the cluster, altering its reactivity. In a system dominated by bulk, the addition of a single atom to the system has (almost by the definition of bulk) no appreciable effect on the chemistry, but at the nanoscale, every atom counts.

Beautiful demonstrations of the atom-by-atom size dependence of the chemistry in small clusters have been reported by Cheng and Wang, who carried out theoretical studies of clusters of chromium (Cr_n, $0 \leq n \leq 15$) [34] and of metallocarbohedranes (M_nC_m) [35,36]. In agreement with experimental observations, they found that the chemistry of chromium clusters is uniquely dependent on the cluster size n. For example, the dissociation energy shows a distinct even-odd oscillation, as there is strong energetic preference for building up the cluster two atoms at a time. Perhaps more interestingly, Wang, Wu, and Cheng [37] report the onset of bulklike behavior in the photoelectron spectrum of these clusters at $n = 11$, the smallest possible value of n for which it is possible to construct a complete body-centered-cubic unit, the preferred structure of Cr bulk.

Even-odd oscillations in the reactivity of gold cluster anions (Au_n^-) with O_2 have been reported by Kim, Fischer, and Gantefor [38]. These oscillations correlate exactly with oscillations in electron affinities of the gold cluster anions. The even n neutral clusters are so-called magic clusters with markedly high stability and an unusually large HOMO-LUMO gap. The theoretical studies show that upon reduction to the corresponding anion, these "magic" clusters become chemically very active because they are excellent electron donor species.

The reductive activation of CO_2 on the surface of semiconducting nanocrystals is a prototypical example of catalytic reduction by nanoparticles. This process holds promise for sequestration of atmospheric CO_2 to combat the accumulation of these greenhouse gases. CdSe nanocrystals have been of particular interest for fixation of CO_2. It is known that only the smallest CdSe nanocrystals catalyze the reduction of CO_2 [32]. One possible explanation

for this effect is that only the smaller nanocrystals possess conduction band orbitals that, when occupied, have a sufficient redox potential to reduce CO_2. This hypothesis is depicted schematically in Fig. 2. On the right side of the figure are the energies of CO_2 molecular orbitals arising from linear combinations of the carbon and oxygen 2s and 2p atomic orbitals. Their occupancy in molecular CO_2 is indicated. Note that the highest occupied molecular orbital (HOMO) is degenerate, composed of nonbonding and essentially unhybridized oxygen $2p_x$ and $2p_y$ orbitals [39]. The addition of an electron or electrons to the lowest unoccupied molecular orbital (LUMO) then reduces the bond order of the C—O bonds, thereby weakening the molecule, the first step in its reduction. On the left side of the figure, the valence and conduction bands of the nanoparticle are shown. Owing to size quantization effects, at least some of the bands increase in energy with decreasing particle size. There is therefore a threshold particle size below which at least some of the conduction band orbitals will, if occupied, have sufficient reducing power to donate electrons into the CO_2 LUMO, initiating its reduction. Electronic structure calculations on the nanocrystals may be used to quantify the reducing power of the nanocrystals. Subsequent calculations of the interaction of CO_2 with the nanocrystal model the reduction process, including electron transfer, bond activation, and molecular dissociation.

Although global electronic structure effects are thought to play a key role in the reductive activation of CO_2 on the surface of semiconducting nanocrystals, other evidence suggests that local surface structures are critical as well. For example, Fujiwara et al., [40] have shown that the interaction of CO_2 with sulfur vacancies on the surface of CdS nanocrystals results in a much stronger reduction of the CO_2 than interactions with nondefect surface sites. Numerous local geometries can be generated simply by removing two, three, or several atoms from the surface. Theoretical calculations have been reported that use this "trick" of generating local geometries by creating a vacancy in an otherwise defect-free surface. In particular, calculations of the activation barrier for absorption of a CO_2 molecule at a surface Se vacancy on CdSe show that the surface vacancy creates a favorable electronic environment for chemisorption [41]. Photoexcitation of the chemisorbed CO_2 is then simulated by

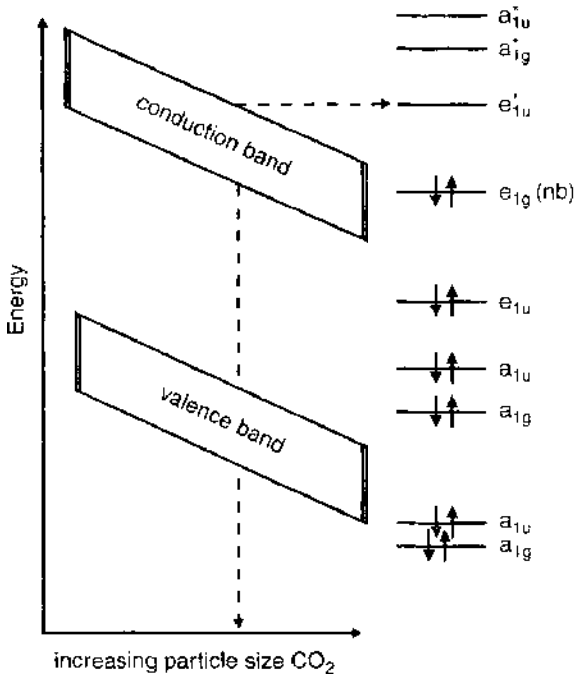


Figure 2. Schematic of CO_2 reduction by nanoparticles. On the right side of the figure are the energies of CO_2 molecular orbitals arising from linear combinations of the carbon and oxygen 2s and 2p atomic orbitals. Their occupancy in molecular CO_2 is indicated. On the left side of the figure, the valence and conduction bands of the nanoparticle are shown. Owing to size quantization effects, at least some of the bands increase in energy with decreasing particle size. For the smaller nanoparticles, electrons promoted to the conduction band have sufficient energy to reduce CO_2.

placing an In atom in the unit cell far from the adsorbed molecule. Because of the low ionization potential of In, the electron transfers to the CO_2, and lowers the activation barrier. Extrapolating to the behavior at a nanocrystal surface, the energy gap increases with decreasing particle size. Therefore, for sufficiently small nanocrystals, eventually it becomes energetically favorable for the CO_2 to desorb as CO_2^-. Such a mechanism capitalizes on both key aspects of nanotechnology. First, only a nanoscale particle of CdSe will have a sufficient concentration of surface atoms to present surface defects with the abundance necessary for appreciable catalysis. Second, the desorption of CO_2 anions becomes energetically favorable only for sufficiently small nanocrystals, because quantum confinement increases the energy of the photoexcited electron for small nanocrystals.

As noted previously, below the size threshold for the onset of bulk behavior, the electronic structure of nanocrystals depends uniquely on the number of atoms in the structure. Given that the electronic structure governs reactivity, this sensitivity suggests the intriguing possibility of selecting particles for specific catalytic activity. If the size-dependent reactivity can be predicted, the "holy grail," catalytic specificity by design, comes within reach.

Local electronic structure effects may be present in clusters even above the size threshold for the onset of bulk behavior because even above this threshold, as long as one is still in the "nano" regime, a significant fraction of the atoms in the cluster are at the surface. A straightforward calculation estimates the fraction of atoms on the surface of a simple Platonic solid. One simply divides the volume of the solid by the effective volume of an atom to determine the number of atoms in the structure. Next, one divides the sum of the surface areas of the faces by the effective cross sectional area of an atom to determine the number of surface atoms and divides the latter by the former. Carrying out this procedure for a range of values of the edge length yields the fraction of surface atoms on a cube of pure Cu as a function of the edge length, as shown in Fig. 3.

When surface atoms become an appreciable fraction of the total, novel structures develop that are dictated by bonding compromises between the surface and bulk. These novel structures force the surface atoms into atypical bonding arrangements and, consequently, give rise to atypical local electronic structures. As highlighted below, atoms forced into bonding arrangements atypical for the specific element can produce unusually reactive surface sites, that is, ideal catalytic centers. Such novel structures have been directly observed by atomic-resolution scanning tunneling microscopy (STM) [42] and are thought to provide highly specific catalytic sites. The manner in which they are terminated may be at least equally important [43]. The coordination of Al cations on γ- and η-alumina surfaces is a case of particular interest because these cations are a source of surface Lewis acidity, central to the catalytic activity of alumina [44].

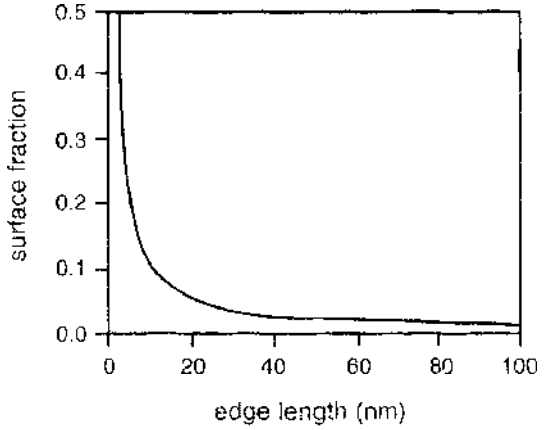


Figure 3. Fraction of atoms on the surface as a function of the edge length for a cube of Cu. Note that surface becomes an appreciable fraction of the total when the edge length drops into the 1–10 nm range.

3.2. Metal Oxide Materials

3.2.1. γ- and η-Alumina

These two forms of alumina (γ and η) are in widespread use in heterogeneous catalysis systems [45], such as those employed in petroleum refining [46], in the production of bulk and fine chemicals [47], and in automotive catalytic converters [48]. The two materials have very similar bulk structures, but the surface acidity of η-alumina is considerably greater than that of γ-alumina [45, 49, 50].

Both γ- and η-alumina are cubic aluminas; that is, they are based on cubic close packing of oxygen anions with Al cations occupying interstitial sites within the *ccp* oxygen lattice. These structures may be described in terms of the structure of the closely related Mg-spinel [51]. Mg-spinel has a crystallographers' cubic unit cell of contents $Mg_8Al_{16}O_{32}$, and a primitive unit cell of contents $Mg_2Al_4O_8$. The Mg atoms occupy T_d interstitial sites within the *ccp* oxygen lattice, and Al atoms occupy O_h interstitial sites. The space group is Fd3m [52]. Alumina contains no Mg, and therefore all cations are Al. This necessitates the introduction of vacancies to maintain the balance of valance. In particular, one of every nine cation sites in the spinel structure must be vacant to produce the Al_2O_3 stoichiometry. The crystallographers' cubic unit cell is $\square_{2\frac{2}{3}}Al_{21\frac{1}{3}}O_{32}$ and the primitive unit cell is $\square_2Al_{16}O_{24}$. Referencing the spinel notation, we note that the primitive cell consists of $(T_d)_6(O_h)_{12}O_{24}$, showing that there are $6T_d$ sites and $12O_h$ sites per primitive unit cell. Obviously, with 16 cations per primitive unit cell, neither the T_d cation sublattice (which has 6 sites per primitive cell) nor the O_h cation sublattice (which has 12 sites per primitive unit cell) is sufficient to hold all of the Al atoms. The Al must therefore be distributed over both T_d and O_h sites and will therefore fall within the range $[25\% \leq X_{T_d} \leq 37.5\%]$, where X_{T_d} is the fraction of Al cations on T_d sites. The extremes of this range are represented by $X_{T_d} = 25\%$ given by $(\square_2Al_4)(Al_{12})O_{32}$ and $X_{T_d} = 37.5\%$ given by $(Al_8)(\square_2Al_{10})O_{32}$. As reviewed at length in Ref. [53], experimental investigations have not resolved the correct distribution of vacancies over the T_d and O_h sublattices, and nearly all computational investigations have addressed this issue.

Some of the earliest serious modern computational studies of γ-alumina were reported by Alvarez et al. [54] and Blonski and Garofalini [55, 56]. Recognizing that stoichiometry is a key issue in γ-alumina, Alvarez et al. performed a classical molecular dynamics simulation based on pair-wise additive empirical potential functions on a 1440-atom supercell, $Al_{576}O_{864}$. This large unit cell allows for considerable variation in the distribution of vacancies, while maintaining the correct Al_2O_3 stoichiometry of aluminum oxide. (Even today, more than a decade later, such a unit cell is too large for treatment with first-principles methods.) One particularly important outcome of the calculations was the finding that cation sites other than the 8(a) and 16(d) sites of the spinel lattice [such as 16(c) and 48(f)] also have non-negligible occupancy. Occupancy of nonspinel sites leads to ambiguity in the specification of X_{T_d}. This finding of nonspinel site occupancy is not definitive, however. During the simulation (15.5 ps total) the populations of the tetrahedral and octahedral sites were observed to change by up to 1.1%. A tendency to retain the initial configuration throughout the entire simulation may result from low cation mobility due to appreciable potential barriers to Al migration within the alumina bulk (for example Ref. [57]), coupled with the 300 K simulated temperature.

Lee et al. [58] carried out a Monte Carlo simulation of the γ-alumina structure in support of their [27] Al NMR study. They employed an empirical potential energy function in the simulation. Parameters for the function were found by fitting to the results density functional theory (DFT) calculations on nine "artificial" crystal structures. Based on the simulation they reported $X_{T_d} = 30 \pm 3\%$.

First-principles DFT studies have been reported by Mo, Zu, and Ching [59], based on structural models determined from empirical pair-potential functions. Based on these calculations, it was reported that there is an energetic preference of 3.7 eV/vacancy for vacancies to exist on octahedral sites. Application of the Boltzmann distribution function to this energy difference yields $X_{T_d} = 37.5\%$. One criticism of these calculations is that the unit cell employed was of the stoichiometry $Al_{21}O_{32}$. This is the nearest integer approximation to the $\square_{2\frac{2}{3}}Al_{21\frac{1}{3}}O_{32}$ cubic unit cell. Despite this approximation, this vacancy distribution has been confirmed with elaborate empirical potential simulations reported by Streitz and

Mintmire [60] on a 160-atom unit cell of the correct stoichiometry. The computational expense of using a large unit cell and the use of an approximate stoichiometry can both be avoided by employing the primitive unit cell or other noncubic cell [13]. The primitive unit cell ($\square_2Al_{16}O_{24}$) contains just 40 atoms. The two vacancies may be distributed both on octahedral sites (O_hO_h), both on tetrahedral sites (T_dT_d), or one on a tetrahedral site and one on an octahedral site (T_dO_h). Gutierrez, Taga, and Johansson have carried out first-principles calculations for all 14 symmetry unique distributions of the two vacancies in the primitive unit cell and again confirmed the preference for vacancies at octahedral sites [61].

A more complete understanding of the cubic aluminas is afforded when one recognizes that cubic aluminas are derived by dehydration of aluminum hydroxide minerals, and there exists the possibility of the presence of hydrogen within the bulk [13]. Experimental evidence suggests that conversion to the cubic structure occurs well before complete dehydration and points to the fact that once the spinel structure is reached, it remains intact throughout the remainder of the dehydration process [62–64]. The presence of hydrogen is compensated for by an Al deficiency, so that an appropriate formula for the sequence is $H_{3m}Al_{2-m}O_3$. The stoichiometry corresponding to $m = 1/8$ is particularly notable. It corresponds to a unique case where all the cation sites and all the anion sites are occupied. An ideal hydrogen-aluminum-spinel formula, HAl_5O_8, first suggested by de Boer and Houben [62], is formed. For $m < 0.125$, the system is H-poor, and there are vacant cation sites. The system is hydrogen rich when $m > 0.125$ and in addition to all of the cation sites being full, there are interstitial H atoms as well. The $m = 0$ limit corresponds to what is commonly described as the defect-spinel structure. In this hydrogen-free limit, the extremes of the physically allowed distribution of vacancies over the two cation sublattices may be nominally assigned to fully dehydrated γ- and η-alumina [13]. Studies employing [27] Al NMR by John, Alma, and Hays [65], found $X_{T_d} = 25 \pm 4\%$ for γ-alumina and $X_{T_d} = 35 \pm 4\%$ for η-alumina.

DFT surface relaxation studies of γ-alumina found that the nominally three-coordinated Al (those at nominal T_d sites on the surface) relax inward to become quasioctahedral. (Hints of this result may be seen in the results of earlier molecular dynamics studies [66–68].) By contrast, calculations of the relaxation of η-alumina surfaces show that Al cations at nominal T_d sites on the surface relax only until they become quasitrihedral [69, 70]. (See Fig. 4.) This remarkable difference in the degree of surface relaxation between γ-alumina and η-alumina is in quantitative agreement with experiments [64].

This result also explains why η-alumina is a more acidic catalyst than γ-alumina [45, 49, 50]. It is known from semiempirical [71] and *ab initio* [72] cluster model calculations that Al atoms in T_d sites, when exposed at the surface (three-coordinated Al), have a lower

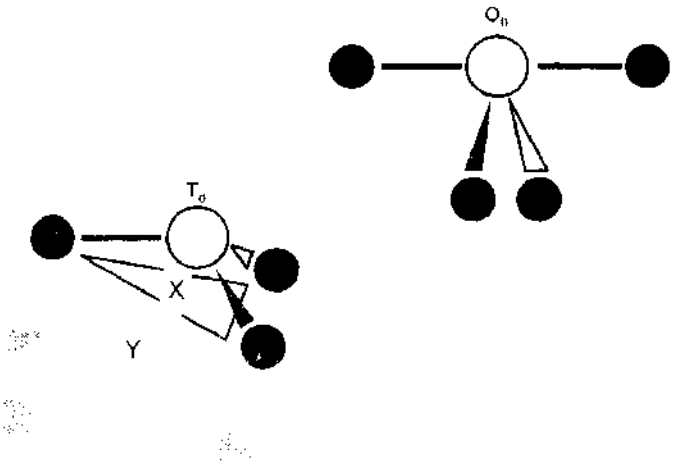


Figure 4. Surface Al in a nominal T_d site (lower left) and surface Al in a nominal O_h site (upper right). Aluminum atoms are shown by open circles and oxygen atoms are shown by filled circles. The gray circles are oxygen atoms within the bulk. In η-alumina, the surface Al in the nominal T_d site will relax until it reaches the quasitrihedral position near the center of the (111) plane marked "X" in the figure. In γ-alumina, however, the Al will relax a distance of 1.7 Å into an empty O_h interstitial site in the first subsurface layer, marked "Y" in the figure.

energy acceptor orbital and therefore serve as stronger Lewis acids than Al atoms in O_h sites exposed at the surface. This is a manifestation of a general principle that Kawakami and Yoshida [73] deduced, based on small basis set *ab initio* studies of small aluminum oxide and hydroxide clusters; Lewis acidity is greater for Al atoms with lower coordination. Because surface Al in T_d sites relax into six-coordinated sites in γ-alumina but remain in three-coordination on the surface of η-alumina, it is expected that η-alumina surfaces should exhibit the greater Lewis acidity, in agreement with the well-known properties of these transition aluminas [45, 49, 50]. A comprehensive review of the bulk and surface structure of the cubic aluminas may be found in Ref. [53].

Reactivity at cubic alumina surfaces has been studied with cluster models. Cai and Sohlberg [74] reported semiempirical electronic structure calculations of the preferred adsorption sites for several simple alcohols on γ-alumina by using an $Al_{48}O_{72}$ cluster model. This is one of the largest cluster models that have been reported and includes all interactions through the second-nearest neighbors of the binding site. It was reported that successful dehydrogenation of the alcohol (a critical initial step in the formation of alkenes from alcohols) may occur at two different sites, but in one site both the alcohol oxygen and hydrogen must interact with the surface, whereas at the second site only $O_{surface}$–$H_{alcohol}$ interaction is required. Similar semiempirical studies of water dissociation on the γ-alumina surface have been reported by Fleisher, Golender, and Shimanskgya [75] but with a much smaller Al_2O_8 cluster model. Again a two-point interaction was found where both the oxygen atom and one of the hydrogen atoms of the water molecule are interacting with the surface. It is likely that many catalysts rely on multipoint interactions because such will accomplish both electronic and geometrical modification of the adsorbed molecule, setting the stage for its subsequent reactivity.

3.2.2. κ-Alumina

One of the best examples of the use of first-principles methods to elucidate the atomic scale structure of a complex metal oxide material is the determination of the κ-Al_2O_3 structure by Yourdshahyan et al. [76]. Samples of κ-Al_2O_3 generally have poor crystallinity, so experimental studies have left some uncertainty about its atomic scale structure. Yourdshahyan et al. [76] present an eight-step procedure for structure determination and apply it to κ-Al_2O_3.

1. The unit cell, symmetry, and lattice parameters are identified from experimental data.
2. As much information as possible regarding the atomic positions is extracted from experimental data.
3. All possible configurations of the unit cell that are consistent with the crystal symmetry are enumerated.
4. The list of possible configurations is pruned with simple rules, such as Pauling's structural rules for ionic crystals.
5. Crude structural optimizations of the reduced list of possible configurations are carried out with first-principles methods to identify those that are locally stable.
6. Detailed structural optimizations, including lattice parameter optimization, are carried out for the locally stable structures.
7. Structures are eliminated that have a material density inconsistent with experimental results.
8. The remaining few structures are analyzed for lowest energy and best agreement with experimental data.

The structure of κ-Al_2O_3 resulting from the application of this procedure is one with close-packed oxygen atoms in an ABAC stacking arrangement. Aluminum atoms occupy 25% of the tetrahedral interstitial sites and 75% of the octahedral interstitial sites within the close-packed oxygen sublattice. One of the most interesting outcomes of the application of this procedure to κ-alumina is that the resulting structural model gives novel insight into the mechanism of CVD growth of this material. Once a single close-packed oxygen layer is formed, the Al ions naturally fill "hollow" sites in the next layer. Electrostatic repulsion of the Al ions is minimized if they form zigzag lines. Once this happens, the pattern is set, and the crystal will grow as κ-alumina in the [001] direction, as observed experimentally.

Of broader significance, the overall procedure is essentially completely general. Because it employs first-principles techniques, it is not restricted by empirical parameterization. It is therefore, in principle, applicable to almost any complex catalytic material.

3.2.3. Ceria

Cerium oxide (ceria) is a critical and multifunctional component of automotive emissions catalysts. One function of ceria is to act as an "oxygen storage" component. Ceria is capable of oxygen uptake during excursions of the air-fuel ratio into the net oxidizing regime and of oxygen release during excursions into the reducing regime, thus facilitating the maintenance of the proper stoichiometric ratio for sustained CO and HC oxidation and simultaneous NO_x reduction, that is, three-way catalysis (TWC). An atomic-scale understanding of the structure and dynamics of ceria would obviously be valuable in the design of improved automotive emission catalysts.

Many of the key properties of ceria that contribute to its success in automotive catalysis are thought to arise from the fact that cerium has two stable oxidation states, Ce^{+3} and Ce^{+4} [77]. As a consequence, cerium oxide may exist over a range of possible stoichiometries CeO_{2-x} [$0 \leq x \leq 0.5$] [78]. The deviation of ceria from its ideal CeO_2 composition has been extensively studied by temperature programmed reduction with hydrogen. These studies have given rise to an important controversy: Is there hydrogen uptake by ceria upon reduction with hydrogen? This question has been addressed with first-principles DFT methods by Sohlberg, Pantelides, and Pennycook [79].

The affinity of ceria for hydrogen is determined from the reaction:

$$HCe_4O_8 \rightarrow Ce_4O_8 + 1/2H_2 \tag{1}$$

Here Ce_4O_8 is used as the reference instead of the empirical formula CeO_2 because the former is the formula for the primitive unit cell of cerium oxide. Employing DFT total energies to estimate ΔH (pressure–volume work is neglected, typically an approximation of little consequence in condensed phase systems) and tabulated third-law entropies to determine ΔS shows that for the above reaction,

$$\Delta G = 0.45 - 6.768 \times 10^{-4} T \tag{2}$$

where the energies are given in eV/molecule. This expression predicts that below 665 K, hydrogen uptake is spontaneous. The calculations also predict that this uptake produces a lattice expansion of about 1.5%. The H taken up forms hydroxide structures within the ceria bulk with a predicted vibrational frequency of 3414 cm^{-1}. These results are in good agreement with experimental reports [77, 80]. This work, therefore, has resolved the issue of the thermochemistry of hydrogen uptake by ceria, finding that hydrogen uptake is spontaneous but that in sufficiently small quantities it should be of minor consequence in catalytic systems.

3.2.4. Zirconia

Like ceria, zirconia is a catalytically important material with good oxygen storage properties. Christensen and Carter [81] have reported an exhaustive first-principles study (DFT, GGA, supercells, plane waves) of the three observed bulk phases of ZrO_2 and their probable exposed surfaces. Cohesive energy curves are presented for cubic, tetragonal, and monoclinic bulk phases. The energy minima are in excellent agreement with the experimentally observed lattice parameters. Their work also leads to an explanation for the observed depression of the tetragonal → monoclinic phase transition temperature for very small particles. The tetragonal (100) surface relaxes to (001), thereby rotating the tetragonal distortion below the surface. Because other surfaces exhibit little relaxation, the tetragonal → monoclinic phase transition must change a stable surface into a less stable one. Because small particles have a larger fraction of surface atoms (see Section 3.1), this effectively increases the activation energy for transformation for small particles.

One of the more valuable aspects of the Christensen and Carter report, with impact beyond zirconia, is a presentation of three simple "rules" for preliminary identification of the most probable stable surfaces. They are: (1) the compactness condition, select surfaces that are the most compact; (2) the bond-breaking condition, select surfaces that minimize coordination loss for the surface atoms; and (3) electrostatic condition, select surfaces that are the least polar. Rules of this type are of considerable value because the number of possible exposures is large, and especially so for complex catalytic oxide materials. Rather than carrying out costly first-principles optimizations for a large number of possible surfaces and terminations, the set of candidates can first be reduced with the Christensen and Carter criteria.

3.2.5. Zeolites

Another class of complex materials important for heterogeneous catalysis is the zeolites. Shah et al. [82] have reported a first-principles study of zeolite catalytic behavior on methanol. The common methanol conversion zeolite catalyst, ZSM-5, has a primitive unit cell sufficiently large that first-principles calculations are impractical, so they chose to model the related material chabazite as a representative structure and, for comparison, sodalite. Because the chabazite structure has tetrahedral sites with 1/3 Al occupancy, there are a large number of possible structures. To simplify the calculations, only one Al atom per unit cell was considered. This is a common issue in modeling complex catalytic materials. The complexity of the material means that there are many possible structures. (The configurational entropy can be large.) To reduce the number of computations to a practical value, one must select representative structures.

The calculation of Shah et al. [82] showed a good ability to distinguish catalytic activity between these structures. In chabazite, methanol undergoes a kinetically spontaneous (no energy barrier along the reaction coordinate) protonation by a Bronsted acid site. In the more open sodalite structure, methanol was only physisorbed. This is one of the first successful applications of first-principles methods to actual catalytic reactions in realistic catalytic materials.

3.2.6. RuO_2 and First-Principles Thermodynamics

One of the great challenges of understanding heterogeneous catalytic materials is to determine the active site. Based on simple electrostatic and valence arguments, such as those discussed in the section about zirconia, it has long been assumed that so-called polar surface terminations and surfaces with highly valence-unsaturated metal atoms are energetically unfavorable. Unsaturated metal atoms, however, are strong Lewis acid sited, and chemical intuition suggests that these would serve as excellent catalytic sites. Indeed, catalytic mechanisms that assume exposed metal Lewis acid sites have often been hypothesized and subsequently found to predict catalytic action in agreement with experimental observations. This apparent conflict is almost certainly a consequence of attempting to explain equilibrium surface conditions over a range of laboratory conditions based on $T = 0$, $P = 0$ arguments. Reuter and Scheffler have studied the prototypical heterogeneous catalyst, RuO_2, and shown that the application of first-principles thermodynamics (as opposed to more conventional static total-energy calculations) closes this gap.

Reuter and Scheffler [83] point out that the appropriate preferred surface structure under a given set of temperature and pressure conditions will depend on the relative free energies of the various possible forms. A typical first-principles total energy calculation considers only the electronic structure contribution to the surface energies. There are also pressure-volume (PV) and vibrational contributions to the surface free energy. Furthermore, the chemical potential (free energy per formula unit) of the oxygen will depend on the temperature and pressure of the reservoir of oxygen gas in contact with the surface. Reuter and Scheffler [83] have shown how to account for all of these contributions. They show that in the pressure range of interest to usual catalytic systems, the PV contribution is negligible in comparison to the expected accuracy with which the total energy can be calculated by using today's first-principles methods, but the vibrational contribution can be appreciable.

The most rigorous technique for determining the vibrational contribution would be to carry out a calculation of the density of phonon states, but a reasonable estimate may be made with the Einstein approximation (see any statistical physics text). Because catalytic systems run at temperatures greatly in excess of the boiling point of O_2, a reasonable approximation of the oxygen chemical potential may be derived from the ideal gas law. Incorporating all of these terms into the determination of the system free energy, Reuter and Scheffler [84, 85] have shown that nonstoichiometric "polar" surface terminations of RuO_2 are actually thermodynamically favorable under certain realistic temperature and pressure conditions. Furthermore, they have found that the free energies for adsorption of other gas species on the surface have similar T and P dependence [84, 85]. These findings are almost certainly more general, and similar studies for other catalytically important oxide surfaces would be of significant value.

The concept of first-principles thermodynamics may, in principle, be extended to studies of phase transitions between solid phases. Consider the reaction

$$\text{Phase A} \rightarrow \text{Phase B} \tag{3}$$

The ultimate objective would be to generate a P/T phase diagram from first-principles calculations. To do so, one starts from the observation that at a phase boundary,

$$\Delta G = \Delta H - T\Delta S = 0 \tag{4}$$

for phase A $\rightarrow$ phase B transition, where ΔG is the change in Gibbs free energy, ΔH is the enthalpy change and ΔS is the entropy change. There are three leading contributions to the enthalpy change: ΔE^e, the internal energy change due to electronic contributions; ΔE^v, the internal energy change due to vibrational contributions; and $P\Delta V$, pressure/volume work. In terms of these contributions,

$$\Delta G = \Delta E^e + \Delta E^v + P\Delta V - T\Delta S = 0 \tag{5}$$

ΔE^e is obtained from the difference in total energies of phases A and B as computed by first-principles total energy calculations. The vibrational contribution ΔE^v may be obtained from computations of the phonon spectra of the two phases, or approximated by the Einstein method using local mode frequencies from finite differences calculations of the gradient of the potential energy [86] for each symmetry unique atom in the unit cell. Regardless of how they are determined, knowledge of the vibrational frequencies affords an estimate of the entropic change as well. Finally, pressure/volume work is a straightforward multiplication of the pressure by the volume change between the two phases. The volumes for the individual phases are easily obtained as a function of pressure by recognizing that $P = -dE/dV$. The total energy as a function of volume $E(V)$ as computed by first-principles total energy calculations may therefore be used to generate $P(V)$. In this way, first-principles calculations may be used to obtain an estimate of every major term in the free energy change associated with the phase change. Pressure-volume phase space is then searched for points on the phase boundary.

3.3. Catalytic Systems Based on Supported Metal Particles

Remarkably, some metal clusters that are not catalytically active independently become active catalysts when deposited on a support [87]. In some cases, doping the metal cluster with another metal further improves its catalytic properties [88]. Subtle changes in the local electronic structure appear to be controlling the catalytic activity [89]. A great many heterogeneous catalysts are of this general nature, metal clusters dispersed on a support. Freund has presented an outstanding review of oxide-supported catalytic clusters [90].

Many heterogeneous catalytic systems are an intricate combination of supported materials. A large surface area support material such as γ-alumina may be used as a catalyst in its own right, but more commonly as a support for another oxide and/or a finely dispersed metal (to impart properties such as improved durability or to tailor the catalytic chemistry) [46].

To guide tailoring and engineering of improved catalysts, it is necessary to have a clear understanding of the catalyst structure, of the dynamics of this structure during operation, and of the role of the support in relation to the catalytic activity of the supported phase [91, 92]. As we will now review, recent theoretical work has yielded very significant advances in this area.

Overall, γ-alumina is one of the quintessential materials of heterogeneous catalysis. The catalytic properties of alumina for alcohol dehydration have been known and exploited for more than 200 years [47, 93]. Today, γ-alumina is more typically used as a support for a finely dispersed metal catalyst. For example, one of the most important industrial applications of alumina-based catalysts is the dehydration of alcohol to form olefins and/or ethers [47, 93]. In addition, an improved alcohol dehydration catalyst could render the conversion of biomass-derived alcohols into transportation fuels a competitive source for such fuels. Much computational work on complex catalytic materials has focused on alumina-supported metal atoms.

3.3.1. Cr/Alumina System

Chromia/alumina catalysts are widely used industrially for dehydrogenation of alkanes to alkenes as given by the reaction,

$$C_nH_{2n+2} \rightarrow C_nH_{2n} + H_2 \tag{6}$$

In this process, different transition aluminas have been used to support the heterogeneous catalyst. At the present time, η-alumina is the preferred support in industrial practice because the catalysts Cr/η-alumina can last several years before the catalyst degrades. Quite remarkably, despite the close structural similarity of γ-alumina to η-alumina, the analogous catalytic system based on γ-alumina degrades within weeks. In the case of the γ-alumina support, the active chromium atoms disappear from the surface and have a tendency to form aluminum chromates. This considerable difference is quite puzzling because the two support materials differ essentially only in the distribution of cation vacancies [64, 94], and were historically sometimes taken to be identical [95, 96].

Rashkeev et al. [97] have applied first-principles DFT calculations to the study of Cr atoms on the surfaces of both γ- and η-aluminas and revealed the origin of the difference in degradation rates. These studies focused on the (110) surface because it is widely believed to be preferentially exposed [46]. The system was modeled by infinitely repeating slabs with an interslab vacuum spacing of 10 Å. The structural relaxations were performed for supercells consisting of slabs five atomic layers thick. Individual chromium atoms were deposited on the relaxed alumina surfaces at several different sites, and the minimum energy positions, as well as energy barriers to migration, were calculated.

Rashkeev et al. [97] found that when Cr is deposited on the relaxed η-alumina surface, it can be accommodated in several different sites, but the minimum energy position corresponds to the Cr atom, forming a strong bond with a three-coordinated Al and two oxygen atoms. Cr can diffuse along the surface, jumping from one three-coordinated Al to another in a "zigzag" trajectory. The activation energy for this process is found to be 2.5 eV. (It was pointed out for comparison that the energy cost to form a vacancy/Al-interstitial Frenkel pair is about 4–5 eV.) Throughout the entire trajectory, the Cr atoms are chemically active. For example, a Cr–H bond of bond energy about 4.7 eV will form when a hydrogen atom is added to the structure. This is larger than the energy cost of breaking a typical C–H bond in a hydrocarbon molecule, so it is not surprising that Cr/η-alumina is a good dehydrogenation catalyst.

On the reconstructed γ-alumina surface, the energetically preferred site places Cr attached to four undercoordinated surface oxygen atoms, with one of the four bonds considerably longer and weaker than the other three. Just as on η-alumina, Cr can also diffuse along the γ-alumina surface, also with an activation barrier of about 2.5 eV. Again, a Cr—H bond of about 4.8 eV binding energy will form in the presence of a hydrogen atom. Clearly, both the Cr/γ- and Cr/η-alumina systems exhibit catalytic activity.

It was also found that in both systems it is energetically favorable for Cr to occupy a nominally vacant position in a subsurface layer where it becomes sixfold coordinated. The positions are depicted in Fig. 5. The activation barrier for Cr to enter such positions, however, is much higher in the case of γ-alumina.

The activation barrier for Cr entering γ-alumina bulk is only 2.2 eV, as compared to ~5 eV in η-alumina. An important consequence of this result is that Cr can sustain its catalytic activity at the η-alumina surface for a very long time, while becoming completely deactivated on the γ-alumina surface within several days.

3.3.2. Pt/Alumina System

Studies of the alumina support have shown that that γ-alumina without bulk hydrogen is merely the terminus of a progression of hydrated forms [13]. The starting point, boehmite has the empirical formula $HAlO_2$, which can be written $Al_2O_3 \cdot 1(H_2O)$. Upon heating, boehmite loses water, progressing through various g-alumina forms $Al_2O_3 \cdot n(H_2O)$ until all of the water is driven off, leaving $Al_2O_3 \cdot 0(H_2O)$. From the observation that γ-alumina loses water upon heating [62, 98], and the variable hydrogen content, it follows that the material behaves as a sponge, in that it can store and subsequently release water, but in a unique, "reactive" way. When a water molecule arrives at the surface, it breaks up, H enters the bulk, and O stays at the surface. Al atoms countermigrate from the bulk to the surface, where they recombine with the new O atoms and extend the crystal matrix. The ratios, determined by valence requirements, work out so that for every three H_2O molecules, six H move in, two Al move out, and the crystal extends by a stoichiometric Al_2O_3 unit. In the reverse process, H comes out from the bulk and combines with surface O to evolve as water while surface Al countermigrates into the bulk. The net result is an etching of the material as both Al and O atoms leave the surface. Of course, the ratios work out the same way, so that the etching occurs by stoichiometric Al_2O_3 units. This unusual chemistry is a natural consequence of the fact that γ-alumina is not a single substance but a *sequence* of hydrogen-containing forms. The possibility that vacancies left behind in this reactive sponge process serve as nucleation sites for Pt clusters was explored by Sohlberg, Pantelides, and Pennycook [99]. More detailed studies by the authors are under way at the time of the preparation of this manuscript.

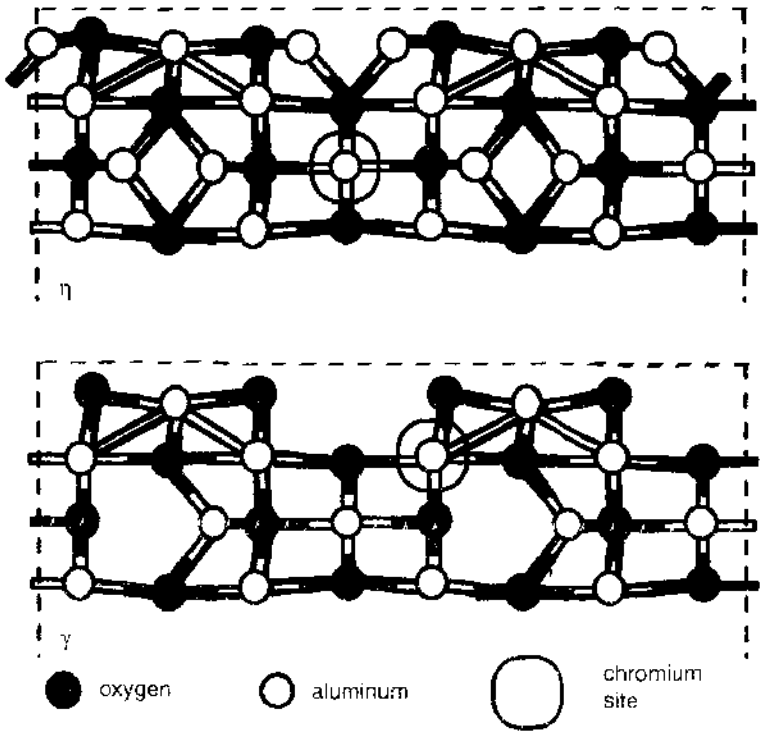


Figure 5. Cr atom adsorbed into a subsurface layer of γ-alumina (bottom) and η-alumina (top). The barrier to adsorption into η-alumina is about two times higher.

3.3.3. Nb/α-Alumina System

First-principles calculations have been reported for niobium on sapphire, studied as a representative metal–ceramic system [4]. Such systems are common in catalytic systems. The calculations show that the system is remarkably stable in that there is very little relaxation. The Nb atoms bind strongly to the "hollow" sites above vacant octahedral sites on the sapphire (0001) surface, effectively replacing an outermost layer of Al. In pure α-alumina, each O atom accepts a total of two Al valence electrons, and each Al donates its three valence electrons into O 2p orbitals. Nb differs from Al in that it has five valence electrons. When Nb replaces Al, three of these five valence electrons are therefore donated into oxygen 2p orbitals in the valence band. Given that O 2p orbitals are nominally 5.7 eV below Nb 4d orbitals, very strong binding can be expected, and indeed the first-principles calculations find the adsorption energy to be 13 eV/Nb atom. Mixing of about 10% O 2p character into the Nb 4d orbitals results in states within the band gap that are occupied by the remaining two Nb valence electrons.

In contrast to the lack of relaxation of the Nb-coated α-alumina surface, on the clean surface (no Nb atoms) the surface Al atoms were found to relax toward the bulk by about 80%, approaching very close to the adjacent oxygen plane. In addition to contributing to our understanding of the stability of the Nb/sapphire interface, the results are important because they show that full structural relaxation is crucial in accurately determining binding site preferences for add-atoms.

3.3.4. Au/Titania System

Although gold is a noble metal and generally chemically inactive, gold clusters and nanoparticles have useful catalytic properties [30]. The preferred support material is TiO_2. Theoretical calculations have been reported for both the rutile and anatase polymorphs of TiO_2 as supports for catalytic gold clusters.

Gold clusters on rutile TiO_2 (110) have been studied with both cluster [100] and periodic slab [101] models. Giordano et al. [100] carried out calculations for the Cu/TiO_2, Ag/TiO_2, and Au/TiO_2 systems. The Cu/TiO_2 calculations were performed with a periodic slab model and on $(TiO_2)_{13}$ and $(TiO_2)_{18}$ clusters. (Such stoichiometric clusters are though to minimize edge effects [102].) For Ti–Cu and O–Cu bond distances, both sets of calculations gave the same qualitative trends. They argued that the results of the cluster calculations were sufficiently "in line" with those employing periodic slabs that they deemed it adequate to report only cluster model calculations for the more computationally demanding Ag/TiO_2 and Au/TiO_2 systems. For the adsorption energies, however, the cluster and periodic slab models did not yield the same qualitative trends. For a single Cu adatom, for example, both cluser and periodic slab models showed that the most favorable site for adsorption is atop a bridging (twofold coordinated = "2f") oxygen on the stoichiometrically terminated surface, but the models disagree over whether adsorption to a fivefold coordinated Ti atom is preferred over adsorption in a fourfold hollow site. This discrepancy is noteworthy because their results for the adsorption of a single Au atom (cluster model) differ to some degree from those of Wahlström et al. [101], who employed a periodic slab model. The Giordano and Wahström groups found for adsorption atop a two-fold coordinated bridging oxygen 0.99 (1.5) eV and for adsorption on a five-fold coordinated Ti 0.51 (1.20) eV. Of greater significance, however, Wahlström et al. found that adsorption in an oxygen vacancy is even more stable (D_e = 2.0 eV), and a single vacancy can support the adsorption of three to five Au atoms. Vacancies therefore serve as nucleation centers for the formation of Au_n clusters. This work is an excellent example of theoretical calculations leveraging microscopy. The STM observations yield the vacancy-cluster ratio, the origin of which becomes clear in light of first-principles total energy calculations.

Theoretical studies of gold clusters on anatase TiO_2 (101) have yielded very similar results to those on the rutile polymorph. Like rutile (110), the stoichiometric termination of anatase (01) exhibits two-fold coordinated "bridging" oxygen atoms and five-fold coordinated Ti atoms. Vittadini and Sellioni [103] found the binding energy for adsorption of an Au adaton to the Ti(5f) on anatase (101) to be 0.39 eV and the binding energy for adsorption atop an O(2f) to be 0.31 eV. The interaction of Au with the Ti(5f) is similar to that

on rutile (110), but apparently its interaction with O(2f) is much weaker on anatase than on rutile. Because it is thought that Au binds to oxygen by transfer of its 6s electron into the conduction band of TiO_2, it may be possible to understand this bonding by scrutinizing the relative placement of the oxygen orbitals within the anatase and rutile band structures. Bonding of Au to Ti(5f) is thought to take place by donation of a pair of Au 5d electrons into empty Ti 3d orbitals. Presumably the Ti 3d orbitals are similarly positioned in the anatase and rutile band structures. Such an analysis could yield significant insight into the catalytic activity of the Au/TiO_2 system.

One striking similarity between adsorption of Au on anatase (101) and rutile (110) is that in both cases, adsorption is highly favored at oxygen vacancies. As long as there are more Au atoms than oxygen vacancies, the vacancies serve as nucleation centers for the formation of Au_n clusters. The Au–vacancy interaction must be stronger than the Au–Au interaction, however, because no such clustering occurs if the concentration of Au atoms is comparable to that of oxygen vacancies [103]. In TiO_2, nominally the four Ti 3d electrons are placed in empty oxygen 2p orbitals, the latter being buried in the valence band [104]. The formation of an oxygen vacancy then leaves behind two electrons that must go into the conduction band. In effect, the Fermi level jumps to the conduction band minimum, at least locally, which is probably related to the very large formation energy for oxygen vacancies (4.7 eV [103]).

3.4. Stability of the Support Material

At elevated temperatures (1000–1100°C), high surface area catalytic support materials, such as γ-alumina, undergo a catastrophic loss of porosity through sintering. This loss of porosity depletes the material of surface area, rendering it useless as a support, a process that places an effective upper limit on the temperature of operation for the catalyst. Stabilization of γ-alumina, therefore, represents an important industrial and commercial problem.

3.4.1. γ-Alumina → θ-Alumina Transformation

The phase transition of γ- to θ-alumina is the critical first step in the sintering process that leads to loss of porosity. It has been studied by first-principles DFT calculations by Cai et al. [57] and Cai and Sohlberg [105]. To model the transformation with first-principles methods requires a unit cell that can be morphed from representing the γ-phase to one that represents the θ-phase. Although the unit cell of cubic γ-alumina looks quite different from that of monoclinic θ-alumina, both of them can be redefined to cells containing $Al_{16}O_{24}$ that are very similar. (See Fig. 6.) This provides the required framework to study the transformation of γ- to θ-alumina. It is found that when some of the aluminum atoms in γ-alumina move to specific sites, a close approximation of θ-alumina is formed.

Two possible schemes for reordering the Al sub-lattice were identified: scheme A, reorder 8 aluminum atoms from 16d/8a sites to 2 16c and 6 48f sites; scheme B, reorder 14 aluminum atoms from 16d/8a sites to 6 16c, 6 48f, and 2 8b sites. The oxygen sublattice remains essentially unchanged. The θ-alumina models that result from these two different reorderings are translationally equivalent and optimize to the same structure as the experimental one within the margins of error of the models. The orientation relationship between γ- and θ-alumina defined by these models agrees with experimental observations.

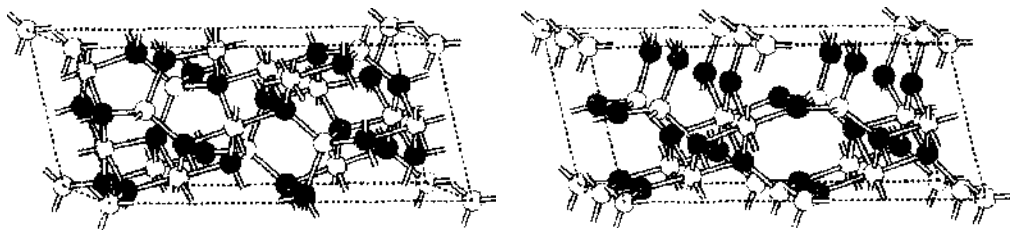


Figure 6. (left) Defect-free cell $\gamma_{\square}$; (right) cell $\theta_{\square}$ with translation of origin (solid spheres: oxygen; empty spheres: aluminum). Note the similarity, especially in the oxygen sublattice.

Based on a comparison of the energy differences obtained from first-principles calculations, the aluminum migration is found to take place first in the vicinity of cation vacancies to reduce strong Al–Al interactions. As 8b sites are involved in scheme B and Al atoms at 8b sites have one more strong Al–Al repulsive interaction than those at 48f or 16c sites, scheme 3 is energetically less favorable. In addition, as six more Al atoms are reordered in scheme 3, scheme B is also statistically less probable. Starting from a lowest energy configuration of γ-alumina, the lowest energy pathway of transformation by scheme A was mapped out. Transition states between adjacent intermediates were identified by successively fixing the position of the migrating atom and one of the atoms far away from it and relaxing all other atoms. (The second atom is chosen to be an atom that does not appreciably change its position throughout the transformation from the initial to final intermediates. This prevents "siding" of the entire unit cell.) The conversion rate for the transformation is then computed based on this pathway. The rate is then given by

$$r = \upsilon f \rho(E > \Delta E) \tag{7}$$

where υ is the vibrational frequency corresponding to small oscillations of the Al atom in its equilibrium position prior to the rate determining step, f is the population of this reactant, and $\rho(E > \Delta E)$ is the probability that the system has an energy greater than ΔE. Vibrational frequencies are typically estimated by finite differences [86]. The latter two factors are temperature dependent. Computing the rate as a function of temperature recovers the experimental transformation temperature to high accuracy. The experimentally observed translational and rotational interfaces in θ-alumina can be explained by different aluminum migration paths (resulting in models A, B, and their variants) in neighboring domains during the γ- to θ-alumina transformation.

3.4.2. La/Alumina System

It is widely known in the catalysis community through decades of empirical studies that doping of catalytic materials can improve thermal stability. The stabilization of γ-Al_2O_3 (or other porous alumina polytype used as a catalytic support) to prevent its transformation to α-Al_2O_3 can be achieved by adding certain chemical elements at different stages of the γ-alumina production process. Patented processes using La, P, Si, Ba, and Ce have been reported. Lanthanum is an expensive chemical, but an efficient industrial process has been devised to introduce it into γ-alumina. Silicon and some of the other elements, on the other hand, are inexpensive chemicals, but efficient industrial processes for incorporating them in γ-alumina are not available. Thus, industry would like to optimize the amount of La needed for stabilization or develop processes that use less expensive chemicals.

In recent work at Alcoa [106], it was demonstrated that stabilization of γ-alumina up to 1200°C can be achieved with as little as 0.3% La. This is an order of magnitude lower than existing practices, and the results suggest that in the industrial process of stabilization, large amounts of La are not active in the desirable mode. Unfortunately, the method of preparation by which alumina can be stabilized with only 0.3% La does not lend itself immediately to industrial use. Nevertheless, reduction of the amount of La used to stabilize γ-alumina is in principle possible and remains a major industrial objective.

It is clear that an atomic-scale understanding of the stabilization issue could impact industrial production. In particular, understanding the role of La might lead to insights as to how to modify the industrial process of incorporating La in order to achieve stabilization with much smaller concentrations. On the atomic scale, a combination of experimental and theoretical investigations can identify the configurations of La in γ-alumina that are responsible for stabilization and those that are inactive. The problem of stabilization is, of course, more complex and involves processes on longer length scales. For example, the transformation from γ to α alumina is likely to occur through the nucleation and growth of α-alumina particles. Experimental [107, 108] and theoretical [55] work has demonstrated that γ-alumina has lower surface energy than α-alumina so that an interplay between bulk volumes and surface areas of the competing phases may be the key to stabilization. What then is the role of La and the other additives? Is La dispersed as single atoms and remain so at high temperatures?

Does it stop the nucleation of the α-phase, or does it inhibit its growth by residing on the surface of nucleated α-phase particles? If so, is the process controlled by surface (interface) or bulk diffusivities of La? Though experiments and theory have provided valuable insights for some of these questions, definitive answers have been lacking. On the basis of very limited information, there have been conflicting suggestions that La forms $LaAlO_3$ or La_2O_3 monolayers on γ-Al_2O_3 surfaces or it is incorporated in the bulk [106, 109–112]. There have been no suggestions about the atomic-scale processes that actually delay the onset of the transformation process.

Wang et al. [113] investigated the role of La in stabilizing γ-alumina using a unique combination of first-principles density-functional total-energy calculations and experimental tools including direct atomic-resolution Z-contrast imaging of La-doped γ-Al_2O_3 samples and extended X-ray absorption fine structure spectra (EXAFS). The three approaches yield complementary information that all leads to the same conclusion: La atoms avoid going into the bulk and adsorb strongly on γ-Al_2O_3 surfaces as *single atoms* without any clustering. Perhaps the most compelling evidence is provided by the Z-contrast scanning transmission electron microscopy (*Z*-STEM), where single La atoms are seen clearly superposed on the crystal lattice planes formed by the substrate. Surprisingly, unlike the undoped case [13, 55], there was no apparent preferential exposure of the [110] surface. Instead, the [100] surface was encountered often. A square arrangement of Al-O columns (representative of the (100) termination) is clearly resolved. In Z-STEM the intensity contributed by an atom is roughly proportional to Z^2, where Z is the atomic number [114], allowing single La atoms to be visible in the form of brighter spots on the background of thicker but considerably lighter γ-Al_2O_3 support. Most of the La atoms are located directly over Al-O columns, but a small fraction also occupies a position shifted from the Al-O column. The Z-STEM images also reveal that there is no apparent correlation in the distribution of dopant atoms.

The presence of La was also confirmed by electron energy loss spectroscopy (EELS), but high-resolution EELS study was not feasible because of rapid beam damage of the substrate at very high magnifications, so it was not possible to use EELS to determine the position of the La atoms relative to the alumina lattice. Further information about the distribution of La was obtained by changing the focus of the microscope electron beam from the top surface to the middle of the sample to the bottom surface. It was found that the La atom images gradually fade and finally reappear, confirming that the La atoms are, in fact, on the two surfaces.

The theoretical calculations were performed within density functional theory, using the pseudopotential method and a plane wave basis set [5–7, 115]. Exchange correlation was included using the generalized gradient-corrected functionals (GGA) given by Perdew and Becke [15, 116]. The Vanderbilt ultrasoft pseudopotentials were used for O and H atoms, a norm-conserving pseudopotential for Al, and a projector-augmented wave (PAW) potential for La [117–119]. A plane wave energy cutoff of 400 eV and two special k-points in the irreducible part of the two-dimensional Brillouin zone of the surfaces were used for calculating both the (100) and (110) surfaces of γ-Al_2O_3. Semi-infinite (100) and (110) surfaces were modeled by repeated slabs (supercells) containing 7 to 12 atomic layers for the (100) surface and 5 to 8 layers for the (110) surface (68–112 and 70–128 atoms, respectively) separated by a vacuum region equivalent to 10 to 12 Å (supercells containing 80 atoms and a 2×2 surface cell were employed to represent the α-Al_2O_3 surface). The cation vacancies that are inherently present in the spinel form of γ-Al_2O_3 were located on the tetrahedral cation sublattice. All the atoms in the supercell except for those in the lower one or two atomic layers (which were kept fixed) were relaxed until the forces on the atoms were smaller than 0.05 eV/Å.

The calculations found that La atoms favor the surfaces over bulk sites by substantial energy differences (~4 eV), with several configurations, including the two observed in Fig. 7.

Figure 7 represents the configurations of the undoped and La-doped (100) surface. In the absence of La, the (100) surface (Fig. 7a) shows only minor relaxation effects; all the surface aluminum atoms are five-coordinated, and the surface oxygen atoms are either three- or four-coordinated. The cation vacancies are located between the first and second oxygen subsurface layer. However, when an La atom is introduced, a significant relaxation of the structure occurs. One of the five-coordinated Al atoms (adjacent to La) is displaced from

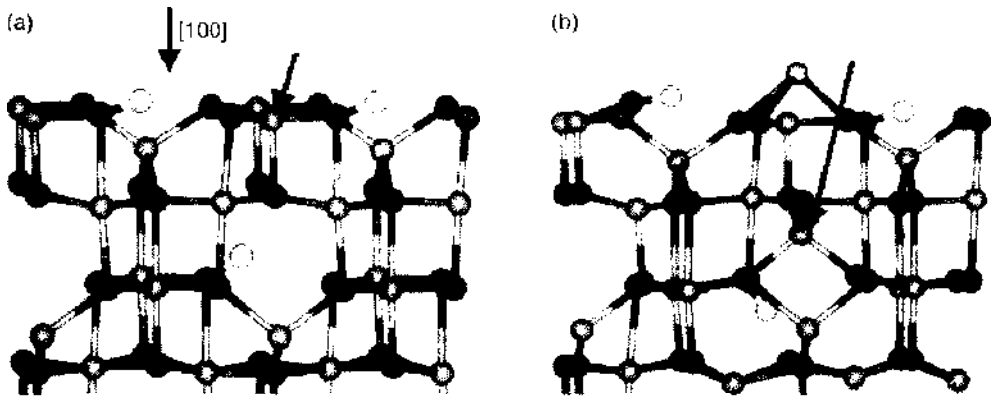


Figure 7. Schematics of the configurations for the (100) surface of γ-Al_2O_3, determined by first-principles calculations: (a) undoped and (b) La-doped. The Al, O, H, and La atoms are shown in light, red, white, and blue, respectively. The position of the aluminum atom that relaxes from the surface into the cation vacancy is indicated by the blue arrow.

the surface into the subsurface tetrahedral vacancy site (Fig. 7b). The La atom occupies the resultant surface fourfold hollow site, which is close to the initial location of the Al atom in planar coordinates but located ~1.2 Å above it, making La—O bond lengths 2.3–2.5 Å. This configuration, obtained by total-energy minimization, is found in the Z-STEM images. Calculations aimed at revealing the structure of the other observed site found that it corresponds to an La atom with one of the four neighboring surface O atoms missing. The asymmetry forces the La atom off the Al-O column, as observed. Its formation is clearly the result of the presence of surface O vacancies.

Additional calculations exploring the incorporation of La atoms in bulk γ-Al_2O_3 found that an La atom, when initially placed at a vacancy, interstitial, or substitutional site in the second or third subsurface layer, would relax up to the surface. When an La atom is initially located in a deeper layer (the eighth or ninth layer of the supercell for the (100) surface, and the fifth or sixth layer for the (110C) surface), which is equivalent to the bulk, the total energy of the system is significantly higher (~4 eV) than that of the configuration with La on the surface (models with La in different bulk sites were examined). The preference for surface sites over the bulk arises primarily from the strong binding of La on the surface and the large difference in ionic size between La^{+3} (1.03 Å) and Al^{+3} (0.54 Å) [120]. The theoretical result corroborates the through-focus imaging analysis. The marked preference for surface sites versus the bulk is an important factor in the inhibition of sintering. Progress of the sintering process would inevitably trap some of the surface La atoms in the bulk, thus forcing the system into highly strained and energetically unfavorable configuration. The resulting inhibition effect helps retain a large surface-to-volume ratio (specific surface area) for γ-Al_2O_3 at higher temperatures, in agreement with experimental observations [109, 121].

The binding energy of La to the (100) surface is very high (8.6 eV), due partly to the removal of the surface Al atom into the subsurface, which enhances the attractive interaction between La and the surface O atoms and reduces the otherwise strong repulsion between La and the Al atom. The strong binding also causes large migration energies (4–5 eV) for typical paths connecting equivalent configurations. Similar calculations for the (110C) surface, which is exposed preferentially in the *undoped* γ-Al_2O_3 [13, 55], also resulted in high values of the binding energy (7.5 eV). In this case, however, La atoms occupy existing surface hollow sites, which are created on the undoped surface [69, 70] by displacement of three-coordinated surface Al atoms into the empty octahedral sites in the first subsurface layer. It is possible that the difference in the two binding energies is responsible for the occurrence of (100) surfaces only after annealing in the presence of La dopant. This interesting possibility needs to be further explored. When the same computational procedure is carried out for α-Al_2O_3 (0001) surface, much lower binding energy (4.3 eV) for La atoms is obtained. This difference means that doping would cause an increase in the enthalpy of α-Al_2O_3 relative to γ-Al_2O_3, thus further stabilizing γ-Al_2O_3.

Finally, independent evidence that La prefers surface sites was provided by EXAFS measurements. These were carried out for γ-alumina samples doped with La at several temperatures below and above the transformation temperature. It was found that La does not have a second well-defined coordination shell, indicating that atomically dispersed distribution of La is favorable.

Wang et al. [113] proposed the following scenario: La atoms coat the surfaces of porous alumina as dispersed single adatoms. As the temperature is increased, their presence on the surface prohibits sintering and transformation to nonporous α-alumina, because any sintering would either trap La atoms in the bulk, which is energetically unfavorable, or expel them to the surface of α-alumina, which is again not favored energetically. La, of course, only delays the onset of the transformation until new pathways become available and lead to the formation of an aluminate. Overall, this study is one of the strongest examples of the successful interplay of theoretical and experimental methods to elucidate atomic-scale details of complex catalytic materials.

3.5. Photocatalysis

3.5.1. TiO_2

Titanium oxide is perhaps best known as a paint pigment, but it is also a photocatalyst for dissociation of water into hydrogen and oxygen. With the growing interest in moving the economy from a petroleum base to a hydrogen base, which was seeded as far back as the "energy crisis" of the early 1970s, there is commensurate interest in photocatalytic dissociation of water. Titanium oxide has become the *de facto* standard system for the study of this process. Its catalytic properties also find use in remediation of pollutants and in the preparation of antimicrobial surfaces.

Early theoretical treatments of the electronic structure of TiO_2 gave insight into the basic band structure [122] and differences in the orbital interactions among the rutile, anatase, and brookite polymorphs [123], but the first in-depth information about the geometric structure of exposed surfaces, which is essential for understanding catalytic action, did not come until more than a decade later. Ramamoorthy, Vanderbilt, and King-Smith [124] reported DFT calculations with pseudopotentials and a plane wave basis on supercell slab models of various stoichiometric exposures of rutile TiO_2 surfaces and identified (110) as having the lowest surface energy, while (100) is also stable with respect to forming facets of (110). Only very minor displacements of the surface atoms from their bulk positions were found upon full structural relaxation. The most significant relaxations were 0.13 displacement of 6f Ti in the direction of the surface vector, 0.17 Å displacement of the 5f Ti in the direction of the surface vector, and 0.13 Å displacements of the bridging oxygen atoms in the direction of the surface vector (6f = sixfold coordinated, 5f = fivefold coordinated, and so on). These displacements were quantitatively confirmed by surface X-ray diffraction by Charlton and coworkers in 1997 [125]. Surface energies reported by Ramamoorthy, Vanderbilt, and King-Smith [124] are about three times larger than those reported experimentally [126]. Ramamoorthy, King-Smith, and Vanderbilt [127] also studied the nonstoichiometric Ti–O termination of the rutile TiO_2 (110) surface. There are two major qualitative differences from the stoichiometric termination. First, upon full structural relaxation, the 3f surface oxygen atoms displace outward from the surface so that they are no longer in a trigonal-planar coordination but instead become the apex of a trigonal pyramid. This result was subsequently confirmed by more detailed spin-polarized calculations reported by Lindan et al. [128]. This relaxation is consistent with simple freshman chemistry ideas about oxygen coordination. The preferred local geometry of oxygen is tetrahedral coordination by electron pairs (bonds or free pairs). Within the rutile bulk, the oxygen atoms are constrained to the trigonal planar local geometry, but when exposed to the surface, they are free to pucker outward, thereby achieving a local geometry closer to the chemically preferred one. The second major feature of the Ti–O terminated surface that distinguishes it from the stoichiometric termination is that in the former case, there are occupied states at the conduction band minimum. These states are localized near surface-exposed Ti atoms, an unsurprising result given that the removal of surface oxygen atoms will deplete acceptor oxygen p orbitals from the valence

band. Absent enough acceptor oxygen p orbitals in the valence band, the Ti valence electrons are forced to fill conduction band orbitals; that is, they remain on Ti. Band structure calculations [122, 129] reveal that the O_{2p} bands are predominantly within the valence band and the Ti_{3d} bands are predominantly within the conduction band. The occupied states at the conduction band minimum almost certainly are active in catalysis, and it has recently been confirmed by an elegant combination of scanning tunneling microscopy and DFT calculations that oxygen vacancies are correlated to active sites when TiO_2 acts in the photocatalytic dissociation of water on the (110) surface [130]. Restructuring around such vacancies has been seen by STM and in first-principles calculations as reported in Ref. [131].

Full potential calculations on TiO_2 include a study of the optical properties of the anatase polymorph [132] and of the rutile (110) surface [133]. The FLAPW calculations on rutile TiO_2 (110) surface were less successful than the DFT/pseudopotential calculations described previously. Displacements of surface atoms upon structural relaxation were reported that are now known to be contrary to experimental observation [125]. In light of a more recent and elaborate numerical convergence study by Muscat, Harrison, and Thorton [134], the reason for this failure is almost certainly the use of an insufficiently thick slab model. (Given that FLAPW calculations are considerably more expensive than pseudopotential calculations, and considering the early date of the FLAPW work, it is understandable that the calculations were restricted to a small supercell.) The convergence study considered the (100) termination. (Muscat and Harrison also reported calculations of the electronic and geometric structure of the rutile TiO_2 (001) termination, which is known to be unstable with respect to faceting [124].) It was found that 90% of the convergence in atom displacements and surface energies is recovered with a slab model that is six atomic layers thick, equivalent to two TiO_2 units. Nine atomic layers (3 TiO_2 units) were found sufficient to yield essentially full convergence. The convergence study also investigated the influence of different treatments of the exchange and correlation on computed properties. Differences in geometric structure between DFT/LDA, DFT/GGA, and Hartree-Fock (HF) treatments were found to be very minor. Differences in formation energy were quantitatively different by as much as 30%, but the ordering of the different surface exposures was found to be invariant with the choice of exchange and correlation treatment.

The anatase polymorph of TiO_2 has been the subject of fewer theoretical investigations than the rutile form, but it is in fact a better catalyst in some applications, and such studies are of great interest. In addition to the FLAPW study noted before, Lazzeri and Selloni [135] have reported DFT/plane wave/pseudopotential calculations on anatase and its surfaces. One particularly interesting feature of anatase is that while its bulk is less stable than the rutile polymorph by about 1.2 kcal/mol, it has generally lower surface energies [104]. A (4 × 1) reconstruction of the anatase (001) surface lowers its surface energy still further [135]. The lower surface energy of the anatase polymorph relative to rutile is important because for sufficiently small particles, the anatase form may actually be favored. Consider the volume V and surface area A of a particle as a function of the quantity of material. Given a density ρ and molar mass M, if we assume a spherical particle of radius r for simplicity (the qualitative result that follows is essentially invariant with particle shape), we may write:

$$V = nM/\rho = 4\pi r^3/3 \tag{8}$$

where n is the number of moles. Solving for r yields

$$r = (3V/4\pi)^{1/3} = (3nM/4\pi\rho)^{1/3} \tag{9}$$

Relating the surface area to the quantity of material yields

$$S = 4\pi r^2 = 4\pi(3nM/4\pi\rho)^{2/3} \tag{10}$$

It follows that the volume increases linearly with the quantity of material but the surface area increases with only the (2/3) power of the quantity of material. If we write the total particle energy E_T as a sum of bulk E_V and surface E_S terms,

$$E_T = E_V + E_S = aV + bS = a\,nM/\rho + b4\pi(3nM/4\pi\rho)^{2/3} \tag{11}$$

where a and b are unspecified constants. It is clear from Eq. (8) that for large n, the volume term will dominate, but for sufficiently small n, the surface term will dominate. It follows that if the b coefficient for anatase is smaller than that for rutile, the former will be more stable at sufficiently small particle size. Evidently, the threshold particle size is about 10 nm.

Compared with the undoped materials, $TiO_{2-x}C_x$ has been shown to exhibit greatly improved catalytic efficiency for water splitting. An atomic-level understanding of the physical origins of this improved efficiency would facilitate the systematic engineering of improved efficiency in other photocatalytic materials. It has observed that the band gap in TiO_2 decreases upon doping with carbon [136]. It is speculated to decrease further with the introduction of more impurities of C. We have carried out all-electron FLAPW band structure calculations of for pure TiO_2, Ti_4O_7C and $Ti_4O_6C_2$. We find that the decrease of the band gap from TiO_2 to Ti_4O_7C is dominated by C p electrons. For $Ti_4O_6C_2$, strong C interaction eliminates the band gap entirely. We have also carried out surface reconstruction calculations for the Ti—O exposure of the (110) surface and find the reconstruction reported earlier by Ramamoorthy, King-Smith, and Vanderbilt [127], whereby the O atoms, trigonally coordinated within the TiO_2 bulk, "pop out" of the surface and seek T_d geometry. The moving out of O atoms on the surface may form a site where absorbed water H_2O is excited into a "ready-for-splitting" state. Although the stoichiometric (oxygen terminated) (110) surface is widely held to be preferentially exposed, combined DFT calculations based on a periodic slab model and STM experiments have established that the catalytically active sites are oxygen vacancies, namely, exposed Ti atoms [130].

4. FUTURE DEVELOPMENTS

4.1. Some Selected Outstanding Problems

4.1.1. Interaction of N_2 with Molybdenum Nanoclusters

It is known that geometrical surface structures, not electronic effects due to cluster size, are a key to understanding trends in the reactivity of N_2 with metal clusters [137]. This is because different-sized nanoclusters expose varying surface areas of different surface panes, and the reactivity of N_2 exhibits a strong dependence on the surface plane with which it reacts. It is thought that when N_2 interacts with the (100) surface of W metal, it penetrates close to the surface where π^* orbital of N_2 interacts with metal d orbitals, activating the N—N bond for dissociation or reaction. When N_2 interacts with the (110) surface of W metal, however, the close packing of the metal atoms on the surface prohibits this strong interaction. Mitchell et al. [137] propose that the same mechanism holds for N_2 reacting with molybdenum nanoclusters. This mechanism has been tested with empirical potential-based methods for N_2 and W surfaces [138] but not at all for Mo surfaces. Testing of this hypothesized mechanism with first-principles quantum mechanical methods, which explicitly model the interactions of the valence electrons, would be a valuable contribution to the science of cluster catalysis.

4.1.2. Dissociative Adsorption of NO and CO

The catalytic treatment of NO and CO from vehicle exhaust remains one of the most important problems in automotive catalysis and one with tremendous potential to benefit from designed catalysts. Nanoclusters of vanadium (V_n) have been shown to have strong size-dependent reactivity toward D_2 (deuterium, used for isotopic labeling) and N_2 [139]. Similar results have been reported for D_2 and nickel clusters [140]. Size-dependent reactivity is the ideal characteristic for designing catalysts with specific activity, but unfortunately the V_n clusters show very weak size-dependent reactivity with CO and NO. Understanding the fundamental properties that govern the size dependence of the electronic structure, and hence reactivity is now a key barrier to designing catalysts with specific activity.

In the most widely advocated mechanism of the reduction of NO in vehicle exhaust, the reduction initiates with the dissociative adsorption of NO [48],

$$NO \rightarrow NO_{ad} \rightarrow N_{ad} + O_{ad} \qquad (12)$$

Here the subscript *a* denotes an adsorbed species. The adsorption site is unknown but often hypothesized to be a metal center or metal nanocluster on the supporting alumina washcoat. This step seems plausible, particularly in light of experimental [141] and theoretical [142] evidence for the dissociative adsorption of CO. One theoretical study, however, has suggested that the adsorption of NO on Pt or Pd actually *strengthens* the NO bond [143]. A coupling of two adsorbed NO molecules was suggested as the initial step.

Two possible steps have been proposed for the final production of N_2. One possibility is that adjacent adsorbed N atoms combine,

$$2N_a \rightarrow N_2 \tag{13}$$

Another possibility is that an adsorbed N atom reacts with an adjacent adsorbed NO,

$$NO_a + N_a \rightarrow N_2 + O_a \tag{14}$$

This case of the reduction of NO is just one example where the computational methods that could be used choose between proposed mechanisms. This kind of application (choosing between proposed mechanisms) is one of the strengths of computational methods.

4.1.3. V_2O_5

Vanadium has numerous stable oxidation states and consequently forms many oxides [144]. Vanadium pentoxide releases oxygen reversibly upon reduction and is used as a catalyst in the mild oxidation of organics [145]. Unlike many metal oxide heterogeneous catalysts in which the coordinatively unsaturated (cus) metal atoms at the surface function as Lewis acid sites, it is thought that in V_2O_5, nucleophilic reaction occurs at surface oxygens [145]. Vanadium pentoxide has a layered structure with weak interlayer coupling. This fact has been used to advantage in designing cluster calculations; a single layer, terminated with hydrogen, gives a reasonable facsimile of the bulk electronic structure. Such cluster calculations of vanadium pentoxide surfaces have revealed that there are at least three different types of surface-exposed oxygen: singly coordinated (O^1), two-coordinated (O^2), and three-coordinated (O^3), hinting at a rich surface chemistry [146]. The singly coordinated vanadal (V=O) species are of particular interest because the vanadal oxygen will be electron-rich. Periodic slab DFT calculations have shown that binding of a single H atom at the surface is energetically most exothermic at the vanadal oxygen and that the O–H bond formed there is shorter than that formed by H binding at either the O^2 or O^3 surface oxygen species [147]. These results represent only preliminary forays into the study of vanadium oxide. This material is ripe for theoretical investigations. Oxides other than V_2O_5 have seen very little theoretical attention. The surface chemistry is of great interest, and future bulk and surface studies could be particularly fruitful.

4.2. Leveraging Infrared Spectroscopy

To study heterogeneous catalyst surfaces and catalytic *reactions*, one needs a technique that will probe surface-adsorbed species and reaction intermediates. HREM is inappropriate because the high-energy electrons will dissociate adsorbed species and destroy the reaction intermediates long before an image can be captured. With good resolution, high sensitivity, and no requirement of a vacuum environment, IR spectroscopy of adsorbed molecules would seem an ideal source of experimental input. Consider, for example, the interaction of an alcohol with the surface of γ-alumina. The catalytic properties of alumina for alcohol dehydration have been exploited for more than 200 years [47, 93], but the mechanism is still unknown! If one knew *how* the alcohol molecule adsorbs to the surface, it would provide important clues for elucidating the mechanism. Several chemically reasonable adsorbed structures are easily envisioned. If one could compute the vibrational frequencies of these structures with sufficient accuracy, the correct structure could be easily identified by simply matching the experimental IR spectrum.

Regrettably, current computational methods do not have adequate accuracy. Today, the most common method of carrying out such a calculation is to find the normal modes of vibration for the adsorbed species by diagonalizing the mass-weighted force-constant matrix for the adsorbed molecule. This technique gives the vibrational frequencies within the harmonic approximations, typically accurate to about ±5% (about ±150–200 cm $^{-1}$ for OH vibrations). By contrast, definitive assignment of OH on γ-alumina, for example, would require an accuracy of about ±10 cm^{-1} [148].

To obtain the desired accuracy, two approximations, which are the primary sources of error in conventional computations of vibrational frequencies for surface-adsorbed species, must be removed. These approximations are: (1) the harmonic approximation and (2) the neglect of dynamical coupling of the adsorbate to the support surface. To remove these approximations, one possibility is to compute the vibrational frequencies for surface-adsorbed species by extracting them directly from classical dynamical trajectories. Quantum information can be extracted from a classical trajectory as a consequence of the Heisenberg correspondence principle [149, 150]. The frequencies from a dynamical trajectory are based not on an approximating set of harmonic oscillators but on the correct molecular Hamiltonian and therefore incorporate all of the couplings and anharmonicities of the system. To avoid errors introduced by a functional fit to the potential energy surface (PES), the trajectories should be computed based on total energy and derivatives thereof (forces) computed "on the fly" with electronic structure methods. These electronic structure calculations must reproduce the *shape* of the PES with quite high accuracy so as not to reintroduce errors of a magnitude comparable with those removed by eliminating the harmonic approximation. Detailed studies from the literature [151] suggest that the best choice for such electronic structure calculations is an all-electron DFT calculation for periodic (slab) systems. This will avoid the overestimation of force constants typical of Hartree-Fock calculations, as well as the finite size effects of cluster calculations.

Overall, advances in the computational modeling of complex catalytic materials are inextricably linked to advances in theoretical methodology.

ACKNOWLEDGMENTS

Our research on complex catalytic materials is sponsored in part by the U.S. Department of Energy under contract DE-FC02-01CH11085, by an NSF GOALI grant DMR-0111841 with Alcoa, Inc., and by a DuPont Young Professor award to K. S. Computations have been partially supported by the National Center for Supercomputing Applications (NCSA) under grant CHE990015N and used the SGI Origin2000, University of Illinois at Urbana-Champaign.

REFERENCES

1. I. E. Maxwell, *Nature* 394, 325 (1998).
2. C. Pisani, *J. Molec. Struct. THEOCHEM* 463, 125 (1999).
3. F. Jensen, "Introduction to Computational Chemistry," Wiley, Chichester, England, 1999.
4. C. Kruse, M. W. Finnis, V. Y. Milman, M. C. Payne, A. D. Vita, and M. J. Gillan, *J. Am. Ceram. Soc.* 77, 431 (1994).
5. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* 64, 1045 (1992).
6. G. Kresse and J. Furthmüller, *Phys. Rev. B* 54, 11169 (1996).
7. G. Kresse and J. Furthmüller, *Comput. Mater. Sci.* 6, 15 (1996).
8. J. S. Lin, A. Qteish, M. C. Payne, and V. Heine, *Phys. Rev. B* 47, 4174 (1993).
9. Molecular Simulations Inc., San Diego, CA.
10. G.-D. Lee, M. H. Lee, and J. Ihm, *Phys. Rev. B* 52, 1459 (1995).
11. A. Takada, C. R. A. Catlow, J. S. Lin, G. D. Prince, M. H. Lee, V. Milman, and M. C. Payne, *Phys. Rev. B* 51, 1447 (1995).
12. B. Winkler, V. Milman, B. Hennion, M. C. Payne, M.-H. Lee, and J. S. Lin, *Phys. Chem. Minerals* 22, 461 (1995).
13. K. Schlberg, S. J. Pennycook, and S. T. Pantelides, *J. Am. Chem. Soc.* 121, 7493 (1999).
14. W. Kohn and L. J. Sham, *Phys. Rev. A* 140A, 1133 (1965).
15. J. P. Perdew, *Phys. Rev. B* 33, 8822 (1986).
16. L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* 48, 1425 (1982).

17. K. Laasonen, A. Pasquarello, C. Lee, R. Car, and D. Vanderbilt, *Phys. Rev. B* 47, 10142 (1993).
18. K. Laasonen, R. Car, C. Lee, and D. Vanderbilt, *Phys. Rev. B (Rapid Communications)* 43, 6796 (1991).
19. H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* 13, 5188 (1976).
20. K. Wefers and C. Misra, "Oxides and Hydroxides of Aluminum," Alcoa, Pittsburgh, 1987.
21. H. Sun, *J. Phys. Chem. B* 102, 7338 (1998).
22. J. C. Boettger, *Phys. Rev. B* 55, 750 (1997).
23. P. Blaha, K. Schwarz, P. Sorantin, and S. B. Trickey, *Comp. Phys. Commun.* 59, 399 (1990).
24. P. Blaha, K. Schwarz, and J. Luitz, Karlheinz Schwarz, "WIEN," Techn. Univ., Vienna, 1999.
25. H. Krakauer, D. D. Koelling, E. Wimmer, M. Weinert, S. H. Wei, D. Singh, C. S. Wang, W. E. Pickett, and M. Posternak, LAPW: based in part on: *J. Phys. C* 10, 2107 (1977); F. B. Hildebrand, "Introduction to Numerical Analysis," McGraw-Hill, New York, 1956, pp. 71–76; A. H. MacDonald, W. E. Pickett, and D. D. Koelling, *J. Phys. C* 13, 2675 (1980); W. E. Pickett, D. D. Koelling, and A. J. Freeman, *Phys. Rev. B* 23, 1266 (1981); M. Weinert, *J. Math. Phys.* (1980); U. Von Barth and L. Hedin, *J. Phys. C: Solid State Phys.* 5, 1629 (1972); L. Hedin and B. I. Lundqvist, *J. Phys. C: Solid State Phys.* 4, 2064 (1971).
26. D. Alfe and M. J. Gillan, *Phys. Rev. B* 58, 5161 (1998).
27. C. Wolverton and K. C. Hass, *Phys. Rev. B* 63, 024102 (2001).
28. M. D. Segall, R. Shah, C. J. Pickard, and M. C. Payne, *Phys. Rev. B* 54, 16317 (1996).
29. M. D. Segall, C. J. Pickard, R. Shah, and M. C. Payne, *Mol. Phys.* 89, 571 (1996).
30. M. Valden, X. Lai, and D. W. Goodman, *Science* 281, 1647 (1998).
31. H.-J. Liu, J. T. Hupp, and M. A. Ratner, *J. Phys. Chem.* 12204 (1996).
32. J. M. Nedeljkovic, M. T. Nenadovic, O. I. Micic, and A. J. Nozik, *J. Phys. Chem.* 90, 12 (1986).
33. Y. H. Pan, K. Sohlberg, and D. P. Ridge, *J. Am. Chem. Soc.* 113, 2406 (1991).
34. H. Cheng and L.-S. Wang, *Phys. Rev. Lett.* 77, 51 (1996).
35. L.-S. Wang and H. Cheng, *Phys. Rev. Lett.* 78, 2983 (1997).
36. L.-S. Wang, X.-B. Wang, H. Wu, and H. Cheng, *J. Am. Chem. Soc.* 120, 6556 (1998).
37. L.-S. Wang, H. Wu, and H. Cheng, *Phys. Rev. B* 55, 12884 (1997).
38. Y. D. Kim, M. Fischer, and G. Gantefor, *Chem. Phys. Lett.* 377, 170 (2003).
39. B. E. Douglas, D. H. McDaniel, and J. J. Alexander, "Concepts and Models of Inorganic Chemistry," Wiley, New York, 1983.
40. H. Fujiwara, H. Hosokawa, K. Murakoshi, Y. Wada, S. Yanagida, T. Okada, and H. Kobayashi, *J. Phys. Chem. B* 101, 8270 (1997).
41. L. G. Wang, S. J. Pennycook, and S. T. Pantelides, *Phys. Rev. Lett.* 89, 075506 (2002).
42. S. Helveg, J. V. Lauritsen, E. Laegsgaard, I. Stensgaard, J. K. Norskov, B. S. Clausen, H. Topsoe, and F. Besenbacher, *Phys. Rev. Lett.* 84, 951 (2000).
43. J. R. Sachleben, V. Colvin, L. Emsley, E. W. Wooten, and A. P. Alivisatos, *J. Phys. Chem. B* 102, 10117 (1998).
44. D. Coster, A. L. Blumenfeld, and J. J. Fripiat, *J. Phys. Chem.* 98, 6201 (1994).
45. C. N. Satterfield, "Heterogeneous Catalysis in Practice," McGraw Hill, New York, 1980.
46. H. Knozinger and P. Ratnasamy, *Catal. Rev.–Sci. Eng.* 17, 31 (1978).
47. B. C. Shi and B. H. Davis, *J. Catal.* 157, 359 (1995).
48. K. C. Taylor, *Catal. Rev.–Sci. Eng.* 35, 457 (1993).
49. D. S. MacIver, H. H. Tobin, and R. T. Barth, *J. Catalysis* 2, 485 (1963).
50. A. J. Leonard, P. N. Semaille, and J. J. Fripiat, *Proc. Br. Ceram. Soc.* 103, 103 (1969).
51. B. C. Lippens and J. J. Steggerda, in "Physical and Chemical Aspects of Adsorbents and Catalysts" (B. G. Linsen, Ed.), Academic, London, 1970.
52. R. W. G. Wyckoff, "Crystal Structures," Interscience, New York, 1963.
53. K. Sohlberg, J. Pennycook, and S. T. Pantelides, *Chem. Eng. Comm.* 181, 107 (2000).
54. L. J. Alvarez, J. F. Sanz, M. J. Capitan, and J. A. Odriozola, *Chem. Phys. Lett.* 192, 463 (1992).
55. S. Blonski and S. H. Garofalini, *Catal. Lett.* 25, 325 (1994).
56. S. Blonski and S. H. Garofalini, *Surf. Sci.* 295, 263 (1993).
57. S.-H. Cai, S. N. Rashkeev, S. T. Pantelides, and K. Sohlberg, *Phys. Rev. Lett.* 89, 235501 (2002).
58. M.-H. Lee, C.-F. Cheng, V. Heine, and J. Klinowski, *Chem. Phys. Lett.* 265, 673 (1997).
59. S.-D. Mo, Y.-N. Xu, and W.-Y. Ching, *J. Am. Ceram. Soc.* 80, 1193 (1997).
60. F. H. Streitz and J. W. Mintmire, *Phys. Rev. B* 60, 773 (1999).
61. G. Gutierrez, A. Taga, and B. Johansson, *Phys. Rev. B* 65, 012101 (2001).
62. J. H. de Boer and G. M. M. Houben, in "Proceedings of the International Symposium on the Reactivity of Solids," Vol. I, p. 237, 1952.
63. V. A. Ushakov and E. M. Moroz, *React. Kinet. Catal. Lett.* 24, 113 (1984).
64. R.-S. Zhou and R. L. Snyder, *Acta Cryst.* B47, 617 (1991).
65. C. S. John, N. C. M. Alma, and G. R. Hays, *Applied Catalysis* 6, 341 (1981).
66. L. J. Alvarez, J. F. Sanz, M. J. Capitan, M. A. Centeno, and J. A. Odriozola, *J. Chem. Soc., Faraday Trans.* 89, 3623 (1993).
67. L. J. Alvarez, L. E. Leon, J. F. Sanz, M. J. Capitan, and J. A. Odriozola, *Phys. Rev. B* 50, 2561 (1994).
68. L. J. Alvarez, A. L. Blumenfeld, and J. J. Fripiat, *J. Chem. Phys.* 108, 1724 (1998).
69. K. Sohlberg, S. J. Pennycook, and S. T. Pantelides, *J. Am. Chem. Soc.* 123, 26 (2001).
70. K. Sohlberg, S. J. Pennycook, and S. T. Pantelides, *J. Am. Chem. Soc.* 121, 10999 (1999).
71. M. B. Fleisher, L. O. Golender, and M. V. Shimanskaya, *J. Chem. Soc. Faraday Trans.* 87, 745 (1991).
72. H. Tachikawa and T. Tsuchida, *J. Molec. Catal. A* 96, 277 (1995).

73. H. Kawakami and S. Yoshida, *J. Chem. Soc., Faraday Trans.* 81, 1129 (1985).
74. S. Cai and K. Sohlberg, *J. Molec. Catal. A* 193, 157 (2003).
75. M. B. Fleisher, L. O. Golender, and M. V. Shimanskqya, *React. Kinet. Catal. Lett.* 46, 173 (1992).
76. Y. Yourdshahyan, C. Ruberto, M. Halvarsson, L. Bengtsson, V. Langer, S. Ruppi, U. Rolander, and B. I. Lundqvist, *J. Am. Ceram. Soc.* 82, 1365 (1999).
77. C. Lamonier, G. Wrobel, and J. P. Bonnelle, *J. Mater. Chem.* 4, 1927 (1994).
78. M. Ozawa and C.-K. Loong, *Catalysis Today* 50, 329 (1999).
79. K. Sohlberg, S. T. Pantelides, and S. J. Pennycook, *J. Am. Chem. Soc.* 123, 6609 (2001).
80. A. Badri, C. Binet, and J.-C. Lavalley, *J. Chem. Soc., Faraday Trans.* 92, 4669 (1996).
81. A. Christensen and E. A. Carter, *Phys. Rev. B* 58, 8050 (1998).
82. R. Shah, M. C. Payne, M.-H. Lee, and J. D. Gale, *Science* 271, 1395 (1996).
83. K. Reuter and M. Scheffler, *Phys. Rev. B* 65, 035406 (2001).
84. K. Reuter and M. Scheffler, *Phys. Rev. B* 68, 045407 (2003).
85. K. Reuter and M. Scheffler, *Phys. Rev. Lett.* 90, 046103 (2003).
86. K., Sohlberg, S. J. Pennycook, and S. T. Pantelides, *Recent Research Developments in Physical Chemistry* 4, 41 (2000).
87. N. Lopez, F. Illas, and G. Pacchioni, *J. Phys. Chem. B* 103, 8552 (1999).
88. O. Swang, K. Faegri, and O. Gropen, *Chem. Phys. Lett.* 207, 397 (1993).
89. P. Mikusik, S. Pick, N. J. Castellani, P. Legare, and C. Demangeat, *Int. J. Quantum Chem.* 57, 887 (1996).
90. H.-J. Freund, *Surf. Sci. Lett.* 500, 271 (2002).
91. M. Shelef and G. W. Graham, *Catal. Rev.-Sci. Eng.* 36, 433 (1994).
92. M. Te, E. E. Lowenthal, and H. C. Foley, *Chem. Eng. Sci.* 49, 4851 (1994).
93. H. Knozinger, *Angew. Chem. Internat. Edit.* 7, 791 (1968).
94. F. Ernst, P. Pirouz, and A. H. Heuer, *Phil. Mag. A* 63, 259 (1991).
95. J. A. Thornton and J. Chin, *Am. Ceram. Soc. Bull.* 56, 504 (1977).
96. A. L. Dragoo and J. J. Diamond, *J. Am. Ceram. Soc.* 50, 568 (1967).
97. S. N. Rashkeev, K. Sohlberg, M. V. Glazoff, J. Novak, S. J. Pennycook, and S. T. Pantelides, *Phys. Rev. B* 67, 115414 (2003).
98. S. Soled, *J. Catalysis* 81, 252 (1983).
99. K. Sohlberg, S. T. Pantelides, and S. J. Pennycook, *MRS Soc. Symp. Proc.* 589, 241 (2001).
100. L. Giordano, G. Pacchioni, T. Bredow, and J. F. Sanz, *Surf. Sci.* 471, 21 (2001).
101. E. Wahlström, N. Lopez, R. Schaub, P. Thostrup, A. Ronnau, C. Africh, E. Laegsgaard, J. K. Norskøv, and F. Besenbacher, *Phys. Rev. Lett.* 90, 026101 (2003).
102. H. Tada, F. Suzuki, S. Ito, T. Akita, K. Tanaka, T. Kawahara, and H. Kobayashi, *J. Phys. Chem. B* 106, 8714 (2002).
103. A. Vittadini and A. Selloni, *J. Chem. Phys.* 117, 353 (2002).
104. M. Lazzeri, A. Vittadini, and A. Selloni, *Phys. Rev. B* 63, 155409 (2001).
105. S.-H. Cai and K. Sohlberg, *MRS Soc. Symp. Proc.* 731, W3.6.1 (2002).
106. M. V. Glazov, J. Novak, and L. G. Hector, Alcoa Technical Center, Pittsburgh, 1999.
107. J. M. McHale, A. Aroux, A. J. Perotta, and A. Navrotsky, *Science* 277, 788 (1997).
108. J. M. McHale, A. Navrotsky, and A. J. Perrotta, *J. Chem. Phys. B* 101, 603 (1997).
109. K.-N. P. Kumar, J. Tranto, J. Kumar, and J. E. Engell, *J. Mater. Sci. Lett.* 15, 266 (1996).
110. B. Beguin, E. Garbowski, and M. Primet, *Appl. Catal.* 75, 119 (1991).
111. A. Vazquez, T. Lopez, R. Gomez, X. Bokhimi, A. Morales, and O. Novaro, *J. Solid State Chem.* 128, 161 (1997).
112. T. Yamamoto, T. Tanaka, T. Matsuyama, T. Funabiki, and S. Yoshida, *J. Synchrotron Rad.* 8, 634 (2001).
113. S. Wang, A. Y. Borisevich, S. N. Rashkeev, M. V. Glazoff, K. Sohlberg, S. J. Pennycook, and S. T. Pantelides, (submitted).
114. S. J. Pennycook and D. E. Jesson, *Phys. Rev. Lett.* 64, 938 (1990).
115. G. Kresse and J. Hafner, *Phys. Rev. B* 48, 13115 (1993).
116. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
117. D. Vanderbilt, *Phys. Rev. B* 41, 7892 (1990).
118. P. E. Blöchl, *Phys. Rev. B* 50, 17953 (1994).
119. G. Kresse and D. Joubert, *Phys. Rev. B* 59, 1758 (1999).
120. P. Burtin, J. P. Brunelle, M. Pijolat, and M. Soustelle, *Appl. Catal.* 34, 225 (1987).
121. J. S. Church, N. W. Cant, and D. L. Trimm, *Appl. Catal. A: General* 107, 267 (1994).
122. R. V. Kasowski and R. H. Tait, *Phys. Rev. B* 20, 5168 (1979).
123. J. K. Burdett, T. Hughbanks, G. J. Miller, J. W. Richardson, and J. V. Smith, *J. Am. Chem. Soc.* 109, 3639 (1987).
124. M. Ramamoorthy, D. Vanderbilt, and R. D. King-Smith, *Phys. Rev. B* 49, 16721 (1994).
125. G. Charlton, P. B. Howes, C. L. Nicklin, P. Steadman, J. S. G. Taylor, C. A. Muryn, S. P. Harte, J. Mercer, R. McGrath, D. Norman, T. S. Turner, and G. Thornton, *Phys. Rev. Lett.* 78, 495 (1997).
126. S. H. Overbury, P. A. Bertrand, and G. A. Somorjai, *Chem. Rev.* 75, 547 (1975).
127. M. Ramamoorthy, R. D. King-Smith, and D. Vanderbilt, *Phys. Rev. B* 49, 7709 (1994).
128. P. J. D. Lindan, N. M. Harrison, M. J. Gillan, and J. A. White, *Phys. Rev. B* 55, 15919 (1997).
129. S.-D. Mo and W. Y. Ching, *Phys. Rev. B* 51, 13023 (1995).

130. R. Schaub, P. Thostrup, N. Lopez, E. Lægsgaard, I. K. Norskov, and F. Besenbacher, *Phys. Rev. Lett.* 87, 266104 (2001).
131. M. Li, W. Hebenstreit, L. Gross, U. Diebold, M. A. Henderson, D. R. Jennison, P. A. Schultz, and M. P. Sears, *Surf. Sci.* 437 (1999).
132. R. Asahi, Y. Taga, W. Mannstadt, and A. J. Freeman, *Phys. Rev. B* 61, 7459 (2000).
133. D. Vogtenhuber, R. Podloucky, A. Neckel, S. G. Steinemann, and A. J. Freeman, *Phys. Rev. B* 49, 2099 (1994).
134. J. Muscat, N. M. Harrison, and G. Thornton, *Phys. Rev. B* 59, 2320 (1999).
135. M. Lazzeri and A. Selloni, *Phys. Rev. Lett.* 87, 266105 (2001).
136. S. U. M. Khan, M. Al-Shahry, and W. B. Ingler, *Science* 297, 2243 (2002).
137. S. A. Mitchell, L. Lian, D. M. Rayner, and P. A. Hackett, *J. Chem. Phys.* 103, 5539 (1995).
138. A. Kara and A. E. DePristo, *J. Chem. Phys.* 88, 2033 (1988).
139. L. Holmgren and A. Rosén, *J. Chem. Phys.* 110, 2629 (1999).
140. L. Zhu, J. Ho, E. K. Parks, and S. J. Riley, *J. Chem. Phys.* 98, 2798 (1993).
141. P. Gelin, J.-F. Dutel, and Y. B. Taarit, *J. Chem. Soc., Chem. Comm.*, 1746 (1990).
142. R. Ramprasad, K. M. Glassford, and J. B. Adams, "Materials Research Society Symposium Proceedings: Materials and Processes for Environmental Protection," Vol. 344, 1994.
143. T. R. Ward, R. Hoffmann, and M. Shelef, *Surf. Sci.* 289, 85 (1993).
144. N. N. Greenwood and A. Earnshaw, "Chemistry of the Elements," Pergamon, Oxford, 1985.
145. K. Hermann, M. Witko, and R. Druzinic, *Faraday Discuss.* 114, 53 (1999).
146. C. Kolczewski and K. Hermann, *J. Chem. Phys.* 118, 7599 (2003).
147. X. Yin, H. Han, A. Endou, M. Kubo, K. Teraishi, A. Chatterjee, and A. Miyamoto, *J. Phys. Chem. B* 103, 1263 (1999).
148. A. A. Tsyganenko and P. P. Mardilovich, *J. Chem. Soc., Faraday Trans.* 92, 4843 (1996).
149. R. B. Shirts, *J. Phys. Chem.* 91, 2258 (1987).
150. W. Heisenberg, "The Physical Principles of the Quantum Theory," Dover, New York, 1949.
151. A. P. Scott and L. Radom, *J. Phys. Chem.* 100, 16502 (1996).
152. "CRC Handbook of Chemistry and Physics" (D. R. Lide, Ed.), CRC, Boca Raton, FL, 1992.
153. C. W. White, C. J. McHargue, P. S. Sklad, L. A. Boatner, and G. C. Farlow, *Mat. Sci. Rep.* 4, 41 (1989).
154. R. J. Hill, *Clays Clay Min.* 29, 435 (1981).

CHAPTER 13

Properties of Superconducting Nanostructures

Rosario Fazio, Fabio Taddei
NEST-INFM and Scuola Normale Superiore, Pisa, Italy

CONTENTS

1. Introduction 647
2. Hybrid Systems 648
   2.1. Andreev Reflection and Proximity Effect 648
   2.2. Derivation of the Bogoliubov–de Gennes Equation . . . 650
   2.3. Solutions of the Bogoliubov–de Gennes Equation 654
   2.4. Scattering Theory 659
   2.5. Two-Probe Differential Conductance and Conductance in More Details 665
   2.6. Quasi-Classical Green's Function Approach 668
   2.7. Panorama on Results in NS Systems 670
3. Coulomb Blockade and Superconductivity 671
   3.1. Tunneling Rates and Parity Effects 671
   3.2. NSN Transistor 674
   3.3. Cooper Pair Tunneling 675
   3.4. Superconducting Nanoparticles 677
   3.5. Josephson Arrays 678
   3.6. Quantum Computation with Superconducting Nanocircuits 681
4. Conclusions 683
   References 684

1. INTRODUCTION

Mesoscopic physics concerns the properties of small systems with sizes in the range of a few nanometers to micrometers and at low temperatures typically below 1 K. The constant progress in nanofabrication techniques allows for a controlled fabrication of these structures and a consequent increasing interest in this physics. Characteristic for superconductivity is the macroscopic phase coherence of the order parameter and the supercurrent flow. On the

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (647–687)

one hand, superconductivity adds new degrees of freedom and makes the description of mesoscopic system richer; this is the case of hybrid (normal/superconducting) systems. On the other hand, superconducting properties are deeply influenced by mesoscopic effects, as it is the case of charging effects in small superconducting junctions. Besides the interesting effects associated with the superconducting coherence in nanodevices, a very appealing direction of research with superconducting nanocircuits has been the study of macroscopic quantum dynamics of small Josephson junctions. This area of research has been fertile and very recently had an astonishing impact also in solid-state quantum information, where they are believed to be among the best candidates for the implementation of a quantum computer. The study of superconducting nanodevices has been proven to have important connections also in the area of statistical mechanics and phase transitions. We refer to Josephson arrays that have been studied for the past two decades to observe, in artificially controlled structures, a variety of collective properties and quantum phase transitions.

This brief introduction makes evident how the field of superconductivity in nanostructures is so vast and diversified that each topic would require a separate review to be adequately introduced. It is our intention to give a general overview of the field pointing to the relevant references for a more detailed analysis. The description is kept as simple as possible. This chapter is divided in three sections (including this "Introduction"), which are almost independent among themselves, and a Conclusions section. We begin, in Section 2, by introducing the field of hybrid systems, which are mesoscopic devices constituted by normal and superconducting parts in electrical contact. The electrical transport through the system is dominated by the conversion of normal current into supercurrent, which takes place at the interface. We give, for educational purposes, a rather detailed introduction to the technique of Bogoliubov–de Gennes equations combined to the scattering approach for the study of transport in hybrid systems. The complementary approach using quasiclassical Green functions is then briefly introduced, however we refer to the existing literature for a more detailed presentation. In Section 3, we begin with the discussion on the superconducting properties in small systems where charging effects are dominant. The study of charging effects in superconductors has important ramifications for the subsequent sections where we describe superconducting properties in ultrasmall grains (Section 3.4), the properties of Josephson arrays (Section 3.5) and, finally, the role of Josephson junctions in quantum computation (Section 3.6), which is surely one of the most important breakthroughs in the field of superconducting nanostructures.

Basics in superconductivity and Josephson physics can be found in the books by Tinkham [1] and by Barone and Paternò [2]. Because many ideas discussed here belong to the field of more traditional mesoscopic physics, we refer for these topics to the books by Beenakker and van Houten [3] and Imry [4] and to the conference proceedings [5, 6]. Excellent introductions to single charge tunneling can be found in Refs. [7–9]. We also refer to the recent reviews on hybrid structures [10], ultrasmall superconductors [11], solid state quantum computation [12] and Josephson arrays [13] for a more complete description of these topics and a more complete list of references.

2. HYBRID SYSTEMS

2.1. Andreev Reflection and Proximity Effect

A fascinating aspect of mesoscopics is the possibility of fabricating hybrid nanostructures formed from combinations of normal conductors (N) and superconductors (S). The interest for these systems stems from the fact that the nature of charge transport is dramatically different in the normal and in the superconducting constituents of the heterostructure, giving rise to a rich variety of effects. While the charge current in N is carried by quasi-particle excitations (electrons in a metal), at sufficiently low energy (e.g., temperature, applied voltage, frequency, and so on) the current in S is carried by the superconducting condensate, which is a many-body ground state property of the system, and flows without dissipation.

The matching between these two different charge transport mechanisms can be well understood for phase-coherent structures in terms of the so-called *Andreev reflection* process [14].

This can be introduced starting from the microscopic Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity [15]. In the presence of boundaries and nonuniformities, it is convenient to perform an approximate Fourier transform of the BCS Hamiltonian. Making use of the Hartree approximation, one arrives to an effective single-particle Hamiltonian that leads to the Bogoliubov–de Gennes equation [16]

$$\begin{pmatrix} H_0 & \Delta \\ \Delta^* & -H_0^* \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = E \begin{pmatrix} u \\ v \end{pmatrix} \tag{1}$$

for the coherence factors u and v, where the energy E is measured from the condensate chemical potential. This is a matrix eigenvalue equation that contains the single-particle Hamiltonian H_0 describing electrons in the superconductor in the absence of the attractive potential, and the time-reversal of such a Hamiltonian $(-H_0^*)$, which describes the hole degree of freedom. The off-diagonal term Δ, which introduces a coupling between electrons and holes, is given by the product of the point-like electron–electron attractive potential and the pairing amplitude describing superconducting correlations among electrons. The pairing amplitude, proportional to the anomalous average $\langle \psi\psi \rangle$, can be understood as being the wave function of the Cooper pairs. Because Δ, which is the superconducting order parameter, is determined by the coherence factors, it should be calculated self-consistently. In the absence of superconductivity ($\Delta = 0$), the system is described in terms of decoupled electrons and holes, the latter being characterized by having group velocity and wave-vector in opposite directions, as well as opposite charge and spin with respect to electrons. When an electronic attractive interaction is present ($\Delta \neq 0$), electron and hole degrees of freedom get mixed, forming particle-like and hole-like quasi-particles. As a result, a gap of amplitude $|\Delta|$ opens up in the energy spectrum (see Fig. 6) forbidding quasi-particles excitations with energy inside the gap to propagate in the superconductor.

Now consider a piece of normal metal in contact with a superconductor and assume that electrons of a given energy $E < |\Delta|$ are injected from the normal part toward the NS interface. As explained above, such electrons cannot propagate through the superconductor, in fact they are allowed to penetrate as an evanescent wave up to a depth of the order of the BCS superconducting coherence length ξ. They can, however, undergo two different processes. In the first one, *normal reflection*, electrons are reflected back into the normal slab and do not contribute to the charge current. The second one is made possible by the electron–hole coupling term of the Bogoliubov–de Gennes Hamiltonian and consists in the coherent evolution of the incoming electron into a retro-reflected hole (Fig. 1). This constitutes the *Andreev reflection* process, which is indeed responsible for charge transport and corresponds to the transfer of a pair of electrons from the normal side of the interface to the superconducting condensate. Noteworthy, each Andreev reflection process contributes

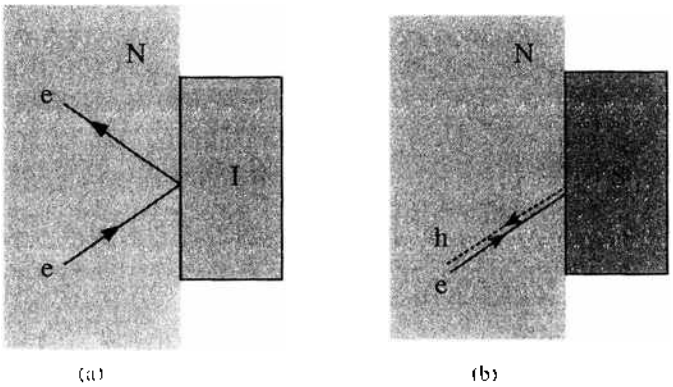


Figure 1. (a) In a normal reflection process, which occurs for example at an interface with an insulator (I), an electron is scattered back into the normal metal, conserving the component of the momentum parallel to the interface. (b) In an Andreev reflection, occurring at a NS interface, the incoming electron is reflected back as a hole. In this case, the component of the momentum parallel to the interface is reversed, so that, at the Fermi energy, the reflected hole retrace back the trajectory of the impinging electron.

$2e$ to the charge current. It is important to notice that Andreev reflection does not contribute to the transport of energy, as no excitations are transferred into the superconductor. While providing good charge conductivity, NS interfaces prove to be good thermal insulators.

Andreev reflection also accounts for the influence that the superconductor exerts on the normal conductor, known as *proximity effect*. In much the same way as the quasi-particle excitations penetrate the superconductor, the superconductor pairing amplitude also leaks out into the normal side of NS contact over a large distance, of the order the superconducting coherence length, even in the absence of an attractive interaction. As a result, partial superconducting properties are induced into the normal conductor.

In the next section 2.2, we will give a derivation (following Ref. [16]) of the Bogoliubov–de Gennes equation, and in Section 2.3 we shall solve such an equation for different systems. In the following sections, we shall describe the two most widely used theoretical approaches applied to study hybrid systems, namely the scattering formalism (Section 2.4), which we will analyze in detail, and the quasi-classical Green's function theory (Section 2.6). In Section 2.7, we shall review the most important results attained in the field.

2.2. Derivation of the Bogoliubov–de Gennes Equation

As mentioned above, the Bogoliubov–de Gennes equation is derived from the BCS theory. In this section, we show how to derive such an equation in the general case where an exchange field is present. This will be useful for studying hybrid structures in which ferromagnets are present, too. It is convenient to make use of the Bogoliubov's self-consistent field method [16, 17]. The starting point is the following Hamiltonian in the second quantization form:

$$\hat{H} = \int d\vec{r}(\hat{\psi}_\uparrow^\dagger(\vec{r}), \hat{\psi}_\downarrow^\dagger(\vec{r}))\mathcal{H}(\vec{r})\begin{pmatrix}\hat{\psi}_\uparrow(\vec{r})\\ \hat{\psi}_\downarrow(\vec{r})\end{pmatrix} - \frac{V}{2}\sum_{\sigma\sigma'}\int d\vec{r}\hat{\psi}_\sigma^\dagger(\vec{r})\hat{\psi}_{\sigma'}^\dagger(\vec{r})\hat{\psi}_{\sigma'}(\vec{r})\hat{\psi}_\sigma(\vec{r}) \qquad (2)$$

where the second term accounts for the electron–electron coupling, which is assumed to be a point-like, two-particle interaction (introduced for the first time by Gorkov [18]) and independent of spin, therefore characterized by a single coefficient V. In Eq. (2),

$$\mathcal{H}(\vec{r}) = \left[\frac{1}{2m}(-i\hbar\vec{\nabla} - \frac{e}{c}\vec{A}(\vec{r}))^2 + V_0(\vec{r}) - \mu\right]\underline{\underline{1}} - \frac{he}{2mc}(\underline{\underline{\vec{\sigma}}}\cdot\vec{M}) \qquad (3)$$

is the single-particle Hamiltonian in the presence of an exchange field (second term). $\vec{A}(\vec{r})$ is the vector potential, $V_0(\vec{r})$ is the normal potential, μ is the chemical potential, $\vec{M}$ is the internal mean field due to the exchange interaction, and $U_0(\vec{r})$ is the periodic potential due to the ions in the crystal, c is the speed of light, $\underline{\underline{1}}$ is the 2×2 unity matrix, and $\underline{\underline{\vec{\sigma}}}$ is the vector of Pauli matrices. Before proceeding, we want to remark that ferromagnetism is introduced according to the Stoner model (see Refs. [19–22]). In this model a molecular field, due to the permanent magnetization present in a ferromagnet, replaces the external magnetic field in the Zeeman energy, and it is added to the single-particle energy. Such a molecular field can be obtained calculating the exchange interaction in a system of spins using the mean field approximation (see Ref. [23]). Such a molecular field can be treated as an adjustable parameter to fit band structure calculations.

The operators $\hat{\psi}_\sigma(\vec{r})$ and $\hat{\psi}_\sigma^\dagger(\vec{r})$ are, respectively, destruction and creation field operators for a particle of spin σ at the point $\vec{r}$, which obey the Fermi commutation relations

$$\hat{\psi}_\sigma(\vec{r})\hat{\psi}_{\sigma'}(\vec{r}') + \hat{\psi}_{\sigma'}(\vec{r}')\hat{\psi}_\sigma(\vec{r}) = 0 \qquad (4)$$

$$\hat{\psi}_\sigma^\dagger(\vec{r})\hat{\psi}_{\sigma'}(\vec{r}') + \hat{\psi}_{\sigma'}(\vec{r}')\hat{\psi}_\sigma^\dagger(\vec{r}) = \delta_{\sigma\sigma'}\delta(\vec{r}-\vec{r}') \qquad (5)$$

We now seek an approximate solution of the many body problem in terms of an effective single-particle Hamiltonian $\hat{H}_{eff}$. According to the Hartree approximation (see Ref. [24]), such a Hamiltonian describes a particle that moves in a mean-field potential produced by all

other particles. Using the variational principle, the best such a Hamiltonian is the one that minimizes the total free energy F. We can try the following guess:

$$\hat{H}_{eff} = \int d\vec{r}(\hat{\psi}_\uparrow^\dagger(\vec{r})\hat{\psi}_\downarrow^\dagger(\vec{r}))\left[\mathcal{H}(\vec{r}) + \begin{pmatrix} \mathcal{H}^\uparrow(\vec{r}) & 0 \\ 0 & \mathcal{H}^\downarrow(\vec{r}) \end{pmatrix}\right]\begin{pmatrix} \hat{\psi}_\uparrow(\vec{r}) \\ \hat{\psi}_\downarrow(\vec{r}) \end{pmatrix}$$
$$+ \int d\vec{r}[\hat{\psi}_\uparrow^\dagger(\vec{r})\hat{\psi}_\downarrow^\dagger(\vec{r})\Delta(\vec{r}) + \hat{\psi}_\downarrow(\vec{r})\hat{\psi}_\uparrow(\vec{r})\Delta^*(\vec{r})] \tag{6}$$

where $\mathcal{H}^\uparrow(\vec{r})$, $\mathcal{H}^\downarrow(\vec{r})$ and $\Delta(\vec{r})$ are effective potentials to be determined in the following. By definition

$$F = \langle \hat{H} \rangle - TS \tag{7}$$

where $\hat{H}$ [Eq. (2)] is the initial Hamiltonian and the angle brackets denote thermal average given by:

$$\langle \hat{H} \rangle = \frac{\sum_\phi \langle \phi | \hat{H} | \phi \rangle e^{-\beta E_\phi}}{\sum_\phi e^{-\beta E_\phi}} \tag{8}$$

T is the temperature and S is the entropy. The matrix elements are taken with respect to eigenstates of $\hat{H}_{eff}$

$$\hat{H}_{eff}|\phi\rangle = E_\phi|\phi\rangle \tag{9}$$

A general method to calculate $\langle \hat{H} \rangle$ is to replace the fields $\hat{\psi}$'s with the fields $\hat{\gamma}$'s, which diagonalize $\hat{H}_{eff}$ through an unitary transformation

$$\begin{aligned} \hat{\psi}_\uparrow &= \sum_n (u_n^\uparrow(\vec{r})\hat{\gamma}_{n\uparrow} - v_n^{\uparrow *}(\vec{r})\hat{\gamma}_{n\downarrow}^\dagger) \\ \hat{\psi}_\downarrow &= \sum_n (u_n^\downarrow(\vec{r})\hat{\gamma}_{n\downarrow} + v_n^{\downarrow *}(\vec{r})\hat{\gamma}_{n\uparrow}^\dagger) \end{aligned} \tag{10}$$

then to use the mean value rules

$$\langle \hat{\gamma}_{n\sigma}^\dagger \hat{\gamma}_{m\sigma'} \rangle = \delta_{nm}\delta_{\sigma\sigma'} f_n \tag{11}$$

$$\langle \hat{\gamma}_{n\sigma} \hat{\gamma}_{m\sigma'} \rangle = 0 \tag{12}$$

where

$$f_n = \frac{1}{e^{\beta E_n} + 1} \tag{13}$$

However, it will not be necessary to perform this entire calculation. We write $\langle \hat{H} \rangle$ in the form

$$\langle \hat{H} \rangle = \int d\vec{r}\left\langle (\hat{\psi}_\uparrow^\dagger(\vec{r}), \hat{\psi}_\downarrow^\dagger(\vec{r}))\mathcal{H}(\vec{r})\begin{pmatrix} \hat{\psi}_\uparrow(\vec{r}) \\ \hat{\psi}_\downarrow(\vec{r}) \end{pmatrix}\right\rangle - \frac{V}{2}\sum_{\sigma\sigma'}\int d\vec{r}\langle \hat{\psi}_\sigma^\dagger(\vec{r})\hat{\psi}_{\sigma'}^\dagger(\vec{r})\hat{\psi}_{\sigma'}(\vec{r})\hat{\psi}_\sigma(\vec{r})\rangle \tag{14}$$

The product $\langle \hat{\psi}^\dagger(1)\hat{\psi}^\dagger(2)\hat{\psi}(3)\hat{\psi}(4)\rangle$ can be simplified thanks to Wick's theorem, which make use only of the fact that $\hat{\psi}^\dagger$ and $\hat{\psi}$ are linear functions of $\hat{\gamma}^\dagger$ and $\hat{\gamma}$. The theorem gives

$$\langle \hat{\psi}^\dagger(1)\hat{\psi}^\dagger(2)\hat{\psi}(3)\hat{\psi}(4)\rangle = \langle \hat{\psi}^\dagger(1)\hat{\psi}(4)\rangle\langle \hat{\psi}^\dagger(2)\hat{\psi}(3)\rangle - \langle \hat{\psi}^\dagger(1)\hat{\psi}(3)\rangle\langle \hat{\psi}^\dagger(2)\hat{\psi}(4)\rangle$$
$$+ \langle \hat{\psi}^\dagger(1)\hat{\psi}^\dagger(2)\rangle\langle \hat{\psi}(3)\hat{\psi}(4)\rangle \tag{15}$$

We now vary the amplitudes $(u^{\uparrow}, v^{\downarrow}, u^{\downarrow}, v^{\uparrow})$ and the occupation number f_n. The free energy (7) then varies by δF:

$$\delta F = \int d\vec{r}\,\delta\left\langle (\hat{\psi}_{\uparrow}^{\dagger}(\vec{r}), \hat{\psi}_{\downarrow}^{\dagger}(\vec{r}))\mathcal{H}(\vec{r})\begin{pmatrix}\hat{\psi}_{\uparrow}(\vec{r})\\ \hat{\psi}_{\downarrow}(\vec{r})\end{pmatrix}\right\rangle - V\sum_{\sigma\sigma'}\int d\vec{r}\langle\hat{\psi}_{\sigma}^{\dagger}(\vec{r})\hat{\psi}_{\sigma}(\vec{r})\rangle\delta\langle\hat{\psi}_{\sigma'}^{\dagger}(\vec{r})\hat{\psi}_{\sigma'}(\vec{r})\rangle$$
$$+ V\sum_{\sigma}\int d\vec{r}\langle\hat{\psi}_{\sigma}^{\dagger}(\vec{r})\hat{\psi}_{\sigma}(\vec{r})\rangle\delta\langle\hat{\psi}_{\sigma}^{\dagger}(\vec{r})\hat{\psi}_{\sigma}(\vec{r})\rangle$$
$$- V\int d\vec{r}[\langle\hat{\psi}_{\uparrow}^{\dagger}(\vec{r})\hat{\psi}_{\downarrow}^{\dagger}(\vec{r})\rangle\delta\langle\hat{\psi}_{\downarrow}(\vec{r})\hat{\psi}_{\uparrow}(\vec{r})\rangle + \text{c.c.}] - T\delta S \tag{16}$$

where we have assumed $\langle\hat{\psi}_{\uparrow}^{\dagger}(\vec{r})\hat{\psi}_{\downarrow}(\vec{r})\rangle = 0$, as no magnetic exchange interaction was left in the interacting part of (2). We also assumed $\langle\hat{\psi}_{\sigma}^{\dagger}(\vec{r})\hat{\psi}_{\sigma}^{\dagger}(\vec{r})\rangle = 0$, as we are considering spin-singlet pairing. Now, notice that the quantity

$$F' = \langle\hat{H}_{eff}\rangle - TS \tag{17}$$

is stationary with respect to $(\delta u^{\uparrow}, \delta v^{\downarrow}, \delta u^{\downarrow}, \delta v^{\uparrow})$ and δf_n, since our excitations diagonalize $\hat{H}_{eff}$ exactly. Using (6), this condition becomes explicitly

$$0 = \delta\langle\hat{H}_{eff}\rangle - T\delta S = \int d\vec{r}\,\delta\left\langle(\hat{\psi}_{\uparrow}^{\dagger}(\vec{r})\hat{\psi}_{\downarrow}^{\dagger}(\vec{r}))\left[\mathcal{H}(\vec{r}) + \begin{pmatrix}u^{\uparrow}(\vec{r}) & 0\\ 0 & u^{\downarrow}(\vec{r})\end{pmatrix}\right]\begin{pmatrix}\hat{\psi}_{\uparrow}(\vec{r})\\ \hat{\psi}_{\downarrow}(\vec{r})\end{pmatrix}\right\rangle$$
$$+ \int d\vec{r}[\Delta(\vec{r})\delta\langle\hat{\psi}_{\uparrow}^{\dagger}(\vec{r})\hat{\psi}_{\downarrow}^{\dagger}(\vec{r})\rangle + \text{c.c.}] - T\delta S \tag{18}$$

By comparing (16) and (18), it is easy to see that F will be stationary if we take as effective potentials

$$u^{\uparrow}(\vec{r}) = -V\langle\hat{\psi}_{\downarrow}^{\dagger}(\vec{r})\hat{\psi}_{\downarrow}(\vec{r})\rangle \tag{19}$$

$$u^{\downarrow}(\vec{r}) = -V\langle\hat{\psi}_{\uparrow}^{\dagger}(\vec{r})\hat{\psi}_{\uparrow}(\vec{r})\rangle \tag{20}$$

and

$$\Delta(\vec{r}) = -V\langle\hat{\psi}_{\downarrow}(\vec{r})\hat{\psi}_{\uparrow}(\vec{r})\rangle = V\langle\hat{\psi}_{\uparrow}(\vec{r})\hat{\psi}_{\downarrow}(\vec{r})\rangle \tag{21}$$

$u^{\uparrow}(\vec{r})$ and $u^{\downarrow}(\vec{r})$ are single particle potentials (standard Hartree result for point-like interactions), and $\Delta(\vec{r})$ is referred to as the pairing potential.

The key feature of the mean field theory of superconductivity is that we admit non-vanishing expectation values for the spin-paired operators $\hat{\psi}_{\uparrow}(\vec{r})\hat{\psi}_{\downarrow}(\vec{r})$ and $\hat{\psi}_{\downarrow}(\vec{r})\hat{\psi}_{\uparrow}(\vec{r})$. If we consider $\hat{H}_{eff}$ in (6), we see that, on the one hand, the terms in $u^{\uparrow}$ and $u^{\downarrow}$ destroy and create one electron and therefore conserve the number of particles. On the other hand, the terms in Δ increases or decreases the number of particles by 2. This is not a problem, as $\hat{H}_{eff}$ operates on the BCS wave function, which is not an eigenfunction of the number operator.

As we shall see, the transformation in Eq. (10) diagonalizes $\hat{H}_{eff}$ in such a way that

$$\hat{H}_{eff} = E_g + \sum_{n\sigma} E_n\hat{\gamma}_{n\sigma}^{\dagger}\hat{\gamma}_{n\sigma} \tag{22}$$

where E_g is the superconducting ground state energy, E_n is the energy of the nth quasi-particle excitation, and $\hat{\gamma}_{n\sigma}$ ($\hat{\gamma}_{n\sigma}^{\dagger}$) is the quasi-particle destruction (creation) operator satisfying Fermi commutation relations:

$$\{\hat{\gamma}_{n\sigma}^{\dagger}, \hat{\gamma}_{m\sigma'}\} = \delta_{nm}\delta_{\sigma\sigma'} \tag{23}$$

$$\{\hat{\gamma}_{n\sigma}, \hat{\gamma}_{m\sigma'}\} = 0 \tag{24}$$

From equations (22), (23) and (24), we obtain

$$[\hat{H}_{eff}, \hat{\gamma}_{n\sigma}] = -E_n \hat{\gamma}_{n\sigma} \tag{25}$$

$$[\hat{H}_{eff}, \hat{\gamma}^{\dagger}_{n\sigma}] = E_n \hat{\gamma}^{\dagger}_{n\sigma} \tag{26}$$

By writing explicitly Eq. (6), setting $\vec{A} = 0$ and $M_x = M_y = 0$, we obtain $\hat{H}_{eff}$ in the following form:

$$\hat{H}_{eff} = \int d\vec{r}[\hat{\psi}^{\dagger}_{\uparrow}(\vec{r})(H_0 + u^{\uparrow}(\vec{r}) - h_z)\hat{\psi}_{\uparrow}(\vec{r}) + \hat{\psi}^{\dagger}_{\downarrow}(\vec{r})(H_0 + u^{\downarrow}(\vec{r}) + h_z)\hat{\psi}_{\downarrow}(\vec{r})]$$
$$+ \int d\vec{r}[\hat{\psi}^{\dagger}_{\uparrow}(\vec{r})\hat{\psi}^{\dagger}_{\downarrow}(\vec{r})\Delta(\vec{r}) + \hat{\psi}_{\downarrow}(\vec{r})\hat{\psi}_{\uparrow}(\vec{r})\Delta^{*}(\vec{r})]$$

where

$$H_0 = \left[-\frac{\hbar^2}{2m}\vec{\nabla}^2 + V_0(\vec{r}) - \mu\right] \tag{27}$$

$$h_z = \frac{he}{2mc}M_z \tag{28}$$

It is now possible to calculate the commutators of $\hat{\psi}_{\uparrow}(\vec{r})$, $\hat{\psi}_{\downarrow}(\vec{r})$ with $\hat{H}_{eff}$:

$$\begin{aligned}[\hat{\psi}_{\uparrow}, \hat{H}_{eff}] &= [H_0 - h_z + u^{\uparrow}(\vec{r})]\hat{\psi}_{\uparrow} + \Delta(\vec{r})\hat{\psi}^{\dagger}_{\downarrow} \\ [\hat{\psi}_{\downarrow}, \hat{H}_{eff}] &= [H_0 + h_z + u^{\downarrow}(\vec{r})]\hat{\psi}_{\downarrow} - \Delta(\vec{r})\hat{\psi}^{\dagger}_{\uparrow}\end{aligned} \tag{29}$$

Substituting the unitary transformation of Eq. (10), between $\hat{\psi}$'s and $\hat{\gamma}$'s operators, in Eq. (29) and assuming that the operators $\hat{\gamma}_{n\sigma}$ are linearly independent with respect to the index n, we obtain two equations

$$\begin{pmatrix} \overline{H}^{\uparrow} - h_z & \Delta(\vec{r}) \\ \Delta^{*}(\vec{r}) & -\overline{H}^{\downarrow} - h_z \end{pmatrix} \begin{pmatrix} u^{\uparrow}_n(\vec{r}) \\ v^{\downarrow}_n(\vec{r}) \end{pmatrix} = E_n \begin{pmatrix} u^{\uparrow}_n(\vec{r}) \\ v^{\downarrow}_n(\vec{r}) \end{pmatrix} \tag{30}$$

and

$$\begin{pmatrix} \overline{H}^{\downarrow} + h_z & \Delta(\vec{r}) \\ \Delta^{*}(\vec{r}) & -\overline{H}^{\uparrow} + h_z \end{pmatrix} \begin{pmatrix} u^{\downarrow}_n(\vec{r}) \\ v^{\uparrow}_n(\vec{r}) \end{pmatrix} = E_n \begin{pmatrix} u^{\downarrow}_n(\vec{r}) \\ v^{\uparrow}_n(\vec{r}) \end{pmatrix} \tag{31}$$

where $\overline{H}^{\sigma}(\vec{r}) = H_0 + u^{\sigma}(\vec{r})$. They are the Bogoliubov–de Gennes equations for, respectively, spin-up electrons and spin-down holes [vector $(u^{\uparrow}_n(\vec{r}), v^{\downarrow}_n(\vec{r}))$] and spin-down electrons and spin-up holes [vector $(u^{\downarrow}_n(\vec{r}), v^{\uparrow}_n(\vec{r}))$]. Note that if $M_z = 0$, then $u^{\uparrow} = u^{\downarrow}$ and the two Eqs. (30) and (31) are equal. The Bogoliubov–de Gennes equations are eigenvalue equations that allow one to calculate the energy spectrum of the quasi-particles and the coherence factors $u^{\sigma}_n(\vec{r})$ and $v^{\sigma}_n(\vec{r})$. It is interesting to note that if $\begin{pmatrix} u^{\uparrow}_n \\ v^{\downarrow}_n \end{pmatrix}$ is the solution to Eq. (30) with eigenvalue E_n, then $\begin{pmatrix} -v^{\downarrow *}_n \\ u^{\uparrow *}_n \end{pmatrix}$ is the solution to Eq. (31) with eigenvalue $-E_n$. We can put together Eqs. (30) and (31) in order to group the particle degrees of freedom in the top-left of the Hamiltonian and the hole degrees of freedom in the bottom-right:

$$\begin{pmatrix} (\overline{H}^{\uparrow} - h_z) & 0 & 0 & \Delta(\vec{r}) \\ 0 & (\overline{H}^{\downarrow} + h_z) & \Delta(\vec{r}) & 0 \\ 0 & \Delta^{*}(\vec{r}) & (-\overline{H}^{\uparrow} + h_z) & 0 \\ \Delta^{*}(\vec{r}) & 0 & 0 & (-\overline{H}^{\downarrow} - h_z) \end{pmatrix} \cdot \begin{pmatrix} u^{\uparrow}_n(\vec{r}) \\ u^{\downarrow}_n(\vec{r}) \\ v^{\uparrow}_n(\vec{r}) \\ v^{\downarrow}_n(\vec{r}) \end{pmatrix} = E_n \begin{pmatrix} u^{\uparrow}_n(\vec{r}) \\ u^{\downarrow}_n(\vec{r}) \\ v^{\uparrow}_n(\vec{r}) \\ v^{\downarrow}_n(\vec{r}) \end{pmatrix} \tag{32}$$

We shall refer to the matrix Hamiltonian in (32) as the Bogoliubov–de Gennes Hamiltonian H_{BG}. If we identify the top-left 2×2 block of such a Hamiltonian with the particle

Hamiltonian H_p and the bottom-right 2×2 block with the hole Hamiltonian H_h, we see that the equality $H_h = -H_p^*$ holds. Note that, by using the transformation (10), one can prove that

$$\hat{H}_{eff} = E_g + \sum_{n\sigma} E_n \hat{\gamma}_{n\sigma}^{\dagger} \hat{\gamma}_{n\sigma} \tag{33}$$

So far, we have diagonalized $\hat{H}_{eff}$ using the transformation (10), which was proposed independently by Bogoliubov [25] and Valatin [26], known as Bogoliubov–Valatin transformation. $\hat{H}_{eff}$ is now written in terms of quasi-particle operators $\hat{\gamma}$, $\hat{\gamma}^{\dagger}$, which greatly simplifies calculations. Although the Bogoliubov–de Gennes equation (32) does not determine the potentials $\mathcal{U}^{\uparrow}(\vec{r})$, $\mathcal{U}^{\downarrow}(\vec{r})$ and $\Delta(\vec{r})$, they can be fixed self-consistently. By replacing the field operators in the definition of the above potentials (19), (20), (21) and using the Bogoliubov–Valatin transformation (10) we obtain

$$\mathcal{U}^{\uparrow}(\vec{r}) = -V \sum_n [|u_n^{\downarrow}(\vec{r})|^2 f_n + |v_n^{\uparrow}(\vec{r})|^2 (1 - f_n)] \tag{34}$$

$$\mathcal{U}^{\downarrow}(\vec{r}) = -V \sum_n [|u_n^{\uparrow}(\vec{r})|^2 f_n + |v_n^{\downarrow}(\vec{r})|^2 (1 - f_n)] \tag{35}$$

and

$$\Delta(\vec{r}) = V \sum_n [u_n^{\downarrow} v_n^{\uparrow *} (1 - f_n) - v_n^{\uparrow *} u_n^{\downarrow} f_n] \tag{36}$$

where we have used the mean-value rules (11) and (12). Equations (34), (35), and (36) constitute the self-consistent equations for $\mathcal{U}^{\uparrow}(\vec{r})$, $\mathcal{U}^{\downarrow}(\vec{r})$ and $\Delta(\vec{r})$. These ensure that if we solve the Bogoliubov–de Gennes equation (32), then the value of $\mathcal{U}^{\uparrow}$, $\mathcal{U}^{\downarrow}$ and Δ, which is calculated from the solutions of (34), (35), and (36), is equal to the initial value of $\mathcal{U}^{\uparrow}$, $\mathcal{U}^{\downarrow}$ and Δ.

There is an important distinction between $\mathcal{U}^{\uparrow}(\vec{r})$ and $\mathcal{U}^{\downarrow}(\vec{r})$, on one side, and $\Delta(\vec{r})$ on the other. The Hartree potentials $\mathcal{U}^{\uparrow}(\vec{r})$ and $\mathcal{U}^{\downarrow}(\vec{r})$ come from a sum involving all states below the Fermi level and hence are nearly temperature independent and can be approximated by the Hartree potentials calculated in the normal state. However, the pair potential $\Delta(\vec{r})$ is a sum of terms of the form $u_n^{\sigma}(\vec{r}) v_n^{\sigma' *}(\vec{r})$ which are nonzero only in the neighborhood of the Fermi surface. For this reason, $\Delta(\vec{r})$ is a strong function of temperature.

2.3. Solutions of the Bogoliubov–de Gennes Equation

In this section, we discuss the solutions of the Bogoliubov–de Gennes equation (32) in heterostructures containing normal metals, ferromagnets, and superconductors. In the following we provide, first, the solution in the case of a homogeneous and clean ferromagnet (where $\Delta = 0$), second, the solution in the case of a homogeneous and clean superconductor (where $\vec{h} = 0$), and finally a solution for a ferromagnet/superconductor interface.

2.3.1. Ferromagnet

In the limit when all potentials are zero apart from $\vec{h}$, the Bogoliubov–de Gennes equation reduces to the Schrödinger equation for a Stoner ferromagnet. Because we are considering collinear ferromagnets and we are ignoring the effect of spin-orbit scattering, the problem can be decoupled in two equivalent Schrödinger equations, the first one relative to spin ↑ and ↓ particles:

$$\begin{pmatrix} H - h_z & h_{xy} \\ h_{xy}^* & H + h_z \end{pmatrix} \begin{pmatrix} u^{\uparrow}(\vec{r}) \\ u^{\downarrow}(\vec{r}) \end{pmatrix} = E \begin{pmatrix} u^{\uparrow}(\vec{r}) \\ u^{\downarrow}(\vec{r}) \end{pmatrix} \tag{37}$$

and the second one relative to spin ↓ and ↑ holes:

$$\begin{pmatrix} -H - h_z & -h_{xy}^* \\ -h_{xy} & -H + h_z \end{pmatrix} \begin{pmatrix} v^{\downarrow}(\vec{r}) \\ v^{\uparrow}(\vec{r}) \end{pmatrix} = E \begin{pmatrix} v^{\downarrow}(\vec{r}) \\ v^{\uparrow}(\vec{r}) \end{pmatrix} \tag{38}$$

One can note that if $(u^{\uparrow}, u^{\downarrow})$ is an eigenvector for Eq. (37) relative to eigenvalue E, then the vector $(v^{\downarrow}, v^{\uparrow}) = (-u^{\uparrow}, -u^{\downarrow})$ is eigenvector for Eq. (38) relative to the eigenvalue $-E$. Assuming plane wave solutions of the form

$$\begin{pmatrix} u^{\uparrow}(\vec{r}) \\ u^{\downarrow}(\vec{r}) \end{pmatrix} = \begin{pmatrix} \psi^{\uparrow} \\ \phi^{\downarrow} \end{pmatrix} e^{i\vec{k}\cdot\vec{r}} \tag{39}$$

Eq. (37) becomes:

$$\begin{pmatrix} (k^2 - \mu - h_z - E) & h_{xy} \\ h_{xy}^* & (k^2 - \mu + h_z - E) \end{pmatrix} \begin{pmatrix} \psi^{\uparrow} \\ \phi^{\downarrow} \end{pmatrix} = 0 \tag{40}$$

which admits nontrivial solutions if the determinant of the matrix is zero. Note that we have assumed $H = k^2 - \mu$, where $k^2 \equiv |\vec{k}|^2$, and we have set $\hbar = 1$ and $2m = 1$. One thus finds that the eigenvalues of (37) satisfy

$$E = (k^2 - \mu) \pm |\vec{h}| \tag{41}$$

The eigenvectors of (37) can be found by writing explicitly Eq. (40) and by dividing the first equation by $\psi^{\uparrow}$ and the second by $\phi^{\downarrow}$. By defining $r = \frac{\phi^{\downarrow}}{\psi^{\uparrow}}$ one obtains

$$[(k^2 - \mu) - h_z - E] + h_{xy} r = 0 \tag{42}$$

$$h_{xy}^* \frac{1}{r} + [(k^2 - \mu) + h_z - E] = 0 \tag{43}$$

and subtracting the second from the first

$$h_{xy} r^2 - 2h_z r - h_{xy}^* = 0 \tag{44}$$

whose soution is

$$r = \frac{h_z \pm |\vec{h}|}{h_{xy}} \tag{45}$$

This makes clear that the eigenvalue, unlike the eigenvectors, do not depend on the direction of $\vec{h}$, but only on its magnitude $|\vec{h}|$. From the dispersion curve, plotted in Fig. 2 for the z component of $\vec{k}$, we can see that the effect of an exchange field is to shift downward the spin $\uparrow$ particle parabola, and shift upward the spin $\downarrow$ particle parabola. In Fig. 3, we present particle and hole dispersion curves together for energies $E > E_F$.

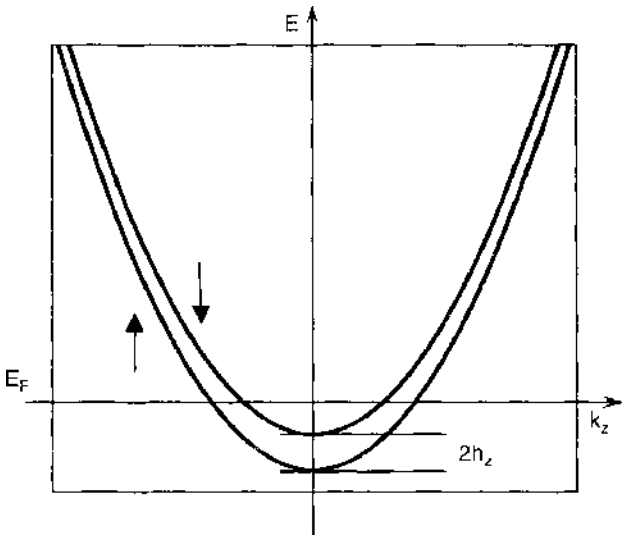


Figure 2. Dispersion curves for spin $\uparrow$ and spin $\downarrow$ electrons in a ferromagnet. Note that the presence of an exchange field has lifted the spin degeneracy by shifting the two bands vertically.

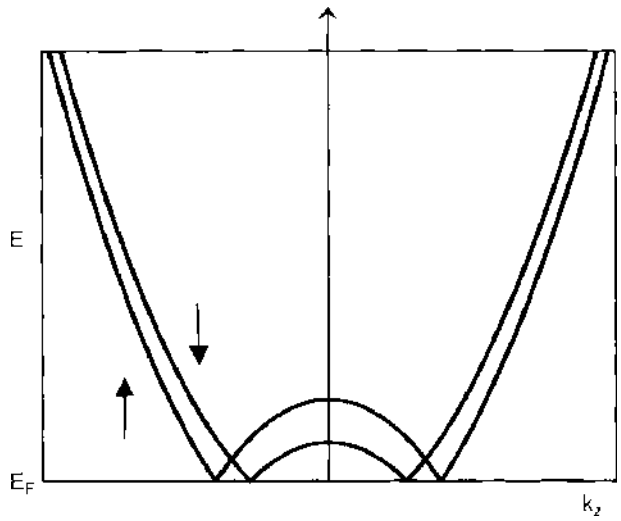


Figure 3. Dispersion curves for spin ↑ and spin ↓ electrons (upward parabolas) and for spin ↑ and spin ↓ holes (downward parabolas) in a ferromagnet.

2.3.2. Superconductor

We now consider another limiting case: all potentials are set to zero apart from $\Delta \neq 0$. The Bogoliubov–de Gennes equation (32) again decouples into two sets of equations, one relative to spin ↑ particles and spin ↓ holes

$$\begin{pmatrix} H & \Delta \\ \Delta^* & -H \end{pmatrix} \begin{pmatrix} u(\vec{r}) \\ v(\vec{r}) \end{pmatrix} = E \begin{pmatrix} u(\vec{r}) \\ v(\vec{r}) \end{pmatrix} \tag{46}$$

and another, equivalent, relative to spin ↓ particles and spin ↑ holes. Considering plane wave solution of the form

$$\begin{pmatrix} u(\vec{r}) \\ v(\vec{r}) \end{pmatrix} = \begin{pmatrix} \psi \\ \phi \end{pmatrix} e^{i\vec{k}\cdot\vec{r}} \tag{47}$$

one finds the following dispersion relation:

$$E = \pm\sqrt{(k^2 - \mu)^2 + |\Delta|^2} \tag{48}$$

which is plotted in Fig. 4 in the direction k_z when $|\Delta| = 1/3\mu$ (solid line) and $|\Delta| = 0$ (gray line). We can notice that there is an energy gap in the energy spectrum for nonzero Δ as expected. It is useful to write down the solution for k_z at a given energy E:

$$k_z = \pm\sqrt{\bar{\mu} \pm \sqrt{E^2 - |\Delta|^2}} \tag{49}$$

where $\bar{\mu} = \mu - k_x^2 - k_y^2$. This tells us that for an energy E above the gap, there are four possible solutions denoted by k_+, k_-, q_+ and q_- in Fig. 4. We shall denote excitations with wave-vectors k_α as particle-like excitations and those with wave-vectors q_α to be hole-like excitations. As we can see from Fig. 4, this choice means that particle-like excitations have their group velocity parallel to their momenta, whereas hole-like excitations have their group velocity antiparallel to their momenta. This choice is made because the wave function associated with wave-vector k_α (q_α) is predominantly particle-like (hole-like). This can be seen by substituting k_α and q_α in the amplitudes ψ and ϕ, solutions of equation (46):

$$\psi = \frac{e^{i\varphi/2}}{\sqrt{2}}\sqrt{1 + \frac{\sqrt{E^2 - |\Delta|^2}}{E}} \tag{50}$$

$$\phi = \frac{e^{i\varphi/2}}{\sqrt{2}}\sqrt{1 - \frac{\sqrt{E^2 - |\Delta|^2}}{E}} \tag{51}$$

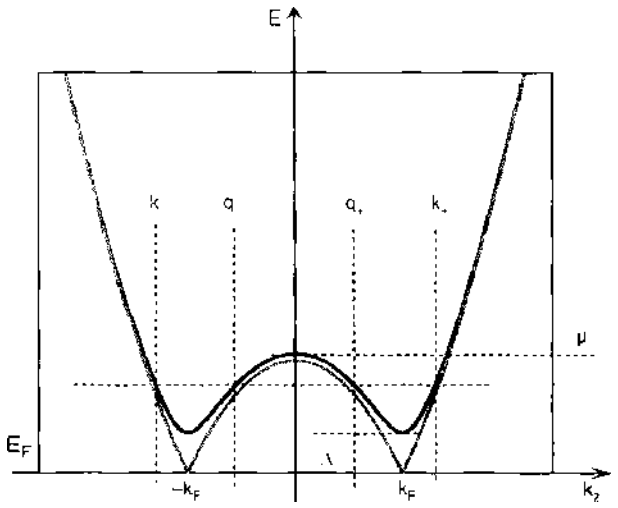


Figure 4. Dispersion curve relative to a uniform superconductor for $\Delta = 1/3\mu$ (solid line) and $\Delta = 0$ (gray line), taking $\Delta = 1$, $\lambda = 3$, and $k_+ = k_- = 2\pi/20$. For $\Delta \neq 0$, an energy gap of amplitude $|\Delta|$ appears in the spectrum.

where φ is the phase of the order parameter. Note also that as only the absolute values of the coherence factors ψ and ϕ are fixed ($|\psi|^2 + |\phi|^2 = 1$), there is an arbitrary choice in where to put the phase of the order parameter.

2.3.3. Ferromagnet/Superconductor Interface

In this section we solve analytically the Bogoliubov-de Gennes equation for a Ferromagnetic/Superconductor (F/S) interface. We shall prove that the Andreev reflection amplitude r_a is suppressed when we are in the presence of an exchange field $\vec{h}$. Before proceeding, note that, for a ballistic N/S interface, r_a was first calculated by Blonder, Tinkham, and Klapwijk (BTK) in Ref. [27]. r_a was determined as a function of the energy of the quasi-particles and as a function of the strength of a barrier potential at the interface. The generalization of the BTK calculation to the case of a F/S interface was first derived in Ref. [28].

For simplicity, let us consider a one-dimensional structure in which all potentials are set to zero apart from $\vec{h} = h\theta(-x)\hat{z}$ and $\Delta(x) = \Delta_0\theta(x)$, where h and Δ_0 are constants. In this case, the Bogoliubov–de Gennes equation (1) can be decoupled into two equivalent equations, the first of which reads

$$\begin{pmatrix} (-\partial_x^2 - \mu - h) & \Delta(x) \\ \Delta^*(x) & (\partial_x^2 + \mu - h) \end{pmatrix} \begin{pmatrix} u^\uparrow(x) \\ v^\downarrow(x) \end{pmatrix} = E \begin{pmatrix} u^\uparrow(x) \\ v^\downarrow(x) \end{pmatrix} \tag{52}$$

We now consider plane wave solutions of Eq. (52) and we solve the scattering problem. In the left-hand ($x < 0$) ferromagnetic region, the wave function $\psi_L(x)$ produced by a source of right-going spin $\uparrow$ particles of unit flux at energy E can be written as

$$\psi_L(x) = \begin{pmatrix} u^\uparrow(x) \\ v^\downarrow(x) \end{pmatrix} = \begin{pmatrix} \dfrac{e^{ikx}}{\sqrt{v_k}} + r_0 \dfrac{e^{-ikx}}{\sqrt{v_k}} \\ r_a \dfrac{e^{iqx}}{\sqrt{v_q}} \end{pmatrix} \tag{53}$$

where $k = \sqrt{E + \mu + h}$, $q = \sqrt{-E + \mu - h}$, $v_k = 2k$ and $v_q = 2q$ are the group velocities relative, respectively, to the particle wave-vector k and to the hole wave-vector q. r_0 is the normal reflection amplitude, whereas r_a is the Andreev reflection amplitude, which corresponds to the reflection of an incoming spin $\uparrow$ particle into a spin $\downarrow$ hole. The Andreev reflection process [14] consists of the coherent evolution of a particle-like excitation into

a hole-like excitation. In the right-hand ($x > 0$) superconducting region, the wave function $\psi_R(x)$ is

$$\psi_R(x) = \begin{pmatrix} u^{\uparrow}(x) \\ v^{\downarrow}(x) \end{pmatrix} = t_0 \begin{pmatrix} \psi^{+} \\ \phi^{-} \end{pmatrix} \frac{e^{ik_+ x}}{\sqrt{v_{k_+}}} + t_a \begin{pmatrix} \psi^{-} \\ \phi^{-} \end{pmatrix} \frac{e^{-iq_+ x}}{\sqrt{v_{q_+}}} \tag{54}$$

where $\psi^{\uparrow}$ and $\phi^{\downarrow}$ are the coherence factors, solutions of the Bogoliubov–de Gennes equation (46) for a superconductor (see Section 2.3.2), given by

$$\begin{pmatrix} \psi \\ \phi \end{pmatrix}^{\alpha} = \frac{\Delta^*}{E \pm \sqrt{E^2 - \Delta_0^2}} \tag{55}$$

and $|\psi^{+}|^2 + |\phi^{-}|^2 = 1$. t_0 and t_a are, respectively, normal and Andreev transmission amplitudes. In (54) we have defined:

$$k_- = \sqrt{\mu + \sqrt{E^2 - \Delta_0^2}} \tag{56}$$

$$q_+ = \sqrt{\mu - \sqrt{E^2 - \Delta_0^2}} \tag{57}$$

$$v_{k_+} = \frac{\partial E}{\partial k_+} = \frac{2k_+(k_+^2 - \mu)}{\sqrt{(k_+^2 - \mu)^2 + \Delta_0^2}} \tag{58}$$

and similarly for v_{q_+}. Because we are interested in the subgap solutions (i.e., when $E < \Delta_0$), both k_- and q_+ are complex and $\psi_R(x)$ is an evanescent wave. The length ξ over which the decay of $\psi_R(x)$ occurs can be defined by $\mathrm{Re}[ik_+\xi] = -1$, so that $\xi = 1/\mathrm{Im}[k_-]$. ξ is known as the superconducting coherence length and for $E = 0$ one has $\xi \sim 1/\Delta$. It is therefore useful to define: $k_- = \sqrt{\mu + i\eta}$ and $q_+ = \sqrt{\mu - i\eta}$ with $\eta = \sqrt{\Delta_0^2 - E^2}$. The scattering problem is solved once r_0, r_a, t_0 and t_a are calculated. This can be done by matching the wave functions ψ_L and ψ_R and their derivatives at the interface ($x = 0$). If we also include a delta function potential barrier of strength H at the interface, those boundary conditions become

$$\psi_L(0) = \psi_R(0) \tag{59}$$

and

$$\psi'_R(0) - \psi'_L(0) = H\psi_L(0) \tag{60}$$

yielding the following system of four equations:

$$\begin{cases} r_a = \beta t_0 v^{+} + \delta t_a v^{-} \\ r_0 = \alpha t_0 u^{-} + \gamma t_a u^{-} - 1 \\ \epsilon r_0 + \alpha t_0 u^{-} k_- - \gamma t_a u^{-} q_+ - \epsilon^* = 0 \\ \phi r_a + \beta t_0 v^{+} k_- - \delta t_a v^{-} q_+ = 0 \end{cases} \tag{61}$$

Here, $\alpha = \left(\frac{v_q}{v_k}\right)^{1/2}$, $\beta = \left(\frac{v_q}{v_{k_-}}\right)^{1/2}$, $\gamma = \left(\frac{v_k}{v_{q_+}}\right)^{1/2}$, $\delta = \left(\frac{v_q}{v_{q_+}}\right)^{1/2}$, $\epsilon = (k + iH)$, and $\phi = (-q + iH)$. Solving (61) with respect to r_a, we find

$$r_a = -\left(\frac{v_q}{v_k}\right) \frac{2k(q_+ + k_+)}{(\epsilon + k_+)(\phi - q_+)\left(\frac{\phi}{\alpha}\right)^{*} - (\epsilon - q_+)(\phi + k_-)\left(\frac{\phi}{\alpha}\right)} \tag{62}$$

In the case where $H = 0$ and $E = 0$, in the limit of small Δ_0/μ, (62) reduces to

$$r_a = 2ie^{i\varphi} \frac{(\mu\sqrt{\mu^2 - h^2})^{1/2}}{\mu + \sqrt{\mu^2 - h^2}} \tag{63}$$

in such a way that, as shown in Fig. 5, the Andreev reflection probability $R_a = |r_a|^2$ is a decreasing function of h. From Fig. 5, we note that in one dimension, R_a is significantly suppressed only for large values of h, close to μ. For small values of h, r_a at $E = 0$ can be approximated by

$$r_a \simeq ie^{i\varphi} \tag{64}$$

like in the N/S case. In real hard ferromagnets, like Co, the value of the exchange field, although high with respect to other ferromagnets, is one order of magnitude smaller than the Fermi energy, and therefore (64) holds. In three dimensions, however, the picture changes as μ has to be replaced by $\mu - E_n$, where E_n is the transverse kinetic energy relative to the nth longitudinal mode. For quasi-particles approaching the interface at large angles, $\mu - E_n$ can become comparable to or smaller than h and, as a result, the suppression of R_a is enhanced with respect to the one-dimensional case.

To conclude we remark, first, that r_a contains the phase of the superconducting order parameter φ. It is this that gives rise to interference phenomena for systems containing more than one superconducting interface with different phases. Second, Andreev reflection is the process by which electrical current can flow across a normal-metal/superconductor interface in the subgap regime. Because the change in momentum required for the Andreev process $[\delta p = \hbar(k - q) \simeq \hbar k_F E/\mu]$ is much smaller than the change in momentum required for a normal reflection process $[\delta p = 2\hbar k \simeq 2\hbar k_F]$, Andreev reflection is strongly favored at a clean normal-metal/superconductor interface for subgap energies. During this process, a charge of $2e$ is deposited into the superconducting condensate in the form of a Cooper pair, which would be carried away by a supercurrent. For energies greater than $|\Delta|$, normal transmission is dominant for clean normal-metal/superconductor interfaces.

2.4. Scattering Theory

Electronic transport through a phase-coherent conductor can be studied in the same way as wave propagation into a scatterer. A conductor can be modeled as a scattering region connected to external leads that act as waveguides carrying a current of quasi-particles originating in external reservoirs. Scattering is assumed to be elastic, and all inelastic processes occur in such reservoirs where quasi-particles are distributed in energy according to the Fermi-Dirac function $f(E) = [\exp[(E - \mu)/k_B T] + 1]^{-1}$ and characterized by a chemical potential $\mu = eV$ and a temperature T, e being the electronic charge. The conductance G of completely normal structures can be expressed in terms of scattering probabilities [29–32] yielding the Landauer–Büttiker formula:

$$G = \frac{2e^2}{h}\mathrm{Tr}[t^\dagger t] \tag{65}$$

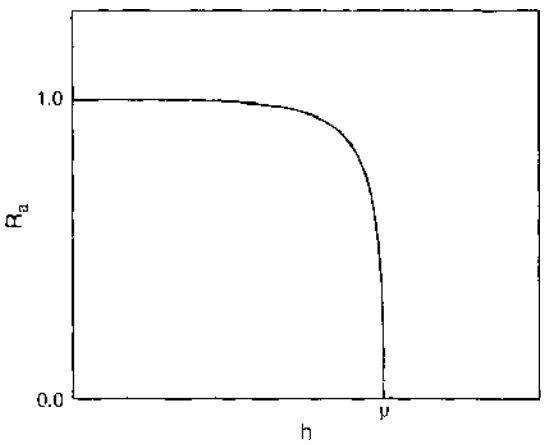


Figure 5. Andreev reflection coefficient R_a as a function of the exchange field h in one dimension. R_a is strongly suppressed only for h close to μ.

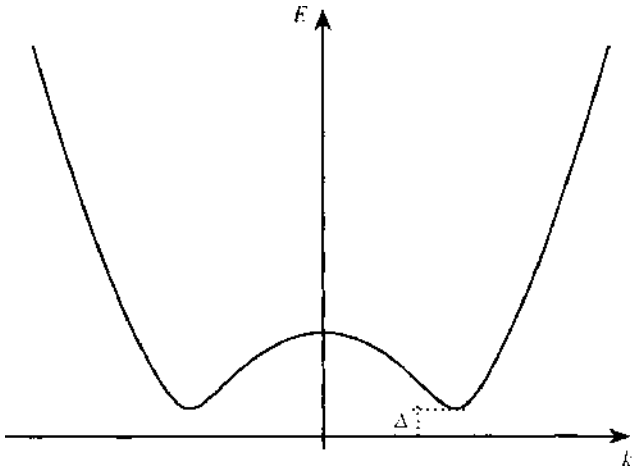


Figure 6. Energy spectrum of quasi-particle excitations in a superconductor: quasi-particle energies E are plotted as a function of their wave-vectors k. Superconducting correlations give rise to a gap of amplitude $|\Delta|$ above the Fermi level set by the horizontal axis. Typical values of Δ are hundreds of μeV.

for a two-terminal system at zero temperature in the linear response regime. In Eq. (65), e^2/h is the quantum of conductance, the factor 2 accounts for the twofold spin degeneracy, and t is the transmission amplitude matrix for an electron to propagate through the one-dimensional conductor.

For hybrid systems, the scattering theory was first applied [33] to a NIS interface (I being an insulating barrier of arbitrary strength) in Ref. [27]. The total current I_{NS} was calculated in the normal side of the junction in terms of the difference between incoming and outgoing quasi-particle distribution functions yielding the so-called BTK formula:

$$I_{NS} = 2N(0)ev_F\mathcal{A}\int_{-\infty}^{\infty}[f(E-eV)-f(E)][1+A(E)-B(E)]\,dE \tag{66}$$

where $A(E)$ and $B(E)$ are the energy-dependent Andreev and normal reflection probabilities, respectively, and V is the bias voltage applied to the junction. In Eq. (66), $N(0)$ is the density of states at the Fermi energy, v_F is the Fermi velocity, $\mathcal{A}$ is the effective cross-sectional area, and f is the Fermi distribution function. A and B were determined by imposing matching conditions to the wavefunctions, solution of the Bogoliubov–de Gennes equation, at the N and S side assuming a δ-like potential located at the interface. For low transmitting interfaces, I_{NS} reproduces the conventional transfer Hamiltonian result for which the current is proportional to the density of states of the superconductor and to the transparency of the interface (Andreev processes are completely neglected). Namely, I_{NS} is vanishingly small for sub-gap voltages and presents a pronounced peak at $eV = \Delta$, which rapidly decays asymptotically reaching the normal state current value. For highly transmitting interfaces, the I/V curve presents a large sub-gap current, due to Andreev reflection, which takes its maximum value when eV approaches Δ and thereafter decreasing to the normal state value. In addition to this, it is also possible to determine that the length at which the quasi-particle evanescent wave penetrates into the superconductor is equal to $\hbar v_F/2\Delta$ at energies close to the Fermi energy (i.e. of the same order of the BCS superconducting coherence length $\xi = \hbar v_F/\pi\Delta$).

A more general treatment generalizing the Landauer–Büttiker formulae to superconducting systems was developed in Refs. [36–38]. It accounts for generic multiterminal, multichannel hybrid structures, even in the presence of ferromagnetism and spin-flip scattering. As mentioned before, in the absence of inelastic scattering (although a generalization to include this is available [39]), dc transport is determined by the quantum mechanical scattering matrix $S(E, \mathcal{H})$, which yields scattering properties at energy E, of a phase coherent structure described by a Hamiltonian $\mathcal{H}$.

Consider a scattering region connected to external leads that carry a current of quasi-particles originating in external reservoirs at chemical potential $\mu_i = ev_i$. In the presence of Andreev scattering, current-voltage relations for a phase-coherent scatterer connected to

normal reservoirs were first derived in Ref. [36]. If the structure is connected to external reservoirs by current-carrying leads with open scattering channels labeled by a set of quantum numbers n, then the S-matrix elements are defined through the relation:

$$\hat{b}_n(E) = \sum_{n'} S_{n,n'}(E, \mathcal{H})\hat{a}_{n'}(E) \tag{67}$$

which connects the second quantization operators $\hat{b}_n$ of quasi-particles leaving the scattering region through channel n to the operators $\hat{a}_{n'}$ of quasi-particles entering the scattering region through channel n'. Both $\hat{a}_n$ and $\hat{b}_n$ satisfy anticommutation relations. The S-matrix satisfies the unitarity condition $S^{-1}(E, \mathcal{H}) = S^{\dagger}(E, \mathcal{H})$, due to quasi-particle conservation, and the time-reversibility condition $S^{t}(E, \mathcal{H}) = S(E, \mathcal{H}^*)$. In the presence of superconductivity, it is useful to label the quasi-particle open channels in the leads using the following set of discrete quantum numbers: $\{i, a, \alpha\}$, where $\alpha = +1$ characterizes particle-like excitations, $\alpha = -1$ hole-like excitations, and a is the open channel index in lead i. In doing so, the quantity $P_{ij}^{\alpha,\beta}(E, \mathcal{H}) = \sum_{a,b} |S^{\alpha,\beta}_{(i,a),(j,b)}(E, \mathcal{H})|^2$ is the probability of reflection (if $i = j$) or transmission (if $i \neq j$) of a quasi-particle of type β in lead j to a quasi-particle of type α in lead i. For $\alpha \neq \beta$, $P_{ij}^{\alpha,\beta}$ is referred to as an Andreev scattering probability, whereas for $\alpha = \beta$, it is a normal scattering probability. In the presence of superconducting leads, we insist that all superconductors share a common condensate chemical potential μ. This is to avoid time-dependent order parameter phases varying at the Josephson frequency, which would render a time-independent scattering approach invalid. When the energy E is measured with respect to μ, the particle-hole symmetry $S^{\alpha,\beta}_{(i,a),(j,b)}(E, \mathcal{H}) = \alpha\beta[S^{-\alpha,-\beta}_{(i,a),(j,b)}(-E, \mathcal{H})]^*$ is satisfied too.

Because unitarity yields

$$\sum_{\beta,\sigma',j,b} |S^{\alpha\sigma,\beta\sigma'}_{(i,a),(j,b)}(E, \mathcal{H})|^2 = \sum_{\alpha,\sigma,i,a} |S^{\alpha\sigma,\beta\sigma'}_{(i,a),(j,b)}(E, \mathcal{H})|^2 = 1 \tag{68}$$

where i and j sum only over leads supporting open channels at energy E, we have

$$\sum_{\beta,\sigma',j} P^{\alpha\sigma,\beta\sigma'}_{i,j}(E, \mathcal{H}) = N_i^{\alpha\sigma}(E) \tag{69}$$

and

$$\sum_{\alpha,\sigma,i} P^{\alpha\sigma,\beta\sigma'}_{i,j}(E, \mathcal{H}) = N_j^{\beta\sigma'}(E) \tag{70}$$

where $N_i^{\alpha\sigma}(E)$ is the number of open channels for α-type σ-spin quasi-particles of energy E in lead i. The number of incident σ-spin quasi-particles per unit time of type β along lead j, with energies between E and $E + \delta E$ is $\tilde{f}_j^{\beta\sigma}(E)\delta E$, where

$$\tilde{f}_j^{\beta\sigma}(E) = \frac{1}{h} N_j^{\beta\sigma}(E) f_j^{\beta}(E) \tag{71}$$

where $f_j^{\beta}(E)$ is the Fermi distribution of incoming quasi-particles of type β from reservoir j at temperature T_j and chemical potential μ_j:

$$f_j^{\beta}(E) = \frac{1}{e^{\frac{E - \beta(\mu_j - \mu)}{k_B T_j}} + 1} \tag{72}$$

and μ is the chemical potential of the condensate, when superconductors are present. When deriving Eq. (71), we used the fact that for each channel, the product of the density of states per unit length and longitudinal group velocity is a constant equal to $1/h$. Note that, in the case of a ferromagnetic reservoir, $f_j^{\beta}(E)$ does not depend on the spin of quasi-particles because the chemical potential μ_j of the reservoir is independent of spin.

The number of outgoing α-type quasi-particles per unit time in lead i, with energies between E and $E + \delta E$ is $J_i^{\alpha\sigma}(E)\delta E$, where

$$J_i^{\alpha\sigma}(E) = \frac{1}{h} \sum_{j,\beta,\sigma'} P_{i,j}^{\alpha\sigma,\beta\sigma'}(E) f_j^{\beta}(E) \tag{73}$$

The electrical current in lead i due to quasi-particles between E and $E + \delta E$ is therefore

$$I_i(E)\delta E = e \sum_{\alpha\sigma} (\alpha)[\tilde{J}_i^{\alpha\sigma}(E) - J_i^{\alpha\sigma}(E)]\delta E \tag{74}$$

and the total current due to all quasi-particles is the integral

$$I_i = \frac{e}{h} \int_0^{\infty} dE \sum_{\alpha\sigma} (\alpha)\left[N_i^{\alpha\sigma}(E) f_i^{\alpha}(E) - \sum_{j\beta\sigma'} P_{i,j}^{\alpha\sigma,\beta\sigma'}(E) f_j^{\beta}(E)\right] \tag{75}$$

If we write down I_i in the form:

$$I_i = \sum_{j\sigma} A_{ij}^{\sigma} \tag{76}$$

then we can define the matrix A_{ij}^{σ} as

$$A_{ij}^{\sigma} = \frac{e}{h} \int_0^{\infty} dE \sum_{\alpha\beta\sigma'} (\alpha)\left[N_i^{\alpha\sigma}(E) f_i^{\alpha}(E)\delta_{ij}\delta_{\sigma\sigma'}\delta_{\alpha\beta} - P_{i,j}^{\alpha\sigma,\beta\sigma'}(E) f_j^{\beta}(E)\right] \tag{77}$$

One should note that if the system is periodic in the plane perpendicular to the longitudinal direction, the total current is the sum of currents associated with each Bloch wave-vector in the first Brillouin zone, so that (76) should be replaced by

$$I_i = \sum_{j\sigma} \sum_{k \in BZ} A_{ij}^{\sigma}(\vec{k}) \tag{78}$$

where the dependence on $\vec{k}$ in (77) is contained in the quantities $N_i^{\alpha\sigma}$ and $P_{i,j}^{\alpha\sigma,\beta\sigma'}$.

At finite voltages, but zero-temperature we have

$$f_i^{+}(E) = \begin{cases} 1 & \text{if} E < (\mu_i - \mu) \\ 0 & \text{if} E > (\mu_i - \mu) \end{cases} \tag{79}$$

and

$$f_i^{-}(E) = \begin{cases} 1 & \text{if} E < -(\mu_i - \mu) \\ 0 & \text{if} E > -(\mu_i - \mu) \end{cases} \tag{80}$$

therefore, (77) reduces to

$$A_{ij}^{\sigma} = \frac{e}{h} \int_0^{\infty} dE \sum_{\sigma'} \left\{\left[N_i^{+\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{+\sigma,+\sigma'}(E) + P_{ij}^{-\sigma,+\sigma'}(E)\right] f_j^{+}(E) - \left[N_i^{-\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{-\sigma,-\sigma'}(E) + P_{ij}^{+\sigma,-\sigma'}(E)\right] f_i^{-}(E)\right\} \tag{81}$$

Now, if $(\mu_i - \mu) > 0$, then the second term in (81) is zero, as $f_i^{-}(E) = 0$ for $E > -(\mu_i - \mu)$ and one finds that

$$A_{ij}^{\sigma} = \frac{e}{h} \int_0^{e(\mu_i - \mu)} dE \sum_{\sigma'} \left[N_i^{+\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{+\sigma,+\sigma'}(E) + P_{ij}^{-\sigma,+\sigma'}(E)\right] \tag{82}$$

If $(\mu_i - \mu) < 0$, then the first term in (81) is zero, but the resulting A_{ij}^{σ} is formally equal to (82).

In the linear response limit, we assume that $(v_i - v) \to 0$ and we write down I_i in the form:

$$I_i = \sum_{j\sigma} a_{ij}^{\sigma}(v_j - v) \tag{83}$$

To obtain the coefficients a_{ij}^{σ}, we expand f_i^{α} in Eq. (77) in powers of $(v_j - v)$:

$$f_j^{\alpha}(E) = f_j^{\alpha}(E)\Big|_{(v_j - v)=0} + e(v_j - v)\left(\frac{\partial f_j^{\alpha}(E)}{e\partial(v_j - v)}\right)\Big|_{(v_j - v)=0} + \cdots \tag{84}$$

It is easy to show that

$$\frac{\partial f_j^{\alpha}(E)}{e\partial(v_j - v)} = -\alpha\frac{\partial f_j^{\alpha}(E)}{\partial E} \tag{85}$$

and

$$\left(\frac{\partial f_j^{\alpha}(E)}{\partial E}\right)\Big|_{(v_j - v)=0} = \frac{\partial f}{\partial E} \tag{86}$$

where

$$f(E) = \frac{1}{1 + e^{E/k_B T}} \tag{87}$$

Therefore, we obtain

$$f_j^{\alpha}(E) \simeq f(E) - \alpha e(v_j - v)\frac{\partial f}{\partial E} \tag{88}$$

Substituting (88) in (77) and calculating I_i, one finds that the terms in $f(E)$ cancel because of the unitarity of the S-matrix expressed by Eq. (69), and we are left with

$$I_i = \sum_{j\sigma} A_{ij}^{\sigma} = \frac{e}{h}\int_0^{\infty} dE \sum_{j\alpha\beta\sigma'}(\alpha)\Big[-(\alpha)N_i^{\alpha\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'}\delta_{\alpha\beta}e(v_j - v)\frac{\partial f}{\partial E} + (\beta)P_{ij}^{\alpha\sigma,\beta\sigma'}(E)e(v_j - v)\frac{\partial f}{\partial E}\Big] \tag{89}$$

Noting that $\alpha^2 = +1$ and exchanging σ' with σ, one finds

$$a_{ij}^{\sigma} = \frac{e^2}{h}\int_0^{\infty} dE\left(-\frac{\partial f}{\partial E}\right)\sum_{\alpha\beta\sigma'}\left[N_i^{\alpha\sigma'}(E)\delta_{ij}\delta_{\sigma\sigma'}\delta_{\alpha\beta} - (\alpha\beta)P_{ij}^{\alpha\sigma',\beta\sigma}(E)\right] \tag{90}$$

In order to simplify further Eq. (90), we write explicitly the sum over α and β:

$$a_{ij}^{\sigma} = \frac{e^2}{h}\int_0^{\infty} dE\left(-\frac{\partial f}{\partial E}\right)\sum_{\sigma'}\{[N_i^{+\sigma'}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{+\sigma',+\sigma}(E) + P_{ij}^{+\sigma',-\sigma}(E)] + [N_i^{-\sigma'}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{-\sigma',-\sigma}(E) + P_{ij}^{-\sigma',+\sigma}(E)]\} \tag{91}$$

It is possible to show that in the general case of spin-flip scattering with the possibility of ferromagnetic leads, the particle-hole symmetry can be written as

$$S_{(i,\alpha),(j,\beta)}^{\alpha\sigma,\beta\sigma'}(E, \mathcal{H}) = \alpha\beta\left[S_{(i,\alpha),(j,\beta)}^{-\alpha\sigma,-\beta\sigma'}(-E, \mathcal{H})\right]^* \tag{92}$$

In particular, we have that $N_i^{-\sigma'}(E) = N_i^{+\sigma}(-E)$, $P_{ij}^{-\sigma',+\sigma}(E) = P_{ij}^{+\sigma',-\sigma}(-E)$, and $P_{ij}^{-\sigma',-\sigma}(E) = P_{ij}^{+\sigma',+\sigma}(-E)$. By using these relations and noting that $\partial f(-E)/\partial E = -\partial f(E)/\partial E$, one can change variables in the second part of the integral (91) obtaining the

same expression given by the first part, but over a different range of energies. As a result (91) becomes

$$a_{ij}^{\sigma} = \frac{e^2}{h}\int_{-\infty}^{\infty} dE\left(-\frac{\partial f}{\partial E}\right)\sum_{\sigma'}\left[N_i^{+\sigma'}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{-\sigma',+\sigma}(E) + P_{ij}^{+\sigma',-\sigma}(E)\right] \tag{93}$$

At zero temperature, where $-\partial f/\partial E = \delta(E)$, (93) reduces to

$$a_{ij}^{\sigma} = \frac{e^2}{h}\sum_{\sigma'}\left[N_i^{-\sigma}(0)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{-\sigma',-\sigma}(0) + P_{ij}^{+\sigma',-\sigma}(0)\right] \tag{94}$$

Starting from the above results, we can obtain the expression for the differential conductance at finite voltages $g_i^j(v_j)$ defined as the derivative of the current in lead i with respect to the voltage v_j, with v and all other potentials held constant. At this point, it is important to notice that all scattering coefficients can be computed in the presence of self-consistently determined order parameter Δ and normal potentials u^{σ}. In such a case, one must take into account the effects on the system of changes in the applied potentials v_js. Such changes affect both the normal potential and the superconducting order parameter. For normal structures, self-consistency in the normal potential has been considered by Christen et al. (Ref. [40]), and it turned out that such effect can become important at large voltages. Calculations taking into account a self-consistent order parameter have been carried out for one dimensional structures in Refs. [41–43]. These demonstrate that provided the currents are low enough with respect to the critical current, the matrix A_{ij}^{σ} can remain unchanged, even by the application of finite voltages of order $|\Delta|$. However, here we are only interested in the effect due to the presence of the order parameter on a system, thus we shall neglect such effects, and we shall restrict ourselves to voltages of the order of $|\Delta|$.

Taking the derivative of I_i with respect to $(v_j - v)$ we obtain

$$\begin{aligned} g_i^j &= \frac{\partial I_i}{\partial(v_j - v)} \\ &= \frac{e}{h}\int_0^{\infty} dE \sum_{\alpha\beta\sigma'\sigma}\left[N_i^{\alpha\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'}\delta_{\alpha\beta}\left(-\frac{\partial f_i^{\alpha}}{\partial E}\right) - (\alpha\beta)P_{ij}^{\alpha\sigma',\beta\sigma}(E)\left(-\frac{\partial f_j^{\beta}}{\partial E}\right)\right] \end{aligned} \tag{95}$$

where we have used the following equality:

$$\frac{\partial f_i^{\alpha}(E)}{\partial(v_i - v)} = -\alpha\frac{\partial f_i^{\alpha}(E)}{\partial E}\delta_{ij}\delta_{\sigma\sigma'} \tag{96}$$

Expanding the sum over α and β and reorganizing the terms, one obtains

$$\begin{aligned} \frac{\partial I_i}{\partial(v_j - v)} = \frac{e}{h}\int_0^{\infty} dE \sum_{\sigma'\sigma}\Bigg\{&\left[N_i^{+\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{+\sigma',+\sigma}(E) + P_{ij}^{-\sigma',+\sigma}(E)\right]\left(-\frac{\partial f_j^{+}}{\partial E}\right) \\ &+ \left[N_i^{-\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{\sigma',-\sigma}(E) + P_{ij}^{+\sigma',-\sigma}(E)\right]\left(-\frac{\partial f_j^{-}}{\partial E}\right)\Bigg\} \end{aligned} \tag{97}$$

Using the particle-hole symmetry relations (92) and

$$\frac{\partial f_i^{\sigma}(E)}{\partial E} = -\frac{\partial f_i^{-\sigma}(-E)}{\partial E} \tag{98}$$

one arrives at the final formula

$$\begin{aligned} g_i^j(v_j) &= \frac{\partial I_i}{\partial(v_j - v)} \\ &= \frac{e}{h}\int_{-\infty}^{\infty} dE \sum_{\sigma\sigma'}\left[N_i^{+\sigma}(E)\delta_{ij}\delta_{\sigma\sigma'} - P_{ij}^{-\sigma,+\sigma}(E) + P_{ij}^{-\sigma',+\sigma}(E)\right]\left(-\frac{\partial f_j^{-}}{\partial E}\right) \end{aligned} \tag{99}$$

We conclude by noting that, at finite voltages, the chemical potential of the condensate μ should be determined self-consistently by imposing the conservation of the current $\sum_i I_i = 0$ which yields, from Eq. (76), $\sum_{i,j,\sigma} A_{ij}^{\sigma} = 0$. From the expression of A_{ij}^{σ} in (77), it is clear that such a self-consistency condition involves integrals over all incident quasi-particle energies, which in turns require a knowledge of the S-matrix over a range of energies. In the case where superconductivity is present in one lead (say j), the solution is trivial, as μ coincides with one of the voltages v_j. In the general case, however, the solution of the self-consistent equations is difficult, but for some simple structures may be determined by symmetry arguments. In a spatially symmetric two-probe system, for example, one can assume that $\mu = 1/2(\mu_1 + \mu_2)$. Finally, we refer the reader to Ref. [44] for issues related to shot noise power.

2.5. Two-Probe Differential Conductance and Conductance in More Details

We now apply the above formalism to derive two-terminal conductance formulas. We first consider the case of two nonsuperconducting leads (from now on, referred to as NN system) and then the case where one of the leads is a superconductor (NS system).

The NN system can be schematically drawn as in Fig. 7, insisting that $I_2 = -I_1 = -I$ (current conservation). In the current case, one can distinguish between two possible definitions for the differential conductance:

$$g_1(v_1) = \left.\frac{\partial I_1}{\partial(v_1 - v)}\right|_{v_2} \quad \text{and} \quad g_2(v_2) = \left.\frac{\partial I_1}{\partial(v_2 - v)}\right|_{v_1} \tag{100}$$

depending on whether one varies v_1 (for fixed v_2), or v_2 (for fixed v_1). g_1, for example, can be written using Eq. (99), to give

$$g_1(v_1) = \frac{e}{h}\int_{\infty} dE\left(-\frac{\partial f_1^+}{\partial E}\right)\sum_{\sigma'\sigma}[N_1^{+\sigma}(E)\delta_{\sigma\sigma'} - R_{11}^{\sigma'\sigma}(E) + R_a^{\sigma'\sigma}(E)] \tag{101}$$

where we have defined $R_{11}^{\sigma'\sigma} = P_{11}^{1\sigma',1\sigma}$ ($R_a^{\sigma'\sigma} = P_{11}^{-\sigma',1\sigma}$) as the normal (Andreev) reflection probability for an injected σ-spin quasi-particle to come back as a σ'-spin quasi-particle (hole).

The conductance at zero-temperature and zero-voltage can be calculated by rewriting Eq. (83) in matrix form as

$$\begin{pmatrix} I \\ -I \end{pmatrix} = \begin{pmatrix} a_{11}^{\uparrow} + a_{11}^{\downarrow} & a_{12}^{\uparrow} + a_{12}^{\downarrow} \\ a_{21}^{\uparrow} + a_{21}^{\downarrow} & a_{22}^{\uparrow} + a_{22}^{\downarrow} \end{pmatrix} \begin{pmatrix} v_1 - v \\ v_2 - v \end{pmatrix} \tag{102}$$

and considering its inverse

$$\begin{pmatrix} v_1 - v \\ v_2 - v \end{pmatrix} = \frac{1}{\tilde{d}} \begin{pmatrix} \tilde{a}_{22} & -\tilde{a}_{12} \\ -\tilde{a}_{21} & \tilde{a}_{11} \end{pmatrix} \begin{pmatrix} I \\ -I \end{pmatrix} \tag{103}$$

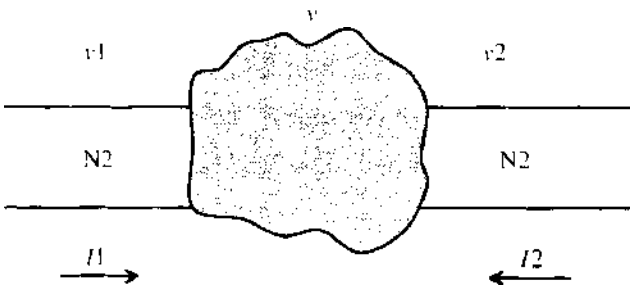


Figure 7. A schematic picture of the NN system. The central gray region is the scatterer, whereas the two lateral objects denoted by N1 and N2 represent the leads. v is the chemical potential of the possible superconducting regions contained within the scatterer.

where $\tilde{a}_{ij} = a^{\uparrow}_{ij} + a^{\downarrow}_{ij}$ and $d = \tilde{a}_{11}\tilde{a}_{22} - \tilde{a}_{12}\tilde{a}_{21}$ is the determinant. Because we are interested in the conductance defined as $G = \frac{I}{v_1 - v_2}$, we want to get rid of the condensate chemical potential v. We can achieve this goal by subtracting the second equation in (103) from the first one, obtaining

$$G = \frac{d}{\tilde{a}_{11} + \tilde{a}_{12} + \tilde{a}_{21} + \tilde{a}_{22}} \tag{104}$$

where the matrix elements $\tilde{a}_{ij}$, evaluated at the Fermi energy, are expressed as follows:

$$\tilde{a}_{11} = \frac{e^2}{h}\left[\left(N^{\uparrow} - R_0^{\uparrow\uparrow} - R_0^{\downarrow\uparrow} + R_a^{\uparrow\uparrow} + R_a^{\downarrow\uparrow}\right) + \left(N^{\downarrow} - R_0^{\uparrow\downarrow} - R_0^{\downarrow\downarrow} + R_a^{\downarrow\downarrow} + R_a^{\downarrow\downarrow}\right)\right] \tag{105}$$

$$\tilde{a}_{21} = \frac{e^2}{h}\left[\left(T_a^{\uparrow\uparrow} + T_a^{\downarrow\uparrow} - T_0^{\uparrow\uparrow} - T_0^{\downarrow\uparrow}\right) + \left(T_a^{\uparrow\downarrow} + T_a^{\downarrow\downarrow} - T_0^{\uparrow\downarrow} - T_0^{\downarrow\downarrow}\right)\right] \tag{106}$$

$$\tilde{a}_{12} = \frac{e^2}{h}\left[\left(T_a^{\prime\uparrow\uparrow} + T_a^{\prime\downarrow\uparrow} - T_0^{\prime\uparrow\uparrow} - T_0^{\prime\downarrow\uparrow}\right) + \left(T_a^{\prime\uparrow\downarrow} + T_a^{\prime\downarrow\downarrow} - T_0^{\prime\uparrow\downarrow} - T_0^{\prime\downarrow\downarrow}\right)\right] \tag{107}$$

and

$$\tilde{a}_{22} = \frac{e^2}{h}\left[\left(N^{\prime\uparrow} - R_0^{\prime\uparrow\uparrow} - R_0^{\prime\downarrow\uparrow} + R_a^{\prime\uparrow\uparrow} + R_a^{\prime\downarrow\uparrow}\right) + \left(N^{\prime\downarrow} - R_0^{\prime\uparrow\downarrow} - R_0^{\prime\downarrow\downarrow} + R_a^{\prime\uparrow\downarrow} + R_a^{\prime\downarrow\downarrow}\right)\right] \tag{108}$$

Here, $T_0^{\sigma\sigma'} = P_{21}^{+\sigma,+\sigma'}$ ($T_a^{\sigma\sigma'} = P_{21}^{-\sigma,+\sigma'}$) is the probability for normal (Andreev) transmission for an injected σ'-spin quasi-particle from the left-lead into a σ-spin quasi-particle (hole) in the right-lead. For the primed quantities, the quasi-particles are injected from the right-lead into the left-lead. Note that $T_0^{\sigma\sigma'}$ and $R_0^{\sigma\sigma'}$, for $\sigma \neq \sigma'$, and $R_a^{\sigma\sigma'}$ and $T_a^{\sigma\sigma'}$ are nonzero only if spin-flip scattering is present. As a check of consistency, it is easy to prove that in the absence of superconductors, one recovers the well-known Landauer formula:

$$G = \frac{e^2}{h}\left[T_0^{\uparrow\uparrow} + T_0^{\downarrow\uparrow} + T_0^{\uparrow\downarrow} + T_0^{\downarrow\downarrow}\right] \tag{109}$$

When superconductors are present, it is interesting to note that their common condensate chemical potential v can be determined imposing the conservation of the current, which yields:

$$v = \frac{(\tilde{a}_{12} + \tilde{a}_{22})v_1 + (\tilde{a}_{22} + \tilde{a}_{21})v_2}{\tilde{a}_{11} + \tilde{a}_{12} + \tilde{a}_{21} + \tilde{a}_{22}} \tag{110}$$

The NS system can be schematically drawn as in Fig. 8, where again we insist that $I_2 = -I_1 = -I$ and that the superconducting lead and the superconductors in the scattering region share the same chemical potential: $v_2 = v$. Note that the relation $I_2 = -I_1$ allows us to avoid the explicit calculation of the current in the superconducting lead (I_2), which will be a combination of quasi-particle current and supercurrent. In this case, there is only one

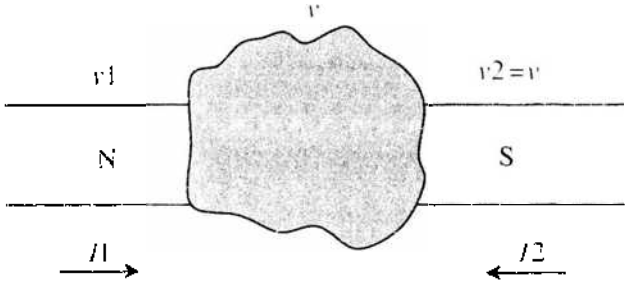


Figure 8. A schematic picture of the NS system. The central gray region is the scatterer, whereas the two lateral objects represent the leads (in light gray is the superconducting one). v is the chemical potential of the possible superconducting regions contained within the scatterer, which equals the chemical potential of the right-hand superconducting lead.

possible definition for the differential conductance, namely $g = \frac{\partial I_1}{\partial(v_1 - v)}$. From Eq. (99), we easily obtain

$$g = \frac{e}{h}\int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_1^+}{\partial E}\right)\sum_{\alpha\alpha'}\left[N_1^{\alpha\alpha'}(E)\delta_{\alpha\alpha'} - R_{11}^{\alpha'\alpha}(E) + R_{a}^{\alpha'\alpha}(E)\right] \tag{111}$$

The conductance at zero-temperature and zero-voltage is also easily derived. Equation (102) becomes trivial because, on the one hand, no transmission of quasi-particles to the superconductor is allowed at the Fermi energy ($\tilde{a}_{12} = 0$) and, on the other, $v_2 - v = 0$. Therefore, (102) reduces to $I = \tilde{a}_{11}(v_1 - v)$, which gives

$$G = \frac{I}{v_1 - v} = \frac{e^2}{h}\left[\left(N_1^{\uparrow} - R_{11}^{\uparrow\uparrow} - R_{11}^{\uparrow\downarrow} + R_a^{\uparrow\uparrow} + R_a^{\uparrow\downarrow}\right) + \left(N_1^{\downarrow} - R_{11}^{\downarrow\downarrow} - R_{11}^{\downarrow\uparrow} + R_a^{\downarrow\downarrow} + R_a^{\downarrow\uparrow}\right)\right] \tag{112}$$

where the probability coefficients are calculated at the Fermi energy.

It is very instructive to consider a structure composed of a superconducting scatterer of length L_S attached to two normal leads, one on the left and the other on the right. From Eq. (75), one can derive the following two-probe linear-regime zero-temperature conductance [45]:

$$G = \frac{2e^2}{h}\left[T + T_a + \frac{2(R_a R_a' - T_a T_a')}{R_a + R_a' + T_a + T_a'}\right] \tag{113}$$

where T (T_a) is the normal (Andreev) transmission probability for quasi-particles injected from the left-lead and arriving on the right-lead. R (R_a) is the normal (Andreev) reflection probability for quasi-particles injected from the left-lead. Similarly T_a' and R_a' are Andreev scattering probabilities for quasi-particles injected from the right-lead. For $L_S \gg \xi$, all transmission probabilities tend to zero, as the quasi-particles penetrate into the superconductor up to a depth of the order of ξ, so that the conductance reduces to the series of two interface resistances, namely $1/2R_a$ and $1/2R_a'$ in units of $h/(2e^2)$. A consequence of this is the fact that the overall conductance of the structure does not depend on L_S, but simply on the microscopic structure of the NS interfaces. In particular, for $L_S \to \infty$, although the resistance of the system attains an asymptotic finite value, the resistance per unit length (*resistivity*) vanishes, as it must be for a superconductor.

According to Eq. (75), the problem of determining the current–voltage characteristic is reduced to the calculation of scattering amplitudes. This can be done in different ways; in simple ballistic systems, for example, by solving the Bogoliubov–de Gennes equation piecewise in the different homogeneous regions composing the scatterer and imposing proper matching conditions to the wavefunctions at the boundaries between these regions. Complex structures such as disordered and diffusive conductors, heterostructures composed of different materials, and multiterminal systems can be dealt with too. Scattering amplitudes can be determined numerically, for example, by discretizing the system in real space within the tight-binding formalism. For diffusive wires and chaotic quantum dots, a different approach can successfully be employed, namely the random-matrix theory (see, for example, the review papers of Refs. [46, 47]). According to this theory, from the statistical properties of a class of matrices with random elements describing a certain physical system, it is possible to extract the properties of the system. This can be applied to scattering matrices: the first problem consists in determining their random-matrix probability distribution (statistical ensemble). The second problem is to find the correlation functions of the transmission eigenvalues from which the transport properties can be derived. The conductance, for example, is simply given by $2e^2/h\int_0^1 dT T\rho(T)$, where $\rho(T)$ is the mean eigenvalue density. A complete solution to this problem has been found for diffusive wires and quantum dots. For hybrid systems consisting of a phase-coherent structure connected to superconducting leads, the random-matrix theory is based on a relationship that links the Andreev reflection to the transmission eigenvalues of the corresponding normal system [48]:

$$\frac{4e^2}{h}\sum_{n=1}^{N}\frac{T_n^2}{(2 - T_n)^2} \tag{114}$$

where T_n is a transmission eigenvalue and N is the number of open channels [49, 50]. From Eq. (114), it is clear that hybrid systems can be studied, making use of the results developed for normal conductors.

2.6. Quasi-Classical Green's Function Approach

In this section, we briefly underline another very important technique for dealing with hybrid systems. This approach applies when the characteristic length scale of the problem is large if compared with Fermi wavelength. It proves to be particularly useful for dirty structures where a sufficient amount of nonmagnetic impurities is present so to make the motion of electrons isotropic.

The quasi-classical theory of superconductivity is developed through quantum field theory methods, and it is formulated in terms of Green's functions in the particle-hole space (known as *Nambu space*). Such Green's functions are denoted by a *hat* and, in addition to standard Green's functions G, they contain anomalous components F that describe the superconducting correlations:

$$\widehat{G}(E, \mathbf{r}_1, \mathbf{r}_2) = \begin{pmatrix} G(E, \mathbf{r}_1, \mathbf{r}_2) & F(E, \mathbf{r}_1, \mathbf{r}_2) \\ F^*(E, \mathbf{r}_1, \mathbf{r}_2) & G^*(E, \mathbf{r}_1, \mathbf{r}_2) \end{pmatrix} \tag{115}$$

The starting point is the Gorkov equation [18] for the Green's function of a bulk superconductor, which was derived directly from the BCS Hamiltonian. The quasi-classical approximation consists in averaging the Green's function $\widehat{G}$ over the rapid oscillations in the relative space coordinates $\boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2$ and over impurities. As a result, one obtains the quasi-classical Green's function $\hat{g}(E, \mathbf{r}, \mathbf{n})$, which depends on the energy E, on the center-of-mass coordinate $\mathbf{r} = (\mathbf{r}_1 + \mathbf{r}_2)/2$, and the versor of the relative momentum $\mathbf{n}$ associated to the relative coordinate $\boldsymbol{\rho}$. The Gorkov equation reduces to the following (Eilenberger) equation for $\hat{g}$:

$$\mathbf{v}_F \partial_r \hat{g} + \left[-iE\hat{\sigma}_3 + \widehat{\Delta} + \frac{\langle \hat{g} \rangle}{2\tau}, \hat{g} \right] = 0, \quad \text{with} \quad \widehat{\Delta} = \begin{pmatrix} 0 & \Delta \\ \Delta^* & 0 \end{pmatrix} \tag{116}$$

derived in Refs. [51, 52]. Here, $\mathbf{v}_F$ is the Fermi velocity, Δ is the pair potential, and τ is the elastic scattering time. Furthermore, the square brackets represent the commutator, whereas the angular brackets $\langle \cdots \rangle$ denote an averaging over the direction $\mathbf{n}$. Note that Eq. (116) determines $\hat{g}$ up to a multiplicative constant, and the following normalization condition must be applied: $\hat{g}^2 = \hat{1}$.

A further simplification can be realized in the presence of an isotropic impurity scattering potential, which makes the motion of electrons diffusive (dirty limit). In this case, $\hat{g}$ can be expanded in spherical harmonics, and Eq. (116) reduces to the Usadel equation [53] for the isotropic function $\hat{g}(E, \mathbf{r}) = \langle \hat{g}(E, \mathbf{r}, \mathbf{n}) \rangle$:

$$D\partial_r(\hat{g}\partial_r\hat{g}) + [iE\hat{\sigma}_3 - \widehat{\Delta}, \hat{g}] = 0 \tag{117}$$

where $D = v_F^2\tau/3$ is the diffusion constant. It is worthwhile to stress that both Eq. (116) and (117) must be supplemented with a self-consistent condition for the pair potential Δ. The electrical current can then be calculated once the Green's function $\hat{g}$ is determined.

It should be noted that the quasi-classical approximation does not allow to take into account nonuniformities that occur on the Fermi wavelength scale, such as boundaries, barriers, and interfaces with other materials. It was shown, however, that this problem can be circumvented by applying proper boundary conditions to the quasi-classical Green's function. For the Eilenberger equation, they were derived in Ref. [54] and for the Usadel equation in Ref. [55]. It is interesting to note that such boundary conditions are obtained by making use of the connection between scattering amplitudes and Green's functions. For a more detailed treatment of the quasi-classical theory, we refer the reader to a number of review papers on the subject (for the most recent see, for example, Refs. [10, 56–58]).

An important progress in the quasi-classical theory was put forward by Nazarov who formulated it, within the dirty limit, in terms of a circuit theory [59–61]. This constitutes a

generalization of the classical Ohm's law to hybrid coherent nanostructures. These can be viewed as coherent networks consisting of *nodes* linked by *connectors*. The starting point consists in observing that when the second term in the Usadel equation (117) can be neglected, the same equation can be written in the form of a conservation law for a matrix current $\check{\mathbf{j}}(\mathbf{r})$ [62]:

$$\partial_{\mathbf{r}}\check{\mathbf{j}}(\mathbf{r}) = 0; \quad \check{\mathbf{j}}(\mathbf{r}) = \sigma\check{G}\partial_{\mathbf{r}}\check{G} \tag{118}$$

where σ is the normal state conductivity proportional to the diffusion constant D. The second equation in (118) resembles the local Ohm's law in a normal metal structure, whereas the first expresses the conservation of the current. The problem is finally defined by imposing additional boundary conditions, which sets the bias voltage established across the structure and describe interfaces in the conductor. By solving those equations, the voltage distribution over the conductor and hence the local current density and total current can be determined. The ordinary circuit theory is formulated by discretizing Eqs. (118) plus boundary conditions through a finite element approximation so that the structure can be separated into resistive elements pairwise connected. Current conservation in each node, and current–voltage relations for each element set the voltage in each node allowing the determination of the total current.

The extended circuit theory cannot be derived in such a simple way because of the matrix form of the equations and the fact that one deals with matrix currents and Green's functions. However, the situation can be simplified due to the symmetry properties of the Green's function at zero energy so that Eqs. (118) can be separated into two parts. The first one determines the equilibrium spectral local properties and the second one the nonequilibrium chemical potentials in each node, which determines the electron propagation. These equations allow to formulate a circuit theory that can be conveniently condensed into a set of rules (see Ref. [59]) for determining currents and potential differences of a coherent network of nodes and connectors. It is worth mentioning that (i) a node is required to present enough disorder to prevent ballistic transmission between the two connectors attached to it. Furthermore, the size of a node must be smaller than the coherence length. (ii) Connectors can be of three types, namely diffusive conductors, tunnel junctions, or ballistic constrictions. The net effect of superconductivity is to induce a renormalization of the conductivity in the last two types of connectors.

It is important to recall that the above circuit theory is valid when the second term in the Usadel equation can be neglected (i.e., for small temperatures and voltages [59]): $k_BT, eV \ll \Delta, D/L^2$, L being the system size. The limitations described above are quite restrictive, namely the voltage and temperature dependence of the transport properties cannot be accessed. However, a circuit theory can still be formulated by discretizing the Usadel equation (117) including the second term. This can be done by discretizing the space into a set of nodes x_i such that the Green's function $\check{G}(x_i) \equiv \check{G}_i$ in neighboring points are close to each other. A resistor is associated with each connection. The Eq. (117) can then be viewed as a conservation law in each node i ($\sum_k \check{I}_{ik} + \check{I}_{\text{leakage},i} = 0$), where now the total current is composed of $\check{I}_{ik}$ [which corresponds to the second expression in Eq. (118)] and an additional *leakage* contribution $\check{I}_{\text{leakage},i} \propto -ie^2[\check{G}_i, \check{H}]$ associated to the second term of the Usadel equation (117). The leakage current describes two processes, the first one, proportional to the energy ϵ, is associated to the decoherence of electrons and holes due to wave-vector mismatch at finite energies ("leakage" of coherence). The second one, proportional to Δ, is responsible for the conversion between quasi-particles and Cooper pairs ("leakage of quasi-particles"). In order to obtain an accurate agreement with the full theory, the distance between the nodes has to be smaller than the coherence length $\xi = \sqrt{D\text{Max}(\Delta, \epsilon)}$. It is important to notice that ballistic point contacts can be considered as a particular kind of connectors, but they need a specific treatment. To conclude, this new circuit theory can be formalized into a set of rules given in Ref. [61].

2.7. Panorama on Results in NS Systems

Superconducting hybrid nanostructures possesses a rich physics related, in particular, to the interplay between coherent transport of electrons and Andreev reflection. Physical systems investigated so far comprise NS junctions, mesoscopic SNS Josephson junctions, structures containing several superconducting islands and, more recently, ferromagnet/superconductor (FS) junctions, where the interplay between spin-dependent transport and superconductivity can be investigated. The vast knowledge available in semiconductor technology has been also exploited by replacing in the above systems normal-metals with semiconductors (Sm). The possibility of using unconventional superconducting materials, such as d-wave superconductors and, more generally, high-T_c superconductors, has been explored too. As a result, a variety of effects and phenomena have been reported so far, the most important of which we shall review in the current section, making use of the various theoretical approaches outlined in the previous sections. It is worth mentioning that these allow one to calculate, in addition to the electrical current, all transport properties of a given system such as thermal current, current noise, and the whole full counting statistics of electronic transport.

We shall start considering NS junctions, which present different properties depending on whether N is diffusive or ballistic and on the transparency of the barrier at the interface. Perhaps the first effect that received much attention is the zero bias anomaly (ZBA), which refers to a conductance peak observed at zero bias voltage in a low-transparency NS point contact [63]. This contrast with the BTK result, where the conductance presents a minimum at zero bias, is particularly striking, as the height of the peak is of the order the normal state conductance. ZBA can be explained as an interplay between Andreev reflection at the interface with the superconductor and disorder scattering in the N side of the junction due to the presence of impurities. When the elastic mean free path in the N side of the junction is smaller than the junction size, electrons have the chance to be scattered back to the S interface many times, finally undergoing an Andreev reflection. The net result is that Andreev reflection processes occur with a much larger probability with respect to the case of absence of impurities. In other words, the low bias conductance is determined by complex interference effects that produce an enhancement by several orders of magnitude. Such an effect dies away for energies larger than the Thouless energy, which sets the scale for particle-hole dephasing. Although ZBA was first understood within the quasi-classical approach, it was then confirmed using scattering methods and also in the tunneling Hamiltonian approach. This effect is also known as *reflectionless tunneling*, as it can be explained in terms of the disorder-induced opening of tunneling channels.

Non-monotonic behaviors of the conductance as a function of voltage and temperature in NS junctions with clean interfaces were also reported. When a diffusive wire is connected to a superconducting reservoir through a highly transparent contact, the zero-temperature, zero-voltage average conductance equals the normal-state conductance [48, 64]. Even though this can be proved rigorously, intuitively it is the consequence of two facts, namely that the Andreev reflection effectively doubles the length of the normal metal conductor and that the conductance in a NS point contact is doubled with respect to the normal-state one. However, it was found [64–69] that a conductance peak appears either for voltages or temperatures of the order of the Thouless energy $\hbar D/L^2$. This phenomenon, known as *reentrance effect*, was experimentally proved in Refs. [70–72].

Another interesting phenomenon is the Andreev interferometry, which is realized in a hybrid structure containing at least two superconducting islands. This is based on the fact that when a quasi-particle Andreev reflects from a NS interface, the phase of the outgoing excitation is shifted by the phase of the superconducting order parameter. In the presence of two superconducting islands with order parameter phases ϕ_1 and ϕ_2, the transport properties are oscillatory functions of the difference $\phi_1 - \phi_2$. The conductance of individual samples was found to be 2π-periodic in the phase difference.

Heat transport through NS interfaces can be applied to microcooling [73–75]. The physical mechanism underlying this electronic cooling is quite simple. When a normal metal is brought in contact with a superconductor, quasi-particle transport is effective only at energies larger than the superconducting gap ($E > \Delta$). This selective transfer of "hot" carriers

leads to the lowering of the effective electron temperature of the N electrode, even in the regime when electrons are thermally decoupled from the lattice. This situation can be experimentally realized with SN *tunnel* junctions where transport is dominated by quasi-particle dynamics. This unique property of such contacts was successfully employed for the realization of microcoolers [73, 74].

During the past few years, the interplay between Andreev scattering and the ferromagnetic order has been addressed in numerous studies of electronic transport properties in nanostructures containing both ferromagnets and superconductors. The interest comes from the fact that the electron and the hole involved in the Andreev process must belong to opposite spin band. This produces a suppression of the current flowing through a FS interface, which depends on the value of the exchange field characteristic of the ferromagnet. This phenomenon has been used for estimating the spin-polarization of different kind of ferromagnets [76, 77].

3. COULOMB BLOCKADE AND SUPERCONDUCTIVITY

The astonishing progress in modern technology has made it possible to fabricate in a controlled way tunnel junctions with capacitances of the order of $C = 10^{-15}$F and below. In this case, even the charging energy associated with a single-electron accumulation at the junction, $E_C \equiv e^2/2C$, can be of the order of several Kelvins. This implies that the properties in the sub-Kelvin regime are strongly affected by the presence of charging effects. Indeed, charging energy allows the control of the electron number of small islands with precision e or $2e$ in the case of superconducting devices. Adding one electron to a small superconducting island puts it into an excited state with an energy exceeding the gap. Only when a second electron is added can both recombine to form a Cooper pair. This leads to *parity effects* (i.e., to different properties of the superconducting systems depending on the number of electrons being even or odd).

The relevant mechanisms for charge transport in superconducting single-electron devices are quasi-particle tunneling, two-particle tunneling, and coherent tunneling of Cooper pairs. All of them are strongly affected by charging. Tunneling of quasi-particles is very similar to the case of normal metals but take into account the modified BCS density of states. Two-particle tunneling is a higher order process related to the Andreev reflection discussed previously. Both two-particle tunneling and coherent oscillations of Cooper pairs dominate transport at very low voltages. Moreover, the interplay of charging and coherent tunneling has a deep significance because it leads to the possibility of observing macroscopic quantum dynamics in these systems. The charge and the phase difference in a Josephson junction, although being macroscopic degrees of freedom, are quantum mechanical conjugate variables. Therefore, the eigenstates are in general superpositions of different charge states. In addition, the combination of coherent Cooper pair tunneling and quasi-particle tunneling leads to a variety of structures in the I–V characteristic.

There are numerous examples of systems where single-electron properties and superconductivity have been studied theoretically and realized experimentally. A prototype example that we will analyze is the normal-superconductor-normal (NSN) transistor where a superconducting island is connected to two normal conducting leads by means of tunnel junctions. This example allows to discuss both quasi-particle and two-particle tunneling. We then discuss some basic properties of the coherent oscillations of Cooper pairs, which will be important for the discussion on the implementation of quantum computers by means of superconducting nanocircuits.

3.1. Tunneling Rates and Parity Effects

As prototype examples to discuss charging effects in superconducting nanostructures, we consider the single-electron (SET) transistor shown in Fig. 9a and the single-electron box of Fig. 9b. Here we concentrate on the transistor, whereas the box will be analyzed in the section devoted to quantum computation. The charging energy of the SET transistor depends on the electron number in the central island and on the applied voltages. The central island is coupled via two tunnel junctions to a transport voltage source, $V = V_L - V_R$, so that a current

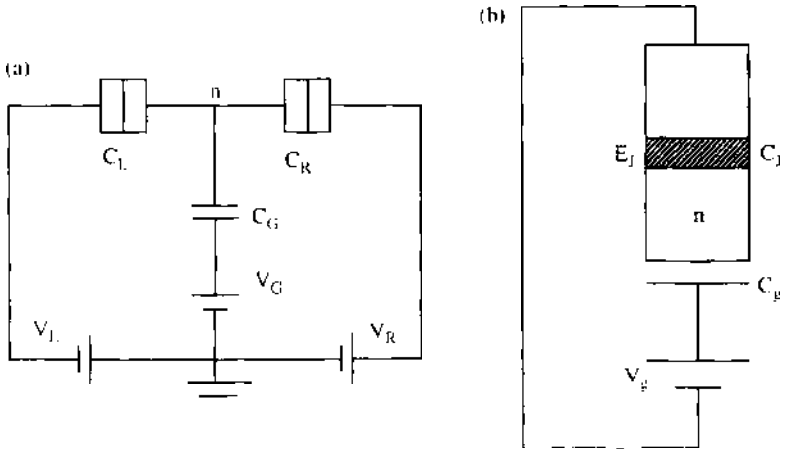


Figure 9. (a) The SET transistor. (b) The Cooper-pair box.

can flow. The island is, furthermore, coupled capacitively to a gate voltage V_G. The charging energy of the system depends on the integer number of excess electrons $n = \pm 1, \pm 2, \ldots$ on the island and on the continuously varied voltages. Elementary electrostatics yields the "charging energy"

$$E_{ch}(n, Q_G) = \frac{(ne - Q_G)^2}{2C} \tag{119}$$

Here, $C = C_L + C_R + C_G$ is the total capacitance of the island. The effect of the voltage sources is contained in the gate-charge $Q_G = C_G V_G + C_L V_L + C_R V_R$. Similar expressions hold for the single-electron box.

In a tunneling process that changes the number of excess electrons in the island from n to $n + 1$, the charging energy changes. Tunneling from the left junction to the island is possible at low temperatures only if the energy in the left-lead, eV_L, is high enough to compensate for the increase in charging energy $eV_L > E_{ch}(n + 1, Q_G) - E_{ch}(n, Q_G)$. Similarly, tunneling from the island (transition from $n + 1$ to n) to the right-lead is possible at low temperatures only if $E_{ch}(n + 1, Q_G) - E_{ch}(n, Q_G) > eV_R$. Both conditions have to be satisfied simultaneously for a current to flow through the transistor. If this is not the case, the current is exponentially suppressed, realizing the so-called *Coulomb blockade*. Varying the gate voltage produces Coulomb oscillations, (i.e. an e-periodic dependence of the conductance on Q_G). To understand further the characteristic of a SET transistor, we need to determine the tunneling rates associated to electron tunneling. An electron tunnels from one of the states k in the left-lead into one of the available states q in the island, thereby changing the electron number from n to $n + 1$, with rate γ_{L1}. Such a rate, which is calculated by means of the Fermi golden rule, can be expressed in a transparent way as

$$\gamma_{L1}(n, Q_G) = \frac{1}{e} I_{qp}\left(\frac{\delta E_{ch}}{e}\right) \frac{1}{\exp(\delta E_{ch}/k_B T) - 1} \tag{120}$$

The function $I_{qp}(V)$ is the quasi-particle tunneling characteristic (see, e.g., Ref. [1]), which is suppressed at voltages below the superconducting gap(s), and δE_{ch} is the charging energy difference between the states with n and $n+1$ electrons in the island. Charging effects reduce the quasi-particle tunneling further. At low temperatures, $k_B T \ll |\delta E_{ch}|$, a tunneling process that would increase the charging energy is suppressed. This phenomenon is called *Coulomb Blockade* of electron tunneling. At zero temperature, the rate is nonzero only if the gain in charging energy compensates the energy needed to create the excitations $\delta E_{ch} + 2\Delta \leq 0$.

The rates describe the stochastic time evolution of the charge of the junction system. For the theoretical analysis, a master equation approach can be used. Several examples of the current–voltage characteristics of normal metal can be found in Refs. [7, 8]; an important characteristic is the e-periodic dependence of the current and conductance on the applied gate charge.

As dictated by Eq. (120), the rate for quasi-particle tunneling are exponentially suppressed below the superconducting gap. In this case, higher order processes involving multi-electron tunneling can play a role. Indeed, for NS interfaces, there still exists a two-electron tunneling process, in many respects similar to the Andreev reflection discussed in the previous section and for this reason denoted as Andreev tunneling. The rate for Andreev tunneling taking into account charging effects is discussed in Ref. [78]. Andreev tunneling is a second-order coherent process. In the first part of the transition, one electron is transferred from an initial state (e.g., $k\uparrow$ of the normal lead) into an intermediate excited state $q\uparrow$ of the superconducting island. In the second part of the coherent transition, an electron tunnels from $k'\downarrow$ into the partner state $-q\downarrow$ of the first electron, such that both form a Cooper pair. The final state contains two excitations in the normal lead and an extra Cooper pair in the superconducting island, and the rates reads:

$$\gamma_{11}^{A}(n, Q_G) = \frac{G^{A}}{4e^2} \frac{\delta E_{ch,2}}{\exp(\delta E_{ch,2}/k_B T) - 1} \tag{121}$$

Note that the functional dependence of this rate coincides with that for single-electron tunneling in a normal junction, Eq. (120), with a linear I–V characteristics. Hence, Andreev reflection is subject to Coulomb blockade like normal-state single-electron tunneling [79] with the exception that the charge transferred in an Andreev reflection is $2e$, and the charging energy changes accordingly. The effective Andreev conductance is of second order in the tunneling conductance and, as shown in Ref. [80], is sensitive to spatial correlations in the normal metal, which can be expressed by the Cooperon propagator.

In a normal-metal electron box (see Fig. 9b), by sweeping the applied gate voltage, the number of electrons on the island in unit of e and, as a consequence, the voltage of the island shows a periodic sawtooth behavior [81]. The periodicity in the gate charge Q_G is e. If the island is superconducting and the gap Δ smaller than the charging energy E_C, the charge and the voltage show at low temperatures a characteristic long-short cyclic, $2e$-periodic dependence on Q_G. This effect arises because single-electron tunneling from the ground state, where all electrons near the Fermi surface of the superconducting island are paired, leads to a state where one extra electron—the "odd" one—is in an excited state [82]. In a small island, as long as charging effects prevent further tunneling, the odd electron does not find another excitation for recombination. Hence, the energy of this state stays above that of the equivalent normal system by the gap energy. Only at larger gate voltages can another electron enter the island, and the system can relax to the ground state. This scenario repeats with periodicity $2e$ in Q_G, as displayed in Fig. 10.

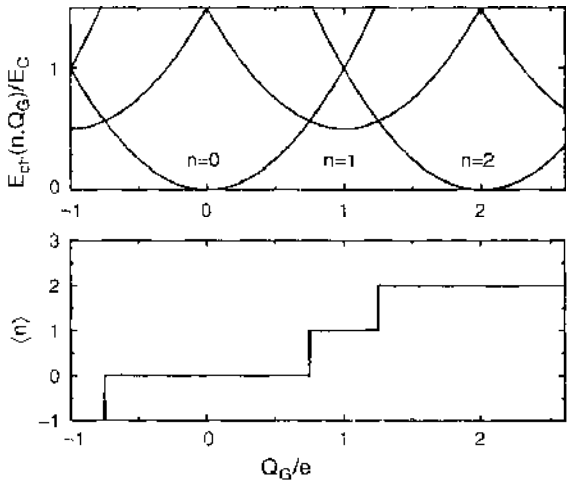


Figure 10. The charging energy of a superconducting single-electron box as a function of the gate voltage shows a difference between even and odd numbers n of electron charges on the island. Accordingly, the average island charge $\langle n \rangle$ is found in a broader range of gate voltages in the even state than in the odd state.

Parity effects appear at temperatures lower than a crossover temperature $T_{\rm eff}$, which, in the experiments, is typically of the order of 10–30% of T_c. At low temperatures, the even–odd asymmetry has been observed in the single-electron box [83] as well as in the I–V characteristics of superconducting SET transistor [84–86]. However, at higher temperatures, above a crossover value $T_{\rm cr} \ll \Delta$, the e-periodic behavior typical for normal-metal systems is recovered.

3.2. NSN Transistor

The analysis sketched above can be used to analyze the I–V characteristics of SET transistors with a superconducting island, the so-called NSN transistor. The interplay of parity effects and Andreev tunneling makes the current–voltage characteristics quite rich. If the energy gap is smaller than the charging energy, the important processes are single-electron tunneling processes in the left and right junctions, causing transitions between even and odd states. In the opposite limit of a superconducting gap bigger than the charging energy, the odd states have a large energy and Andreev tunneling, with rate given by Eq. (121), becomes important [78].

At low temperatures, and superconducting gap larger than the charging energy, a set of parabolic current peaks is found centered around the degeneracy points $Q_G = \pm e, \pm 3e, \ldots$ [78]

$$I^{\wedge}(\delta Q_G, V) = G^{\wedge}\left(V - 4\frac{\delta Q_G^2}{VC^2}\right)\Theta\left(V - 4\frac{\delta Q_G^2}{VC^2}\right) \tag{122}$$

Here, $\delta Q_G = Q_G - e$ for Q_G close to e, and similarly near the other degeneracy points. At larger transport voltages, single-electron tunneling sets in, even if $\Delta > E_C$, and Andreev reflection gets "poisoned" [78]. This occurs for

$$V \geq V_{\rm poison} = \frac{2}{e}\left(E_C - \frac{eQ_G}{C} + \Delta\right) \tag{123}$$

Figure 11 shows the current–voltage characteristic of a NSN transistor with $\Delta > E_C$ [87]. At small transport voltages, the $2e$-periodic peaks due to Andreev reflection dominate; they get poisoned above a threshold voltage. The peaks at larger transport voltages arise from a combination of single-electron tunneling and Andreev reflection. The shape and size of the even–even Andreev peaks and some of the single-electron tunneling features at higher transport voltages agree well with those observed in the experiments of Hergenrother et al. [86].

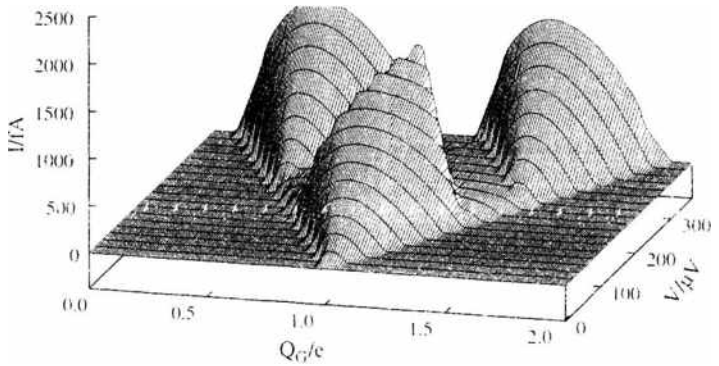


Figure 11. The current $I(Q_G, V)$ through a NSN transistor with $\Delta > E_c$. The parameters correspond to those of the experiments [86]. $E_c = 100\ \mu eV$, $\Delta = 245\ \mu eV$, $R_{1,R} = 43\ k\Omega$, $1/G^{\wedge} \approx 1.2(2.4)10^{9}\ \Omega$ for the left (right) junction. With permission from [87], G. Schön et al., in "Mesoscopic Superconductivity," (F. W. J. Hekking, G. Schön, and D. V. Averin, Eds.). Proceedings of the NATO ARW, p. 340 *Physica B* 203 (1994)

3.3. Cooper Pair Tunneling

In large Josephson junctions, Cooper pair tunneling leads to dissipationless transport between the superconducting electrodes. The coupling is described by the Josephson energy $-E_J \cos\varphi$, with φ being the phase difference between the superconducting order parameters of the two electrodes. The energy scale $E_J = \hbar I_{cr}/2e$ is related to the critical current of the junction I_{cr}, which in turn can be expressed by the tunneling resistance of the junction R_t and the energy gap of the superconductor Δ: $I_{cr}(T = 0) = \pi\Delta/(2eR_t)$.

Charging effects introduce quantum dynamics: The phase difference φ and the charge on the electrodes, Q, are quantum mechanical conjugate variables. An ideal Josephson junction is governed by the Hamiltonian

$$H_0 = \frac{Q^2}{2C} - E_J \cos\varphi \tag{124}$$

The interplay between the phase φ of an island and number of charge carriers Q on it plays a crucial role. Together they determine the properties of small Josephson junctions. The competition between these two canonically conjugated variables is captured by the following Heisenberg relation:

$$[\varphi, Q] = 2ei$$

This relation indicates that suppression of the fluctuations of one of the two variables leads to enhanced fluctuations in the other. The magnitude of the fluctuations is controlled by the E_C/E_J ratio. An important question, addressed in Refs. [88–91], is how to account for dissipation due to the flow of normal currents and/or quasi-particle tunneling. The so-called "macroscopic quantum effects" like macroscopic quantum tunneling of the phase or quantum coherent oscillations are derived from the Hamiltonian (124). Macroscopic quantum tunneling has been observed in tunnel junctions with small capacitances of the order 10^{-12} F. These values are still orders of magnitude too large for single-electron effects to play a role.

An elegant illustration of this competition is presented by what became known as the Heisenberg transistor [92, 93]. The aim of the experiment was to control and measure the quantum phase fluctuations through a modulation of the critical current of the system. In Fig. 12, the layout of the device is shown. Two junctions in series (indicated by crosses) are connected to a current source. The junction parameters are such that $E_J \sim E_C$ (i.e., quantum mechanical fluctuations of the number of Cooper pairs and of the phase of the central island are comparable). A large superconducting reservoir is coupled to the island through a superconducting quantum interference device (SQUID). In the experiment, the critical current was measured as a function of the applied flux through the SQUID ring. It shows a periodic modulation with a period equal to the superconducting flux quantum ($\Phi_0 = h/2e$). The role of the SQUID is to provide a tunable coupling to the reservoir of Cooper pairs. When the flux equals an integer times half a flux quantum ($n\Phi_0/2$), the coupling is turned off and fluctuations in the number of Cooper pairs are suppressed. At the same time, phase fluctuations reach their maximum as indicated by the Heisenberg relation. At fields equal to zero or an integer number of flux quanta, the coupling is maximum so that the amount of charge fluctuations reaches a maximum as well. In the experiment, the scale of charge fluctuations was probed by a measure of the critical current; large charge fluctuations would correspond to favorable Cooper-pair tunneling and a high critical current. Thus, for zero applied field, a high critical current is measured because charge fluctuations are at their maximum. At half a flux quantum applied, the critical current reaches its minimum value because phase fluctuations are at their maximum.

We now turn to mesoscopic Josephson junctions or junction systems, where the number of electrons or Cooper pairs in small islands is the relevant degree of freedom. The charging energy has been discussed above. The Josephson coupling describes the transfer of Cooper-pair charges forward or backward, and can be written in a basis of charge states as

$$\langle n|E_J \cos\varphi|n'\rangle = \frac{E_J}{2}(\delta_{n',n+2} + \delta_{n',n-2}) \tag{125}$$

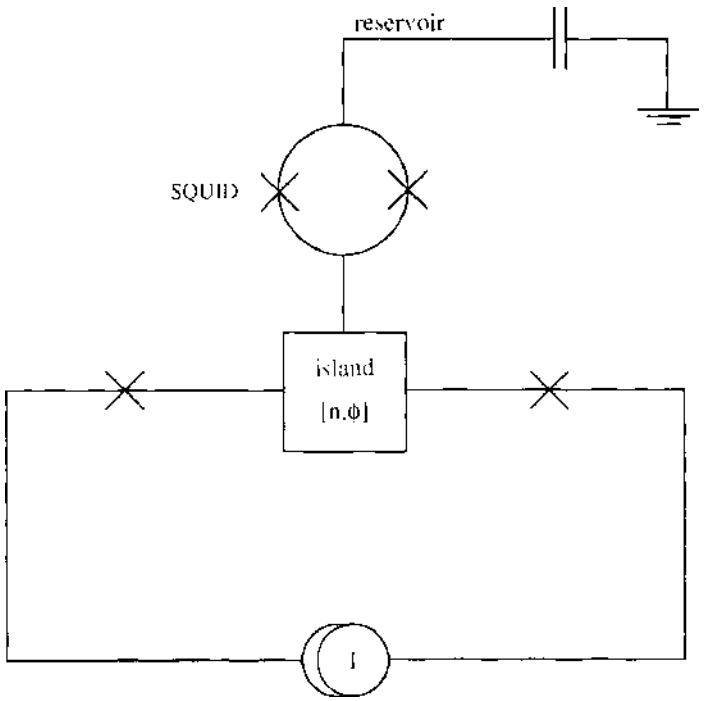


Figure 12. A schematic drawing of the Heisenberg transistor. The phase and the charge on the island are quantum mechanical conjugated variables. By varying the flux through the SQUID ring, the effective Josephson energy is tuned and, as a result, phase fluctuations on the central island can be varied.

We first consider an electron box (see Fig. 9b) with superconducting island and lead with large energy gap at low temperatures, $\Delta > E_C \gg k_B T$. In this case, at low voltages, quasi-particle tunneling is suppressed, and the island charge can change only by Cooper-pair tunneling in units of $2e$ as described by Eq. (125). The tunneling is strong near points of degeneracy. For instance, for $Q_G \approx e$, the charging energies of the states with $n = 0$ and $n = 2$ are comparable, and we can restrict our attention to these two charge states. The coherent tunneling between both is described by the 2×2 Hamiltonian

$$H = \begin{pmatrix} E_{ch}(0) & -\dfrac{E_J}{2} \\ -\dfrac{E_J}{2} & E_{ch}(2) \end{pmatrix} \tag{126}$$

This Hamiltonian is easily diagonalized: the eigenstates and eigenenergies are

$$\begin{aligned} &\psi_0 = \alpha|0\rangle + \beta|2\rangle, \qquad \psi_1 = \beta|0\rangle - \alpha|2\rangle \\ &\alpha^2 = \frac{1}{2}\left[1 + \frac{\delta E_{ch}}{\sqrt{\delta E_{ch}^2 + E_J^2}}\right] = 1 - \beta^2 \\ &E_{0,1} = \frac{1}{2}\left[E_{ch}(0) + E_{ch}(2) \mp \sqrt{\delta E_{ch}^2 + E_J^2}\right] \end{aligned} \tag{127}$$

Here, we have introduced the difference in charging energy $\delta E_{ch} \equiv E_{ch}(2) - E_{ch}(0) = 4E_C(Q_G/e - 1)$. The coefficient α approaches unity if the charging energy of the state $|2\rangle$ lies far above that of $|0\rangle$ (i.e. for $\delta E_{ch} \gg 0$) and vanishes in the opposite limit.

The expectation value of the charge on the island in the ground state is given by

$$\langle \psi_0 | n | \psi_0 \rangle = 2\beta^2 \tag{128}$$

It changes near $Q_G = e$ from 0 to 2 over a width of order $\delta Q_G \approx E_J/E_C$. This has been confirmed experimentally in Ref. [94].

The combination of coherent Cooper-pair tunneling and dissipative quasi-particle tunneling leads to a dissipative I–V characteristic of superconductor-superconductor-superconductor (SSS) transistors [84, 95–99]. We consider the case of an SSS-SET transistor with superconducting electrodes and island below the characteristic temperature for observing parity effects. The combination of coherent Cooper pair tunneling and single-electron tunneling leads to a dissipative I–V characteristic. Results are shown in Fig. 13 with parameters corresponding to those of experiments reported in Ref. [84]. We note that the I–V characteristic is $2e$-periodic and shows a rich structure deep in the subgap region. For transport voltage: $eV \gtrsim 2.5E_C$, the $2e$-periodic features disappear, and the current becomes e-periodic in Q_G again.

A very important example where the interplay between coherent tunneling of Cooper pairs and charging effect is the the gate-voltage dependence of the critical current of SSS or superconductor-normal-superconductor (SNS) transistors [100, 101]. As discussed for the Heisenberg tansistor, the critical current can be modulated by means of a gate that favors (or suppresses) charge fluctuations in the central island.

There are many other interesting questions that we could not touch; we mention for example the role of single-electron effects in electrical noise [102] or the question of quantum pumping [103].

3.4. Superconducting Nanoparticles

In the previous section, we saw how the superconducting properties can be modified by the presence of charging effects. In this situation, the hallmarks characteristics of superconductivity in bulk samples (i.e., zero electrical resistance and Meissner effect) are not visible. Dissipationless current is masked by Coulomb blockade and the dimensions of the samples are small compared to the penetration depth so that the field penetrates over the whole superconducting particle. Nevertheless, we can still have a well defined superconducting gap. In nanosized particles, the situation becomes radically different. For samples of few nanometers in diameter, the level spacing can become of the same order of the BCS gap, and superconductivity disappears [104]. These systems have been investigated in a series of experiments by Ralph, Black, and Tinkham (RBT) [105] on small dots (linear size $\simeq 10$ nm down to 2 nm) at low temperatures. RBT have shown in larger grains the existence of a spectroscopic gap (defined as the energy difference between the first excited state and the ground state), which is driven to zero by applying a suitable magnetic field. Reducing the size of the samples down to ~3 nm, no trace of the gap is found. In these samples, the suppression of the spectroscopic gap is due to the spin paramagnetism (the Zeeman effect competing with the pairing interaction), as the orbital effects are negligible.

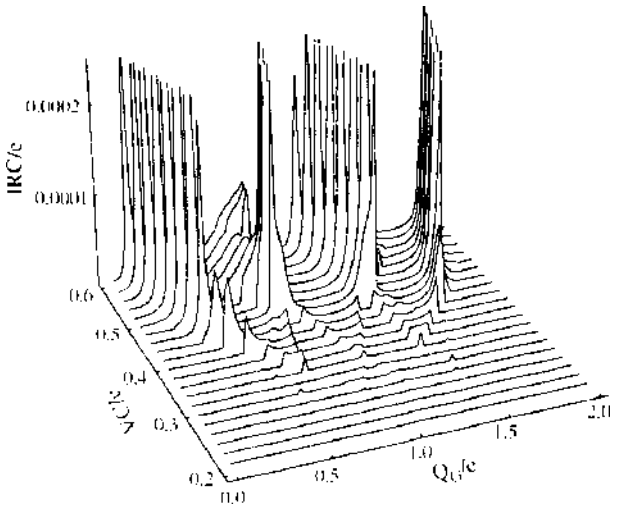


Figure 13. I–V characteristic of an SSS transistor. The parameters are $\Delta = 1.3E_C$, $E_J = 0.17E_C$, $R_1 = R = R_K$, $\gamma = 2.5 \cdot 10^{-5}(RC^{-1})$. Reprinted with permission from [98], J. Siewert and G. Schön, *Phys. Rev. B* 54, 7421 (1996). © 1996, American Physical Society.

The experiments by RBT posed a number of important questions related to the disappearance of superconductivity in nanoparticles and on the definition of superconductivity in this case (where the canonical ensemble should be used). As we already discussed, the properties of an isolated, mesoscopic superconducting grain are quite different from those of a bulk sample. First of all, because such a grain carries a fixed number N of electrons, its behavior depends strongly on whether N is even or odd. Second, fluctuation effects become important as the size of the grain decreases. The interplay between parity and fluctuation effects crucially depends on the ratio δ/Δ of two characteristic energies: the mean level spacing δ and the bulk superconducting gap Δ. As long as the grain is not too small, $\delta \ll \Delta$, the fluctuation region ΔT around the critical temperature T_c is narrow, $\Delta T/T_c \sim \sqrt{\delta/\Delta} \ll 1$, and the mean field description of superconductivity is appropriate.

The Hamiltonian usually adopted [106] for these systems is:

$$H = \sum_{n=1,\sigma=\pm}^{\Omega} (\epsilon_n - \sigma\mu_B H) c^{\dagger}_{n,\sigma} c_{n,\sigma} - g \sum_{m,n=1}^{\Omega} c^{\dagger}_{m,+} c^{\dagger}_{m,-} c_{n,-} c_{n,+} \tag{129}$$

The indexes m and n label the single particle energy levels with energy ϵ_m and annihilation operator $c_{m,\sigma}$. The energy levels ϵ_m fluctuate according to the Gaussian orthogonal ensemble, which describes systems with rotational invariance and time-reversal symmetry. (The time-reversal symmetry is preserved in the case of a vanishing magnetic field). The quantum number $\sigma = \pm$ labels time-reversed electron states. The external magnetic field is H and μ_B the Bohr magneton, and the external field contribution is due only to the Zeeman term, as orbital effects are negligible. The number of interacting levels is Ω, which is twice the Debye frequency ω_D in units of the mean level spacing $\delta \sim 1/N(0)V$ [$N(0)$ is the density of states at the Fermi energy and V the volume of the grain]. Finally, $g = \lambda\delta$ is the BCS coupling constant, λ being a dimensionless coupling, given by the BCS energy scale $\Delta = \Omega\delta/[2\sinh(\delta/\lambda)]$.

The finite level spacing suppresses the BCS gap in a parity-dependent way [107, 108]. When δ becomes of the order Δ, $\Delta T \sim T_c$, and the BCS description of superconductivity breaks down even at zero temperature [104]. The regime $\delta \gtrsim \Delta$ is dominated by strong pairing fluctuations [109–113]. Clear signatures of superconductivity can be detected in the parity gap proposed by Matveev and Larkin [110] and defined as

$$\Delta_p = E^{(0)}_{2N+1} - \frac{1}{2}\left(E^{(0)}_{2N+2} + E^{(0)}_{2N}\right)$$

where $E^{(0)}_N$ is the ground state energy of a dot with N electrons.

The presence of the pairing correlation has further been characterized by studying the behavior of some thermodynamical observables, such as the specific heat [114], the spin susceptibility [114, 115], and field dependent magnetization [116].

3.5. Josephson Arrays

The ability to assemble in ordered networks Josephson junctions, described in Section 3.3, has opened the way toward the study of collective phenomena in artificially fabricated systems. The first artificially fabricated Josephson-junctions arrays (JJAs) were realized 20 years ago at IBM as part of their effort to develop an electronics based on superconducting devices. At the beginning, Josephson arrays were intensively studied to explore a wealth of classical phenomena [117–119]. JJAs proved to be an ideal model system in which classical phase transitions, frustration effects, classical vortex dynamics, nonlinear dynamics, and chaos could be studied in a controlled way. All classical phenomena have been explained by studying the classical (thermo)dynamics of the phases of the superconducting order parameter on each island. Under these conditions, classical JJAs are a physical realization of the two-dimensional XY-model and, above a transition temperature, phase fluctuations destroy the global phase coherence preventing the system to reach the superconducting state. Global

phase coherence is only restored below temperatures corresponding to the Josephson coupling energy E_J, which is the energy scale associated with Cooper pair tunneling between neighboring islands.

Quantum JJAs already allow investigation of quantum (zero-temperature) phase transitions [120, 121]. In recent years, this field of research has attracted the attention of many physicists. Experiments on thin, superconducting films, high-temperature superconductors, spin systems, and two-dimensional electron gases have all shown the existence of quantum critical points. In arrays made of submicrometer junctions, the quantum fluctuations drive the system through a variety of quantum phase transitions. A quantum JJA may be insulating at zero temperature even though each island is still superconducting. In the classical limit $E_J \gg E_C$, the system turns superconducting at low temperatures, as the fluctuations of the phases are weak and the system is globally phase coherent. In the opposite limit, $E_J \ll E_C$, the array becomes a Mott insulator, as the charges on each island are localized, and an activation energy of the order E_C is required to transport charges through the system (Coulomb blockade of Cooper pairs). Strong quantum fluctuations of the phases prevent the system from reaching long-range phase coherence in this regime.

Granular superconducting thin films are closely related to arrays. In granular films, superconducting islands of various sizes and with various coupling energies are connected together. Therefore, disorder plays a crucial role in these granular materials, whereas it is virtually absent in JJAs (in fact, it can be introduced in a controlled way).

Another important field of investigation addressed with JJAs is the study of the quantum dynamics of macroscopic objects. In the classical limit ($E_J \gg E_C$), vortices are the topological excitations that determine the (thermo)dynamic properties of JJAs. In the opposite situation ($E_J \ll E_C$), the charges on each island are the relevant degrees of freedom. Vortices and charges play a dual role, and many features of JJAs can be observed in the two limits if the role of charges and vortices are interchanged. By fabricating arrays with different geometries, vortices can be manipulated to a great extent. Quantum dynamics of macroscopic objects requires knowledge of the coupling to the surrounding environment. To a certain degree, the dissipative environment can be modeled, and therefore JJAs are prototype systems to study macroscopic quantum mechanics as well.

A detailed account of the research work done on quantum Josephson arrays can be found in Ref. [13].

The relevant physics is captured by a model, frequently defined as the quantum phase model (QPM), characterized by the following Hamiltonian

$$H_{\mathrm{QPM}} = \sum_{i,j} (q_i - q_x) U_{ij} (q_j - q_x) - E_J \sum_{\langle i,j \rangle} \cos(\phi_i - \phi_j - A_{ij}) \tag{130}$$

where i, j are array indexes, and q_i and ϕ_i are, respectively, charge and phase of the superconducting island i. The Coulomb interaction is described by the matrix $U_{ij} = e^2 C_{ij}^{-1}$. The simplest, sufficiently realistic, model for the capacitance matrix C_{ij} includes only the ground capacitance C_0 (accounting for the capacitance between the island and the ground) and the junction capacitance C, with the corresponding energy scales $E_C = e^2/2C$ and $E_0 = e^2/2C_0$. The range of the electrostatic interaction between Cooper pairs is, in units of the lattice spacing, $\lambda = \sqrt{C/C_0}$. A control (external) voltage V_x applied to the ground plane enters via the induced charge $Q_x = 2eq_x = \sum_j C_{ij} V_x$. When tuning V_x, different charge configurations minimize the electrostatic energy. It suppresses tunneling (Coulomb blockade) except at degeneracy points. A perpendicular magnetic field with vector potential $\mathbf{A}$ enters the QPM in the standard way through $A_{ij} = 2e \int_i^j \mathbf{A} \cdot d\mathbf{l}$. The relevant parameter that describes the magnetic frustration is $f = (1/2\pi) \sum A_{ij}$, where the summation runs over an elementary plaquette.

The QPM accounts only for Cooper pair tunneling; in some cases, however, one has to take into account the tunneling of quasiparticles and/or the flow of Ohmic current through the substrate or between the junctions. In the case of strong *on-site* Coulomb interaction $U_{ii} = U_0$ and very low temperatures, only few charge states are important. If the gate voltage is tuned close to a degeneracy, the relevant physics is captured by considering only two

adjacent charge states of each island, and the QPM is equivalent to an anisotropic XXZ spin-1/2 Heisenberg model [122]:

$$H_S = -h\sum_i S_i^z + \sum_{i,j} S_i^z U_{ij} S_j^z - E_J \sum_{\langle i,j\rangle} (e^{iA_{ij}} S_i^+ S_j^- + e^{-iA_{ij}} S_j^- S_i^+) \tag{131}$$

The operators S_i^z, S_i^+, S_j^- are the spin-1/2 operators, S_i^z being related to the charge on each island ($q_i = S_i^z + 1/2$), and the raising and lowering $S_i^{\pm}$ operators corresponding to the "creation" and "annihilation" operators $e^{\pm i\phi_i}$ of the QPM. The "external" field h is related to the external charge by $h = (q_x - 1/2)\sum_j U_{ij}$. Various magnetic ordered phases of the XXZ Hamiltonian correspond to the different phases in the QPM. Long-range order in $\langle S^-\rangle$ indicates superfluidity in the QPM, whereas long-range order in $\langle S^z\rangle$ describes order in the charge configuration. There is yet another closely related model that is mostly used in the context of superconductivity in ultrathin films, the Bose–Hubbard model [123]

$$H = \frac{1}{2}\sum_i n_i U_{ij} n_i - \mu \sum_i n_i - \frac{t}{2}\sum_{\langle ij\rangle} (b_i^\dagger b_j + h.c.) \tag{132}$$

where $b^\dagger$, b are the creation and annihilation operators for bosons, and $n_i = b_i^\dagger b_i$ is the number of bosons. Again, U_{ij} describes the Coulomb interaction between bosons, μ is the chemical potential, and t the hopping matrix element. The connection between the Bose–Hubbard model and the QPM is easily seen by writing the field b_i in terms of amplitude and phase and then approximating the amplitude by its average (i.e., $b_i \sim e^{i\phi_i}$). The hopping term is then associated with the Josephson tunneling while the chemical potential plays the same role as the external charge in the QPM. This mapping becomes more accurate as the average number of bosons per sites increases.

The three models are equivalent in the sense that they belong to the same universality class (they lead to the same Ginzburg–Landau effective free energy). However, the nonuniversal features like the location of the phase transitions depend quantitatively on the specific choice of the model.

A simple way to establish various properties of the phase diagram is to employ the mean field approximation, which consists in approximating the Hamiltonian of Eq.(130) by [124]

$$H_{\rm MF} = \frac{1}{2}\sum_{i,j} q_i U_{ij} q_i - zE_J\langle\cos(\phi)\rangle \sum_j \cos(\phi_i)$$

where z is the coordination number in the lattice and $\psi \equiv \langle\cos(\phi)\rangle$ is the order parameter. This has to be calculated self-consistently according to

$$\langle\cos(\phi)\rangle = {\rm Tr}\{\cos(\phi_i)\exp(-\beta H_{\rm MF})\}/{\rm Tr}\{\exp(-\beta H_{\rm MF})\}$$

Close to the transition point, the thermal average on the right-hand side can be evaluated by expanding in powers of ψ. To third order, a Ginzburg–Landau type equation arises:

$$\left[1 - zE_J\int_0^\beta d\tau \langle\cos\phi_i(\tau)\cos\phi_i(0)\rangle_{ch}\right]\psi + \left(\frac{zE_J}{U_0}\right)^3 f_3\psi^3 = 0 \tag{133}$$

Here, the average $\langle\cdots\rangle_{ch}$ is performed over the Eigenstates of the charging part of the Hamiltonian only, and the quantity f_3 entails the four-point phase correlation. If the charging term is absent, the phase-phase correlator in Eq. (133) is equal to one, and we recover the classical result $\beta_c zE_J = 2$. Due to the charging effects, the phase starts to fluctuate and the critical temperature is depressed. The correlator is easy to evaluate. For instance, in the self-charging limit $U_{ij} = U_0\delta_{ij}$ at $T = 0$, it is ($U_0 \equiv 8E_0$ only if the junction capacitance is zero)

$$\langle\cos\phi_i(\tau)\cos\phi_i(0)\rangle_{ch} = \left(\frac{1}{2}\right)\exp\left\{-\left(\frac{U_0}{2}\right)\tau(1-\tau/\beta)\right\}$$

As a result, the superconductor/insulator transition at zero temperature occurs at

$$zE_J = U_0$$

For larger values of the charging energy, the array does not acquire phase coherence even at zero temperature.

Similar types of mean field approaches can be used to study the effect of frustration in these systems.

3.6. Quantum Computation with Superconducting Nanocircuits

Quantum computation (QC) [125] has recently excited many scientists from various different areas of physics. The increasing interest in this field is certainly related to the fact that some problems that are intractable with classical algorithms can be solved much faster with QC. Factorization of large numbers, as proposed by Shor, is probably the best known example in this respect. The elementary unit of any quantum information process is the *qubit*. The two values of the classical bit are replaced by the ground state ($|0\rangle$) and excited ($|1\rangle$) state of a two-level system (it is common to adopt the spin-1/2 language as we will do here). While information is stored either in 0 or in 1 in a classical bit, any state $|\psi(t)\rangle = a(t)|0\rangle + b(t)|1\rangle$ can be used as a qubit. Controlled evolution between the two states $|0\rangle$ and $|1\rangle$ is obtained by applying resonant external radiation to the system or by a fast DC pulse of high amplitude. By choosing appropriate pulse widths, the NOT operation

$$\begin{aligned} |0\rangle &\longrightarrow |1\rangle \\ |1\rangle &\longrightarrow |0\rangle \end{aligned} \tag{134}$$

or the Hadamard transformation

$$\begin{aligned} |0\rangle &\longrightarrow \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |1\rangle &\longrightarrow \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned} \tag{135}$$

can be established. These unitary operations alone do not make a quantum computer yet. Together with one-bit operations, it is of fundamental importance to perform two-bit quantum operations (i.e., to control the unitary evolution of entangled states). One example of a two-qubit gate is the control-NOT operation:

$$\begin{aligned} |00\rangle &\longrightarrow |00\rangle \\ |01\rangle &\longrightarrow |01\rangle \\ |10\rangle &\longrightarrow |11\rangle \\ |11\rangle &\longrightarrow |10\rangle \end{aligned} \tag{136}$$

The unitary single-bit operations and this control-NOT operation are sufficient for performing all tasks of a quantum computer. Therefore, quantum computers can be viewed as programmable quantum interferometers. Initially prepared in a superposition of all the possible input states, the computation evolves in parallel along all its possible paths, which interfere toward the desired output state. It is this intrinsic parallelism in the evolution of quantum systems that allows for exponentially more efficient ways of performing computation.

On the one hand, it is of crucial importance that qubits are protected from the environment (i.e., from any source that could cause decoherence). This is a very difficult task because at the same time, one also has to control the evolution of the qubits, which inevitably means that the qubit is coupled to the environment. On the other hand, large-scale integration (needed to make a quantum computer useful) seems to be impossible. Qubits realized with solid-state devices may offer a great advantage in this respect, because fabrication techniques allow for scalability to a large number of coupled qubits. Josephson qubits are among

the most promising devices to implement solid-state quantum computation [126, 127]. Quantum manipulations of individual [128–134] and coupled [135] qubits has been demonstrated experimentally.

At present, different proposals have been put forward to use superconducting nanocircuits [136–140] for the implementation of quantum algorithms. Depending on the operating regime, they are commonly referred to as charge [136, 137, 140, 141] and flux [138, 139] qubits.

In the case of charge qubits, the building block for quantum computation is realized by the two nearly degenerate charge states of a single electron box as shown in Fig. 9b. They represent the states $|0\rangle$, $|1\rangle$ of the qubit. In the computational Hilbert space, the ideal evolution of the system is governed by the Hamiltonian

$$\mathcal{H} = -\Delta E_{\rm ch}(|0\rangle\langle 0| - |1\rangle\langle 1|) - \frac{E_J}{2}(|0\rangle\langle 1| + |1\rangle\langle 0|) \tag{137}$$

where $\Delta E_{\rm ch} = E_{\rm ch}(n_x - 1/2)$. Any one bit operation can be realized by varying the external charge n_x and, in the proposal of Ref. [136], by varying the Josephson coupling E_J as well. Modulation of E_J is achieved by placing the Cooper-pair box in SQUID geometry. The advantage of this choice is that during idle times, the Hamiltonian can be "switched off" completely eliminating any trivial phase accumulation that should be subtracted for computational purposes.

As discussed before, a quantum computer can be realized once two bit gates are implemented. The Karlsruhe group has proposed an inductive coupling between qubits that lead to a coupling of the type

$$H_c = -E_J \sigma_x^{(1)} \sigma_x^{(2)} \tag{138}$$

This type of coupling is very close in spirit to the coupling used in the ion-trap implementation of QC. The main advantage of this choice is that qubits are coupled via an infinite range coupling and that they can easily be isolated. A different scheme has been proposed in Ref. [137], where the adiabatic aspect of conditional dynamics is emphasized and the use of capacitive coupling between gates is suggested as to reduce unwanted transitions to higher charge states. The coupling reads $H_c = -E_c \sigma_z^{(1)} \sigma_z^{(2)}$, and the qubit is now defined as a finite one-dimensional array, of junctions. By means of gate voltages applied at different places in the array, the qubit–qubit coupling can be modulated in time and a control-NOT can be realized.

A qubit can also be realized with superconducting nanocircuits in the opposite limit: $E_J \gg E_c$. An rf-SQUID (a superconducting loop interrupted by a Josephson junction) provides the prototype of such a device. The Hamiltonian of this system reads

$$\mathcal{H} = -E_J \cos\left(2\pi \frac{\Phi}{\Phi_0}\right) + \frac{(\Phi - \Phi_x)^2}{2L} + \frac{Q^2}{2C} \tag{139}$$

Here, L is the self-inductance of the loop, and the phase difference across the junction $(2\pi\Phi/\Phi_0)$ is related the flux Φ in the loop. The externally applied flux is denoted by Φ_x. The charge Q is canonically conjugated to the flux Φ. In the limit in which the self-inductance is large, the two first terms in the Hamiltonian form a double-well potential near $\Phi = \Phi_0/2$. Also in this case, the Hamiltonian can be reduced to that of a two-state system. The term proportional to σ_z measures the asymmetry of the double well potential, and the off-diagonal matrix elements depend on the tunneling amplitude between the wells. By controlling the applied magnetic field, all elementary unitary operations can be performed.

In order to fulfill various operational requirements, more refined designs should be used. In the proposal of Mooij et al. [138], qubits are formed by three junctions (as in Fig. 14). Flux qubits are coupled by means of flux transformers, which provide inductive coupling between them. Any loop of one qubit can be coupled to any loop of the other, but to turn off this coupling, one would need to have an ideal switch in the flux transformer. This switch is to be controlled by high-frequency pulses, and the related external circuit can lead to decoherence effects.

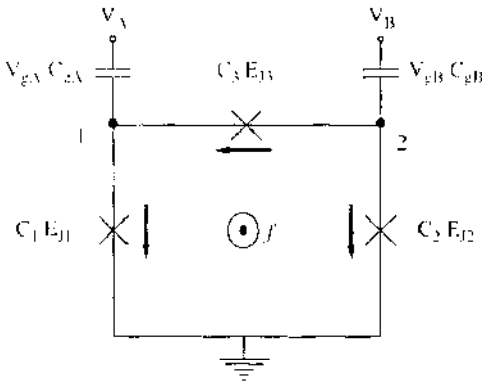


Figure 14. The three-junction flux qubit. Josephson junctions 1 and 2 have the same Josephson energies $E_{J1} = E_{J2} = E_J$ and capacitance $C_1 = C_2 = C$, whereas Josephson junction 3 has a Josephson energy and capacitance that are α times larger. The islands are coupled by gate capacitors $C_{gA} = C_{gB} = \gamma C$ to gate voltages V_A and V_B. The arrows define the direction of the currents.

An intermediate situation in which the charging and Josephson energies are comparable has been very successfully exploited in Ref. [131]; in this case, it is more appropriate to talk about phase qubits.

A high degree of quantum coherence is a crucial requirement for the implementation of quantum logic devices. Decoherence can be a serious limitation, due to the presence of many types low energy excitations in the "internal" environment and of external sources due to the control circuitry. A review of issues related to decoherence in superconducting qubits can be found in Ref. [142].

4. CONCLUSIONS

In this review, we tried to present an overall description of mesoscopic superconductivity. The field is so vast that the choice of the topics was surely biased by our personal taste and by our field of investigations. To summarize, we have started this review by illustrating the transport properties of hybrid systems formed by putting into contact superconductors with normal metals and ferromagnets. In Section 2, we have introduced the Bogoliubov–de Gennes equation in order to discuss transport properties in terms of Andreev processes. We have then provided a derivation of the Bogoliubov-de Gennes equation starting from the BCS theory and gave the solutions of such an equation for three different systems, namely a homogeneous ferromagnet, a homogeneous superconductor, and a ferromagnet/superconductor junction. After that, we have introduced the two most important theoretical approaches to the study of transport properties in hybrid systems. We chose to give a more detailed treatment of the scattering theory, deriving general conductance formulas in terms of scattering matrices and giving explicit expressions for the case of two-probe systems. The quasi-classical Green's function approach has also been discussed, although in less details, and a review of the most important results has been provided. Section 3 has been devoted to the interplay between superconductivity and Coulomb blockade, which is present in small metallic grains where charging effects are important. We have discussed parity effects in single-electron transistors containing a superconducting island and their consequences on the quasi-particle tunneling rates. I–V characteristics have been illustrated. Cooper-pair tunneling has also been addressed in large Josephson junctions where the consequences of quantum mechanical conjugation of phase and charge variables has been considered. Section 3.4 has been devoted to superconducting nanoparticles where the hallmark characteristic of superconductivity of bulk samples are not visible. Parity and fluctuation effects have been illustrated in the presence of a finite level spacing. In Section 3.5, we have discussed Josephson arrays that consist of ordered networks of Josephson junctions. These systems are important because they allow one to investigate a variety of phenomena, such as classical phase transitions, frustration effects, and so forth, in a controlled way. Furthermore, three different models have briefly been reviewed for the study of such systems, and a mean field approximation

has been also introduced. Section 3.6 is devoted to describe how it is possible to implement quantum computation with superconducting nanocircuits. After having introduced the most important notions of the subject, we have discussed how a qubit can be realized with Josephson junctions, either in the charge or flux operating regime.

There are several important topics that were left out; the interested reader may find additional references in the reviews quoted in the "Introduction."

It is quite hard to draw conclusions and even harder to advance ideas about possible future directions of the field. Nevertheless, we take this risk and briefly outline what we believe (remaining conservative) will be the future development of the field. Quantum computation with superconducting nanocircuits is certainly one of the most active areas in this respect. The possibility of building a solid-state computer is very appealing, and the possibility to combine the macroscopic superconducting coherence with the ability of designing nanocircuits will certainly give interesting results also in the future.

A direction that is not fully studied up to now is the role of superconductivity in molecular electronics. Transport through molecules using superconducting electrodes may give additional insight due to the fact that one can perform both single-particle and Cooper pair spectroscopy.

Finally, we would like to mention that all the knowledge obtained in Josephson arrays may be important in the study of condensate in optical lattices. The experimental realization of cold atoms trapped in optical lattices has indeed paved the way for the discovery of a wealth of new phenomena in Bose–Einstein condensates. For example, the superconductor-insulator transition, predicted for optical lattices in Ref. [143], can be described in the framework of the Bose–Hubbard model (or equivalently the quantum phase model) employed also in the study of the properties of Josephson junctions arrays.

ACKNOWLEDGMENTS

We would like to acknowledge, on the topics covered in this review, a very fruitful collaboration with C. Bruder, G. Falci, F. W. J. Hekking, C. J. Lambert, G. M. Palma, R. Raimondi, G. Schön, and J. Siewert. This work was supported by the EU (IST-SQUBIT, HPRN-CT-2002-00144).

REFERENCES

1. M. Tinkham, "Introduction to Superconductivity." McGraw-Hill, New York, 1996.
2. A. Barone and G. Paterno, "Physics and Applications of the Josephson Effect." John Wiley, New York, 1982.
3. C. W. J. Beenakker and H. van Houten, *Solid State Physics* 44, 1 (1991).
4. Y. Imry, "Introduction to Mesoscopic Physics." Oxford University Press, Oxford, 1997.
5. F. W. J. Hekking, G. Schön, and D. V. Averin (Eds.), "Mesoscopic Superconductivity." *Physica B* 203, 201 (1994).
6. L. P. Kouwenhoven, L. L. Sohn and G. Schön (Eds.), "Mesoscopic Electron Transport." Vol. 345, NATO ASI Series, E. Kluwer, Dordrecht, 1997.
7. D. V. Averin and K. K. Likharev, in "Mesoscopic Phenomena in Solids" (B. L. Altshuler, P. A. Lee, and R. A. Webb, Eds.), p. 173. Elsevier, Amsterdam, 1991.
8. H. Grabert and M. H. Devoret (Eds.), "Single Charge Tunneling" Vol. B 294, NATO ASI Series, New York, Plenum Press, 1992.
9. G. Schön and A. D. Zaikin, *Europhys. Lett.* 26, 695 (1994).
10. C. J. Lambert and R. Raimondi, *J. Phys. Condens. Matter* 10, 901 (1998).
11. J. von Delft and D. C. Ralph, *Phys. Rep.* 345, 61 (2001).
12. Yu. Makhlin, G. Schön, and A. Shnirman *Rev. Mod. Phys.* 73, 357 (2000).
13. R. Fazio and H. van der Zant, *Phys. Rep.* 355, 237 (2001).
14. A. F. Andreev, *Sov. Phys. JETP* 19, 1228 (1964)
15. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* 108, 1175 (1957).
16. P. G. de Gennes, "Superconductivity of Metals and Alloys." Addison-Wesley, New York, 1989.
17. N. N. Bogoliubov, V. V. Tolmachev, and D. V. Shirkov, "A New Method in the Theory of Superconductivity." Consultants Bureau, New York, 1959.
18. L. P. Gorkov, *Sov. Phys. JETP* 7, 505 (1958).
19. J. M. Ziman, "Principles of the Theory of Solids." Cambridge University Press, New York, 1964.
20. E. C. Stoner, *Proc. Roy. Soc.* A165, 372 (1938).
21. E. C. Stoner, *Proc. Roy. Soc.* A169, 339 (1939).

22. C. P. Enz, "A Course on Many-body Theory Applied to Solid State Physics," World Scientific, Singapore, 1992.
23. A. Aharoni, "Introduction to the Theory of Ferromagnetism," Oxford University Press, Oxford, U.K., 2001.
24. E. K. U. Gross, E. Runge, and O. Heinonen, "Many-particle Theory," IOP Publishing, Bristol, U.K., 1991.
25. N. N. Bogoliubov, *Nuovo Cimento* 7, 794 (1958).
26. J. G. Valatin, *Nuovo Cimento* 7, 843 (1958).
27. G. E. Blonder, M. Tinkham and T. M. Klapwijk, *Phys. Rev. B* 25, 4515 (1982).
28. M. J. M. de Jong and C. W. J. Beenakker, *Phys. Rev. Lett.* 74, 1657 (1995).
29. R. Landauer, *Philos. Mag.* 21, 863 (1970).
30. M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, *Phys. Rev. B* 31, 6207 (1985).
31. M. Büttiker, *Phys. Rev. Lett.* 57, 1761 (1986).
32. M. Büttiker, *IBM J. Res. Dev.* 32, 317 (1988).
33. We wish to remark that the Bogoliubov–de Gennes equation was applied to SNS Josephson junctions much earlier [34, 35]. However these works cannot be considered to fall within the scattering theory of transport we are concerned with in the present section. The reason is that they deal with equilibrium properties, like the Josephson current, calculated through a thermodynamic approach. Unlike in the Landauer–Büttiker scattering theory, no scattering amplitudes describing a quasi-particle flow were defined.
34. I. O. Kulik, *Zh. Eksp. Teor. Fiz.* 57, 1745 (1969) [*Sov. Phys. JETP* 30, 944 (1970)].
35. J. Bardeen and J. L. Johnson, *Phys. Rev. B* 5, 72 (1972).
36. C. J. Lambert, *J. Phys. Condens. Matter* 3, 6579 (1991).
37. Y. Takane and H. Ebisawa, *J. Phys. Soc. Jpn.* 61, 2858 (1992).
38. C. J. Lambert, *Physica B* 203, 201 (1994).
39. M. Büttiker, *Phys. Rev. B* 33, 3020 (1986).
40. T. Christen and M. Büttiker, *Europhys. Lett.* 35, 523 (1996).
41. A. M. Martin and C. J. Lambert, *Phys. Rev. B* 51, 17999 (1995).
42. J. Sanchez-Canizares and F. Sols, *J. Phys.: Condens. Matter* 7, L317 (1995).
43. A. M. Martin and C. J. Lambert, *J. Phys.: Condens. Matter* 8, L731 (1996).
44. Ya. M. Blanter and M. Büttiker, *Phys. Rep.* 336, 1 (2000).
45. C. J. Lambert, *J. Phys.: Condens. Matter* 5, 707 (1993).
46. A. D. Stone, P. A. Mello, K. A. Muttalib, and J.-L. Pichard, in "Mesoscopic Phenomena in Solids" (B. L. Altshuler, P. A. Lee, and R. A. Webb, Eds.), p. 369, North-Holland, Amsterdam, 1991.
47. C. W. J. Beenakker, *Rev. Mod. Phys.* 69, 731 (1997).
48. C. W. J. Beenakker, *Phys. Rev. B* 46, 12841 (1992).
49. See however a recent paper [50] where the Andreev reflection eigenvalue density was derived.
50. P. Samuelsson, W. Belzig, and Yu. V. Nazarov, *Phys. Rev. Lett.* 92, 196807 (2004).
51. G. Eilenberger, *Z. Phys.* 214, 195 (1968).
52. A. I. Larkin, and Yu. N. Ovchinnikov, *Zh. Eksp. Teor. Fiz.* 55, 2262 (1968) [*Sov. Phys. JETP* 28, 1200 (1969)].
53. K. D. Usadel, *Phys. Rev. Lett.* 25, 507 (1970).
54. A. V. Zaitsev, *Zh. Eksp. Teor. Fiz.* 86, 1742 (1984) [*Sov. Phys. JETP* 59, 1015 (1984)].
55. M. Yu. Kupriyanov, and V. F. Lukichev, *Zh. Eksp. Teor. Fiz.* 94, 139 (1988) [*Sov. Phys. JETP* 67, 1163 (1988)].
56. J. Rammer, and H. Smith, *Rev. Mod. Phys.* 58, 323 (1986).
57. W. Belzig, F. K. Wilhelm, C. Bruder, G. Schön, and A. D. Zaikin, *Superlattices and Microstructures* 25, 1251 (1999).
58. C. W. J. Beenakker, *Rev. Mod. Phys.* 69, 731 (1997).
59. Yu. V. Nazarov, *Phys. Rev. Lett.* 73, 1420 (1994).
60. Yu. V. Nazarov, in "Quantum Dynamics of Submicron Structures" (H. Cerdeira, B. Kramer, and G. Shön, Eds.), p. 687, Kluwer Academic Publishers, 1995.
61. Yu. V. Nazarov, *Superlattices and Microstructures* 25, 1221 (1999).
62. The Usadel equation can be written in terms of Green's function in the Nambu–Keldysh space (denoted by a check) as $D\partial_x(\check{G}\partial_x\check{G}) - i[\check{H},\check{G}] = 0$, where $\check{H} = E\hat{\sigma}_3 + i\hat{\sigma}_2\Re\Delta + i\hat{\sigma}_2\Im\Delta$.
63. A. Kastalsky, A. W. Kleinsasser, L. H. Greene, R. P. Milliken, and J. P. Harbison, *Phys. Rev. Lett.* 67, 3026 (1991).
64. S. N. Artemenko, A. F. Volkov, and A. V. Zaitsev, *Solid State Commun.* 30, 771 (1979).
65. Yu. V. Nazarov and T. H. Stoof, *Phys. Rev. Lett.* 76, 823 (1996).
66. A. A. Golubov, F. K. Wilhelm, and A. D. Zaikin, *Phys. Rev. B* 55, 1123 (1997).
67. S. Yip, *Phys. Rev. B* 52, 15504 (1995).
68. A. Volkov, N. Allsopp, and C. J. Lambert, *J. Phys. Condens. Matter* 8, L45 (1996).
69. G. B. Lesovik, A. L. Fauchère, and G. Blatter, *Phys. Rev. B* 55, 3146 (1997).
70. B. J. van Wees, A. Dimoulas, J. P. Heida, T. M. Klapwijk, W. V. D. Graaf, and G. Borghs, *Physica B* 203, 285 (1994).
71. H. Courtois, Ph. Gandit, D. Mailly, and B. Pannetier, *Phys. Rev. Lett.* 76, 130 (1996).
72. P. Charlat, H. Courtois, Ph. Gandit, D. Mailly, A. F. Volkov, and B. Pannetier, *Phys. Rev. Lett.* 77, 4950 (1996).
73. M. Nahum, T. M. Eiles, and J. M. Martinis, *Appl. Phys. Lett.* 65, 3123 (1994).
74. M. M. Leivo, J. P. Pekola, and D. V. Averin, *Appl. Phys. Lett.* 68, 1996 (1996).
75. J. P. Pekola, D. V. Anghel, T. I. Suppula, J. K. Suoknuuti, A. J. Manninen, and M. Manninen, *Appl. Phys. Lett.* 76, 2782 (2000).
76. R. J. Soulen, J. M. Byers, M. S. Osofsky, B. Nadgorny, T. Ambrose, S. F. Cheng, P. R. Broussard, C. T. Tanaka, J. Nowak, J. S. Moodera, A. Barry, and J. M. D. Coey, *Science* 282, 85 (1998).

77. S. K. Upadhyay, A. Palanisami, R. N. Louie, and R. A. Buhrman, *Phys. Rev. Lett.* 81, 3247 (1998).
78. F. W. J. Hekking, L. I. Glazman, K. A. Matveev, and R. I. Shekhter, *Phys. Rev. Lett.* 70, 4138 (1993).
79. F. Guinea and G. Schön, *Physica B* 152, 165 (1988).
80. F. W. J. Hekking and Yu. V. Nazarov, *Phys. Rev. Lett.* 71, 1625 (1993); *Phys. Rev. B* 49, 6847 (1994).
81. P. Lafarge, H. Pothier, E. R. Williams, D. Esteve, C. Urbina, and M. H. Devoret, *Z. Phys. B* 85, 327 (1991).
82. D. V. Averin and Yu. V. Nazarov, *Phys. Rev. Lett.* 69, 1993 (1992).
83. P. Lafarge, P. Joyez, D. Esteve, C. Urbina, and M. H. Devoret, *Phys. Rev. Lett.* 70, 994 (1993).
84. M. T. Tuominen, J. M. Hergenrother, T. S. Tighe, and M. Tinkham, *Phys. Rev. Lett.* 69, 1997 (1992); *Phys. Rev. B* 47, 11599 (1993).
85. T. M. Eiles, J. M. Martinis, and M. H. Devoret, *Phys. Rev. Lett.* 70, 1862 (1993).
86. J. M. Hergenrother, M. T. Tuominen, and M. Tinkham, *Phys. Rev. Lett.* 72, 1742 (1994); J. M. Hergenrother et al., page 327 in Ref. [4].
87. G. Schön, J. Siewert, and A. D. Zaikin, in "Mesoscopic Superconductivity," (F. W. J. Hekking, G. Schön, and D. V. Averin, Eds.). Proceedings of the NATO ARW, p. 340. *Physica B* 203 (1994).
88. A. O. Caldeira and A. J. Leggett, *Ann. Phys. (N.Y.)* 149, 374 (1983).
89. U. Weiss, "Quantum Dissipative Systems," Vol. 2. Series in Modern Condensed Matter Physics, World Scientific, Singapore, 1993.
90. U. Eckern, G. Schön, and V. Ambegaokar, *Phys. Rev. B* 30, 6419 (1984).
91. G. Schön and A. D. Zaikin, *Phys. Rep.* 198, 237 (1990).
92. W. J. Elion, M. Matters, U. Geigenmüller, and J. E. Mooij, *Nature* 371, 594 (1994).
93. M. Matters, W. J. Elion, and J. E. Mooij, *Phys. Rev. Lett.* 75, 721 (1995).
94. P. Lafarge, P. Joyez, D. Esteve, C. Urbina, and M. Devoret, *Nature* 365, 422 (1993).
95. T. A. Fulton, P. L. Gammel, D. J. Bishop, L. N. Dunkleberger, and G. J. Dolan, *Phys. Rev. Lett.* 63, 1307 (1989).
96. A. Maassen van den Brink, G. Schön, and L. J. Geerligs, *Phys. Rev. Lett.* 67, 3030 (1991); A. Maassen van den Brink, A. A. Odintsov, P. A. Bobbert, and G. Shön, *Z. Phys. B* 85, 459 (1991).
97. D. B. Haviland, Y. Harada, P. Delsing, C. D. Chen, and T. Claeson, *Phys. Rev. Lett.* 73, 1541 (1994).
98. J. Siewert and G. Schön, *Phys. Rev. B* 54, 7421 (1996).
99. D. V. Averin and V. Y. Aleshkin, Pis'ma *Zh. Eksp. Teor. Fiz.* 50, 331 (1989) [*JETP Lett.* 50, 367 (1989)].
100. K. A. Matveev, M. Gisselfält, L. I. Glazman, M. Jonson, and R. I. Shekter, *Phys. Rev. Lett.* 70, 2940 (1993).
101. P. Joyez, P. Lafarge, A. Filipe, D. Esteve, and M. H. Devoret, *Phys. Rev. Lett.* 72, 2458 (1994).
102. J. H. Davies, P. Hyldgaard, S. Hershfield, and J. W. Wilkins, *Phys. Rev. B* 46, 9620 (1992); S. Hershfield, J. H. Davies, P. Hyldgaard, C. J. Stanton, and J. W. Wilkins, *Phys. Rev. B* 47, 1967 (1993); U. Hanke, Yu. M. Galperin, K. A. Chao, and N. Zou, *Phys. Rev. B* 48, 17209 (1993); A. N. Korotkov, *Phys. Rev. B* 49, 10381 (1994); D. V. Averin, cond-mat/0010052 (arXiv.org e-Print archive); E. V. Sukhorukov, G. Burkard, and D. Loss, *Phys. Rev. B* 63, 125315 (2001); M.-S. Choi, F. Plastina, and R. Fazio, *Phys. Rev. Lett.* 87, 116 601 (2001).
103. L. J. Geerligs, S. M. Verbrugh, P. Hadley, J. E. Mooij, H. Pothier, P. Lafarge, C. Urbina, D. Esteve, and M. H. Devoret, *Z. Phys. B: Condens. Matter* 85, 349 (1991); J. P. Pekola, J. J. Toppari, M. Aunola, M. T. Savolainen, and D. V. Averin, *Phys. Rev. B* 60, 9931 (1999).
104. P. W. Anderson, *J. Phys. Chem. Solids* 11, 28 (1959).
105. D. C. Ralph, C. T. Black, and M. Tinkham, *Phys. Rev. Lett.* 74, 3241 (1995); *Phys. Rev. Lett.* 76, 688 (1996); *Phys. Rev. Lett.* 78, 4087 (1997).
106. For a review of results and further references, see J. von Delft and D.C. Ralph, *Physics Reports*, 345, 61 (2001).
107. J. von Delft, A. D. Zaikin, D. S. Golubev, and W. Tichy, *Phys. Rev. Lett.* 77, 3189 (1996).
108. R. A. Smith, V. Ambegaokar, *Phys. Rev. Lett.* 77, 4962 (1996).
109. R. W. Richardson and N. Sherman, *Nucl. Phys.* 52, 221 (1964); R. W. Richardson, *J. Math. Phys.* 6, 1034 (1965); R. W. Richardson, *Phys. Rev.* 141, 949 (1966).
110. K. A. Matveev and A. I. Larkin, *Phys. Rev. Lett.* 78, 3749 (1997).
111. A. Mastellone, G. Falci, and Rosario Fazio, *Phys. Rev. Lett.* 80, 4542 (1998); S. D. Berger and B. I. Halperin, *Phys. Rev. B* 58, 5213 (1998).
112. F. Braun and J. von Delft, *Phys. Rev. Lett.* 81, 4712 (1998).
113. J. Dukelsky and G. Sierra, *Phys. Rev. Lett.* 83, 172 (1999).
114. G. Falci, A. Fubini, and A. Mastellone, *Phys. Rev. B* 65, 140507 (2002).
115. A. Di Lorenzo, R. Fazio, F. W. J. Hekking, G. Falci, A. Mastellone, and G. Giaquinta, *Phys. Rev. Lett.* 84, 550 (2000); G. Falci et al., *J. Low. Temp. Phys.* 118, 355 (2000).
116. M. Schechter, Y. Imry, Y. Levinson, and J. von Delft, *Phys. Rev. B* 63, 214518 (2001).
117. M. Goldman and S. A. Wolf (Eds.), "Percolation, Localization, and Superconductivity," NATO ASI series, Series B, Physics, Vol. 109, Plenum Press, New York, 1984.
118. J. E Mooij and G. Schön (Eds.), "Coherence in Superconducting Networks," *Physica B* 152, 1 (1988).
119. E. Simanek, "Inhomogeneous Superconductors, Granular and Quantum Effects," Oxford University Press, Oxford, 1994.
120. S. L. Sondhi, S. M. Girvin, J. P. Carini, and D. Shahar, *Rev. Mod. Phys.* 69, 315 (1997).
121. S. Sachdev, "Quantum Phase Transition," Cambridge University Press, Cambridge, 1999.
122. K. Liu and M. Fisher, *J. Low. Temp. Phys.* 10, 655 (1973).
123. M. P. A. Fisher, B. P. Weichman, G. Grinstein, and D. S. Fisher, *Phys. Rev. B* 40, 546 (1989).
124. K. B. Efetov, *Sov. Phys. JETP* 51, 1015 (1980).

125. M. Nielsen and I. Chuang, "Quantum Computation and Quantum Communication." Cambridge University Press, Cambridge, 2000.
126. D. V. Averin, *Fortschrit. der Physik* 48, 1055 (2000).
127. Yu. Makhlin, G. Schön, and A. Shnirman, *Rev. Mod. Phys.* 73, 357 (2001).
128. Y. Nakamura, Yu. A. Pashkin, and J. S. Tsai, *Nature* 398, 786 (1999).
129. J. R. Friedman, V. Patel, W. Chen, S. K. Tolpygo, and J. E. Lukens, *Nature* 406, 43 (2000).
130. C. H. van der Wal, A. C. J. ter Haar, F. K. Wilhelm, R. N. Schouten, C. Harmans, T. P. Orlando, S. Lloyd, and J. E. Mooij, *Science* 290, 773 (2000).
131. D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Esteve, and M. H. Devoret, *Science* 296, 886 (2002).
132. Y. Yu, S. Y. Han, X. Chu, S. I. Chu, and Z. Wang, *Science* 296, 889 (2002).
133. J. M. Martinis, S. Nam, J. Aumentado, and C. Urbina, *Phys. Rev. Lett.* 89, 117901 (2002).
134. I. Chiorescu, Y. Nakamura, C. J. P. M. Harmans, and J. E. Mooij, *Science* 299, 1869 (2003).
135. Yu. A. Pashkin, T. Yamamoto, O. Astafiev, Y. Nakamura, D. V. Averin, and J. S. Tsai, *Nature* 421, 823 (2003).
136. A. Shnirman, G. Schön, and Z. Hermon, *Phys. Rev. Lett.* 79, 2371 (1997); Y. Makhlin, G. Schön, and A. Shnirman, *Nature* 398, 305 (1999).
137. D. A. Averin, *Solid State Commun.* 105 659 (1998).
138. J. E. Mooij, T. P. Orlando, L. Tian, C. van der Wal, L. Levitov, S. Lloyd, and J. J. Mazo, *Science* 285, 1036 (1999).
139. L. B. Ioffe, V. B. Geshkenbein, M. V. Feigelman, A. L. Faucher, and G. Blatter, *Nature* 398, 679 (1999).
140. R. Fazio, G. M. Palma, and J. Siewert, *Phys. Rev. Lett.* 81, 5385 (1999).
141. G. Falci, R. Fazio, G. M. Palma, J. Siewert, and V. Vedral, *Nature* 403, 869 (2000).
142. G. Falci, E. Paladino, R. Fazio, Proceedings of the International School of Physics "Enrico Fermi," Course CLI, Varenna (Italy), July 2002. In "Quantum Phenomena of Mesoscopic Systems" (B. Altshuler, V. Tognetti, and A. Tagliacozzo, Eds.), Vol. 151. IOS Press, Amsterdam, 2003; cond-mat/0312550.
143. D. Jaksch, C. Bruder, C. W. Gardiner, and P. Zoller, *Phys. Rev. Lett.* 81, 3108 (1998).

CHAPTER 14

Strain Field Calculations in Embedded Quantum Dots and Wires

R. Maranganti, P. Sharma
Department of Mechanical Engineering, University of Houston, Houston, Texas, USA

CONTENTS

1. Introduction 689
2. Strain Field Calculations in Quantum Dots and Wires 691
   2.1. Simple Illustrative Example 692
   2.2. Effect of Shape 694
   2.3. Effect of Presence of a Free Surface in Near Vicinity . . 708
   2.4. Effect of Material Anisotropy and Nonlinearity 711
   2.5. Effect of Coupled Fields: Piezoelectricity 715
   2.6. Effect of Size 718
   2.7. Some Results from Inclusion Theory Useful for Strain Calculations in Quantum Dot Structures . . . 726
3. Summary and Open Issues 727
   References 728

1. INTRODUCTION

Quantum dots (QDs) are tiny, dimensionally confined (typically semiconductor) objects where quantum effects become obvious, for example, energy spectra become discrete (see Fig. 1). QDs are characterized by a sharp density of states (DOS)—reminiscent of "atoms." To be more precise, a semiconductor material cluster may be termed a quantum dot if its characteristic dimensions become comparable to the exciton Bohr radius. For example, in case of GaAs, the exciton Bohr radius is roughly 10 nm. They are of immense technological importance and (while several technological barriers remain) are often considered as basis for several revolutionary nanoelectronic devices and applications (Fig. 2), for example, next-generation lighting [1–2], lasers [3–4], quantum computing, information storage and quantum cryptography [5–7], biological labels [8], sensors [9] and many others [10–12, 13–15]. QDs and quantum wires (QWRs) are typically embedded in another material with differing

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (689–732)

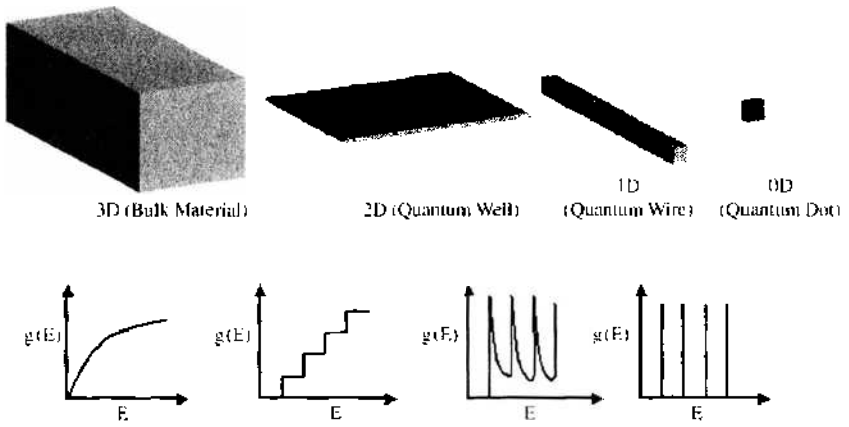


Figure 1. As the dimensionality of the structure reduce DOS, $g(E)$, becomes sharper and pronouncedly discrete.

elastic constants and lattice parameter. Because of the lattice mismatch, both the QD and host matrix strain and relax elastically to accommodate this mismatch and thus admit a state of stress (Fig. 3). As is well known, the electronic structure and the consequent optoelectronic properties of QDs are severely impacted due to this lattice mismatch induced strain [16–19]. In this article, we review pertinent literature on various methods to calculate the

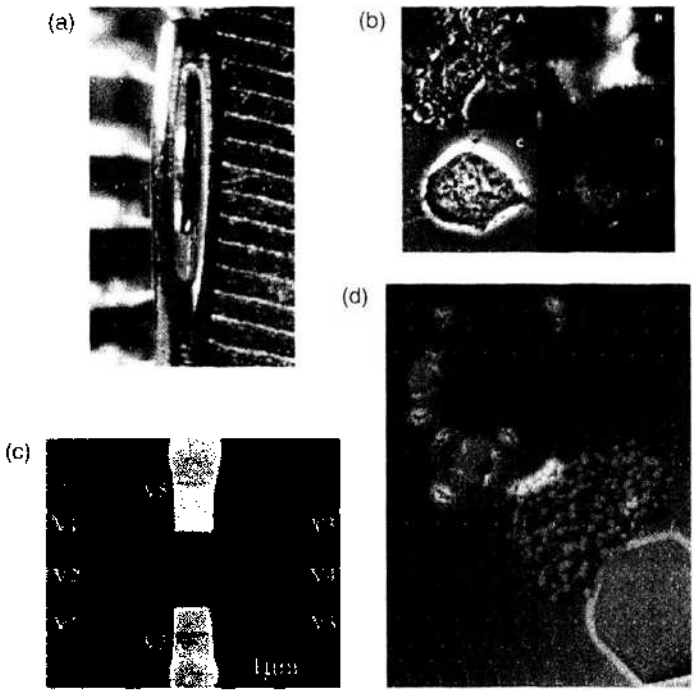


Figure 2. Different applications of quantum wires and dots. (a) Image of a quantum wire laser operating through the eye of a needle. Reproduced with permission from C. Jagadish, Research School of Physical Sciences and Engineering, Australian National University, © (2005). Increased recombination probability and discrete nature of DOS results in efficient lasing characteristics at low threshold currents along with the promise of tunable wavelength. (b) Magnified view of quantum dot attachment to neurons. Reproduced with permission from [18], Winter et al., *Adv. Mater.* 13, 1673 (2001). © 2001, Wiley VCH, Verlag GmbH & Co. Pictured here is the magnification of quantum dot attachment to neurons using antibody (A, B) and peptide (C, D) binding techniques. In B and D, the blue color is the self-fluorescence of the cell's cytoplasm, and the yellow/orange color is the quantum dot luminescence. Using nanostructure sensors such as these, cellular-level target specificity at biological-electronic interfaces can be achieved as against conventional silicon-based electrodes where interfaces are at the tissue level. (c) Image of a quantum dot bases transistor. Reproduced with permission from [19], Chen et al., *Phys. Rev. Lett.* 92, 176801 (2004). © 2004, American Physical Society. Transistors form vital switching components in computers. With quantum-dot based transistors as the one shown in the figure, "quantum computers," can be realized. The memory of a quantum computer can simultaneously be both in "0" (off) and "1" (on) states (in general a superposition of these states) compared to a classical computer's memory which is made up of either a "0" bit or a "1" bit. (d) Illustration of a quantum dot–based sensor for analyte detection. Reproduced with permission from [20], Medintz et al., *Nat. Mater.* 2, 630 (2003). © 2003, Nature Publishing Group.

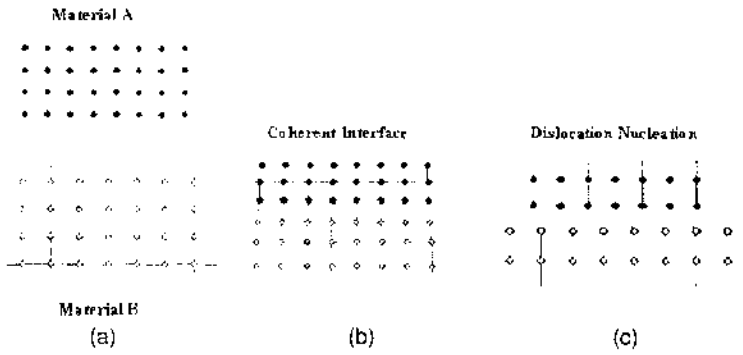


Figure 3. (a) Illustrates two materials "A" and "B" with different lattice constants; (b) formation of a "coherent" interface between A and B, requires that the lattice mismatch is completely accommodated by elastic deformation; (c) depicts the situation where the lattice mismatch is so high that further elastic deformation is energetically too costly and it is preferable to nucleate misfit dislocations. Such an interface is called a semicoherent interface.

state of strain in embedded *coherent* QDs and QWRs; That is, interface dislocations are absent. The impact of dislocations, if formed, on optoelectronic properties is so severe that an entirely different sort of discussion is required. In this review article, we mainly focus on the scenario when misfit dislocations are absent in the very near vicinity of the QD. This is indeed practical for small QDs [20].

The chapter is semitutorial in nature to make it accessible to readers from a broad range of disciplines. The rest of this chapter is organized in two broad sections. In Section 2, we provide, intermixed with a literature review, details on various methods and issues in calculation of strains in embedded QDs. Our particular focus is on cataloging the known analytical expressions. Physical insights obtained from both analytical results and numerical simulations of various researchers (including our own) are also discussed. The effect of various parameters such as material anisotropy, QD shape, surface image forces, elastic nonlinearity and piezoelectricity are addressed (Section 2.2–2.6). Typically classical continuum mechanics, which is intrinsically size independent, is employed for strain calculations. To address this, we have also included a brief discussion on the effect of QD size on strain calculations since most QD are in the sub-20-nm range, and one indeed expects a departure from classical mechanics at such small sizes (Section 2.7). In Section 2.8, we briefly review some relevant results in the elastic theory of inclusions, which while well known in the solid mechanics community, may be less known in the QD literature and are thus likely to be useful to engineers and scientists working in the latter research area. We conclude with a summary in Section 3 where we also present a rather terse personal viewpoint on some future research avenues in this research topic.

2. STRAIN FIELD CALCULATIONS IN QUANTUM DOTS AND WIRES

QDs occur or are fabricated in a variety of shapes, sizes, and material combinations. Classical continuum elasticity (and less frequently, atomistic methods) are typically employed to calculate the strains in these structures, which are then linked to suitable band structure calculation methods to estimate the strained optoelectronic behavior. A review of band structure calculation methods for strained quantum structures is beyond the scope of the present article and the reader is referred to some excellent resources in the literature [12, 21, 22].

Classical linear elastostatics is governed by the following set of partial differential equations [Eq. (1)] along with appropriate boundary conditions at the interfaces/surfaces [Eq. (2)]:

$$\boldsymbol{\varepsilon} = \frac{1}{2}[(\nabla \otimes \mathbf{u})^T + (\nabla \otimes \mathbf{u})] \tag{1a}$$

$$\mathbf{div}\ \boldsymbol{\sigma} + \mathbf{f} = 0 \tag{1b}$$

$$\boldsymbol{\sigma} = \mathbf{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{m}}) \tag{1c}$$

Equation (1a) relates the second-order strain tensor ($\boldsymbol{\varepsilon}$) to the displacement vector ($\mathbf{u}$). ∇ is the gradient operator and superscript "T" indicates transpose operation. Equation (1b) written in terms of the second-order stress tensor ($\boldsymbol{\sigma}$) is the equilibrium equation with external body forces ($\mathbf{f}$) while Eq. (1c) relates the stresses and strain (note that ε^m is the lattice-mismatch strain) through a fourth-order elasticity tensor ($\mathbf{C}$). We shall use both boldface and index notation as convenient. Unless noted otherwise, all tensors are Cartesian and conventional summation rules for repeated indices apply. Boundary conditions can be either prescribed tractions or displacements. In the context of embedded QD/QWRs (typically), the following conditions at the interfaces and free surfaces are obeyed:

$$[\boldsymbol{\sigma}]\mathbf{n} = \mathbf{0}; \quad [\mathbf{u}] = \mathbf{0} \tag{2}$$

The vector $\mathbf{n}$ is the normal at a point on the interface. The symbol $[\bullet]$ denotes jump in the field quantities across an interface. In particular for free surface, the traction-free condition $\boldsymbol{\sigma} \cdot \mathbf{n} = 0$ is satisfied.

Considering the easy and economic availability of commercial numerical packages (e.g., finite element tools) that can solve Eq. (1) for arbitrary geometry, material symmetry, and boundary conditions, discussion of analytical solutions may seem somewhat redundant and antiquated. However, much effort has been expended to develop such analytical models that have proved to be extremely useful in obtaining explicit physical insights and often, reasonable accuracy. Further, one must note that eventually, the strain calculations must be coupled to quantum mechanical equations for band structure calculations, which often proceeds numerically. Analytical expressions for strain can significantly simplify that process. In Section 2.1, we provide a simple illustrative example for calculation of strain in the idealized case of a deeply buried, embedded spherical quantum dot.

2.1. Simple Illustrative Example

Consider a spherical quantum dot ($\boldsymbol{\Omega}$) of radius R (Fig. 4), located in an infinite amount of host material ($\mathbf{D}$). The assumption of infinite host medium corresponds to the situation where the QD is deeply buried and the host matrix boundaries do not impact its strain state. In other words, the distance of the QD from any free boundary is significantly larger than the QD radius (typically $> 3R$).

Let the lattice parameter of the QD be a_{QD} and that of the matrix be a_M. The lattice mismatch strain tensor is then:

$$\varepsilon^m_{ij} = \varepsilon^m \delta_{ij}; \quad \varepsilon^m = \frac{a_{\text{QD}} - a_M}{a_M} \tag{3}$$

Some authors choose to divide $a_{\text{QD}} - a_M$ by the average of the two lattice parameters. We assume for the purpose of this simple example that the material properties of both QD and matrix are isotropic; that is, the elastic tensor can be expressed in terms of the two Lamè constants (λ, μ):

$$C_{ijkl} = \lambda \delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{4}$$

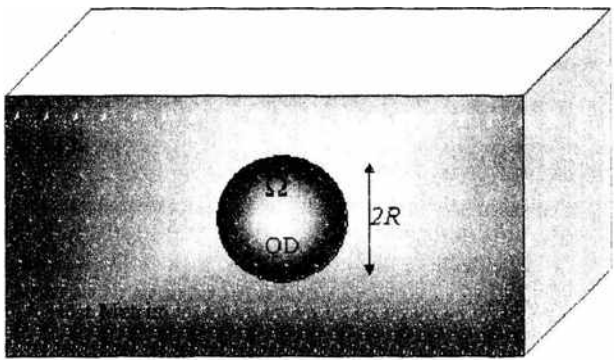


Figure 4. Schematic of the spherical quantum dot ($\boldsymbol{\Omega}$) embedded in an infinite host matrix ($\mathbf{D}$).

In this simple example, it is easier to work in terms of displacement. Equations (1) can be rewritten in terms of displacement (Navier's equations):

$$\mu\nabla^2\mathbf{u} + \frac{\mu}{1-2\nu}\,\text{div grad}\,\mathbf{u} + \mathbf{f} = 0 \tag{5}$$

Here, ν, is the Poisson ratio. To calculate stress, any nonelastic contribution (such as the mismatch strain) must be subtracted from the actual strain before employing Hooke's law. That is, for isotropic materials, the constitutive law in Eq. (1) can be rewritten as

$$\boldsymbol{\sigma} = \lambda_{QD}(\mathbf{Tr}\boldsymbol{\varepsilon} - \mathbf{Tr}\boldsymbol{\varepsilon}^m)\mathbf{I} + 2\mu_{QD}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^m) \tag{6}$$

Here $\mathbf{I}$ is the identity tensor. No subtraction of the mismatch strain is necessary when calculating the matrix stresses since according to convention we adopted in Eq. (3), the mismatch strain is calculated with respect to zero reference strain in the matrix.

The problem under consideration with all its attendant assumptions is manifestly spherically symmetric and hence must admit a displacement field that is purely radial, that is, $\mathbf{u} = u(r)\mathbf{e}_r$. The corresponding infinitesimal strain tensor in spherical polar basis $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ can be written as

$$\boldsymbol{\varepsilon} = \frac{\partial u}{\partial r}\mathbf{e}_r \otimes \mathbf{e}_r + \frac{u}{r}\mathbf{e}_\theta \otimes \mathbf{e}_\theta + \frac{u}{r}\mathbf{e}_\phi \otimes \mathbf{e}_\phi \tag{7}$$

Equation (5) or the Navier's equation then reduces to

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} - \frac{2u}{r^2} = 0 \tag{8}$$

The general solutions to the differential equation of Eq. (8) are simply, r and $1/r^2$, that is,

$$u(r) = Ar + Br^{-2} | r < R \tag{9a}$$

$$Cr + Dr^{-2} | r > R \tag{9b}$$

Here, A, B, C, and D are constants to be determined from the boundary conditions. Two boundary conditions are immediately obvious: (1) since the problem is a purely dilatational problem with spherical symmetry, at the center of the quantum dot u(r) must approach zero and (2) at points infinitely far away from the quantum dot, the displacement must decay to zero. These restrictions render $B = C = 0$. Further, as per the boundary conditions in Eq. (2), the displacements must be continuous, thus $u^+(r \to R)| = u^-(r \to R)$, while the traction continuity condition ensures that $\sigma_{rr}^+ - \sigma_{rr}^- = 0$. The final solution is obtained as

$$u(r) = \begin{cases} \dfrac{3K_{QD}\varepsilon^m}{4\mu_M + 3K_{QD}} r; r \le R & \text{(10a)} \\ \dfrac{3K_{QD}\varepsilon^m}{4\mu_M + 3K_{QD}} \dfrac{R^3}{r^2}; r > R & \text{(10b)} \end{cases}$$

Here the subscript M refers to matrix properties. Strain components and in particular the dilation (which has the dominant effect in electronic calculations) can be trivially obtained to be

$$\varepsilon_{rr}(r) = \varepsilon_{\theta\theta}(r) = \varepsilon_{\phi\phi}(r) = \frac{3K_{QD}\varepsilon^m}{3K_{QD} + 4\mu_M} \Big| r < R \tag{11a}$$

$$\varepsilon_{rr}(r) = \left[\frac{3K_{QD}\varepsilon^m}{3K_{QD} + 4\mu_M}\right]\frac{R^3}{r^3} \Big| r > R \tag{11b}$$

$$\varepsilon_{\theta\theta} = \varepsilon_{\phi\phi}(r) = -\left[\frac{3K_{QD}\varepsilon^m}{3K_{QD} + 4\mu_M}\right]\frac{R^3}{2r^3} \Big| r > R \tag{11c}$$

Yang et al. [23] performed similar analytical calculations on $Si_{0.8}Ge_{0.2}$ sphere, cylinder and a rectangular slab embedded in an infinite Si host and linked these results to band-structure calculations. Numerical results were also obtained for V-groove $Si_{0.8}$ $Ge_{0.2}$ quantum wires under anisotropy considerations [23].

An important point to note is that for band-gap calculations, *the mismatch strain is subtracted from the actual compatible elastic strain*. As a result while in the solid mechanics community the compatible elastic strain is normally expressed and plotted (as we have done so in Fig. 5), the QD research community often illustrate the subtracted strain (see, e.g., Yang et al. [23]). This can potentially cause confusion and care must be exercised in interpreting results from the solid mechanics literature. The simple idealized example in Fig. 5 illustrates and underscores several rather general features: (1) the strain state is uniform inside the quantum dot—as shall be seen, this is a general feature for all ellipsoidal shaped quantum dots; (2) the dilation too is uniform inside the QD; we emphasize this trivial point here since, as will be discussed in the next section, the dilation is uniform for *all quantum dot shapes* (provided certain assumptions such as unbounded host material, etc., are not violated); (iii) the dilation is zero outside the QD. This also is true for *all QD shapes*.

2.2. Effect of Shape

Some of the commonly occurring configurations of quantum dots and wires are illustrated in Fig. 6.

The reader is referred to these references [24–29] that report evidence of a wide variety of shapes, including pyramidal, truncated pyramidal, lens shaped, hemispherical, multifaceted domes, and so on, for the widely studied $In_xGa_{1-x}As/InAs$ quantum dot system. Owing to the inherent size independence of classical continuum elasticity, strain state calculations depend exclusively on inclusion shape—to be more precise, self-similar structures if scaled larger or smaller yield identical results for strains and stresses.

Several of the methodologies that emerged in the context of QD/QWR strain calculations can be considered to be off-shoots or modifications of the now classical work of Eshelby

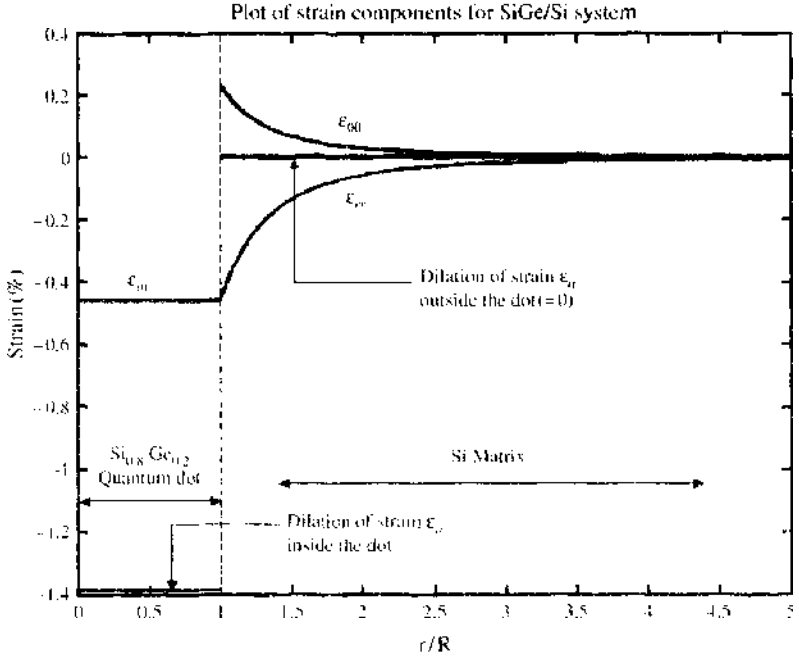


Figure 5. Plot of the strain components for a $Si_{0.8}$ $Ge_{0.2}$ spherical quantum dot embedded in an infinite Si matrix versus the ratio (r/R). Uniform strain ε_{00} exists inside the quantum dot. Outside the dot, ε_{00} is one of the tangential components of strain and ε_{rr} is the radial component of strain. As is noticeable, the dilation of strain ε_{rr} outside the dot is zero. One can notice the rapid decay of the strain outside the dot to a zero value at values of r/R as low as 3.5.

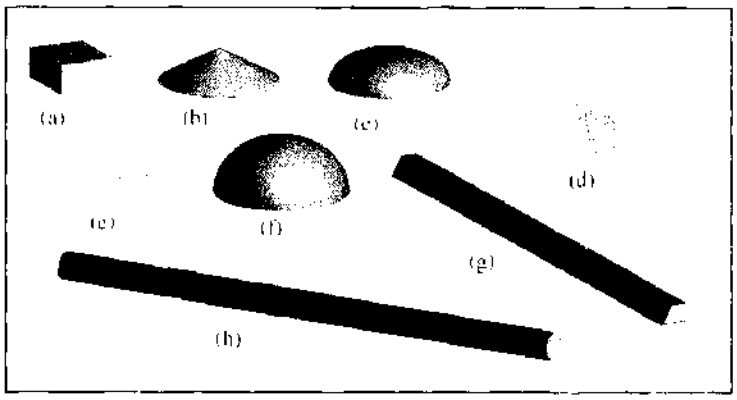


Figure 6. Common configurations of QDs and QWRs. (a) Cuboidal, (b) conical, (c) lens shaped, (d) pyramidal, (e) truncated pyramidal, (f) hemispherical, (g) trapezoidal QWR, (h) cylindrical QWR.

[30–32] on inclusions and inhomogenities. His elegant formalism provides rather straightforward means to evaluate the effect of shape for, in principle, arbitrary shaped quantum dots and wires. We first discuss his formalism briefly before reviewing the specific literature on QD/QWRs. In addition to providing a historical setting, Eshelby's work provides an excellent perspective on inclusion problems and more specifically, then the related works on QD/QWRs can be discussed in appropriate context. This discussion will greatly facilitate our presentation (Section 2.8) of some results on the theory of inclusions (based on Eshelby's approach) that though widely known in the micromechanics community, do not appear to have been fully exploited in the context of quantum dots and wires.

2.2.1. Eshelby's General Formalism for Shape Effects

Consider an arbitrary shaped inclusion (quantum dot) embedded in an unbounded material (the impact of finiteness of the surrounding material will be discussed in Section 2.3). Let a stress-free uniform transformation strain be prescribed within the domain of the QD (Fig. 7). Lattice mismatch induced strain is an example of such a stress free transformation strain and so are, for example, thermal expansion mismatch strains, phase transformation strains among others. For the moment, we assume identical material properties for the QD and the matrix.

Although Eshelby's general approach is not restricted to isotropic materials, analytical results are generally not tractable in the fully anisotropic case. Issues related to anisotropy are fully discussed in Section 2.4.1. For now, we assume isotropic elastic behavior. By definition, the transformation strain is only nonzero within the QD domain ($\mathbf{x} \in \Omega$), and thus we can write the constitutive law for the QD-matrix as follows:

$$\sigma_{ij} = 2\mu(\varepsilon_{ij} - \varepsilon_{ij}^{m} H) + \lambda\delta_{ij}(\varepsilon_{ll} - \varepsilon_{ll}^{m} H) \tag{12}$$

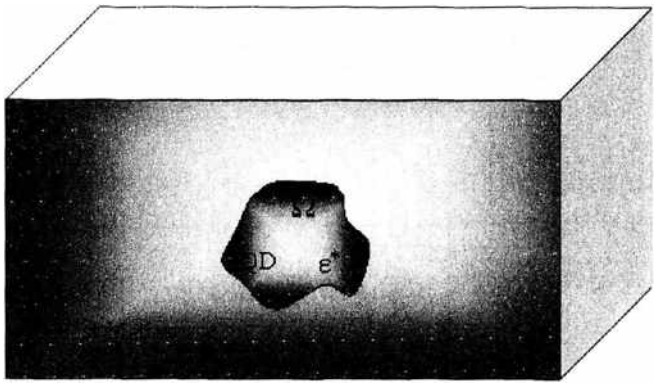


Figure 7. A quantum dot of arbitrary shape (Ω) inside an infinite matrix (D).

Here "H" is the step function defined as $H(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \mathbf{\Omega} \\ 0 & \mathbf{x} \notin \mathbf{\Omega} \end{cases}$. We can, analogous to Eq. (5), rewrite a Navier's type equation:

$$\mu u_{i,jj} + (\mu + \lambda) u_{j,ji} = -[(\lambda \varepsilon_{jj}^{m} \delta_{ik} + 2\mu \varepsilon_{ik}^{m}) H(\mathbf{x})]_{,k} \tag{13}$$

Clearly, the divergence of the mismatch stain $\varepsilon_{ik} H(\mathbf{x})$ defined over the inclusion volume act as delta functions across the inclusion surface i.e., $\varepsilon_{ik}^{m} \delta(S)$ (e.g., Refs. [30, 33]). A comparison with the Navier's Eq. 5 indicates that the mismatch strain terms can be simulated as body forces. The displacement vector can be obtained using the Green's function (of Eq. 5 or 13) as

$$\begin{aligned} u_i(\mathbf{x}) &= (\lambda \varepsilon_{jj}^{m} \delta_{jk} + 2\mu \varepsilon_{jk}^{m}) \int_S G_{ij}(\mathbf{x} - \mathbf{x}') dS_k(\mathbf{x}') \\ &= -(\lambda \varepsilon_{jj}^{m} \delta_{jk} + 2\mu \varepsilon_{jk}^{m}) \int_V G_{ij,k}(\mathbf{x} - \mathbf{x}') dV(\mathbf{x}') \end{aligned} \tag{14}$$

Here Gauss theorem has been used to convert the surface integral into a volume integral. The Green's function for Navier's equation is available in most books on elasticity (e.g., Ref. [34]) and can be written as

$$\begin{aligned} G_{ij}(\mathbf{r}) &= \frac{1}{8\pi(\lambda + 2\mu)} \left[(\lambda + 3\mu) \frac{\delta_{ij}}{|\mathbf{r}|} + (\lambda + \mu) \frac{x_i x_j}{|\mathbf{r}|^3} \right] \\ \mathbf{r} &= \mathbf{x} - \mathbf{x}'; \; r = |\mathbf{r}| \end{aligned} \tag{15}$$

Upon substituting the Green's function in Eq. (14) and invoking the strain-displacement law, we readily obtain [30]:

$$\varepsilon_{ij}(x) = \frac{1}{8\pi(1 - \nu)} [\psi_{kl,klij} - 2\nu \phi_{kk,ij} - 2(1 - \nu)(\phi_{ik,kj} + \phi_{jk,ki})] \tag{16}$$

where ψ and ϕ are biharmonic and harmonic potentials of the inclusion shape ($\mathbf{\Omega}$). They are given as:

$$\psi_{ij}(\mathbf{x}) = \psi \varepsilon_{ij}^{m} = \varepsilon_{ij}^{m} \int_{\omega} |\mathbf{x} - \mathbf{x}'| d^3\mathbf{x}' \tag{17}$$

$$\phi_{ij}(\mathbf{x}) = \phi \varepsilon_{ij}^{m} = \varepsilon_{ij}^{m} \int_{\omega} \frac{1}{|\mathbf{x} - \mathbf{x}'|} d^3\mathbf{x}' \tag{18}$$

Equation (16) is usually cast in the following form:

$$\begin{aligned} \boldsymbol{\varepsilon}(\mathbf{x}) &= \mathbf{S}(\mathbf{X}) \colon \boldsymbol{\varepsilon}^{\mathrm{m}} \; \mathbf{x} \in \mathbf{\Omega} \\ \boldsymbol{\varepsilon}(\mathbf{x}) &= \mathbf{D}(\mathbf{x}) \colon \boldsymbol{\varepsilon}^{\mathrm{m}} \; \mathbf{x} \notin \mathbf{\Omega} \end{aligned} \tag{19}$$

Here **S** and **D** are the so-called Eshelby tensors for interior and exterior points, respectively:

$$\{S_{ijkl}, D_{ijkl}\} = \frac{1}{8\pi(1 - \nu)} \{\psi_{klij} - 2\nu \delta_{kl} \phi_{ij} - (1 - \nu)[\phi_{kj} \delta_{il} + \phi_{ki} \delta_{jl} + \phi_{lj} \delta_{ik} + \phi_{li} \delta_{jk}]\} \tag{20}$$

One obtains **S** or **D** depending upon whether the vector **x** in Eqs. (17 and 18) is located within the quantum dot or outside it. Eshelby's tensors for various shapes (spheres, cylinders, ellipsoids, discs, and cuboids) are well documented in Mura [35]. For example, in the case of a spherical inclusion or quantum dot, we have

$$\psi(x) = \begin{cases} -\frac{1}{60}(r^4 - 10R^2 r^2 - 15R^4) & r \in \mathbf{\Omega} \\ \frac{R^3}{15}\left(5r + \frac{R^2}{r}\right) & r \notin \mathbf{\Omega} \end{cases} \tag{21}$$

$$\phi(x) = \begin{cases} -\frac{1}{6}(r^2 - 3R^2) & r \in \Omega \\ \frac{R^3}{3r} & r \notin \Omega \end{cases} \tag{22}$$

Using Eqs. (20)–(22) Eshelby's interior tensor can be then written explicitly as [35]

$$S_{ijkl} = \frac{5\nu - 1}{15(1-\nu)}\delta_{ij}\delta_{kl} + \frac{4 - 5\nu}{15(1-\nu)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{23}$$

Here, ν is the Poisson ratio. The reader can trivially verify that use of Eshelby tensor expression above for strain calculation in a spherical quantum dot leads to the same results derived in a different way in Eq. (11) of Section 2.2 (provided the matrix properties are set to be the same as those of the quantum dot).

The implications of Eshelby's formalism are manifest. An evaluation of the harmonic and biharmonic potentials (Eqs. [17] and [18]) for various shapes, in principle, allows calculation of Eshelby's tensor and hence the complete strain state inside and outside the embedded quantum dot. An interesting outcome of Eshelby's analysis [30] is that for inclusions of ellipsoidal shape, S is uniform. This implies that any quantum dot belonging to the ellipsoidal family immersed in an unbounded matrix subject to a uniform mismatch strain will admit a uniform strain! This is rather useful since the ellipsoidal shape is very versatile and can be used to mimic and approximate a variety of shapes (see Fig. 8). For the simple case of a spherical quantum dot, this fact has already been noted in the context of Eq. (11) and Fig. 5. This notion remains true even in the case of arbitrary anisotropy. This uniformity of strain rule does not in general hold true for nonellipsoidal shapes (e.g., polyhedral, pyramidal), in absence of linearity or if the matrix is not unbounded. As Eshelby [30] has pointed out, the peculiar property that ellipsoidal inclusions admit a uniform strain state under certain conditions is also very useful for taking into account the mismatch between the elastic moduli of the quantum dot and the surrounding matrix.

If only the dilatation is of interest, matters simplify considerably and we obtain:

$$Tr(\varepsilon) = -\frac{9K\varepsilon^m}{4\mu + 3K}\nabla^2\phi \tag{24}$$

Thus only the harmonic potential needs to be evaluated. Further, the general properties of the harmonic potential [30, 36–37] ensure that the dilation inside the quantum dot is *shape independent*! Again caution must be exercised in using this notion when, for example, the surrounding medium is not unbounded. Within the strict assumptions of the derivation

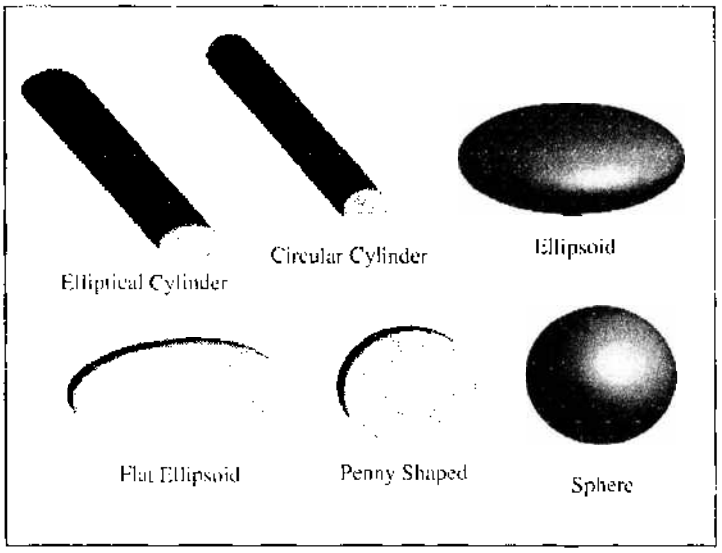


Figure 8. An illustration of the variety of shapes, which fall under the gamut of the general ellipsoid.

given by Eshelby [30] and partly reproduced here, this rather interesting fact suggests that in an isotropic unbounded medium, shape effects are of secondary importance insofar as *dilation* is concerned.

Since the strain state within the inclusion is uniform for ellipsoids, Eshelby [30] was able to devise an elegant method to mimic an inhomogenity, that is, an inclusion with a mismatch strain and material properties differing from the host matrix, by an inclusion containing a fictitious transformation strain. Mura [35] defines an inclusion to be a subdomain in which a stress-free strain is prescribed but the material properties are same everywhere. An inhomogeneity is defined as an inclusion that has differing material properties from those of the surrounding matrix. The so-called equivalent inclusion method simply entails equating the elastic state of an inhomogenity to that of an inclusion albeit with the aforementioned fictitious transformation strain that is,

$$\left.\begin{aligned} \mathbf{C}^M\colon \{\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^m - \boldsymbol{\varepsilon}^f\} &= \mathbf{C}^H\colon \{\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^m\} && \text{(25a)} \\ \boldsymbol{\varepsilon} &= \mathbf{S}\colon \{\boldsymbol{\varepsilon}^m + \boldsymbol{\varepsilon}^f\} && \text{(25b)} \end{aligned}\right\} \mathbf{x} \in \boldsymbol{\Omega}$$

Equations (25) are simply algebraic equations that allow explicit determination of the fictitious transformation strain and hence the elastic moduli mismatch in the case of ellipsoidal quantum dots. For more general shapes, Eshelby's tensor is no longer uniform and hence these equations become rather complicated integral equations. This is will be further discussed when we address arbitrary shaped QDs in the next section.

2.2.2. Quantum Dots

In the specific context of QDs, fair amount of effort has been expended on general solutions (for example, see [38–44]). Some of the earlier works in regards to quantum dots strain calculations are due to Grundmann et al. [48] (see also Refs. [45–47]). They presented some simple analytical results for shapes such as slabs, circular cylinder and spherical. One notes that these results can, of course, also be recovered from Eshelby's formalism discussed in the previous section. For the specific case of InAs pyramidal-shaped quantum dots embedded in a GaAs matrix, Grundmann et al. [48] presented numerical simulations for the strain distributions. This shape is of course a bit difficult to handle analytically although closed-form solution can indeed be derived as will be discussed shortly. Although it is beyond the scope of the present article, their computation of the strain-linked electronic properties of InAs QDs is also of interest and they succeed in obtaining reasonable agreement with experimental data on luminescence and absorption.

Downes et al. [49] devised a simple method for calculating the strain distribution in deeply buried QD structures. Six vectors $\mathbf{A}_{ij}$ are defined such that $\nabla\mathbf{A}$ yields the stress components $\boldsymbol{\sigma}_{\text{sph}}$ for a point spherical inclusion:

$$\mathbf{A}_{xx} = Dx\,\mathbf{i}; \quad \mathbf{A}_{xz} = -Dx\mathbf{k}; \quad \mathbf{A}_{xy} = -Dx\mathbf{i} \tag{26a}$$

$$\mathbf{A}_{yy} = Dy\mathbf{j}; \quad \mathbf{A}_{yz} = -Dz\mathbf{j}; \quad \mathbf{A}_{zz} = Dz\mathbf{k} \tag{26b}$$

$$D = \frac{A}{(x^2 + y^2 + z^2)^{3/2}}; \quad A = \frac{E\varepsilon^m}{4\pi(1-\nu)} \tag{26c}$$

Here, E is the Young's modulus. The stress tensor for a given QD shape can then be written as

$$\boldsymbol{\sigma}(\mathbf{x}) = \int_V \nabla \cdot \mathbf{A}(\mathbf{x} - \mathbf{x}')\, dV(\mathbf{x}') \tag{27}$$

Divergence theorem can then be invoked to convert the volume integral in Eq. 27 to a surface integral to obtain the strain field in and around an arbitrarily shaped QD.

$$\boldsymbol{\sigma}(\mathbf{x}) = \int_S \mathbf{A}(\mathbf{x} - \mathbf{x}') \cdot d\mathbf{S}(\mathbf{x}') \tag{28}$$

The QD surface may be discretized appropriately to convert the surface integral into a summation. This approach facilitates analytic solutions in the case of simple structures like

cuboidal QDs. Formally, this approach by Downes et al. [49] can be related to computing Eshelby's tensor. As example, the stress state of a parallelepiped dot can be easily related to its geometry, that is, lengths and the angles. For such a parallelepiped dot, centered on the origin and aligned along the Cartesian axes, with dimensions, $2a \times 2b \times 2c$. Eq. (28) provides the following stress components [60]:

$$\sigma_{xx} = \sum -\Lambda \tan^{-1}\left(\frac{(b \pm y)(c \pm z)}{(a \pm x)\sqrt{(a \pm x)^2 + (b \pm y)^2 + (c \pm z)^2}}\right) \tag{29a}$$

$$\sigma_{xz} = \sum q\Lambda \log\left\{\sqrt{(x \pm a)^2 + (y \pm b)^2 + (z \pm c)^2} - (y \pm b)\right\} \tag{29b}$$

Here the summation is over the various combinations of + and −. In Eq. (29b), q is +1 for one or 3 plus signs and −1 for 0 or 2 plus signs. Identical results can be obtained using Eshelby's approach and indeed have been derived by Chiu [50] and are also documented in Mura [35]. When the dimension of the parallelepiped along one direction is very large compared with the other two dimensions, it degenerates to the case of a rectangular QWR and the results for the buried strained layer, both of which will be discussed later in this section. As noted by Downes et al. [49], the integral that is used to evaluate σ_{xx} is similar to the integral for the solid angle subtended by the two (100) faces of a parallelepiped aligned along the Cartesian axes (see Fig. 9). The hydrostatic stress inside such a dot $(\sigma_{xx} + \sigma_{yy} + \sigma_{zz})$ is thereby proportional to the solid angle subtended at an interior field point by all the faces of the cube (4π). This fact automatically ensures the constancy of the dilation within the dot. The solid angle is zero for an outside point and so is the hydrostatic strain.

One of the more common shapes for self-assembled quantum dots is the pyramidal geometry (Fig. 10). Finite-element methods have been employed to treat the strain distribution problem for pyramidal quantum dots [48, 51]. However, analytical solutions (although somewhat tedious) can also be readily obtained.

A generic point, (x_1^0, x_2^0, x_3^0) inside the volume of the pyramidal dot illustrated in Fig. 10 can be mathematically represented by

$$-a\left(1 - \frac{x_3^0}{h}\right) \le x_1^0 \le a\left(1 - \frac{x_3^0}{h}\right) \tag{30a}$$

$$-b\left(1 - \frac{x_3^0}{h}\right) \le x_2^0 \le b\left(1 - \frac{x_3^0}{h}\right) \tag{30b}$$

$$0 \le x_3^0 \le hf \tag{30c}$$

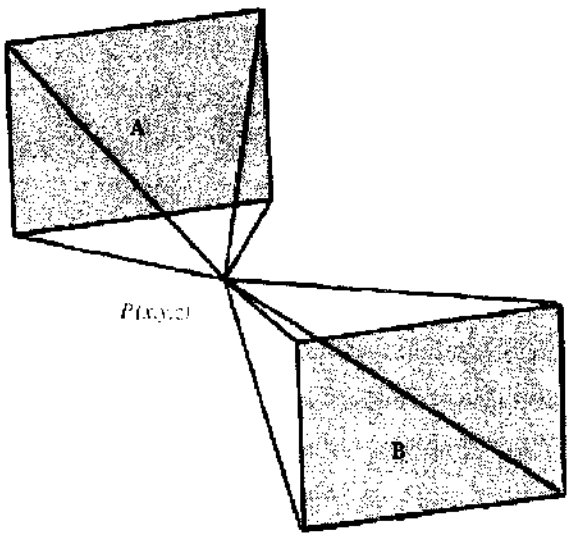


Figure 9. Illustration of the solid angle subtended at point $P(x, y, z)$ by two opposing faces, "A" and "B" of a cuboid. Adapted from Downes et al. [49].

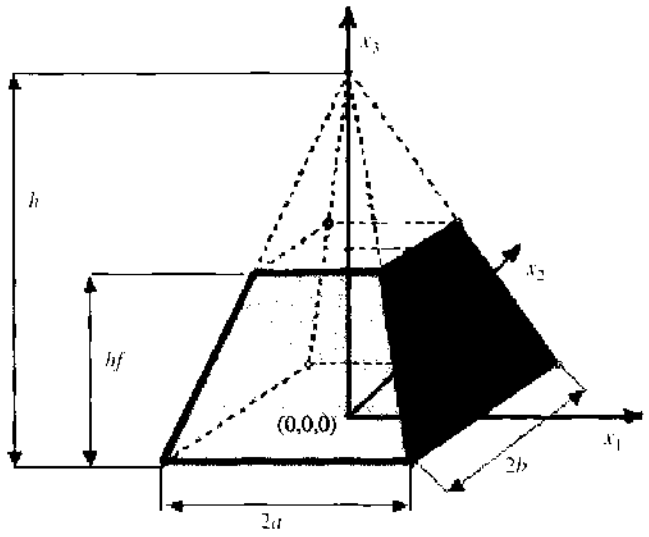


Figure 10. The geometry of the pyramidal quantum dot considered by Pearson and Faux [52] is illustrated; "f" represents the degree of truncation of the pyramid. Adapted from Pearson and Faux [52].

Here, f is the degree of truncation of the pyramid (see Fig. 10). With the volume of integration given by Eq. (29), and using the method developed in Ref. [60] (described in the preceding pages), Pearson and Faux [52] arrived at complex but, nevertheless, closed-form solutions for the stress distribution of pyramidal shaped and illustrated their results for the InAs-GaAs system (relying on the experimental data from Fry et. al. [53] for the geometry in Fig. 10). They also investigated QDs with nonuniform composition. In such a case, the pyramidal QD is sliced into a large number of smaller truncated pyramids, and the small slices are assumed to have uniform composition. Several aspects of their results are of interest. As the reader will gather from Fig. 11 (specific to InAs-GaAs), the magnitudes of the strain components were found to be the largest at the QD/matrix interface (at $x_3 =$ 55 nm), especially at the vertices with the strain attenuating rapidly in the barrier material. This is consistent with the notion that classical continuum elasticity admits a singular solution close to the vertices. Also, the strain distribution in the matrix material immediately below the square face remains unchanged for different truncations (f) considered because the base of the pyramidal dot remains unchanged. ε_{11} and ε_{22} are compressive through out the dot though lesser in magnitude than the initial misfit strain of -6.7% owing to strain relaxation. Outside the dot, ε_{33} is compressive within the barrier material owing to the QD relaxing outward along x_3, compressing the local host material in the process. Interestingly, as a

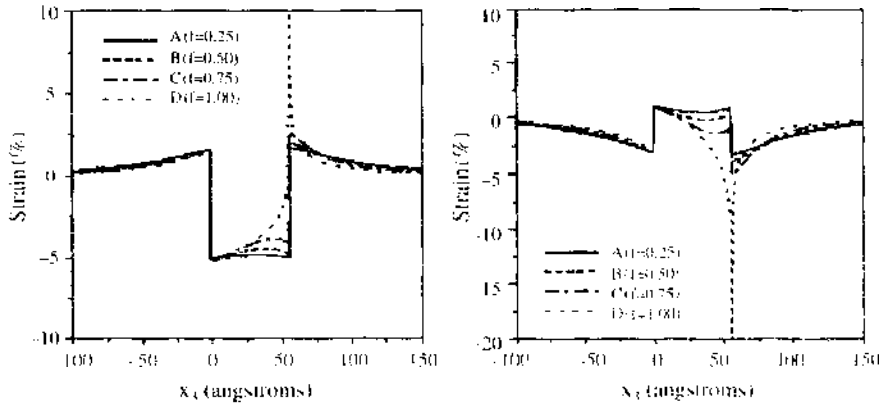


Figure 11. Illustration of the variation of ε_{11} (left) and ε_{33} (right) traced along x_3 obtained by analytical calculations on a pyramidal quantum dot (based on actual InAs-GaAs quantum dot geometry because of Fry et al. [53]) by Pearson and Faux [52]. The height of the dot is 55 A. (The base of the dot is located at $x_3 = 0$ and the apex at $x_3 = 55$ angstroms). Lines A, B, C, and D represent the different truncation factors "f" considered in the above analysis. Reproduced with permission from [52]. Pearson et al., *J. Appl. Phys.* 88, 730 (2000). Figs. 2, 6, © 2000, American Institute of Physics.

result of this, ε_{33} tends to overrelax inside the dot with tensile strains existing in QDs with low truncation factors (compare with Fig. 11). However, Pearson et al. [52] note, as the truncation factor is made higher, the Poisson effect of ε_{11} and ε_{22} is much reduced which results in a compressive ε_{33} in the QD.

In the analysis of Pearson and Faux [52], dots with graded composition exhibited smaller strains at the base because the mismatch strain is lower at the base for this case (compared against dots with uniform composition of InAs). The other trends follow patterns similar to the instance of QDs with uniform InAs composition discussed hitherto. In yet another case of an array of QDs discussed in Pearson et al. [52], superposition of tensile stresses of adjacent material results in further relaxation of the compressive strain component ε_{11} of the QD in the center of the array. ε_{33} is always compressive within the dot in an array (compare this with the isolated quantum dot) due to the compressive stresses of the surrounding material superposing with the small tensile stress.

Davies [54] drew an analogy with the Poisson equation of electrostatics and the lattice mismatch induced strain problem in quantum dots. In a treatment that can be easily related to that of Eshelby [30] as well as Downes et al. [49], he formulated solutions of spherical, cuboidal and pyramidal dots. In this analogy, lattice mismatch plays the role of charge density. Alternatively, as discussed by Eshelby [30], the mismatch transformation strain may be considered the density of an astronomical body while the dilation is related to the Laplacian of the "gravitational" potential of the quantum dot or inclusion shape.

Several works [55–61] have resolved strain distribution in different shapes via numerical methods. For example, Shin et al. [61] used the finite element method and analyzed structures similar to those in Faux and Pearson's work [52]. Additionally, Shin et al. [61] also reported the change in strain distribution with change in dot truncation as a function of stacking period. Some representative pictures and results from Shin et al. [61] are shown in Figs. 12 and 13.

2.2.3. Atomistic versus Continuum Calculations of Strain Distributions in QD/QWRs

A few of the works have resorted to atomistic calculations for strain distributions [61–64]. Since by the very nature and size scale of the lattice mismatch embedded quantum dot problem, millions of atoms have to be considered, *ab initio* methods are computationally too intensive. Empirical force-field molecular dynamics must be resorted to. The accuracy and the value of empirical force field atomistic calculations lie in the choice of a suitable potential. The latter is either fit to available experimental data for the material under consideration or alternatively may be developed using *ab initio* methods. As pointed out by Pryor et al. [58], atomistic methods are faithful to the true point symmetry of the material while analogous continuum models may not necessarily reflect this in their coarse-grained sense. Further nonlinearities (anharmonic effects) are automatically embedded in an atomistic formulation. (However, to be accurate, the potential must have been parameterized appropriately to account for anharmonic effects.)

The valence force field (VFF) [65] provided by Keating [66] and Martin [67] is perhaps the simplest. Anharmonicity is inadequately accounted for in this approach and only nearest neighbors are accounted for. Stillinger-Weber [68–71] potentials and Tersoff potentials are more involved. Stillinger-Weber potentials have been applied to Si/Ge QDs [69–71]. The Tersoff potential has been used for InGaAs/GaAs quantum dots [72]. Kikuchi et al. [190] have compared the VFF and Stillinger-Weber potential for the SiGe/Si system and concluded that for smaller quantum dots, the Stillinger-Weber potential may be better as VFF yields some physical results in the strain profiles (see also Ref. [191]).

Notably, Pryor et al. [58] have presented an interesting comparison between continuum elasticity and atomistic simulations (Fig. 14). For large mismatch strains (e.g., 7% strain mismatch for InAs/GaAs quantum dots) only minor discrepancies from continuum elasticity were found (Fig. 14). Other groups have similarly found that continuum elasticity for most cases provides reasonable answers, for all practical purposes, even at the monolayer level [63, 73].

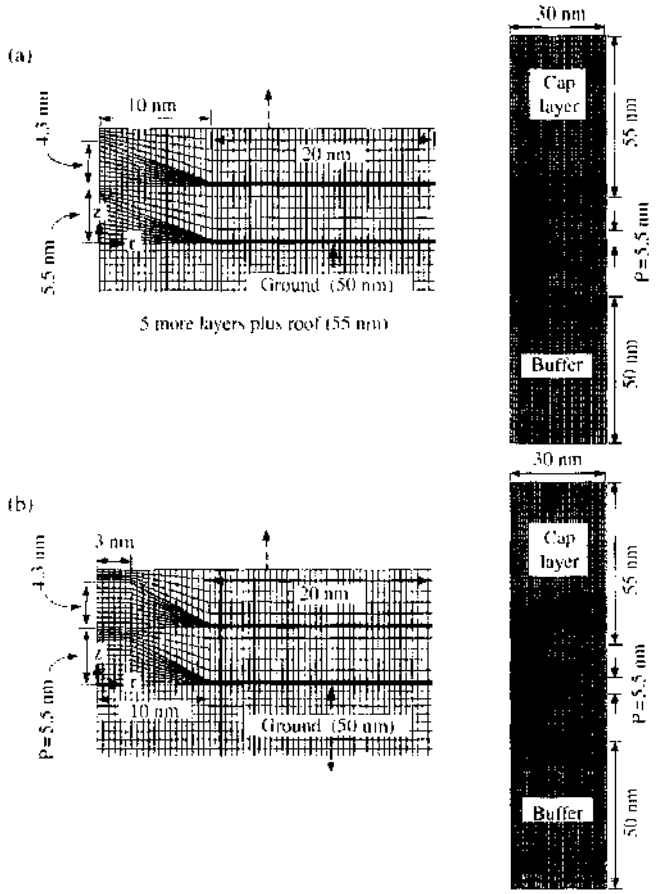


Figure 12. Shin et al.'s finite element model of the multilayer-stacked InAs/GaAs structure with (a) truncated (b) un-truncated quantum dots. Reproduced with permission from [61], Shin et al., *J. Phys.* 15, 3689 (2003), Fig. 1. © 2003, IOP Publishing Limited.

Tadić et al. [59] have also made a comparative study of strain distribution in cylindrical InAs/GaAs and InP/InGaP QDs as obtained from isotropic elasticity theory, anisotropic elasticity and atomistic simulations [42]. Davies' approach [54] outlined above was used for the isotropic case, while finite element method was employed for the anisotropic case. For the

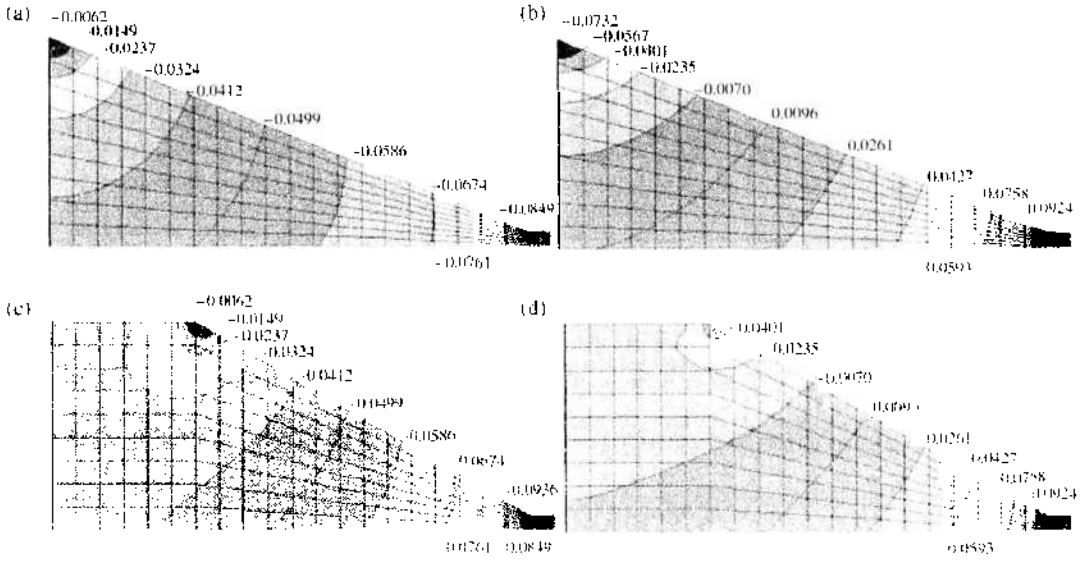


Figure 13. Comparison of the radial and axial strains along the midplane for (a) ε_{rr} contours for untruncated pyramidal quantum dost. (b) ε_{zz} contours for untruncated pyramidal quantum dots. (c) ε_{zz} contours for truncated quantum dots. (d) ε_{zz} contours for truncated quantum dots in the midlayers (of the array) with a stacking period of 7 nm. Reproduced with permission from [61], Shin et al., *J. Phys.* 15, 3689 (2003), Fig. 1. © 2003, IOP Publishing Limited.

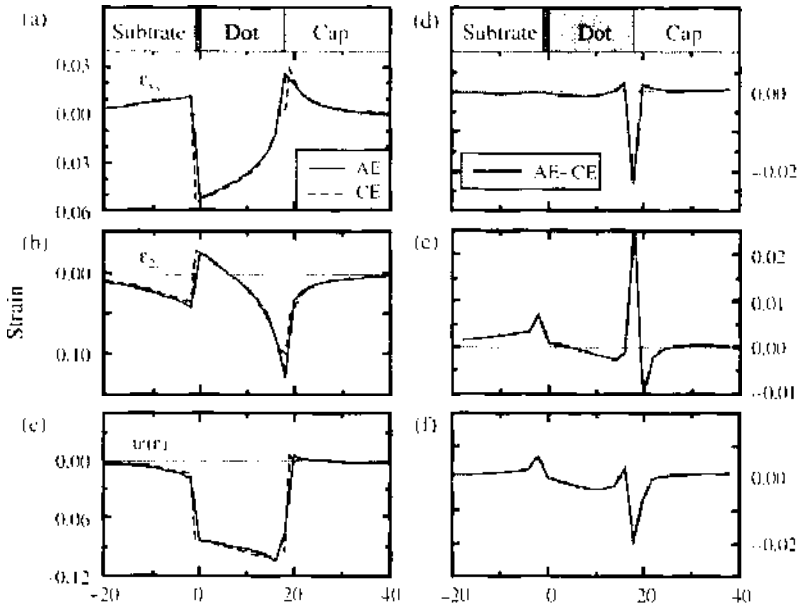


Figure 14. Strain components along the axis of a square InAs pyramidal quantum dot with [110] faces buried in GaAs host. Solid lines are obtained by valence force field (depicted as AE) and dashed lines are from continuum elasticity methods. (a)–(c) Strain distributions; (d)–(f) the difference between these two approaches. The scale for a–c differs from that for d–e. Reproduced with permission from [67], Pryor et al., *J. Appl. Phys.* 83, 2548 (1998), Fig. 5. © 1998, American Institute of Physics.

atomistic simulations, both VFF and Stillinger-Weber potentials were used and compared. Better agreement was found between the strain contours between the four methods employed for InP/InGaP QDs than in InAs/GaAs QDs. This observation was attributed to the smaller lattice mismatch in the InP/InGaP system. In particular, differences are only observed in regions where strain changes very rapidly, that is, the edges and corners of the structures.

2.2.4. Quantum Wires

Some of the early literature on shape dependence effects on quantum dots strain focused on quantum wires (QWR), which are one-dimensional analogues of quantum dots (zero dimensional). In particular, much emphasis was placed on solving problems for specific QWR or QD shapes. A strained QWR (Fig. 15) is a region of material that has two of its dimensions very small in comparison to the third. The cross-sectional size of the wire is of the order of a few tens of nanometers giving rise to quantum confinement of electrons in these two spatial dimensions. Eshelby's formalism [30] can be readily employed to tackle various shaped QWRs. A QWR with a uniform circular cross section (see Fig. 15) surrounded by an infinite host matrix happens to be a special case of an ellipsoidal inclusion in an infinite medium and Eshelby's well-known solution [30] for such ellipsoidal inclusions, discussed earlier, can be invoked to find the strain and displacement fields.

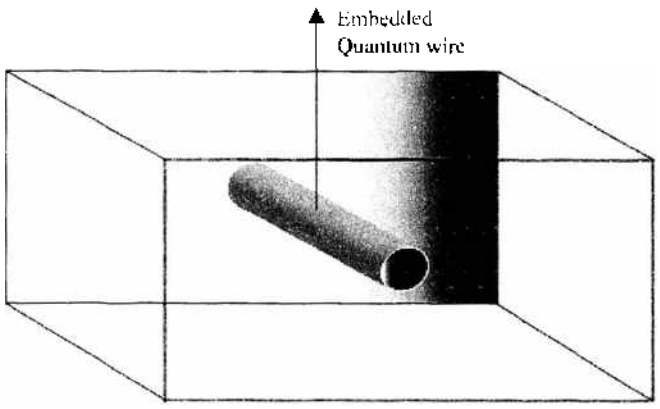


Figure 15. A quantum wire in the form of an infinite elliptical cylinder embedded in an infinite medium.

Notwithstanding early analytical work on thin films or islands present in the vicinity of a free surface [38–44], Downes and Faux [74] were among the first to explicitly deal with the problem of analytical strain-distribution calculation for buried strained layers; the case of a quantum wire with square cross section degenerating to a special instance of this more general problem. The problem addressed by Downes and Faux [74] is schematically depicted in Fig. 16, which illustrates a buried strained layer ($\boldsymbol{\Omega}$) surrounded by an infinite host matrix ($\mathbf{D}$). The dimension of the structure in the z-direction is assumed to be very large so that plane strain conditions apply.

Downes and Faux [74] proposed an approach similar to one already outlined for the three dimensional case (Downes et al. [49]). (However, we note that, chronologically, the two-dimensional work preceded the three-dimensional formulation even though we have discussed the latter in an earlier section.) They obtained (for the problem in Fig. 16)

$$\boldsymbol{\sigma}^{\text{layer}}(x, y) = \int_{l}^{+l} \int_{-c}^{+c} \boldsymbol{\sigma}^{\text{cyl}}(x - x_0, y - y_0)\, dy_0\, dx_0 \tag{31}$$

The stress fields for "point" cylindrical inclusion serve as Green's function for this problem. They are

$$\sigma_{xx}^{cyl} = \Upsilon(x^2 - y^2) \tag{32a}$$

$$\sigma_{yy}^{cyl} = \Upsilon(y^2 - x^2) \tag{32b}$$

$$\sigma_{xy}^{cyl} = \Upsilon(xy) \tag{32c}$$

$$\text{where,} \quad \Upsilon = \frac{\Lambda_2}{(x^2 + y^2)^2}; \quad \Lambda_2 = \frac{E\varepsilon^m}{2\pi(1 - \nu)}$$

Explicit analytical expressions (Eqs. [33]) have been provided by Downes and Faux [74] for a single strained layer after evaluating the integral in Eq. 31

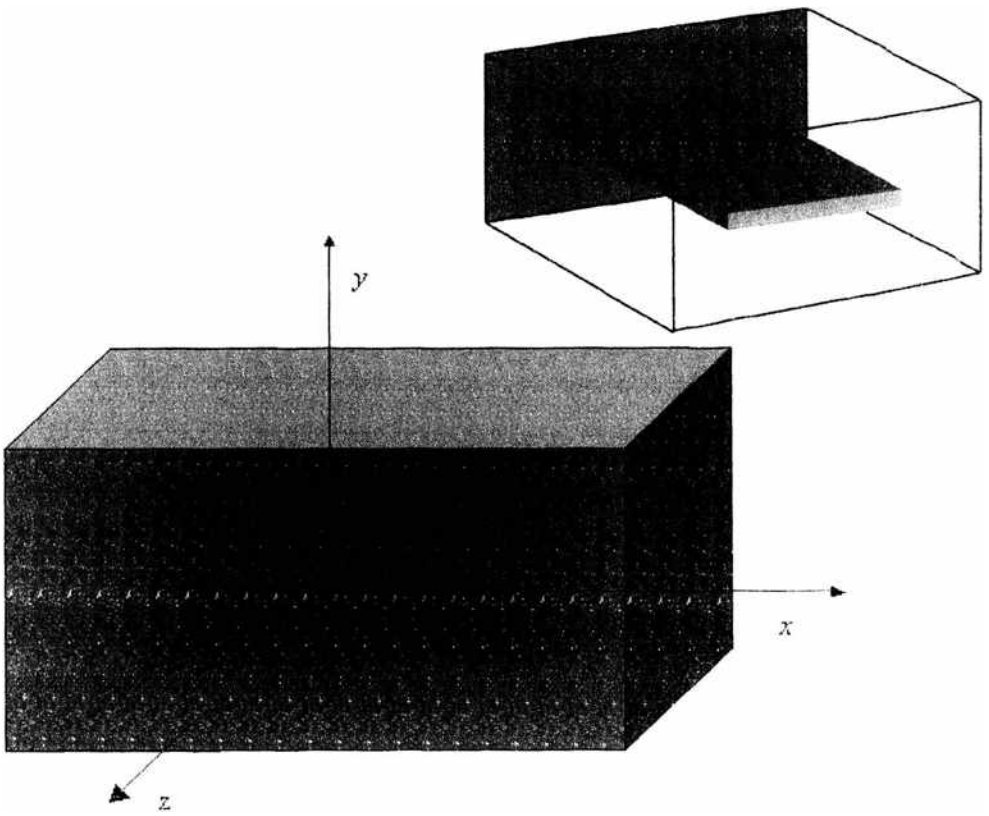


Figure 16. Illustration of a cross section through a buried strained layer. (Inset three-dimensional rendering of a strained layer). Adapted from Downes and Faux [74].

$$\sigma_{xx} = \Lambda_2\left[\sum \tan^{-1}\left(\frac{l \pm x}{c \pm y}\right)\right] \tag{33a}$$

$$\sigma_{xy} = \Lambda_2\left[\sum \tan^{-1}\left(\frac{c \pm y}{l \pm x}\right)\right] \tag{33b}$$

$$\sigma_{yy} = \frac{\Lambda_2}{2}\left[\sum p \ln\{(l \pm x)^2 + (c \pm y)^2\}\right] \tag{33c}$$

Once again, the summation is over various combinations of + and −. In Eq. (33c), p is +1 for odd number of plus signs and −1 for even number of plus signs.

The strained layer reduces to a quantum wire with square cross-section when l equals c (Fig. 16). From the expressions for the strain field in and around a quantum wire with a square cross section, one can deduce that only a small fraction of the lattice mismatch strain is accommodated across the cross-section of the wire, with considerable strain relaxation occurring along these shorter dimensions [75].

Quantum wires are often synthesized *in situ* as arrays and also occur in device structures (say, e.g., lasers, see Fig. 17) as periodic arrangements [76–79].

Considering a periodic array of trapezium-shaped wires buried in an infinite medium (Fig. 18), Gosling and Willis [80] tackled the array problem by simulating each of the wires to be an Eshelby-like inclusion. Modeling trapezium-shaped QWRs as inhomogenities (with elastic constants different than the surrounding host material) can considerably complicate calculations; consequently, same elastic constants are chosen for the quantum wires and the matrix. Further assuming all materials to be isotropic, and employing the elastic Green function and Fourier transformations, exact analytical expressions for the stress field were provided [80].

As can be inferred from Fig. 18, the case of a buried strained epitaxial layer can be recovered when one allows "w" to equal "p." Also by allowing $p \to \infty$, the case of an isolated wire can be examined. Gosling and Willis [80] applied these expressions to compare the structural stability and electronic properties of an array of wires with that of an isolated wire and found that for array periods of five wire widths or more the stability (see Ref. [81–87] for literature on stability issues in nanostructures) and band-gap characteristics of the periodic arrangement were similar to those of isolated wires.

Faux et al. [88] have employed the stress field because of a point cylindrical inclusion (similar to the way the point sphere stress solution was applied in the three-dimensional case [49]) as a "Green function" to calculate the stress field about an arbitrarily shaped QWR buried in an infinite medium. Analogous to the reduction of the three-dimensional volume integral to a surface integral in the case of arbitrary-shaped quantum dots [49], their scheme reduces the problem to the evaluation of a path integral around the boundary of

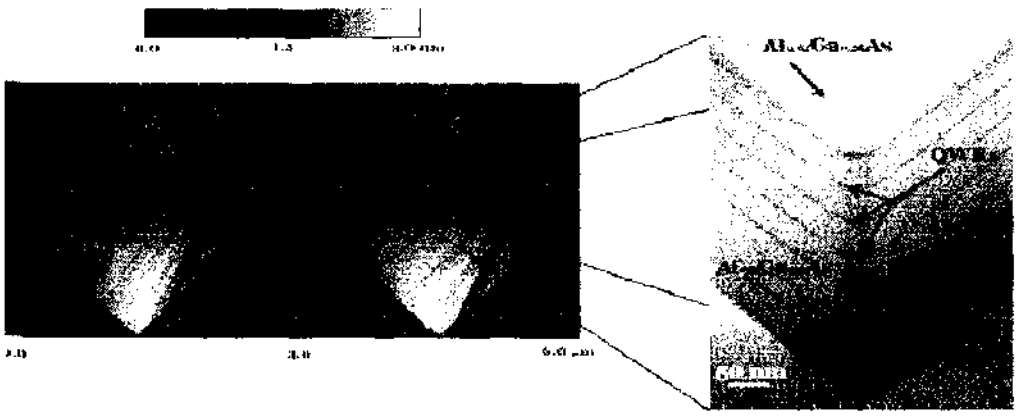


Figure 17. AFM cross-section image (left) of two V-groove QWR lasers separated by a distance of 3 μm. TEM cross-section image (right) of a single V-groove QWR laser with five vertically stacked QWRs in the waveguide core. Reproduced with permission from [79]. Weman and Kapon, www.ifm.liu.se/matephys/nanopto/QWires.html. © 2005, H. Weman.

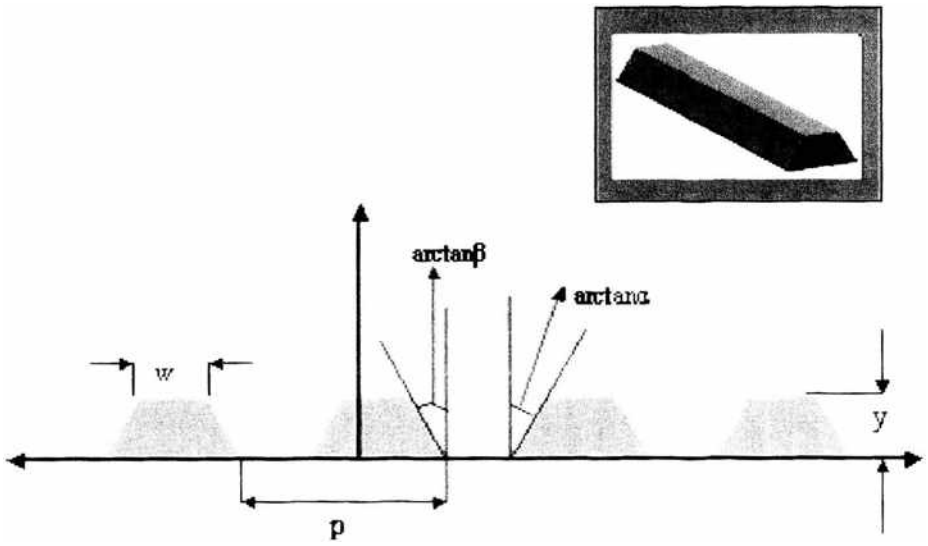


Figure 18. Schematic of a periodic array of trapezoidal quantum wires illustrating the geometry of the wire. "p" is the period of the array, "w" is the width of each wire at the apex, "y" is the height of the wire. The characteristic angles are also shown. (Inset: a three-dimensional rendering of the trapezoidal quantum wire). Adapted from Gosling and Willis [80].

the quantum wire. Consider the three vectors [88]:

$$\mathbf{A}_{xx} = -D_2 y\mathbf{i}; \quad \mathbf{A}_{yy} = D_2 x\mathbf{j}; \quad \mathbf{A}_{xy} = \frac{D_2}{2}(x\mathbf{i} - y\mathbf{j})$$
$$D_2 = \frac{\Lambda_2}{x^2 + y^2}; \quad \Lambda_2 = \frac{E\varepsilon'''}{2\pi(1-\nu)} \tag{34}$$

Similar to the three-dimensional case [49], the vectors of Eq. (34) are chosen such that $\nabla \times \mathbf{A}$ yields the cylindrical Green's function components σ_{xx}, σ_{yy}, and σ_{xy} which are given by Eq. (32).

With the aid of Eqs. (32) and (34), the stress distribution due to a quantum wire with arbitrary cross section was obtained by evaluating the following surface integral [88]:

$$\boldsymbol{\sigma}(\mathbf{r}) = \int \nabla \times \mathbf{A}(\mathbf{r} - \mathbf{r}') \cdot \mathbf{dS}(\mathbf{r}') \tag{35}$$

The integration is performed over the area of the QWR. With the application of Stokes theorem, this surface integral (Eq. [35]) was reduced to a line integral around the boundary of the wire [88]:

$$\sigma(\mathbf{r}) = \oint \mathbf{A}(\mathbf{r} - \mathbf{r}') \cdot d\mathbf{r}' \tag{36}$$

For simple geometries such as rectangular wires, the line integral in Eq. (36), can be performed analytically providing, for example, similar results to the strained layer case visited earlier. For QWRs with a rectangular cross section, Faux et al. [88] have noted that with a proper choice of the integrand, only the horizontal boundaries of the rectangular QWR can be made to contribute to σ_{xx} and only the vertical boundaries to σ_{yy}, which simplifies the computation further. Faux et al. [88] apply their methodology to an InP wire of triangular cross section embedded in GaAs host with a lattice mismatch strain, $\varepsilon''' = -3.7\%$. Nishi et al. [89] have also performed strain calculations on the exact same wire configuration using the finite element method. For the triangular cross section, some results are illustrated in Fig. 19 [88]. They are found to agree well with Nishi et al.'s [89] results.

As Faux et al. [88] have noted from Fig. 19, significant strain relaxation is found to occur over most of the area of the wire with the magnitude of the strain components ε_{xx} and ε_{yy} being less than about $\varepsilon'''/2$ at all points in the QWR, except at the apex of the triangle. Further, ε_{xx} is least relaxed at the base resulting in over-relaxation of ε_{yy}, which is manifest

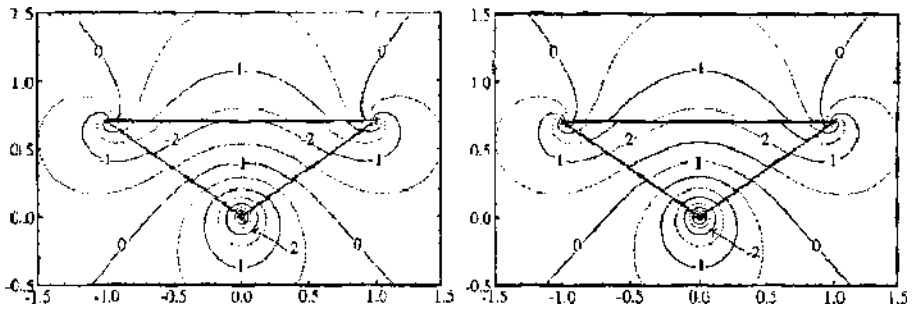


Figure 19. Strain contours (in % units) ε_{xx} (left) and ε_{yy} (right) for an InP triangular wire within an infinite GaAs matrix. The initial misfit strain is assumed to be 3.7%. Reproduced with permission from [88], Faux et al., *J. Appl. Phys.* 80, 2515 (1996). Fig. 1, © 1996, American Institute of Physics.

in the positive values that ε_{yy} assumes. In agreement with inclusion theory, the trace of the strain tensor is constant for all points in the inclusion.

For complicated shapes the QWR boundary may be discretized and the contour integral can be converted into a summation. Arbitrary shapes can be approximated to be composed of a combination of line elements and circular arcs. In some cases, this decomposition may be exact. By employing the simple expressions for line elements, and so forth, contributions due to a line element and a circular arc, stress/strain evaluations can be obtained. In fact, precisely on these lines, Downes et al. [90] in a subsequent work have provided analytic expressions for the strain field due to a lattice-mismatched QWR whose cross section is composed of any combination of line elements and circular arcs. The authors applied this methodology to GaAs/AlGaAs crescent-shaped wire [90]. Typically, crescent-shaped quantum wires are grown *in situ* using organometallic chemical vapor deposition of thin layers on V-grooved substrates. The position of the QWR is germinated at the position of the initial groove on the patterned substrate, leading to highly self-ordered wire arrays.

As illustrated in Fig. 20, the wire was approximated as consisting of two lines and two arcs [90] (the thin quantum wells on the valley sides (not shown in Fig. 20) are usually ignored). For a given radius of curvature of the lower arc and a given thickness of the QWR (this data obtained from Kapon et al. [76] for the GaAs/AlGaAs system), the radius of curvature of the upper arc can be arrived at (using geometry) and the strain contours inside and outside the QWR were obtained (Fig. 21).

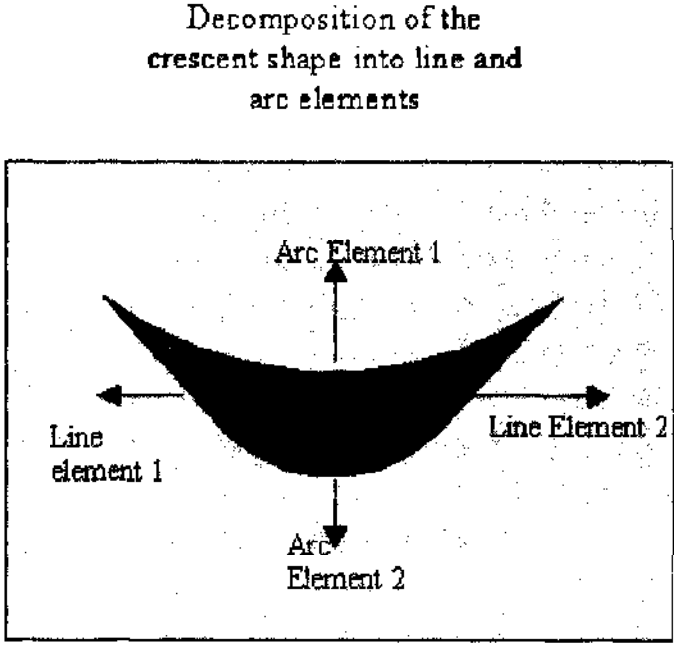


Figure 20. Decomposition of the boundary of a crescent shaped wire into line and arc elements. Adapted from Downes et al. [90].

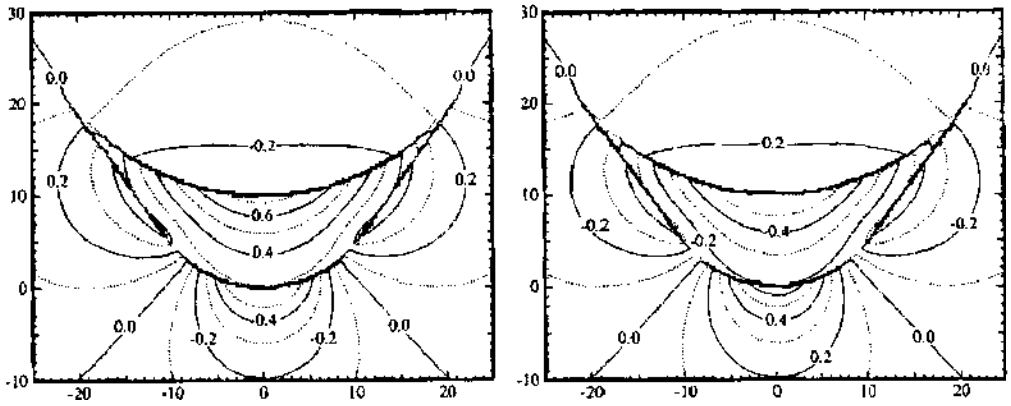


Figure 21. Strain contours in units of percent indicating ε_{xx} (left) and ε_{yy} (right) for a GaAs/AlGaAs crescent-shaped wire (geometry of wire based on Kapon et al. [76]) with initial misfit strain of 1%. Reproduced with permission from [90], Faux et al., *J. Appl. Phys.* 82, 3754 (1997), Fig. 6. © 1997, American Institute of Physics.

As can be observed in Fig. 21, Downes et al. [90] noted that, for an initial misfit strain of 1% (the wire is initially under tensile strain), the maximum strain in the x-direction is about 0.75% (at the center and near the top of the wire, where the width is the largest). ε_{yy} is negative within the QWR. Expectedly, the dilation is constant inside the wire (zero outside the wire). In a rectangular QWR with a high-aspect ratio, the strain relaxation occurs predominantly in one dimension (ε_{yy} would be about -0.9% and ε_{xx} would tend to ε_0). The crescent-shaped QWR has smaller peak strain values because strain relaxation occurs in two dimensions.

Provided wires of the same thickness are considered, on incorporation of thin lateral quantum wells, the strain fields have been observed to be similar [90] except at the corners of the structure. This similarity finds reason in the fact that as the QWs are thin, the contour integral contribution along the lower line of the well is almost cancelled by the contribution because of the upper line.

We have seen that almost all the analyses discussed far assume identical elastic constants for the quantum dot and the barrier material, which begets the question whether this kind of treatment is justified. Clearly, for the ellipsoidal shape quantum dots embedded in an unbounded matrix, Eshelby's equivalent inclusion method [30] provides an easy recipe to account for modulus mismatch. More generally, this is not easy. So are we justified in using the same elastic constants for both the materials? Downes and coworkers [90] have outlined the following issues to consider: (1) Two different materials, at least, are used. (2) The materials are generally semiconductor alloys. (3) At least one of the materials is in a state of strain (It is difficult to estimate the elastic constants for alloys in a state of strain.) Downes and co-workers [90] argue that Keyes scaling relationship [91–92] for III–V semiconductors suggests that the it is appropriate to choose the elastic constants of the barrier material for all materials in the system because all materials have the same lattice spacing before misfit strain relaxation takes place. As an illustration, the elastic constants for a strained InAs QD in a GaAs matrix will be closer to the GaAs values than those for unstrained InAs. In the present authors opinion, this issue requires further investigation. If necessary, the modulus mismatch can be accounted for as a perturbation (see, for example, Andreev et al. [93]).

2.3. Effect of Presence of a Free Surface in Near Vicinity

Because of their mode of fabrication and operation, QDs/QWRs are most often buried fairly close to a free surface rather than being embedded in an "infinite" matrix where the host material boundaries play no role. Such "shallowly" buried quantum dots act like stressors and the resulting elastic fields coerce vertical dot ordering during growth of subsequent layers of self-assembled quantum dots. (Lateral ordering of dot layers is known to occur in the presence of regular dislocation arrays or buried strained layers [94–95].) Furthermore, both qualitatively and quantitatively, the strain behavior of embedded QD/QWRs that are located close to a free surface is quite different from one embedded in an infinite medium For example, the famous Eshelby rule that an ellipsoidal shaped inclusion with a uniform

misfit will admit a uniform elastic state no longer holds true. For some simple shapes and cases, closed-form results are available although given the complications in satisfying the traction free boundary conditions at the host material boundaries, resort must be made toward numerical methods. As already indicated, typically, if the distance of a quantum dot from the free surface is larger than $3R$ (in case of a spherical dot), an infinite space solution may be invoked with reasonable accuracy. Otherwise, the so-called half-space solution is recommended. The schematic of the semi-infinite domain (defined by $x_3 \geq 0$) is shown in Fig. 22 below:

In the context of half-space, Mindlin and Cheng's seminal work [96–97] addressed a center of dilatation. This solution can be elegantly expressed in terms of the thermoelastic potentials of an infinite solid [98]. Following Mindlin and Cheng [96–97], we briefly discuss the displacement field of a spherical dot buried at depth, h, from the free surface. Isotropicity is assumed and both the QD and the matrix are assumed to have identical elastic moduli.

The relation for displacement $\mathbf{u}(\mathbf{r})$ in a semi-infinite region $0 \leq z$ with a free surface at $z = 0$ as provided by Mindlin and Cheng is [97]:

$$4\pi\mathbf{u}(\mathbf{r}) = -\nabla\phi - \nabla_2\phi_2 \tag{37}$$

where ϕ is the harmonic potential of the infinite system given by Eq. (24) and ϕ_2 is the reflection transform of ϕ in the plane $z = 0$. For an *infinite* solid Goodier [98–99] showed that (for purely dilatation axisymmetric problems) the displacement potential obeys the following Poisson's equation:

$$\nabla^2\phi = -4\pi\frac{1+\nu}{1-\nu}\varepsilon^m = -4\pi\frac{9K}{3K+4\mu}\varepsilon^m \tag{38}$$

The operator ∇_2 is given by [97]

$$\nabla_2 = (3-4\nu)\nabla + 2\nabla z\frac{\partial}{\partial z} - 4(1-\nu)\hat{e}_z\nabla^2 z \tag{39}$$

The radii to a point of interest (x, y, z) from $(0, 0, h)$ and its image point $(0,0, -h)$ are, $r = [x^2 + y^2 + (z-h)^2]^{1/2}$ and $r_2 = [x^2 + y^2 + (z+h)^2]^{1/2}$, respectively. Thus ϕ and ϕ_2 are [97]

$$\begin{aligned} \phi &= \frac{4\pi a^3\varepsilon^m}{3R}, \quad \phi_2 = \frac{4\pi a^3\varepsilon^m}{3R_2} \quad \text{for } r \geq a \text{ (exterior to the dot)} \\ \phi &= \frac{2\pi\varepsilon^m(3a^2 - R^2)}{3}, \quad \phi_2 = \frac{4\pi a^3\varepsilon^m}{3R_2} \quad \text{for } r < a \text{ (points inside the dot)} \end{aligned} \tag{40}$$

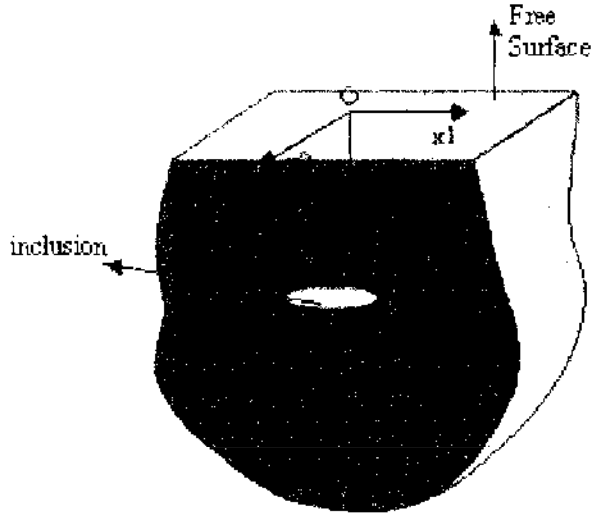


Figure 22. Illustration of a semi-infinite space with a plane free surface ($z < 0$). An embedded inclusion can also be seen.

From Eq. (26), we have the solutions for displacements as [97]

$$\mathbf{u}_e = \frac{a^3\varepsilon^m}{3}\left[\frac{\mathbf{r}}{r^3} + \frac{(3-4\nu)\mathbf{r}_2}{r_2^3} - \frac{6z(z+h)\mathbf{r}_2}{r_2^5} - \frac{2\mathbf{k}}{r_2^3}\{(3-4\nu)(z+c)-z\}\right] \quad \text{for } r > a$$
$$\mathbf{u}_i = \mathbf{u}_e + \frac{a^3\varepsilon^m\mathbf{r}}{3}\left(\frac{1}{a^3} - \frac{1}{r^3}\right) \quad \text{for } r < a \tag{41}$$

Where $\mathbf{u}_e$ is the displacement field for an exterior point, and $\mathbf{u}_i$ is the displacement field for a point inside the quantum dot. The strain field can be recovered in the usual manner from strain-displacement relationships.

Davies [100] has extended Mindlin and Cheng's [97] approach. He showed that the displacements and strains for the semi-infinite solid can be, in a rather simple manner, expressed in terms of those for an infinite solid and their corresponding derivatives normal to the surface. He finds the following expression for displacement [100]:

$$u = u^\infty + (3-4\nu)\bar{u}^\infty + 2z\frac{\partial}{\partial z}(\bar{u}_x^\infty, \bar{u}_y^\infty, -\bar{u}_z^\infty) \tag{42}$$

where

$$\bar{u}^\infty(x, y, z) = u^\infty(x, y, -z) \tag{43}$$

The displacement at the free surface, which can be obtained by setting $z = 0$, is found to be increased by a factor of $4(1-\nu)$ compared to the displacement of the same plane in an infinite solid. The same factor relates the strains at the free surface: ε_{xx}, ε_{yy}, and ε_{xy} to their counterparts in the infinite sample. As noted by Davies [100], part of this enhancement occurs because the QD/QWR is less contained by a free surface than in an infinite solid but the remaining part is due to less obvious effects arising out of distortion of the surrounding which pushes the inclusion toward the surface. A dilation of $-4(1-2\nu)\bar{\varepsilon}_{zz}^\infty$ arises due to the presence of the free surface. Unlike the infinite medium case, the dilation is neither uniform within the QD/QWR nor zero outside it.

Alternative (although essentially equivalent) methods are also available to solve such semi-infinite problem for shallowly embedded QD/QWRS. Many problems of elasticity can be simplified by reducing to force distribution problems over some region, which then can be solved through use of Green function techniques. The Green function for the half-space (Eq. [44]) (corresponding to Fig. 22) was derived by Mindlin [101]:

$$G_{ij}(\mathbf{x}, \mathbf{x}') = \frac{1}{16\pi\mu(1-\nu)}\left[\begin{array}{l} \dfrac{3-4\nu}{r}\delta_{ij} + \dfrac{1}{r^2}\delta_{ij} + \dfrac{(x_i-x_i')(x_j-x_j')}{r^3} + \\ \dfrac{(3-4\nu)(x_i-x_i')(x_j-x_j')}{r_2^3} + \dfrac{2x_3x_3'}{r_2^3}\left\{\delta_{ij} - \dfrac{3(x_i-x_i')(x_j-x_j')}{r_2^2}\right\} + \\ \dfrac{4(1-\nu)(1-2\nu)}{r_2+x_3+x_3'}\left\{\delta_{ij} - \dfrac{(x_i-x_i')(x_j-x_j')}{r_2(r_2+x_3+x_3')}\right\} \end{array}\right] \tag{44}$$

In Eq. (44), the effect of the free surface occurs through the terms containing r_2. Subsequently, Seo and Mura [102] and Chiu [50] employed the half-space Green's function to obtain the Eshelby tensor for dilating ellipsoidal and cuboidal inclusions respectively (in a manner similar to Eq. [14]). For the case of an ellipsoid, closed form expressions are possible only when two of the semiaxes of the ellipsoid are equal [35].

On the lines of the preceding discussion, an interesting work is due to Glas [44]. Consider a uniformly thick layer of a semiconducting material coherently deposited on a substrate and covered by a capping layer of the same substrate material (Fig. 23). Such embedded layers frequently develop nonuniform mismatch strains either due to spatial distribution of temperature or variations in the composition of the quantum well material.

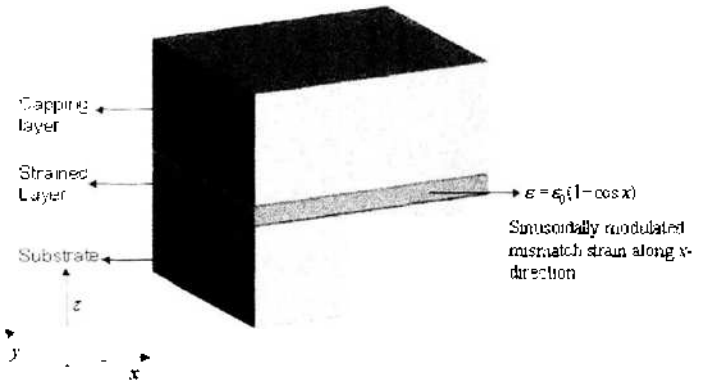


Figure 23. Illustration of a buried strained layer, with a sinusoidally modulated dilational mismatch strain, in the presence of a free surface as considered by Glas [44].

Glas [44] using an Eshelby-like cutting and welding *gendaken* combined with the method of images (to account for the boundary conditions) obtained an analytical solution to a sinusoidally modulated dilatational mismatch strain problem. The relaxation of a general misfit dilational inhomogenity was obtained in the form of an integral, which can be evaluated when the stress-free strain admits a simple Fourier transform. Glas [44] applied his method to a coherent parallelepipedic inclusion (rectangular quantum wire) and to a step formed at the interface between a substrate and a coherent capping layer to arrive at fully analytical solutions.

Further, in a later work, Glas [103] also provided a closed-form solution for the strain fields of truncated fourfold pyramidal QDs and trapezoidal QWRs buried in half-space. To solve the problem, the pyramidal QDs were decomposed into elementary cuboids of infinitesimal height and varying areas. The displacements obtained from each of them were then summed up to obtain the net displacements by superposition principle. Glas [104] extended his solution of a single QD to the study of buried QD arrays and later periodic arrays of QWRs in half-space [105]. Starting with the cases of single and periodic trapezoidal wires in half space, the calculation is extended to QWRs of arbitrary polygonal section. Analytical formulae for strain fields have also been found for a right-angled triangular section lying in half-space using results for those of a semi-infinite rectangular slab and a semi-infinite beveled slab.

Barabash and Krivoglaz [106] employed Fourier transform of the displacement filed due to a random distribution of point inclusions submerged under a plane with a given density-depth profile. The relaxation of the displacement was determined as a one-dimensional Laplace transform of the inclusion density.

Romanov et al. [107] compared the dilation of strain obtained for cuboidal and trapezoidal QDs using finite element method with analytical expressions provided for an spheroidal inclusion (with the same volume as the trapezoidal and cuboidal inclusion) [92] buried at identical depths. Except at the interface between the dot and the matrix (i.e., in the immediate vicinity of the dot), the spheroidal inclusion model provided good approximation to the finite element models of the cuboidal and trapezoidal inclusions (for both isolated and periodic inclusion models) even at significantly shallow levels of submergence, following which Romanov et al. [107] have suggested using the ellipsoidal inclusion model to obtain reasonable (if not very accurate) estimates of stresses for most geometries of dots.

2.4. Effect of Material Anisotropy and Nonlinearity

2.4.1. Anisotropic Effects

Most semiconductor compounds crystallize to cubic crystallinity. Against the isotropic value of 1, the anisotropy coefficient (defined as C_{11}-$C_{12}/2C_{44}$) for most III–IV semiconductors is around 0.5. For some cases, the assumption of elastic isotropy may be justified or alternatively the uncertainty in other material and configurational parameters (e.g., lattice parameters, dimensions, etc.) may far exceed the error due to neglect of the anisotropic effects. In

fact, some authors suggest that the isotropic approximation is reasonable for some cases of the two-dimensional problems [108]. (Caution should be exercised in taking this statement too literally and broadly. The reader is referred to Ref. [108] for further details.)

Explicit analytical expressions for the anisotropic Green's functions are unavailable (except for hexagonal symmetry) so the oft-used technique of reducing an elastic problem to a force-distribution problem cannot be employed readily. For the cubic anisotropic case, approximate solutions for the point force equation can be realized perturbatively and thus are applicable only for weakly anisotropic materials. Dederichs and Liebfried [109] have investigated different approaches for the estimation of approximate cubic Green's function [see also Ref. 110–112].

Starting with the point-force formalism of Green's functions and assuming identical elastic constants for both the QD and the host material, Andreev et al. [93] suggested a general analytical method for the Fourier transform of the anisotropic Green function and presented explicit results for cubic crystals (in Fourier space). With this analytical result in hand, the expression of the Fourier transform of the strain tensor was also obtained. In their results, the shape effect of the quantum dot structure appears as the Fourier transform of a characteristic shape function. Equation (45) provides the Fourier-space strain solution [93]:

$$\tilde{\varepsilon}^{\,\prime}_{ij}(\boldsymbol{\xi}) = \varepsilon^{m}\tilde{\chi}_{QD}(\boldsymbol{\xi})\left\{\begin{array}{l}\delta_{ij} - \dfrac{(C_{11}+2C_{12})\xi_i\xi_j/\xi^2}{1+(C_{12}+C_{44})\sum \xi_p^2/C_{44}\xi^2 + C_{an}\xi_p^2} \\ \times\dfrac{1}{2}\left[\dfrac{1}{C_{44}+C_{an}\xi_i^2/\xi^2}+\dfrac{1}{C_{44}+C_{an}\xi_j^2/\xi^2}\right]\end{array}\right\} \tag{45}$$

$\tilde{\chi}_{QD}(\boldsymbol{\xi})$ is the transformed characteristic shape function of the quantum dot structure defined by

$$\begin{aligned}\chi_{QD}(\mathbf{r}) &= 1, \quad r \text{ being a point inside the inclusion} \\ &= 0, \quad r \text{ being a point outside the inclusion}\end{aligned} \tag{46}$$

and C_{an} is given by

$$C_{an} = C_{11} - C_{12} - 2C_{44} \tag{47}$$

In the isotropic limit, $C_{an} = 0$ and the expression for the transformed isotropic strain simplifies to [93]:

$$\tilde{\varepsilon}^{\,iso}_{ij}(\xi) = \varepsilon_0\tilde{\chi}_{QD}(\xi)\left(\delta_{ij} - \frac{3\lambda+2\mu}{\lambda+2\mu}\xi_i\xi_j/\xi^2\right) \tag{48}$$

Transforming this strain back into real space and taking the trace, we recover the isotropic result discussed earlier (Andreev et al. [93] calculate the subtracted strain; that is, the mismatch strain is subtracted from the actual elastic strain. We have modified their expression to show the latter.):

$$Tr(\boldsymbol{\varepsilon}) = \varepsilon^{m}\frac{9K}{3K+2\mu}\chi_{QD}(\mathbf{r}) \tag{49}$$

Equation (49) emphasizes the constancy of dilation within the QD and that it vanishes for points outside it. As evident from Eq. (45), this may not hold true for all shapes in the cubic anisotropic case.

Andreev et al. [93] also provide a comparison between isotropic and anisotropic models for cubic crystals. As mentioned earlier the degree of anisotropy can be characterized by the anisotropic coefficient, which is about 0.5 for III–IV semiconductors. As an example, the effective Poisson's ratio varies between 0.333 in the (001) direction to about 0.2 in the (111) direction. This variation in Poisson's ratio might be expected to result in a strong dependence of strain on the space direction. However, this was not found to be the case. In most cases where the "anisotropy" in shape of the QD is more than or equal to the anisotropy in the elastic properties, the dominant contribution to the strain distribution is caused by the

"anisotropy" in shape rather than the anisotropy in elastic properties. Figure 24 illustrates how the anisotropic solution of the radial component of the strain tensor ε_{rr} for a spherical InAs/GaAs quantum dot system (misfit strain of −3.7%) plotted along three directions, compares with the isotropic case.

As Andreev et al. [93] have pointed out, the influence of elastic anisotropy is small within the dot with both models yielding nearly the same constant value. Outside the sphere, however, anisotropy considerations are observed to produce significant deviations from the isotropic model. As a matter of fact, the isotropic model effectively gives the strain distribution averaged over the different directions outside the dot. For a cubic dot, with the symmetry of the elastic properties being the same as that of the geometry, the effect of elastic anisotropy is even smaller than for the spherical dot. Unlike the spherical quantum dot, the difference between the two models for a pyramidal QD, exhibiting lesser cubic symmetry than the elastic properties, was observed to be less outside the dot than inside it.

While an exact expression for Green's function for the cubic case is not possible, it can be expressed in terms of a perturbation series where the first term is the isotropic Green's tensor and the subsequent terms are correction terms of increasing order, converging to the full anisotropic result [113–114]. Faux and Pearson [115] exploited this notion and presented explicit (but approximate) Green's tensors to permit rapid calculation of strain distribution in and around QDs. Computational effort is significantly reduced as excellent agreement was found with results of Andreev et al. [93] even with a first-order correction.

Pan and Yuan [116] expressed the half-space anisotropic Green's function, for traction-free surface, as a sum of the infinite space Green function and an image part based on the extended Stroh formalism. The image part can be expressed as a one-dimensional integral. Using these Green functions, the elastic fields of a quantum dot can be expressed as simple integrals over the surface of the dot employing Betti's reciprocal theorem. If the QD is a point source, then the strain fields can be expressed in analytical form. Other notable works on Green's functions (assuming traction-free surface conditions) are due to Refs. [122, 126–128].

Pei et al. [122] investigated the elastic fields due to an array of lens-shaped anisotropic QDs buried under a capping layer of finite thickness using a three dimensional finite element method. The various geometrical parameters used in the model have been shown in Fig. 25. Values of anisotropy ratio $A(= 2C_{44}/(C_{11}-C_{12}))$ ranging from 0.25 to 4.0 were

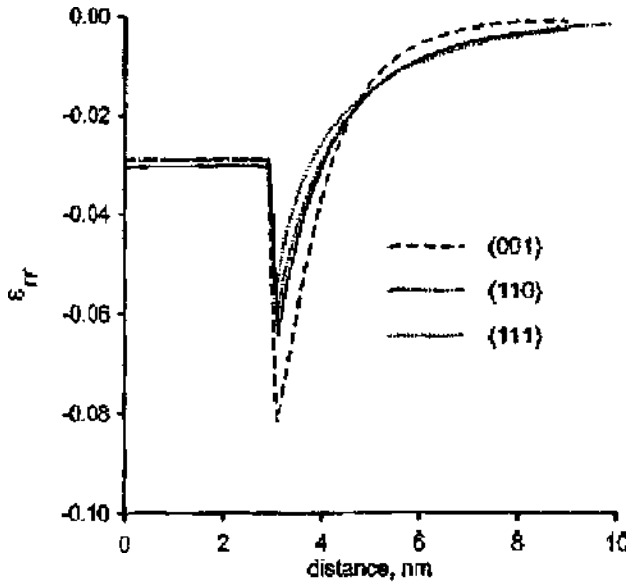


Figure 24. Radial strain tensor component ε_{rr} for different directions through a spherical dot. The radius of the dot considered is 3nm and the center of the QD is the origin of the coordinate system used. The misfit strain is 6.7% and parameters for GaAs have been used. The solid line represents the isotropic model while the dotted dashed lines represent anisotropic models. Reproduced with permission from [93], Andreev et al., *J. Appl. Phys.* 86, 297 (1999). © 1999, American Institute of Physics.

considered both for the QDs and the matrix. The ratios of the cap-layer thickness to dot height considered were 2.0, 3.0, 4.0, 5.0, and 6.0, respectively.

From Fig. 26, Pei et al. [122] noted that the value of ε_{xx} contour increases with A changing from 1.0 to 4.0 while it decreases with A changing from 1.0 to 0.25. Also, when $A > 1$, [100] and [$\bar{1}$00] are the "elastically soft" directions with the strain decaying rapidly in these directions. These same directions become the "elastically hard" directions when $A < 1$. For different values of A, ε_{xx} has a positive value in the matrix and increases from zero to a local maximum at the interface between the matrix and the wetting layer. In the QD region, it changes sign and its absolute value reduces till one reaches the interface between the QD and the capping layer. In the cap layer, ε_{xx} changes its sign again to being positive. Compressive strain increases significantly with increase in A from 0.25 to 4.0. The ε_{xx} profile is also significantly affected due to anisotropy considerations.

2.4.2. Nonlinear Effects

Generally, elastic material properties are assumed to be independent of strain. Ellaway and Faux [123] examined the accuracy of this assumption by investigating the behavior of elastic stiffnesses of InAs under uniform strain using atomistic methods. By using the three distortions as advised by Mehl [124–126] the three independent elastic stiffnesses C_{11}, C_{12}, and C_{44}, for a cubic crystal, were determined. Small distortions were applied to the crystal over a range of volumetric strains to evaluate the effective elastic stiffness at each strain level. For the atomistic simulation, Stillinger-Weber potentials [68] were employed using the parameters for InAs provided by Ichimura [127]. The elastic stiffnesses C_{11} and C_{12} increased with volumetric strain demonstrating that materials tend to become harder on being compressed. C_{44} was however found to hardly vary with volumetric strain [123]. Clearly, the strain dependence of elastic parameters renders the elastic boundary value problem nonlinear. After proper accounting for strain dependent elastic parameters, Ellaway and Faux [123] found significant difference in the hydrostatic strain calculation for a spherical InAs QD embedded in an GaAs matrix.

Ellaway and Faux [128] also investigated the effect of volumetric strain on the degree of anisotropy and on the two-dimensional (biaxial) Poisson's ratio. The anisotropy coefficient was found to have very weak dependence on the volumetric strain while the two-dimensional Poisson's ratio was found to increase with pressure showing good agreement with the estimate provided by Frogley et al. [129–130]. Lepkowski and Majewski [131–132] investigated the pressure dependence of elastic constants in zinc-blende InN, GaN, and AlN using

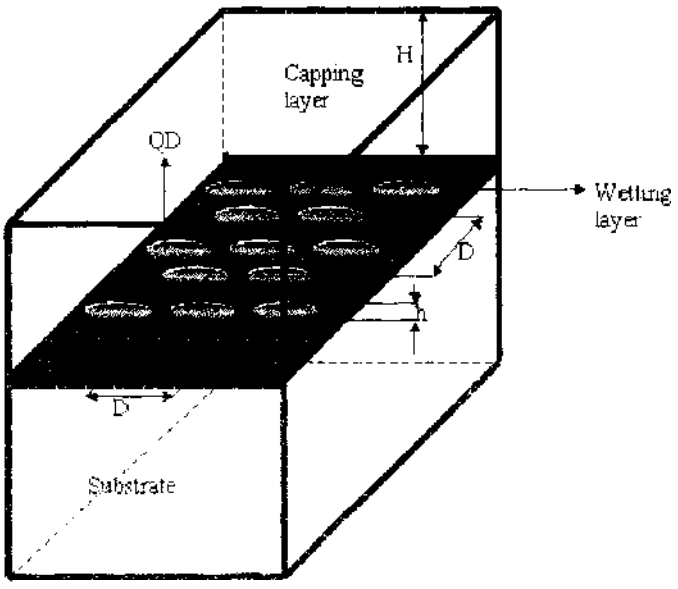


Figure 25. Schematic of an array of lens-shaped QDs. Distance (D) is 45 nm; thickness of the wetting layer (WL) is 1 nm. The base diameter (d) is 24 nm, while the height of the QD (h) is taken to be 6 nm. A lattice mismatch strain of 4 % is assumed. Adapted from Pei et al. [122].

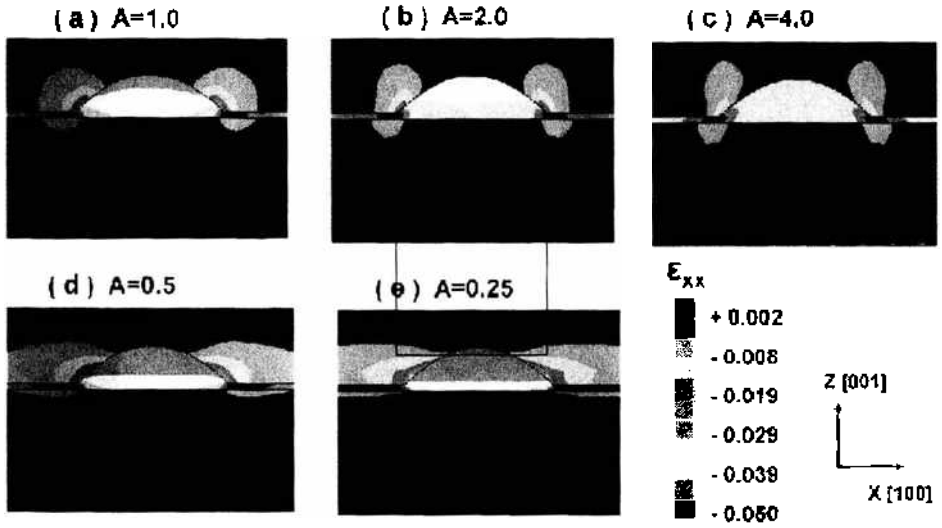


Figure 26. Contours in the midplane of one of the lens-shaped dots from Fig. 25 for different values of anisotropy ratio A. Reproduced with permission from [122]. Pei et al., *J. Appl. Phys.* 93, 1487 (2003). © 2003, American Institute of Physics.

density functional formulism and demonstrated results similar to previous works discussed [123, 128].

2.5. Effect of Coupled Fields: Piezoelectricity

In addition to the coupling with the band structure, the stress arising out of a misfit quantum dot in a piezoelectric semiconductor material, additionally couples with the carriers through the piezoelectric effect. Like strain-band structure coupling, piezoelectric effect too can significantly alter the electronic and optical properties of semiconductor devices [22, 133–135].

2.5.1. Piezoelectric Effect Around a Spherical Dot

Once again, we consider first the simplest example: a spherical QD of radius, R, located in an infinite medium. The piezoelectric polarization, $\mathbf{P}$, induced by strain is given by

$$P_i = e_{ijk}\varepsilon_{jk} \tag{50}$$

where e_{ijk} is the piezoelectric tensor and ε_{jk} is the usual strain tensor,

$$\begin{aligned} e_{ijk} &= e_{14} \text{ for } \{ijk\} = \{123\} \text{ or permutations} \\ &= 0 \text{ otherwise} \end{aligned} \tag{51}$$

where e_{14} is the piezoelectric constant [136].

The polarization vanishes inside the spherical dot because the purely dilational strain present inside the sphere does not cause piezoelectric effect in a material with $\bar{4}3m$ symmetry. As a result there is no piezoelectric potential within a spherical quantum dot. For the region outside the dot, the polarization, $\mathbf{P}$, can be related to the charge density per unit volume, $\rho(\mathbf{r})$ as follows:

$$\rho(\mathbf{r}) = -\nabla \cdot \mathbf{P} \tag{52}$$

which when combined with Eqs. (50), (51), and (11), results in [54]

$$\rho(\mathbf{r}) = \frac{-270Ke_{14}}{3K + 4\mu}\varepsilon^m R^3 \frac{xyz}{r^7} \tag{53}$$

At the surface, the charge density σ is given by

$$\sigma = \mathbf{P} \cdot \hat{\mathbf{n}} \tag{54}$$

which can be written as

$$\sigma = 6e_{14}\left(\frac{9K}{3K+4\mu}\right)\varepsilon^{m}\frac{xyz}{R^3} \tag{55}$$

Because of the polarization described by Eq. (54–55), the piezoelectric potential $\varphi(\mathbf{r})$ inside the dot can be written as [54]

$$\varphi(\mathbf{r}) = -\frac{3e_{14}\varepsilon^{m}R}{\epsilon_0\epsilon_r}\frac{9K}{3K+4\mu}\left(\frac{R}{r}\right)^2\left[1-\left(\frac{R}{r}\right)^2\right]\frac{xyz}{r^3} \tag{56}$$

where ϵ_r, is the dielectric constant of the semiconductor and ϵ_0 is the dielectric constant of free space. The maximum potential occurs at, $r = a\sqrt{2}$. This potential would penetrate inside an inclusion possessing a less symmetric shape and one would expect it to shift the energy levels of electrons inside the dot. However, because of the presence of the extremely small length factor of R (the radius of the quantum dot) in Eq. (56), this potential tends to be very small and exercises little if any influence on the energies of electrons [54].

2.5.2. Piezoelectric Effect Around Any Dot

As highlighted by Davies [54], the elastic displacement, $\mathbf{u}$, in the presence of a lattice mismatch strain, can be described by a scalar potential χ in a manner similar to the electric field-elastic potential relationship:

$$\mathbf{u} = \nabla\chi \tag{57}$$

This potential obeys Poisson's equation:

$$\nabla^2\chi = \frac{9K}{3K+4\mu}\varepsilon^{m} \tag{58}$$

Using the usual Green's function for the Laplace equation the following relationship can be obtained for χ, at a generic point in space [54],

$$\chi(\mathbf{r}) = -\frac{1}{4\pi}\frac{9K}{3K+4\mu}\int\frac{\varepsilon^{m}(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} \tag{59}$$

The piezoelectric charge density is given in terms of χ by

$$\rho(\mathbf{r}) = -e_{ijk}\chi_{ijk} \tag{60}$$

We know that

$$\nabla^2\varphi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0\epsilon_r} \tag{61}$$

Using Eqs. (61), (60), and (58), we have [54]

$$\nabla^4\varphi(\mathbf{r}) = \frac{9K}{3K+4\mu}\left(\frac{e_{ijk}}{\epsilon_0\epsilon_r}\varepsilon(\mathbf{r})^{m}_{ijk}\right) \tag{62}$$

Davies has [100] defined a "piezoelectric pseudopotential" Φ such that

$$\nabla^2\Phi(\mathbf{r}) = \chi(\mathbf{r}) \tag{63}$$

From Eq. (51) we obtain

$$\nabla^4\Phi(\mathbf{r}) = \frac{9K}{3K+4\mu}\varepsilon^{m}(\mathbf{r}) \tag{64}$$

thereby expressing φ in terms of Φ in a biharmonic equation. The piezoelectric potential can then be obtained after some further manipulation as a surface integral of the following form [100]:

$$\varphi(\mathbf{r}) = \frac{9e_{14}\varepsilon'''}{4\pi\,\epsilon_0\epsilon_r}\left(\frac{9K}{3K+4\mu}\right)\oint \frac{(x-x')(y-y')(z-z')(\mathbf{r}-\mathbf{r}')\cdot d\mathbf{S}'}{|\mathbf{r}-\mathbf{r}'|^5} \tag{65}$$

In the analytical models previously described (and some finite-element models notably [137]), the elastic field induced is first determined subject to given mechanical boundary conditions. The elastic solution is then used to estimate the polarization, which induces the electric potential and field. The electric field is then solved for by imposing purely piezoelectric boundary conditions. Such an approach is often called the "semi-coupled model".

A fully coupled model was proposed by Pan [138] predicated on earlier works involving the determination of displacement Green's functions in a fully coupled and generally anisotropic piezoelectric half space [139]. In a fully coupled model, the elastic stress field, σ_{ij}, and the electric displacement field, D_i, are coupled by the following constitutive relations:

$$\sigma_{ij} = C_{ijlm}\varepsilon_{lm} - e_{kij}E_k \tag{66a}$$

$$D_i = e_{ijk}\varepsilon_{jk} + \epsilon_{ij}E_j \tag{66b}$$

where ϵ_{ij} is the dielectric tensor and E_k is the electrical field. The strain ε_{ij} and the electric field E_i are related to the elastic displacement u_i and the electric potential ϕ by their corresponding constitutive relations.

Use of the Barnett-Lothe notation [140] allows one to recast Eqs. (66a) and (66b) into a unified single equation. For a finite-sized QD, Pan [138] then expressed the induced elastic and piezoelectric fields in terms of boundary integrals on the surface of the QD and the point-force/point-charge Green's function solutions.

To facilitate comparison between the semi-coupled and the fully coupled model, Pan [138] presented the cases of GaAs and AlN quantum dots for both the infinite-space and half-space situations. Under the assumptions of elastic isotropy, analytical solutions have been arrived at, using the semi-coupled model in the full- and half-space GaAs and in the full-space AlN.

Some of the results of Pan [138] are now discussed. In his work, a lattice mismatch ε''' of 7% was assumed while modeling the quantum dot system for both GaAs and AlN. The quantum dots themselves are point spheres with an equivalent volume of $\frac{4}{3}\pi a^3 (a = 1)$ nm and are located at the origin $\mathbf{r} = (0, 0, 0)$ for the full-space case and at $\mathbf{r} = (0, 0, h)(h = 2$ nm) for the half space case; $z = 0$ being the free surface. The electromechanical coupling factor g for a piezoelectric material is defined as

$$g = \frac{e_{\max}}{\sqrt{\epsilon_{\max}/C_{\max}}} \tag{67}$$

where $e_{\max}$, $\epsilon_{\max}$, and $C_{\max}$ are the maximum absolute values of the piezoelectric coefficients, dielectric constants, and elastic constants. For the specific case of GaAs, with a weak electromechanical coupling ($g = 0.04$), the semicoupled model was found by Pan [138] to yield results similar to the fully coupled model for the elastic and piezoelectric fields (for full- and half-space). However, in the case of AlN, which has a rather strong electromechanical coupling ($g = 0.32$ opposed to 0.04 for GaAs), the semicoupled model rendered substantially different results from the fully coupled model. This is well illustrated in Fig. 27, where both fully coupled and semicoupled models predict nearly identical electric fields in the case of the GaAs, while significant differences are observed between the two-models while estimating the electric-field for the AlN case [138]. In fact at $r = (1, 1, 1)$ nm, the electric field predicted by the semi-coupled model in the GaAs case falls short of that predicted by the fully coupled model by nearly a factor of 2.

On a related note, Pan in a subsequent work [141] applied the three-dimensional Green's function solution based on the fully coupled model to four substrates, namely, GaAs (001), GaAs (111), Iso (001), and Iso (111) to show that the isotropic model fails to predict the induced elastic and piezoelectric fields for piezoelectric semiconductors. The elastic constants

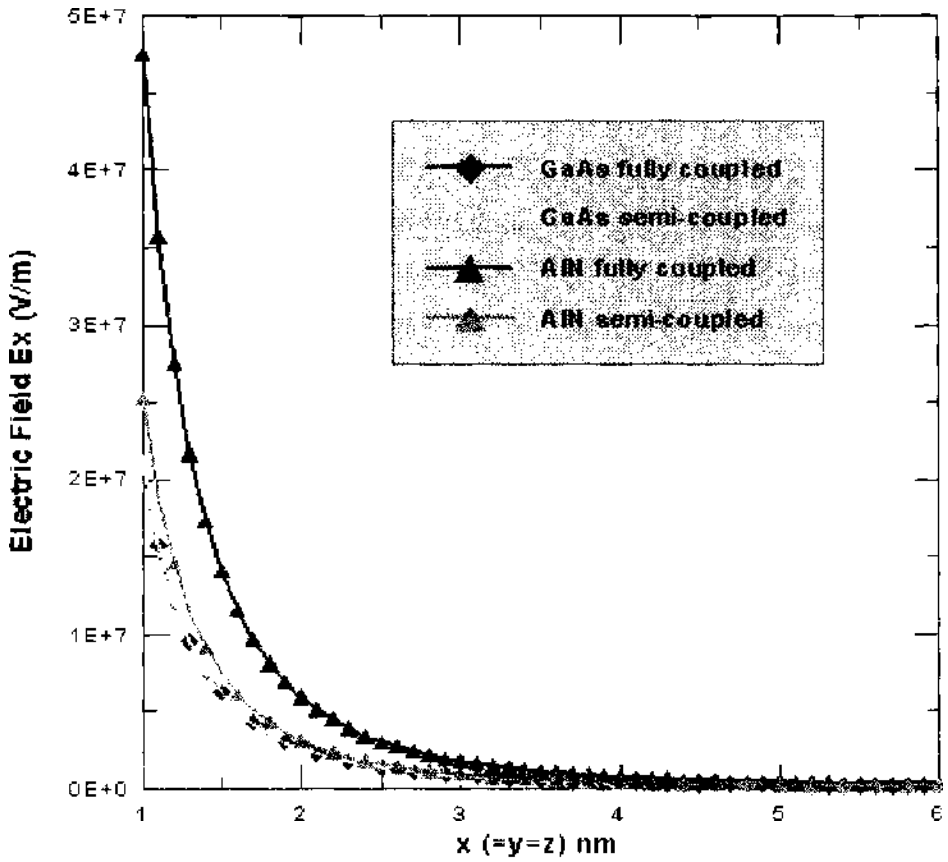


Figure 27. A comparison between the electric-field E_x predicted by the semi-coupled and fully coupled models for two different quantum dot systems (a) AlN and (b) GaAs in infinite space. The dot is point spherical with an equivalent radius of $a = 1$ nm and is located at $r = (0, 0, 0)$. The variation in the above diagram is traced along $r = (x, x, x)$. Reproduced with permission from [138], E. Pan, *J. Appl. Phys.* 91, 3785 (2002), Fig. 3, © 2002, American Institute of Physics.

for Iso (001) and Iso (111) were assumed to be isotropic. The elastic and piezoelectric fields on the surface of these substrates, due to a buried point quantum dot with an initial dilational misfit strain, were also studied (Figs. 28–30). In the examples that Pan illustrates, the QD is a point sphere buried at a depth h ($h = 10$ nm), with an equivalent radius of a ($a = 3$ nm); an initial mismatch strain of $\varepsilon^m = 7\%$ is assumed.

From Fig. 28, as Pan [141] noted, it is clear that the isotropic model is clearly not suitable to estimate the elastic fields in anisotropic GaAs. While the isotropic hydrostatic strain has complete rotational symmetry about the z-axis, GaAs (001) displays a C_4 symmetry and GaAs (111) exhibits a C_3 symmetry. The values of strain relaxation achieved at the surface for the three cases are also markedly different with high-hydrostatic strain prevailing in the GaAs (111) case (10% of misfit strain in the GaAs (111) case as against 7% and 3% for Iso (001) and GaAs (001), respectively).

For similarly orientated GaAs and Iso semiconductors (for example, GaAs (001) or (111) and Iso (001) or (111), respectively), Pan [141] have reported some similarities between the contours of piezoelectric potential (Fig. 29) (even the horizontal electric field has been demonstrated to display the said similarities) at the surface. However, as Fig. 30 show, the vertical electric field E_z, at the surface in GaAs (001) or (111) does not exhibit any similarity to the isotropic case. Pan [141], therefore, have suggested use of caution while applying the isotropic assumption to the GaAs semiconductor case to avoid arriving at erroneous results.

2.6. Effect of Size

As evident so far in our review, generally speaking, researchers have used the well established continuum elasticity theory (both numerically and analytically) to estimate mechanical strains. Those then are coupled to some suitable band-structure calculation method (such

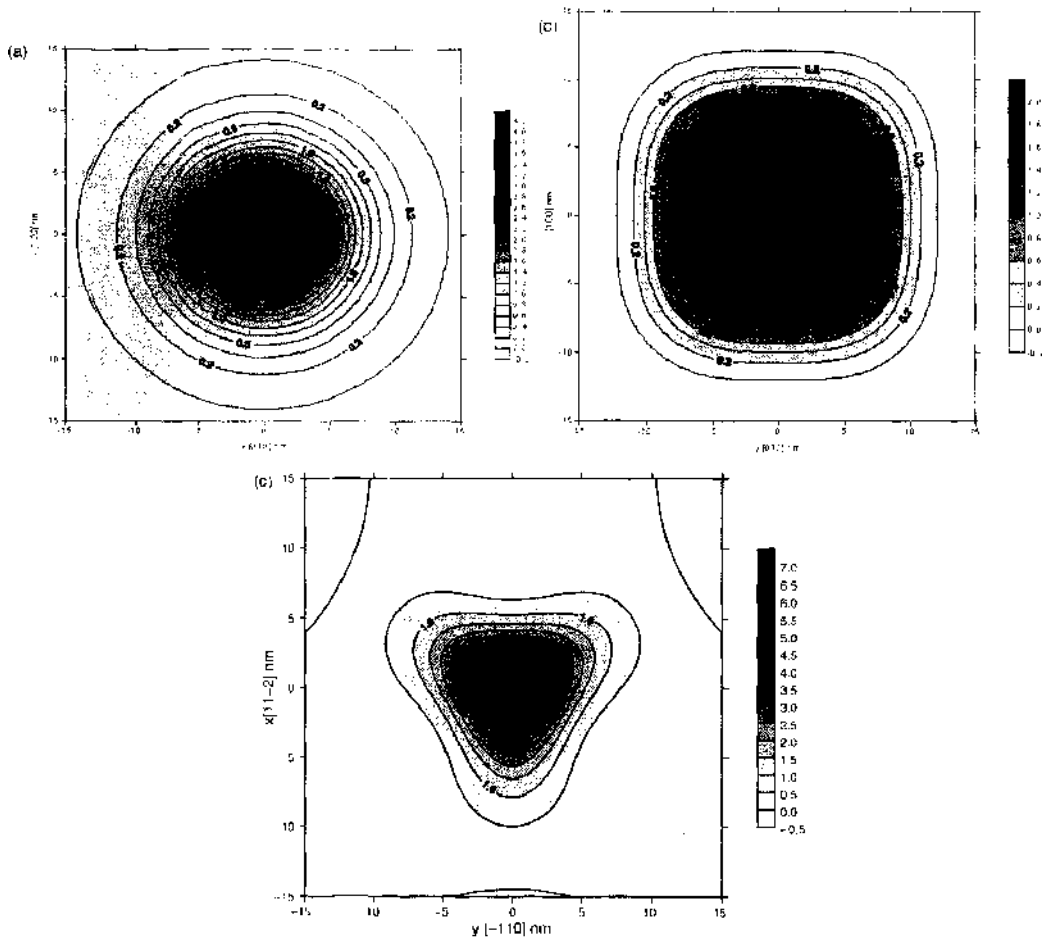


Figure 28. Contours of $\varepsilon_{ii}(10^{-3})$, the hydrostatic strain due to a point QD of volume $v_a = 4\pi a^3/3$ buried at a depth of $h = 10$ nm , (a) on the surface of the isotropic crystal (the contours are similar for Iso(001) and Iso(111)); (b) on the surface off GaAs(001); (c) on the surface of GaAs (111). Reproduced with permission from [141], Pan, *J. Appl. Phys.* 91, 6379 (2002), Fig. 2(a–c). © 2002, American Institute of Physics. These figures clearly indicate that the isotropy model is not suitable to determine strain distributions in anisotropic GaAs semiconductor. Other elasticity fields (like stress and displacement) show similar variation to the strain shown in the diagram.

as tight binding or the **k.p** approach) to estimate the impact of strain on the optoelectronic properties. Classical continuum mechanics, is however, intrinsically size independent. This is in contradiction to the physical fact that at the size-scale of a few nanometers, deformations and elastic state are size dependent and a qualitative departure from classical mechanics is

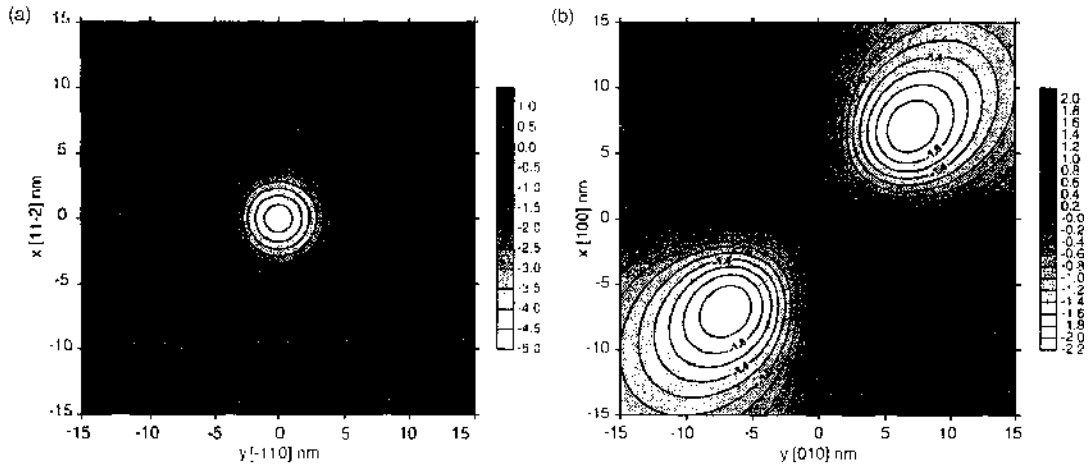


Figure 29. Contours of piezoelectric potential due to a point quantum dot of volume $v_a = 4\pi a^3/3$, and buried under a depth $h = 10$ nm (a) at the surface of GaAs (111); (b) at the surface of GaAs (001). Reproduced with permission from [141], Pan, *J. Appl. Phys.* 91, 6379 (2002), Fig. 3(a–b). © 2002, American Institute of Physics.

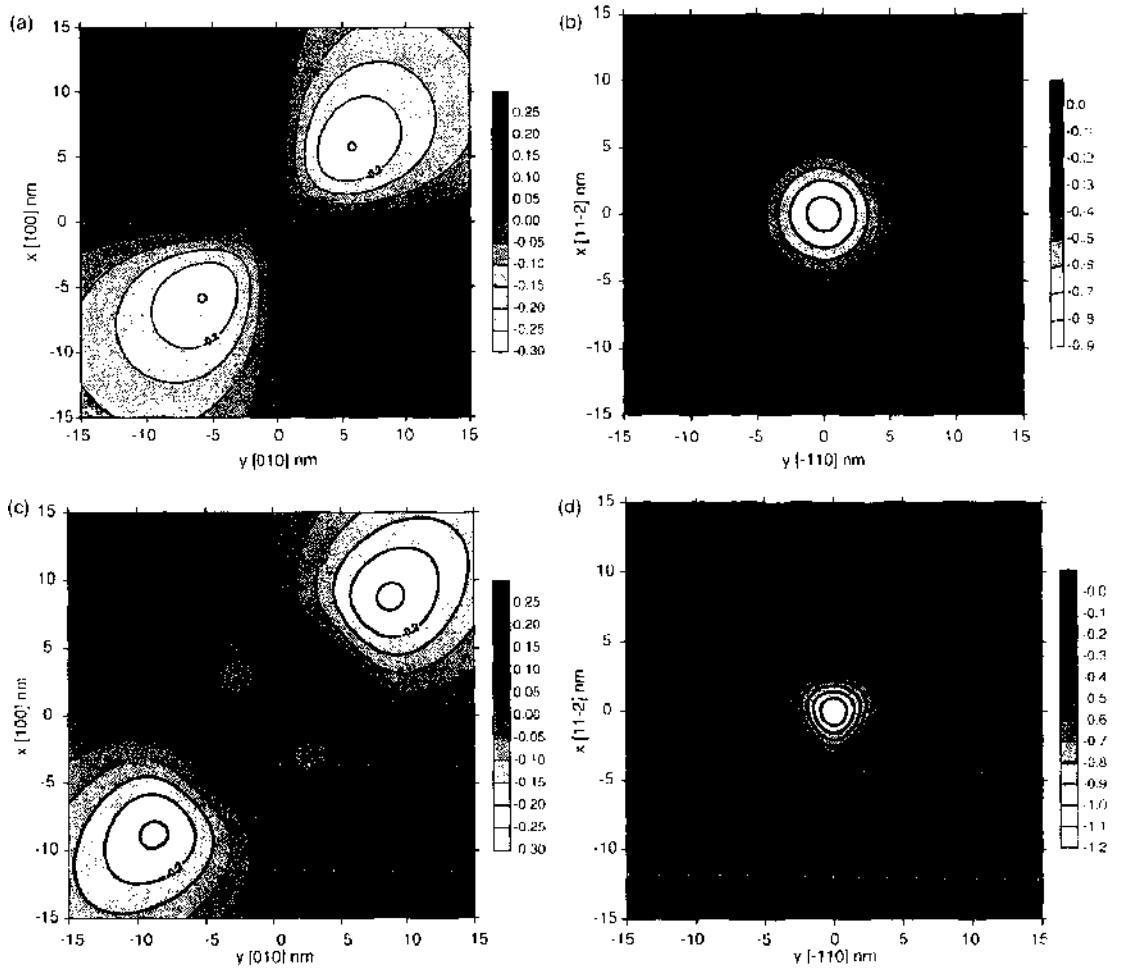


Figure 30. Contours of vertical electric field $E_z(10^7 V/m)$ due to a point quantum dot of volume $v_0 = 4\pi a^3/3$ and buried under a depth $h = 10$ nm: (a) at the surface of Iso (001); (b) at the surface of Iso (111); (c) at the surface of GaAs (001); (d) at the surface of GaAs (111). Reproduced with permission from [141], Pan, *J. Appl. Phys.* 91, 6379 (2002), Fig. 5(a–d). © 2002, American Institute of Physics.

expected. The obvious alternative method to compute strain is the use of atomistic simulations. Nevertheless, a field theoretic method is highly desirable (in the same vein several of the models discussed so far) albeit that does also account for the scaling or size effects in strain likely to be prevalent at these small length scales.

There are chiefly two physical mechanisms that may alter continuum elasticity predictions and result in scaling or size effects in the strain calculations in quantum dots: (1) surface or interfacial energy effects and (2) nonlocal elastic interactions. Both are likely to be small and only of importance for exceedingly small quantum dots. nevertheless these effects can be important in certain cases.

2.6.1. Surface/Interface Energy Effects

For structures with sizes >50 nm, typically, the surface-to-volume ratio is negligible and the deformation behavior is governed by classical volume strain energy. However, at submicron-length scales the properties of the quantum dot surface/interface are expected to play a role in the determination of its elastic state. Simply from dimensional considerations we expect that strain should scale $\sim 1/R$ (where R is some characteristic length of the quantum dot and that the proportionality constant is related to surface or interface energy). In the context of quantum dots and embedded inclusions, this effect has been discussed in some recent publications by one of the authors [142–144]. The latter works, presented the size-dependent elastic state of QDs based on the involvement of surface/interfacial energies at the nanoscale. For example, errors in strain calculation as high as 12% were reported in the determination of hydrostatic strain in a buried spherical QD (in the size range of 2 nm).

Surface elastic effects have been considered by a variety of authors in various contexts, for example, nanostructures [145], nanocomposites, thin films [146], surface steps [147–148], quantum dots, and soforth. The reader is referred to the following review articles and references therein for comprehensive details. From the point of view of a mechanics, Gurtin and Murdoch [149], Murdoch [150], and Gurtin et al. [151] appear to have established a fairly rigorous framework to incorporate surface or interface energies which is what we following in the present article to tackle the quantum dot strain scaling problem.

Consider an arbitrary shaped smooth interface between an embedded inclusion and surrounding host matrix, characterized by a unit normal $\mathbf{n}$. Let this interface be "attached" to the bulk (i.e., both inclusion and matrix) without slipping or any other discontinuity of displacements across it. This implies that we consider only a coherent interface. In contrast to the classical case where surface energies are neglected, we now require that the interface of the inclusion and the matrix be endowed with a deformation dependent interfacial energy, Γ. The interfacial or surface energy is positive definite. This quantity is distinct from the bulk deformation dependent energy due to the different coordination number of the surface/interface atoms, different bond lengths, angles, and a different charge distribution [152]. Within the assumptions of infinitesimal deformations and a continuum field theory, the concept of surface stress and surface tension can be clarified by the (assumed *linearized*) relation between the interface/surface stress tensor, $\boldsymbol{\sigma}^s$, and the deformation dependent surface energy, $\Gamma(\boldsymbol{\varepsilon}^s)$:

$$\boldsymbol{\sigma}^s = \tau_o \mathbf{P}^s + \frac{\partial \Gamma}{\partial \boldsymbol{\varepsilon}^s} \tag{68}$$

Where applicable, superscripts B and S indicate bulk and surface, respectively. Here, $\boldsymbol{\varepsilon}^s$ is the strain tensor for surfaces that will result from the projection of the conventional bulk strain tensor on to the tangent plane of the surface or interface, while τ_o is the deformation independent surface/interfacial tension. The surface projection tensor, $\mathbf{P}^s$ which maps tensor fields from bulk to surface and vice versa is defined as:

$$\mathbf{P}^s = \mathbf{I} - \mathbf{n} \otimes \mathbf{n} \tag{69}$$

Consider an arbitrary vector $\mathbf{v}$. The surface gradient and surface divergence, then, take the following form [151]:

$$\begin{aligned} \nabla_s \mathbf{v} &= \nabla \mathbf{v} \mathbf{P}^s \\ \mathrm{div}_s(\mathbf{v}) &= \mathrm{Tr}(\nabla_s \mathbf{v}) \end{aligned} \tag{70}$$

Here, we have also defined the surface gradient operator (∇_s) and the surface divergence, which we shall shortly employ. We repeat here the equilibrium and isotropic constitutive equations of bulk elasticity:

$$\begin{aligned} \mathrm{div}\, \boldsymbol{\sigma}^B &= 0 \\ \boldsymbol{\sigma}^B &= \lambda \mathbf{I}^3 \mathrm{Tr}(\boldsymbol{\varepsilon}) + 2\mu\boldsymbol{\varepsilon} \end{aligned} \tag{71}$$

At the interface, the concept of surface or interface elasticity [149, 151] is introduced, which is excluded in the classical elasticity formulation:

$$\begin{aligned} [\boldsymbol{\sigma}^B \cdot \mathbf{n}] + \mathrm{div}_s\, \boldsymbol{\sigma}^S &= 0 \\ \boldsymbol{\sigma}^S &= \tau_0 \mathbf{P}^s + 2(\mu^s - \tau_0)\boldsymbol{\varepsilon}^S + (\lambda^s + \tau_0)\mathrm{Tr}(\boldsymbol{\varepsilon}^S)\mathbf{P}^s \end{aligned} \tag{72}$$

Isotropic interfaces or surfaces can be characterized by surface Lamé constants λ^s, μ^s, and surface tension, τ_0. The square brackets indicate a jump of the field quantities across the interface. Only certain strain components appear within the constitutive law for surfaces due to the 2×2 nature of the surface stress tensor (i.e., only the tangential projection of the strains on the interface are included consequently, $\mathbf{P}^s \cdot \mathbf{n} = 0$). In absence of surface terms, Eq. (72) reduce to the usual normal traction continuity equations of classical elasticity.

Thus, while the infinitesimal strain tensor in the bulk (both inclusion and matrix) is defined as usual, the surface strains involve the use of projection tensor (Eq. 69).

$$\boldsymbol{\varepsilon}^{s} = \frac{1}{2}(\mathbf{P}^{s}\nabla_{s}u + \nabla_{s}u^{\mathrm{T}}\mathbf{P}^{s}) \tag{73}$$

Implicit in Eq. (73) is our assumption of a coherent interface. Using these basic equations, Sharma and Ganti [144, 153–155] have explicitly derived the dilatation strains for embedded quantum dots. To be more explicit, the dilatational strain in an isotropic spherical lattice mismatched embedded quantum dot that correctly incorporates size effect due to interfacial energies can be written as:

$$\mathrm{Tr}(\varepsilon) = 3\frac{3K\varepsilon^{m} - 2\tau_{o}/R}{4\mu + 3K + 2K^{s}/R} \tag{74}$$

where K_s is the surface elastic modulus defined as $2(\lambda^s + \mu^s)$. Note that for large radius of QD ($R \to \infty$) or zero surface energy, the result reverts to the classical solution used by several authors (see, e.g., Yang et al. [23]). Equation (74), of course implies that even if the QD is not embedded, there is a finite strain (which of course is very small except in the smallest possible QDs):

$$\mathrm{Tr}(\varepsilon) = -\frac{3}{3K_{\mathrm{QD}} + 2K_{s}/R}\left(\frac{2\tau_{o}}{R}\right) \tag{75}$$

In a collaborative work, one of the authors presented numerical density function theory (DFT) simulations of Si clusters that illustrated the impact of surface energy induced strains (Peng et al. [156]). *Ab initio* simulations of quantum dots, for the purpose of ascertaining scaling laws associated with their optoelectronic properties, have also been carried out by other researchers [157–158]. These previous works, however, did not notice the aforementioned additional scaling effects since they (as is often done conventionally) constrained the surface atoms to remain configurationally fixed thus effectively precluding the manifestation of influence of surface energies on strain. In the simplest possible picture, for the case of unembedded quantum dots where there is no apparent source of strain (i.e., no lattice mismatch), the effect of surface tension is excluded (see Eq. [75]). In the case of embedded quantum dots, the correction to the lattice mismatch-induced strain due to interfacial tension and elasticity is excluded. This form of simulations, where surface atoms are not allowed to relax, is primarily employed for savings in computational time with the unfortunate consequence of effectively voiding the strain-induced scaling. For example, the DFT computation time for a surface relaxed cluster $Si_{59}H_{60}$ is 1,694 minutes, while the time for the same cluster without atomic relaxation is 125 minutes.

We now proceed to present a test of the assertions and implications of Eq. (75) through DFT calculations of unembedded Si clusters of various sizes. We note here the work of Delley [159] who performed such a study without incorporating the surface effects outlined in Eq. (75). The surface atoms were configurationally fixed in his analysis, thus excluding the scaling effect we predict. To retain the tetrahedral configuration of silicon (refer to Fig. 31), all the dangling bonds of the surface silicon atoms were terminated by hydrogen at initial bond length of 1.47 Å. The clusters were varied from Si_5H_{12} to $Si_{230}H_{196}$ i.e., from 5.8 to 21.0 Å).

The surface atoms were allowed to relax to their equilibrium state thus triggering the strain in Eq. (75). The main results are depicted in Fig. (32), where the band gap is plotted as a function of cluster size and compared with Delley's [159] unrelaxed cluster study. As already anticipated, we note a significant shift in band gap at small sizes where surface energy-induced strains make their presence felt. A somewhat surprising result that emerges from Fig. (32) is that this scaling effect disappears for extremely small sizes!

A maximum band-gap shift of nearly 0.51 eV is observed at a cluster size of 17 atoms. Plausible explanations for the disappearance of this new scaling effect at sizes below 17 atoms appear to suggest that surface energy parameterized by surface tension and surface

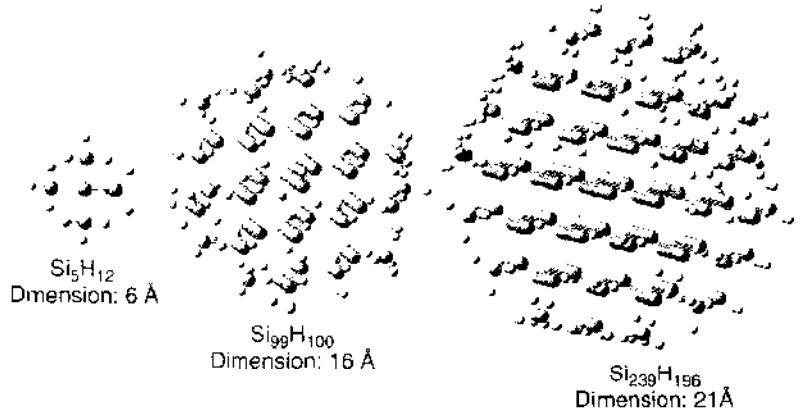


Figure 31. Snapshots of various-sized Si clusters used in the DFT simulations.

elastic modulus must itself be size dependent, that is,

$$\{\tau_s, K_s\} \propto \frac{\{\tau_s, K_s\}^{\text{at "large" size}}}{1 + k/R} \tag{76}$$

where k is a constant. Obviously, as $R \to 0$, the surface energy smoothly vanishes, explaining the disappearance of concerned scaling effect at extremely small sizes. Physically and in hindsight this is easy to justify. The continuum field concept of surface energy, tension, and surface elastic modulus signifies the difference between the surface atom properties with the corresponding ones in the bulk material (due to different coordination number, charge distribution, bond length, etc.; a mismatch strain of 3.5% and K ~ 100 Gpa and $\mu \sim 60$ Gpa. However, at extremely small sizes, where only few atoms remain, this difference and the distinction between surface and bulk atoms becomes very tenuous or, in other words, the continuum notions of surface energy (if one insists on using them) must become zero.

2.6.2. Nonlocal Effects

Nonlocal interactions are another mechanism that is only of importance when the quantum dot size is comparable to the lattice parameter. At small length scales (approaching a few nanometers comparable to the discrete structure of matter) the implicit long-wavelength assumption of classical elasticity breaks down. This break down is caused partially by the long-range interactions between atoms, which are inadequately represented by classical elasticity. As one would expect, several phenomena at the level of a few lattice spacing are inadequately captured by classical elasticity and researchers often see enriched continuum

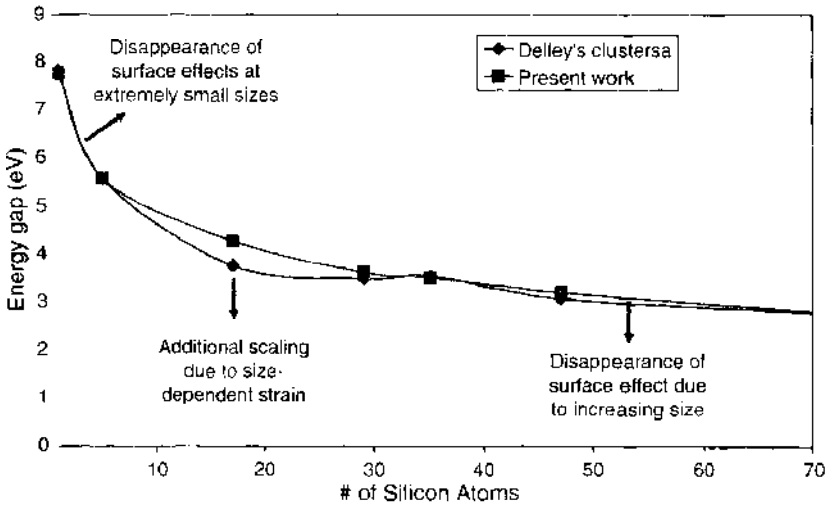


Figure 32. Comparison of the present work incorporating the size-dependent strain due to surfaces and Delley's [159] results. Adapted from Peng et al. [156].

theories such as nonlocal elasticity as a replacement for atomistic simulations (or, alternatively, a bridge between atomistic and conventional continuum mechanics). For example, the ubiquitous singularities ahead of crack tips and dislocation cores (as predicted by classical mechanics) are indeed a break down of traditional elasticity at short wavelengths [157]. One possible means to circumvent the long wavelength assumption of classical elasticity is to invoke higher order gradients in the elastic energy expansion. In a recent work [160], one of the authors derived the complete Eshelby's tensor for an embedded quantum dot.

Consider the classical isotropic elastic material. The strain energy function is quadratic in strains:

$$W(x) = \mu \varepsilon_{ij}\varepsilon_{ij} + \frac{1}{2}\lambda \varepsilon_{kk}^2 \tag{77}$$

Here, $\boldsymbol{\varepsilon}$ is the usual strain tensor related to $\mathbf{u}$ (displacement vector) as $\mathrm{sym}(\nabla \mathbf{u})$ and $[\bullet]_{,j}$ and ∂_j will be used interchangeably to indicate differentiation with respect to spatial variable x_j. Note that the antisymmetric part of the deformation gradient that is, $\boldsymbol{\omega}$ $(= \mathrm{asym}\ \nabla \mathbf{u})$ is absent from Eq. (77) because the quadratic term in ω is not rotationally invariant—a necessary requirement for the energy function in Eq. (77). For small quantum dot sizes, additional gradient terms (absent in Eq. [77]) may also contribute and are considered to phenomenologically representative of nonlocal interactions [157–158]. The latter is achieved by suitably adding higher-order terms containing gradients of strain and rotation. (Indeed, the gradients of ω are admissible because those fields are invariant with respect to the Euclidean group of transformations SO(3)⊳T(3) unlike ω itself.) The general form of the elastic energy involving first gradients of strain and rotation is

$$W(x) = W(\partial_i u_j, \partial_j u_i, \partial_i \partial_j u_l, \partial_l \partial_j u_i) \tag{78}$$

In the isotropic case, the energy density that is invariant to SO(3)⊳T(3) group then takes the form [158]

$$\begin{aligned} W(x) = {} & \frac{\mu}{2}(\partial_i u_j)^2 + \frac{\mu+\lambda}{2}(\partial_j u_i)^2 + \frac{2\mu+\lambda}{2} l'^2 \partial_i \partial_j u_i \partial_i \partial_j u_j \\ & + \frac{\mu l^2}{2}(\partial_l^2 u_i \partial_l^2 u_i - \partial_i \partial_j u_l \partial_i \partial_j u_l) \end{aligned} \tag{79}$$

Two new coupling constants (in addition to the Lamé parameters) now appear namely l' and l. Both have units of length. For band-gap calculations in quantum dots, one typically requires only the dilatation and in the isotropic case (as it turns out), the last term in Eq. (79) plays no role, and hence, in the following, we set $l = 0$. Further, using a variational argument (by appealing to the Euler-Lagrange equations), the governing field equation can be derived as well as the response quantities (i.e., "stresses"). The single Navier-like equation that emerges is

$$-\mu \partial^2 u_i - (\mu + \lambda)\partial_i \partial_j u_j + \underline{(2\mu + \lambda) l'^2 \nabla^2 \partial_i \partial_j u_j} = 3K\delta_{ik}\partial_k[\varepsilon^m H(x)] \tag{80}$$

The underlined portion of Eq. (80) indicates the extra terms absent in *size-independent* classical elasticity. Zhang and Sharma [160] have derived the solution to the problem in Eq. (80) for the case of a misfitting spherical quantum dot. They obtain:

$$tr(\varepsilon) = \begin{cases} \dfrac{9K\varepsilon^m}{3K + 4\mu}\left[1 - (l' + R)e^{-R/l'}\dfrac{1}{l'}\dfrac{\sinh r/l'}{r/l'}\right] & r \in \boldsymbol{\Omega} \\ \dfrac{9K\varepsilon^m}{3K + 4\mu}\left(R\cosh\dfrac{R}{l'} - l'\sinh\dfrac{R}{l'}\right)\dfrac{1}{l'}\dfrac{e^{-r/l'}}{r/l'} & r \notin \boldsymbol{\Omega} \end{cases} \tag{81}$$

In Fig. 33, the normalized dilations strain as a function of position, various inclusion sizes and nonlocal coupling constant l' are plotted. The location $x/R = 1$ indicates the boundary of the spherical quantum dot. The size effect of the nonlocal solution is manifest. We note that, unlike both the classical and interfacial energy-based solution, the dilation incorporating

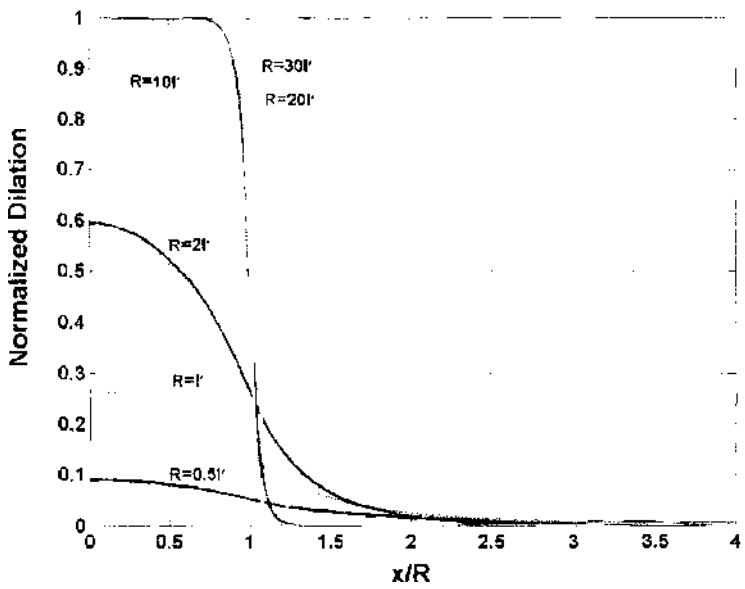


Figure 33. Strain dilatation as a function of position and size. The flat line with abscissa = 1 for $x/R < 1$ and = 0 for $x/R > 1$ represents the classical size-independent solution. Adapted from Zhang and Sharma [160].

nonlocal effects is inhomogeneous within the inclusion. Asymptotically, the nonlocal results converge to that of classical elasticity for large quantum dot size. Further, note that while the classical results predict the well-known zero dilatation outside the spherical quantum dot, in the case of nonlocal results we observe a small nonzero dilation. Conforming to physical intuition, artificial jumps in stresses are removed in nonlocal results and the strains vary smoothly across the interface (or in other words the "continuum" sharp interface of the quantum dot/matrix acquires a "diffuse boundary layer" to the order of the characteristic length scale parameter.

To emphasize on the size dependency of our solution, the dilatation as a function of size (for a fixed position, i.e., $r = 0$) is also plotted in Fig. 34. We observe that while for large quantum dot size, roughly, $R > 7l'$, the nonlocal/strain gradient solution is indistinguishable from the classical one, the dilatation decreases significantly below this threshold and exhibits a marked departure from classical solution.

We now proceed to draw a comparison between surface energy effects (discussed in Eq. [74]) and nonlocal results. Unfortunately, while the nonlocal results can be adequately

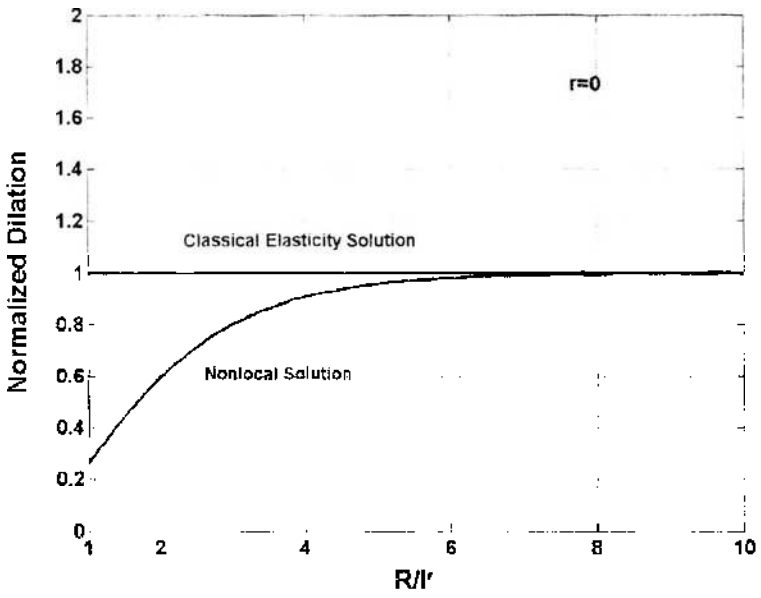


Figure 34. Dilatational strain as a function of size for fixed position ($r = 0$). The results are normalized with respect to classical elasticity solution. Adapted from Zhang and Sharma [160].

normalized (i.e., independent of mismatch strain), the surface/interfacial results canno and are highly material dependent (both on the actual values of interfacial energy parameters as well elastic moduli). We provide some general estimates here. The nonlocal coupling constant l' roughly corresponds to the lattice parameter of the quantum dot which is around ~0.4 nm for most materials. For the specific case of GaAs, in a rather interesting work, DiVincenzo [161] has determined the higher-order constants for the nonlocal elasticity theory. For the particular case of GaAs, we deduce l' to be ~0.82 nm (which, in fact, is nuch larger than the lattice parameter). So, for the purposes of general comparisons, it is safe to adopt l' to be roughly ~ lattice parameter. The interfacial tension (for coherent sysems) is roughly to the order of ~1 J/m^2. Because (unlike both the classical and surface energy based result) the nonlocal solution is nonuniform, we compare results at the quantum dot center ($r = 0$). For diameters of {2, 4, 6, 8, 10} nm, respectively, the nonlocal eevation from classical elasticity result is {22, 2.2, 0.18, 0.014, ~0}%, while we obtain a eevation of {19, 9.5, 6.3, 4.8, 3.8}% for interfacial energy-based results. Because of the exponential decay in the nonlocal solution (with respect to size), these effects decay far more rapidly than the interfacial effects but are likely to dominate for very small sizes (whic though for most materials may be impossibly small and would be irrelevant anyway as we do not expect nonlocal theory to work for sizes smaller than $2l'$—classical elasticity is expected to breakdown at even larger sizes, roughly, $6l'$).

2.7. Some Results from Inclusion Theory Useful for Strain Calculations in Quantum Dot Structures

Since the original appearance of Eshelby's paper [30] several works have extended, nodified, and applied the concept of Eshelby's tensor to a diverse set of physical problems. We provide a brief review of this body of literature here in the hope that researchers addressing quantum dots may find them of use.

1. **Bonding conditions of inclusion.** The original assumption in Eshelby's work is that the inclusion is perfectly bonded to the matrix; that is, the normal tractions are continuous and so are the displacements. All works on quantum dots (that the present authors are aware of) make this same assumption, that is, matrix-quantum dot interface is perfectly bonded. Under certain conditions these conditions must be relaxed; jumps in displacement or tractions may be allowed. Various researchers have considered the imperfectly bonded inclusion, for example, Furuhashi et al. [162], Ru and Schiavone [163], Zhong and Meguid [164], Qu [165, 166], and Kouris et al. [148]. Of course, physically, imperfectly bonded quantum dots will correspond to the case where defects are present at the interface. It is quite unclear whether then it is worthwhile to investigate the stain state in such dislocated quantum dots since most likely in comparison the mere presence of defects will overwhelm the electronic effects.
2. **Coated inclusions.** Frequently for technological reasons inclusions are embedded in a matrix with a coating (or which may be developed due chemical interaction with the matrix). A few representative works in this area are: Walpole [167], Luo and Weng [168], and Cherkaoui et al. [169], among many others. This scenario is very realistic especially in multialloy quantum dots where the outer rim of a quantum dc may be preferentially rich in one phase.
3. **Nonuniform mismatch strains.** Sendeckyj [170] and Moschovidis [171] considered general polynomial transformation strain thus nonuniform lattice mismatch or thermal expansion strains can be mimicked. Their work is also useful for taking into account interactions between inhomogeneities. Asaro and Barnett [172] and Mura and Kinoshita [173] addressed polynomial eigenstrains in an anisotropic media. Note also must be made of the recent work of Rahman [174] who presents simplified ciculations of Eshelby type tensors for polynomial eigenstrains.
4. **Enriched elasticity.** The classical theory of elasticity itself has been modified n several ways. Micromorphic elasticity takes into account additional micro-degrees e freedom such as independent rotations, dilations and shears. An extensive account of tese theories can be found in Eringen [175]. As far as inclusion problems are concerned, t appears

that the only two solutions that exist are due to Cheng and He [176, 177] who, respectively, solve the spherical and cylindrical inclusion problem. On the basis of the latter work, Sharma and Dasgupta [178] have formulated the overall properties of micropolar composites. Another enrichment of elasticity is whereby nonlocal interactions are introduced. Again, an up-to-date survey is given by Eringen [157]. A popular version of nonlocal elasticity is the strain gradient theory. Zhang and Sharma [160] have recently provided solutions of embedded inclusions in the strain gradient elasticity formalism.

5. **Inclusion-inclusion interactions**. As already noted above, the work of Moschovidis [171] provides and interesting discussion on interactions between two inclusions. Rodin and Hwang [179] provide a nice analysis of this problem and also contain several useful references. Interestingly, two dilating inclusions (same properties as matrix) have zero interaction energy in the isotropic limit. This is somewhat obvious since the external dilation in the isotropic case for all inclusion shapes is zero. Hence, if we are to use some of the isotropic results on quantum dots for dot–dot interaction, we may conclude that an interaction is absent. Anisotropy or half-space solution must be invoked to correctly account for this interaction. Interestingly, because the external dilation is nonzero in nonlocal elasticity (see Zhang and Sharma [160]), a finite interaction between dots should exist even in the isotropic case. However, this interaction is likely to be weak and only of importance for dots very closely spaced to each other. This issue requires further investigation.
6. **Anisotropicity and shape**. Several works have modified Eshelby's (originally isotropic) formulation to incorporate anisotropic behavior. Progress has largely been made only in the plane case. An excellent, but somewhat dated, account of these aspects is given in the now classic monograph by Mura [35]. Some more recent works, that also contain an extensive list of references on this subject are Ru [180] who discusses arbitrary-shaped inclusions in anisotropic half and full plane, Li and Dunn [181] address coupled field anisotrope inclusion problems, and Pan and Yang [182] who present a semianalytical method for application to embedded quantum dots. Inclusion shapes: Chiu [50] has considered parallelepiped inclusion. Rodin [183] considers the general polyhedral inclusion and provides and elegant algorithmic approach to determine the elastic state of arbitrary povhedral. So do Nozaki and Taya [184].

In addition to the aforementioned group of papers, several other works exist in the context of nonlinear behavior and of course in application areas (such as effective medium theories, phase transformations, stability among others). A review of those works is beyond the scope of this chapter. The following monographs, review articles, books, and references therein are recommended for the interested reader: Mura [35], Nemat-Nasser and Hori [33], Markov and Preziosi [185], Weng et al. [186], Bilby et al. [187], Mura et al. [188], Mura [189], and Kikuchi et al. [190].

3. SUMMARY AND OPEN ISSUES

In conclusion, we have reviewed and discussed several works and issues in calculation of strains in quantum dots and wires under myriad contexts of shape effects, presence of a free surface, anisotropy considerations, nonlinear effects, and presence of coupled effects (piezoelectricity). We have also provided a discussion on some of the novel size-dependent effects that may manifest themselves at the length scales at which these nanostructures exist.

Several avenues of research remain open and inadequately addressed. We highlight some of our own personal perspectives here.

Size effects in quantum dots appear to be the least explored. The use of classical continuum mechanics for strain calculations is typical which of course is size independent. As mentioned in the preceding Section 2.6.1 (in the context of surface energy effects), if care is not taken, even the atomistic studies may inadvertently exclude or underemphasize size effects. Some rather interesting scenarios are possible here. For example, we also discussed the size effects arising out of nonlocal interactions in a previous section. In such a theory, a characteristic length scale appears that is roughly in the neighborhood of the lattice parameter, that is, size effects in strain may become appreciable for dots that are close in

size to the lattice parameter. Obviously, this effect is small and only applicable for very tiny dots. In a recent work, Drugan and Willis [191] show that material with multiple inclusions behaves as a nonlocal solid with a characteristic length that may be much larger than the lattice parameter (and roughly scales with the inclusion size)! This implies that an array of quantum dots (depending upon the volume fraction or density) may cause a screening effect inducing nonlocal and hence noticeable size effects. This notion, apart from clarifying size effects, may also be a valuable tool in modeling the many body strain effects.

The discussion of possible screening effects of arrays of quantum dots on the scaling of strain leads to another relatively less studied topic: quantum dot–quantum dot interaction. In the isotropic limit, two dilating quantum dots in an infinite medium do not interact mechanically. Perhaps this has led many to conclude that strain-mediated interactions between quantum dots are of relatively less importance. There are, however, several exceptions to this notion. If quantum dots are sufficiently close to a free surface or interface, they will interact even in the isotropic limit through their image forces. Of course, proper accounting for the typically cubic anisotropy of most quantum dot materials will also remove this isotropic degenerate behavior. In particular, we note that the nonlocal solution of a single quantum dot exhibited nonzero dilation in the barrier. This implies that nonlocal interactions are yet another mechanism that may mediate dot–dot interaction (even in the isotropic case: the interactions are expected to be yet stronger in the anisotropic case). Given the computational advances, these many body interaction effects can accounted for easily (and indeed have implicitly done in several numerical works). To the best of the authors' knowledge, however, a systematic study is not yet available.

Frequently, quantum dots are assumed to have identical materials properties as the barrier. Computationally, there is no difficulty in accounting for the modulus mismatch (and indeed is routinely done in numerical simulations). The question, however, remains as to what truly are the material properties of the quantum dot that is often an alloy and under strain. Arguments may be advanced from both points of view. Indeed, some works have justified using identical material properties based on Keyes scaling relation [90–92]. Further, in a binary or ternary mixture, phases are not often uniformly intermixed [192]. For example, InGaN quantum dots one often sees a preferential phase segregation into InN [193]. How uniform is the mixing and hence the strain state in quantum dots? These aforementioned issues certainly could benefit from more detailed experimentation and theoretical work.

Although we have hardly touched on the strain-band structure coupling in this chapter, we mention here that several works have studied the viability and applicability of various approaches to take into account strain-band structure coupling. An oft-quoted work is that due to Pikus and Bir [194]. Other works have extended and modified this in various fashions, which we do not discuss here (e.g., Ref. [195] and the eight-band model. [196–197]). The particular reference that we would like to cite is due to Zhang [198] who presents formalism for taking into account inhomogeneous strain. This is an important advance however, largely predicated on the strain potential (in the quantum mechanical Hamiltonian) being local, that is the deformation potential at a point solely on the strain at that point. As acknowledged by Zhang himself (which in turn was communicated to him by C. Herring), the true strain potential is nonlocal. The implications of altering his formulation to reflect this fact remains unresolved. Modification of the Zhang Hamiltonian [198] along with possibly *ab initio* computation to investigate this issue is an interesting research avenue.

ACKNOWLEDGMENTS

The second author acknowledges fruitful collaborations with Xinyuan Zhang, Dr. Surya Ganti and Dr. Azar Alizadeh. We are grateful to the researchers who kindly provided us with permissions and original pictures from their research for inclusion in this article.

REFERENCES

1. Y. Arakawa, *IEEE, J. Sel. Top. Quantum Electron.* 8, 823 (2002).
2. S. Nakamura, S. Pearton, and G. Fasol, "The Blue Laser Diode: The Complete Story," Springer, Berlin, 2000.
3. D. G. Deppe and D. L. Huffaker, *Appl. Phys. Lett.* 77, 3325 (2000).

4. P. Bhattacharya and J. Ghosh, *Proc. SPIE* 5361, 29 (2004).
5. T. Lundstrom, W. Schoenfeld, H. Lee, and P. M. Petroff, *Science* 286, 2312 (1999).
6. Y. Chye, M. E. White, E. Johnston-Halperin, B. D. Gerardot, D. D. Awschalom, and P. M. Petroff, *Phys. Rev. B* 66, 201301 (2002).
7. P. M. Petroff, *Top. Appl. Phys.* 90, 1 (2003).
8. P. Alivisatos, *TOPS* 40, 86 (2000).
9. P. Bhattacharya, A. D. Stiff-Roberts, S. Krishna, and S. Kennerly, *Int. J. High Speed Electron. and Syst.* 12, 969 (2002).
10. J. Tersoff, C. Teichert, and M. G. Lagally, *Phys. Rev. Lett.* 76, 1675 (1996).
11. D. Bimberg, *Semiconductors* 33, 951 (1999).
12. D. Bimberg, M. Grundmann, and N. N. Ledentsov, "Quantum Dot Heterostructures," Wiley, New York, 1998.
13. J. O. Winter, T. Y. Liu, B. A. Korgel, and C. E. Schmidt, *Adv. Mater.* 3, 1673 (2001).
14. J. C. Chen, A. M. Chang, and M. R. Melloch, *Phys. Rev. Lett.* 92, 176801 (2004).
15. I. L. Medintz, A. R. Clapp, H. Mattoussi, E. R. Goldman, B. Fisher, and J. M. Mauro, *Nat. Mater.* 2, 630 (2003).
16. L. B. Freund and H. T. Johnson, *J. Mech. Phys. Solids* 49, 1925 (2001).
17. H. T. Johnson and L. B. Freund, *Int. J. Solids. Struct.* 38, 1045 (2001).
18. E. P. O'Reilly, *Semicond. Sci. Technol.* 4, 121 (1989).
19. E. P. O'Reilly and A. R. Adams, *IEEE J. Quantum Electron.* 30, 366 (1994).
20. A. Alizadeh, P. Sharma, S. Ganti, S. F. LeBoeuf, and L. Tsakalakos, *J. Appl. Phys.* 95, 8199 (2004).
21. P. Harrison, "Quantum Wells, Wires and Dots: Theoretical and Computational Physics," Wiley, New York, 2000.
22. J. Singh, "Physics of Semiconductors and Their Heterostructures," McGraw-Hill, London, 1992.
23. M. Yang, J. C. Strum, and J. Provost, *Phys. Rev. B* 56, 1973 (1997).
24. M. S. Miller, J.-O. Malm, M. E. Pistol, S. Jeppesen, B. Kowalski, K. Georgsson, and L. Samuelson, *J. Appl. Phys.* 80, 3360 (1996).
25. G. S. Solomon, J. A. Trezza, J. A. Marshall, and J. A. Harris, Jr., *Phys. Rev. Lett.* 76, 952 (1996).
26. Q. Xie, A. Madhukar, P. Chen, and N. P. Kobayashi, *Phys. Rev. Lett.* 75, 2542 (1995).
27. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 52, 11969 (1995).
28. M. Moreno, A. Trampert, B. Jenichen, L. Daweritz, and K. H. Ploog, *J. Appl. Phys.* 92, 4672 (2002).
29. A. Lenz, R. Timm, H. Eisele, Ch. Hennig, S. K. Becker, R. L. Sellin, U. W. Pohl, D. Bimberg, and M. Dahne, *Appl. Phys. Lett.* 81, 5150 (2002).
30. J. D. Eshelby, *Proc. R. Soc. London. Ser. A* 241, 376 (1957).
31. J. D. Eshelby, *Proc. R. Soc. London. Ser. A* 252, 561 (1959).
32. J. D. Eshelby, in "Progress in Solid Mechanics 2" (I. N. Sneddon and R. Hill, Eds.), North Holland, Amsterdam, 1961.
33. S. Nemat-Nasser and M. Hori, "Micromechanics: Overall Properties of Heterogeneous Solids," Elsevier, New York, 1999.
34. L. D. Landau and E. M. Lifshitz, "The Classical Theory of Fields," Butterworth-Heinemann, Oxford, 1980, Vol. 2.
35. T. Mura, "Micromechanics of Defects in Solids," 2nd Edn. Kluwer Academic, Boston, 1998.
36. O. D. Kellogg, "Foundations of Potential Theory," Dover, New York, 1954.
37. D. H. Armitage and S. J. Gardiner, *Classical Potential Theory*, Springer, New York, 2000.
38. A. Blech and E. S. Meieran, *J. Appl. Phys.* 38, 2913 (1967).
39. S. M. Hu, *J. Appl. Phys.* 50, 4661 (1979).
40. A. Fischer and H. Richter, *Appl. Phys. Lett.* 61, 2656 (1992).
41. P. A. Kirkby, P. R. Selway, and L. D. Westbrook, *J. Appl. Phys.* 50, 4567 (1979).
42. S. Christiansen, M. Albrecht, H. P. Strunk, and H. J. Meier, *Appl. Phys Lett.* 64, 3617 (1994).
43. J. Massies and N. Grandjean, *Phys. Rev. Lett.* 71, 1411 (1993).
44. F. Glas, *J. Appl. Phys.* 70, 3556 (1991).
45. J. Y. Marzin and G. Bastard, *Solid State Commun.* 92, 437 (1994).
46. C. Priester, I. Lefebvre, G. Allan, and M. Lannoo, in "Mechanisms of Thin Film Evolution" (S. M. Yalisove, C. V. Thompson, and D. J. Eagle Sham, Eds.), p. 131, CMRS, Pittsburg, 1994.
47. S. Christiansen, M. Albrecht, H. P. Strunk, and H. J. Maier, *Appl. Phys. Lett.* 64, 3617 (1994).
48. M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev. B* 52, 11969 (1995).
49. J. R. Downes, D. A. Faux, and E. P. O'Reilly, *J. Appl. Phys.* 81, 6700 (1997).
50. Y. P. Chiu, *J. Appl. Mech.* 45, 302 (1978).
51. T. Bennabas, P. Francios, Y. Androussi, and A. Lefebvre, *J. Appl. Phys.* 80, 2763 (1996).
52. G. S. Pearson and D. A. Faux, *J. Appl. Phys.* 88, 730 (2000).
53. P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafiji, J. P. R. David, A. G. Cullis, G. Hill, and J. C. Clark, *Phys. Rev. Lett.* 84, 733 (2000).
54. J. H. Davies, *J. Appl. Phys.* 84, 1358 (1998).
55. K. Nishi, A. A. Yamaguchi, J. Ahopelto, A. Usui, and H. Sakaki, *J. Appl. Phys.* 76, 7437 (1994).
56. C. Pryor, M. E. Pistol, and L. Samuelson, *Phys. Rev. B* 56, 10404 (1997).
57. C. Pryor, *Phys. Rev. B* 57, 7190 (1998).
58. C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger, *J. Appl. Phys.* 83, 2548 (1998).

59. M. Tadic, F. M. Peeters, K. L. Janssens, M. Korkusinski, and P. Hawrylak, *J. Appl. Phys.* 92, 5819 (2002).
60. M. Korkusinski and P. Hawrylak, *Phys. Rev. B* 63, 195311 (2001).
61. H. Shin, W. Lee, and Y. H. Yoo, *J. Phys.* 15, 3689 (2003).
62. S. H. Bellaiche, L. Wei, and A. Zunger, *Phys. Rev. B* 54, 18568 (1996).
63. J. Bernard and A. Zunger, *Appl. Phys. Lett.* 65, 165 (1994).
64. M. Cusack, P. Briddon, and M. Jaros, *Phys. Rev. B* 54, R2300 (1996).
65. H. Jiang and J. Singh, *Phys. Rev. B* 56, 4696 (1997).
66. P. N. Keating, *Phys. Rev.* 145, 637 (1966).
67. R. M. Martin, *Phys. Rev. B* 1, 4005 (1970).
68. F. H. Stillinger and T. A. Weber, *Phys. Rev. B* 31, 5262 (1985).
69. W. Yu and A. Madhukar, *Phys. Rev. Lett.* 79, 905 (1997).
70. W. Yu and A. Madhukar, *Phys. Rev. Lett.* 79, 4939 (1997).
71. M. A. Makeev and A. Madhukar, *Phys. Rev. B* 67, 073201 (2003).
72. M. A. Migliorato, A. G. Cullis, M. Fearn, and J. H. Jefferson, *Phys. Rev. B* 65, 115316 (2002).
73. O. Brandt, K. Ploog, R. Bierwolf, and M. Hosenstein, *Phys. Rev. Lett.* 68, 1339 (1992).
74. J. R. Downes and D. A. Faux, *J. Appl. Phys.* 77, 2444 (1995).
75. D. A Faux, S. G. Howells, U. Bangert, and A. J. Harvey, *Appl. Phys. Lett.* 64, 1271 (1994).
76. E. Kapon, M. Walther, J. Christen, M. Grundmann, C. Caneau, D. M. Hwang, E. Colas, R. Bhat, G. H. Song, and D. M. Bimberg, *Superlattices Microstruct.* 12, 491 (1992).
77. T. Arakawa, S. Tsukamoto, Y. Nagamune, M. Nishioka, J.-H. Lee, and Y. Arakawa, *Jpn. J. Appl. Phys.* 32, L1377 (1993).
78. N. Usami, T. Mine, S. Fukatsu, and Y. Shiraki, *Appl. Phys. Lett.* 64, 1126 (1994).
79. H. Weman and E. Kapon, EPFL. www.ijm.liu.se/mate/phys/nanopto/QWires.html.
80. T. J. Gosling and J. R. Willis, *J. Appl. Phys.* 77, 5601 (1995).
81. T. Yamauchi, Y. Arakawa, and J. N. Schulman, *Surf. Sci.* 267, 291 (1992).
82. Y. Arakawa, *IEEE Trans. Fundamentals* E75-A, 20 (1992).
83. T. Yamauchi, T. Takahashi, J. N. Schulman, and Y. Arakawa, *IEEE J. Quantum Electron.* QE-29, 2109 (1993).
84. T. J. Gosling and L. B. Freund, *Acta. Mater.* 44, 1 (1996).
85. G. Bilalbegovic, *Phys. Rev. B* 58, 15412 (1998).
86. I. A. Ovid'ko, *Phys. Rev. Lett.* 88, 046103 (2002).
87. M. E. Sherwin, T. J. Drummond, and D. J. Srolovitz, *J. Appl. Phys.* 69, 717 (1991).
88. D. A. Faux, J. R. Downes, and E. P. O'Reilly, *J. Appl. Phys.* 80, 2515 (1996).
89. K. Nishi, A. A. Yamaguchi, J. Ahopelto, A. Usui, and H. Sakaki, *J. Appl. Phys.* 76, 7437 (1994).
90. J. R. Downes, D. A. Faux, and E. P. O'Reilly, *J. Appl. Phys.* 82, 3754 (1997).
91. R. W. Keyes, *J. Appl. Phys.* 33, 3371 (1962).
92. A. D. Prins and D. J. Dunstan, *Philos. Mag. Lett.* 58, 37 (1988).
93. A. D. Andreev, J. R. Downes, D. A. Faux, and E. P. O'Reilly, *J. Appl. Phys.* 86, 297 (1999).
94. A. E. Romanov, P. M. Petroff, and J. S. Speck, *Appl. Phys. Lett.* 74, 2280 (1999).
95. H. Lee, J. A. Johnson, J. S. Speck, and P. M. Petroff, *J. Vac. Sci. Technol.* 18, 2193 (2000).
96. R. D. Mindlin and D. H Cheng, *J. Appl. Phys.* 21, 926 (1950).
97. R. D. Mindlin, and D. H Cheng, *J. Appl. Phys.* 21, 931 (1950).
98. J. N. Goodier, *Philos. Mag.* 23, 1017 (1937).
99. S. P. Timoshenko and J. N. Goodier, "Theory of Elasticity," 3rd Edn. McGraw-Hill, New York, 1970.
100. J. H. Davies, *J. Appl. Mech.* 70, 655 (2003).
101. R. D. Mindlin, *Midwestern Conf. Solid Mech.* 56 (1953).
102. K. Seo and T. Mura, *J. Appl. Mech.* 46, 568 (1979).
103. F. Glas, *J. Appl. Phys.* 90, 3232 (2001).
104. F. Glas, *Appl. Surf. Sci.* 188, 9 (2002).
105. F. Glas, *Philos. Mag. A* 82, 2591 (2002).
106. R. I. Barabash and M. A. Krivoglaz, *Fiz. Tverd. Tela* (Leningrad) 29, 3081 (1987).
107. A. E. Romanov, G. E. Beltz, W. T. Fischer, P. M. Petroff, and J. S. Speck, *J. Appl. Phys.* 89, 4523 (2001).
108. D. A Faux and J. Haigh, *J. Phys.* 2, 10289 (1990).
109. P. H. Dederichs and G. Leibfried, *Phys. Rev.* 188, 1175 (1969).
110. J. W. Flocken and J. R. Hardy, *Phys. Rev. B* 1, 2447 (1970).
111. R. A. Masamura and G. Sines, *J. Appl. Phys.* 41, 3930 (1970).
112. P. H. Dederichs and J. Pollman, *Z. Phys.* 255, 315 (1972).
113. T. Mura and N. Kinoshita, *Phys. Status Solidi B* 47, 607 (1971).
114. I. M. Lifshits and L. N. Rosentsverg, *Zh. Eks. Teor. Fiz* 17, 9 (1947); [*Sov. Phys. JETP*].
115. D. A. Faux and G. S. Pearson, *Phys. Rev. B* 62, R4798 (2000).
116. E. Pan and F. G. Yuan, *Int. J. Solid Struct.* 37, 5329 (2000).
117. D. M. Barnett and J. Lothe, *Phys. Norv.* 8, 13 (1975).
118. J. R. Barber and F. A. Sturla, *J. Mech. Phys. Solids* 40, 17 (1992).
119. T. C. T. Ting, "Anisotropic Elasticity," Oxford University Press, Oxford, UK, 1996.
120. K. C. Wu, *J. Elast.* 51, 213 (1998).
121. Y. C. Pan and T. W. Chou, *Int. J. Eng. Sci.* 17, 545 (1979).
122. Q. X. Pei, C. Lu, and Y. Y. Wang, *J. Appl. Phys.* 93, 1487 (2003).
123. S. W. Ellaway and D. A. Faux, *J. Appl. Phys.* 92, 3027 (2002).

124. M. J. Mehl, J. E. Osburn, D. A. Papaconstantopoulos, and B. M. Klein, *Phys. Rev. B* 41, 10311 (1990).
125. M. J. Mehl, J. E. Osburn, D. A. Papaconstantopoulos, and B. M. Klein, *Phys. Rev. B* 42, 5362 (1990).
126. M. J. Mehl, *Phys. Rev. B* 47, 2493 (1993).
127. M. Ichimura, *Phys. Status Solidi A* 153, 431 (1996).
128. S. W. Ellaway and D. A. Faux, *Phys. Status Solidi B* 235, 437 (2003).
129. M. D. Frogley, J. L. Sly, and D. J. Dunstan, *Phys. Rev. B* 58, 12579 (1998).
130. M. D. Frogley, J. R. Downes, and D. J. Dunstan, *Phys. Rev. B* 62, 13612 (2000).
131. S. P. Lepkowski and J. A. Majewski, *Acta Phys. Pol. A* 105, 559 (2004).
132. S. P. Lepkowski and J. A. Majewski, *Solid State Commun.* 131, 763 (2004).
133. B. W. Kim, *J. Appl. Phys.* 89, 1197 (2001).
134. S. Sanguinetti, M. Gurioli, E. Grilli, M. Guzzi, and M. Henini, *Thin Solid Films* 380, 198 (2000).
135. A. Patane, A. Levin, A. Polimeni, F. Schindler, P. C. Main, L. Eaves, and M. Henini, *Appl. Phys. Lett.* 77, 2979 (2000).
136. J. F. Nye, "Physical Properties of Crystals." Clarendon, Oxford, 1985.
137. B. Jogai, *J. Appl. Phys.* 90, 699 (2001).
138. E. Pan, *J. Appl. Phys.* 91, 3785 (2002).
139. E. Pan, *Proc. R. Soc. London, Ser. A* 458, 181 (2002).
140. D. M. Barnett and J. Lothe, *Phys. Norv.* 8, 13 (1975).
141. E. Pan, *J. Appl. Phys.* 91, 6379 (2002).
142. P. Sharma and S. Ganti, *Phys. Status. Solidi B* 234, R10 (2002).
143. A. Alizadeh, P. Sharma, S. Ganti, S. LeBoeuf, and L. Tsakalakos, *J. Appl. Phys.* 95, 8199 (2004).
144. P. Sharma and S. Ganti, *J. Appl. Mech.* 71, 663 (2004).
145. R. E. Miller and V. B. Shenoy, *Nanotechnology* 11, 139 (2000).
146. R. C. Cammarata, K. Sieradzki, and F. Spaepen, *J. Appl. Phys.* 87, 1227 (2000).
147. R. Kukta, A. Peralta, and D. Kouris, *J. Appl. Mech.* 69, 443 (2002).
148. D. A. Kouris, E. Tsuchida, and T. Mura, *J. Appl. Mech.* 53, 724 (1986).
149. M. E. Gurtin and A. I. Murdoch, *Arch. Ration. Mech. Anal.* 59, 389 (1978).
150. A. J. Murdoch, *J. Mech. Phys. Solids* 24, 137 (1976).
151. M. E. Gurtin, J. Weissmuller, and F. Larche, *Philos. Mag. A* 78, 1093 (1998).
152. H. Ibach, *Surf. Sci. Rep.* 29, 193 (1997).
153. P. Sharma and S. Ganti, *Philos. Mag. Lett.* 83, 745 (2003).
154. P. Sharma, S. Ganti, and N. Bhate, *Appl. Phys. Lett.* 82, 535 (2003).
155. P. Sharma and S. Ganti, *J. Appl. Mech.* 71, 663 (2004).
156. X. Peng, S. Ganti, P. Sharma, A. Alizadeh, S. Nayak, and S. Kumar, 2005, in review.
157. C. Eringen, "Nonlocal Continuum Filed Theories." Springer, New York, 2002.
158. H. Kleinert, "Gauge Fields in Condensed Matter." World Scientific, Singapore, 1989.
159. B. Delley and E. F. Steigmeier, *Appl. Phys. Lett.* 67, 2370 (1995).
160. X. Zhang and P. Sharma, *Int. J. Solids Struct.* 42, 3833 (2005).
161. D. P. DiVincenzo, *Phys. Rev. B* 34, 5450 (1986).
162. R. Furuhashi, J. H. Huang, and T. Mura, *J. Appl. Mech.* 59, 783 (1992).
163. C. Q. Ru and P. Schiavone, *Proc. R. Soc. London A* 453, 2551 (1997).
164. Z. Zhong and S. A. Meguid, *J. Appl. Mech.* 63, 877 (1996).
165. J. Qu, *Mech. Mater.* 14, 269 (1993).
166. J. Qu, *J. Appl. Mech.* 60, 1048 (1993).
167. L. J. Walpole, *Math. Proc.* Cambridge 83, 495 (1978).
168. H. A. Luo and G. J. Weng, *Mech. Mater.* 8, 77 (1989).
169. M. Cherkaoui, H. Sabar, and M. Berveiller, *J. Eng. Mater.* 116, 274 (1994).
170. G. P. Sendeckyj, Ph.D. Dissertation, Northwestern University, Evanston, 1967.
171. Z. A. Moschovidis, Ph.D. Dissertation, Northwestern University, 1975.
172. R. J. Asaro and D. M. Barnett, *J. Mech. Phys. Solids* 23, 77 (1975).
173. T. Mura and N. Kinoshita, *Phys. Status Solidi A* 48, 447 (1978).
174. M. Rahman, *J. Appl. Mech.* 69, 593 (2002).
175. A. C. Eringen, "Microcontinuum Field Theories I: Foundations and Solids." Springer, New York, 1999.
176. Z. Q. Cheng and L. H. Le, *Int. J. Eng. Sci.* 33, 389 (1995).
177. Z. Q. Cheng and L. H. Le, *Int. J. Eng. Sci.* 35, 659 (1997).
178. P. Sharma and A. Dasgupta, *Phys. Rev. B* 66, 224110 (2002).
179. G. J. Rodin and Y.-L. Hwang, *Int. J. Solids Struct.* 27, 145 (1991).
180. C. Q. Ru, *Acta. Mech.* 160, 219 (2003).
181. J. Y. Li and M. L. Dunn, *Philos. Mag. A.* 77, 1341 (1998).
182. E. Pan and B. Yang, *J. Appl. Phys.* 93 , 2435 (2003).
183. G. J. Rodin, *J. Mech. Phys. Solids* 44, 1977 (1996).
184. H. Nozaki and M. Taya, *J. Appl. Mech.* 68, 441 (2001).
185. K. Markov and L. Preziosi, "Heterogeneous Media: Micromechanics Modeling Methods and Simulations." Birkhauser, Basel, Switzerland, 2000.
186. G. J. Weng, M. Taya, and H. Abe, (Eds.), "Micromechanics and Inhomogenity: The Toshio Mura Anniversary Volume." Springer, Singapore, 1990.

187. B. A. Bilby, K. J. Miller, and J. R. Willis, IUTAM/IFC/ICM Symposium on Fundamentals of Deformation and Fracture, Sheffield, England, Eshelby Memorial Symposium. Cambridge University Press, Cambridge, 1984.
188. T. Mura, H. M. Shodja, and Y. Hirose, *Appl. Mech. Rev.* 49, S118 (1996).
189. T. Mura, *Mater. Sci. Eng. A* 285, 224 (2000).
190. Y. Kikuchi, H. Sugii, and K. Shintani, *J. Appl. Phys.* 89, 1191 (2001).
191. W. J. Drugan and J. R. Willis, *J. Mech. Phys. Solids* 44, 497 (1996).
192. J. Stangl, V. Holy, and G. Bauer, *Rev. Mod. Phys.* 76, 725 (2004).
193. A. Alizadeh, (Personal communication, 2005).
194. G. E. Pikus and G. L. Bir, *Sov. Phys. Solid State* 1, 1502 (1960).
195. G. E. Pikus and G. L. Bir, "Symmetry and Stain Induced Effects in Semiconductors," Wiley, New York, 1974.
196. T. B. Bahder, *Phys. Rev. B* 41, 11992 (1989).
197. T. B. Bahder, *Phys. Rev. B* 46, 9913 (1992).
198. Y. Zhang, *Phys. Rev. B* 49, 14352 (1994).

CHAPTER 15

Optical Properties of Silicon Quantum Wires and Dots

Xanthippi Zianni
Technological Educational Institution of Chalkida, Chalkida, Greece

Androula G. Nassiopoulou
Institute of Microelectronics, NCSR, "Demokritos," Athens, Greece

CONTENTS

1. Introduction 733
2. Electronic Structure Calculations 735
   2.1. Effective Mass Approximation 736
   2.2. Atomistic Approaches 742
   2.3. Quantum Confinement 746
   2.4. Optical Properties 748
3. Conclusion 765
   References 765

1. INTRODUCTION

This chapter is devoted to theoretical calculations of the optical properties of silicon quantum wires and dots. The understanding of these properties is essential for their application in a variety of different fields, including nanoelectronic and optoelectronic devices, physical and chemical sensors, biosensors, waveguides, photonic band gap devices, and others. Semiconductor quantum wires and dots are especially attractive not only for nanotechnology applications but also for basic investigations in nanoscience studies. Nanowires, compared to other low-dimensional systems, have two quantum-confined directions while still leaving one unconfined direction for electrical conduction. They are so useful in applications, which require electrical conduction rather than tunneling transport. In contrast, semiconductor quantum dots constitute a three-dimensional (3D) confined system, in which quantum confinement and Coulomb blockade effects are dominant.

In both cases of quantum wires and dots, their properties are quite different from their bulk 3D counterparts. A size-dependent band-gap, a different density of electronic states, an enhanced exciton binding energy, and an increased surface scattering for electrons and phonons are some of the properties that differ significantly from those of the bulk, resulting

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (733–769)

in different optical and electrical properties. However, even down to sizes as small as approximately 1 nm in the quantum confined directions, quantum wires and dots maintain a local crystal structure closely related to the corresponding bulk material, thereby allowing theoretical calculations of their properties using the bulk parameters.

Because of their large surface, compared to their volume, both quantum wires and dots show properties that are very sensitive to the surface termination and passivation. This is why different systems containing silicon quantum wires or dots may show quite different optical and electronic properties.

Silicon quantum wires may be found in different forms, as is illustrated in Fig. 1. These include nanostructured porous silicon (PS), synthesized Si wires, Si quantum pillars on a silicon substrate fabricated by lithography and etching techniques, or silicon wires on SiO_2 between electrodes. Silicon quantum dots may be fabricated in different insulating matrices on a silicon substrate for use in nanoelectronic devices, as illustrated in Fig. 2: a silicon nanocrystal layer between SiO_2, randomly distributed nanocrystals in an insulator, and multilayers or superlattices of Si nanocrystals and SiO_2. Silicon nanocrystals also may be fabricated in the form of a nanostructured powder.

For the fabrication of Si quantum wires and dots, different techniques were used. The mostly investigated Si nanostructured material is PS, for which intensive studies have been undertaken since the discovery of its intense luminescence at room temperature in 1990 [1]. PS is a complex system, composed of a Si skeleton of interconnected wires or dots in a network of pores [2, 3]. When the material is highly porous, there is experimental evidence that the size of wires or dots is in the nanometer range. This is why most of its properties that differ from those of the bulk (as, e.g., the efficient photoluminescence [PL] and light absorption in the visible range at room temperature), are attributed to quantum size effects.

To understand and explain the properties of PS, other well-defined systems composed of Si wires or dots were fabricated. Si quantum wires on a silicon substrate were fabricated in the mid-1990s, using advanced lithographic and etching techniques [4–5], and it was demonstrated that similar photoluminescence [6, 7] or electroluminescence [8] properties to those of PS may be obtained. More recently, silicon nanowires in the form of pillars or whiskers on a silicon substrate are fabricated by using the so-called vapor–liquid–solid (VLS) mechanism of anisotropic crystal growth [9, 10]. The fabricated silicon wires consist of a crystalline Si core coated by a relatively thin amorphous oxide layer (2–3 nm thick). No PL has been

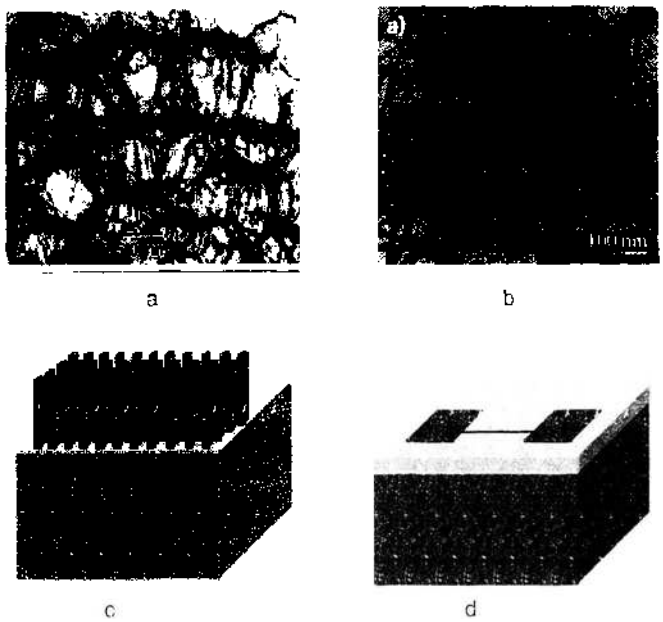


Figure 1. Different forms of silicon nanowires include (a) nanostructured porous silicon, (b) synthesized nanowires, (c) Schematic representation of Si quantum pillars on a silicon substrate, and (d) Quantum wires on insulator.

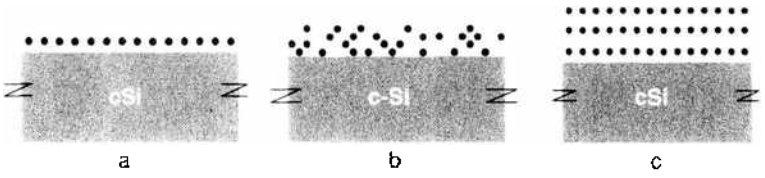


Figure 2. Si quantum dots embedded in thin SiO_2 layers. (a) 1-D layer of dots in SiO_2; (b) dots randomly distributed in SiO_2; (c) multilayers of dots/SiO_2.

reported so far from this kind of wires, mainly because their minimum diameter achieved was not below 2–3 nm.

Si quantum dots may be fabricated either in the form of a powder or in the form of a composite material, composed of silicon nanocrystals in an insulating matrix. For application in silicon nanoelectronic devices, a promising material is a thin SiO_2 layer with silicon nanocrystals embedded therein. Different approaches are used to fabricate this kind of composite film on Si. One of the techniques used is ion implantation of Si into SiO_2, followed by annealing at high temperature [11–13], which is done either at high ion energy, resulting in nanocrystals of different sizes that are randomly distributed in the SiO_2 matrix, or at very low energy and using ultra thin oxides, resulting in a two-dimensional (2D) layer within SiO_2 [14]. Another technique is low pressure chemical vapor deposition (LPCVD) of silicon on an oxidized silicon surface, followed by annealing and oxidation or silicon dioxide deposition [15–18].

The optical and charging properties of Si quantum dots in SiO_2 were intensively investigated over the last years, in view of their application in silicon optoelectronics [16, 19–24] and in silicon nanocrystal nonvolatile memories [18, 25–27]. It is now well established that quantum size effects are found at the origin of efficient PL in the visible range at room temperature from low-dimensional silicon, as they result from both experiments and theoretical calculations. A variety of computational techniques have been used for the theoretical calculations: tight-binding, empirical pseudo-potential, Hartree–Fock approximation, and density-functional theory calculations within effective mass approximation (EMA). From these calculations, it is deduced that in both quantum wires and quantum dots there is widening of the band gap from the near-infrared wavelength region to and beyond the visible range. Furthermore, an enhancement of the dipole matrix element responsible for the radiative transitions is found. Both direct and indirect gaps have been reported, depending on wire or dot size, shape, and crystallographic orientation.

Quantum confinement (QC) of carriers in Si nanowires and dots are at the origin of intense PL at room temperature. QC causes enlargement of the band gap and relaxation of the momentum conserving rule, resulting in enhanced PL and size dependence of the PL energy. Experimental evidence for this has been obtained for both porous silicon and other forms of silicon quantum wires and dots. Hydrogen-terminated nanocrystals confirm the tunability of PL [28, 29], whereas oxidized nanocrystals or nanowires show limitations in tunability, attributed to surface states [15, 28].

The measured PL lifetimes of Si quantum wires and dots strongly depend on the emitted wavelength and the temperature. Typical values are on the order of few microseconds at room temperature and milliseconds at cryogenic temperatures. PL decay is multiexponential in porous silicon [30], whereas both multiexponential [31] and single [20] or double [22] exponential decay was measured from silicon nanocrystals in SiO_2.

In this chapter, the theoretical methods used for electronic structure calculations and the optical properties of silicon quantum wires and dots are reviewed.

2. ELECTRONIC STRUCTURE CALCULATIONS

Continuum models, with the use of the envelope function approximation (EFA), have been widely used in the physics of electronic devices with active elements in the range of micrometers. In modern devices, the active elements are in the nanometric regime, where EFA reaches its limits of application and atomistic approaches are more suitable. *Ab initio*

(e.g., density-functional theory within the local density approximation, *ab initio* pseudopotential) and semiempirical (e.g., empirical pseudopotential, empirical tight binding [ETB]) atomistic methods have been used in the study of Si nanostructures. *Ab initio* approaches are limited by the small number of atoms that can be accounted for. Semiempirical techniques have proved to be successful in describing the optical properties of Si nanostructures, being simultaneously computationally efficient. In this section, we describe the fundamentals of the effective mass approximation, of the TB method, and of the pseudopotential method, focusing on the ETB and empirical pseudopotential methods.

2.1. Effective Mass Approximation

EMA is a continuum approach in which only the envelope of the nanostructure wave function is described, regardless of atomic details. EMA had great success because it is simple to handle and in many cases, although the involved assumptions, it provides reliable description.

In EMA, the wave function is in general written as a product of the Bloch periodic function and an envelope function. The envelope function and the energy eigenvalues are determined by the solution of an effective mass equation, to which are applied the appropriate boundary conditions. This method is described below for electrons and holes in Si quantum wires and dots.

2.1.1. Electrons

In bulk Si, electrons occupy at the minimum of the conduction band three pairs of equivalent valleys along the three main crystallographic directions. These anisotropic valleys are ellipsoids with two transverse masses $m_t = 0.19\ m_e$ and a longitudinal mass $m_l = 0.98\ m_e$. As becomes apparent, the anisotropic character of the conduction band of bulk Si reduces to a rich electronic structure in Si nanostructures that is responsible for their distinct optical behavior compared to, for example, the III–IV semiconductors.

We consider free-standing and homogeneous wires and dots. The distance of the center of the ellipsoids along the [001], [010], and [100] directions from the Γ-point is denoted by α. We denote these six valleys in the x, y, and z axes as [0, 0, $\pm\alpha$], [0, $\pm\alpha$, 0], and [$\pm\alpha$, 0, 0], respectively. We define the nanostructure crystallographic direction in a system of coordinates (X, Y, Z) as follows: the Z axis is along the Z direction, and the X and Y axes are rotated anticlockwise by an angle θ relative to the x and y directions, respectively. Electron eigenstates are obtained by solving Schrödinger's equation. An infinitely deep confining potential is assumed for electrons in the nanostructure.

2.1.1.1. Quantum Wires

The conduction band wave function is written as

$$\psi_c(\mathbf{r}) = \varphi_c(\mathbf{r}) u_c(\mathbf{r}) \tag{1}$$

where $u_c(\mathbf{r})$ is the Bloch function at the bottom of the bulk conduction band and $\varphi_c(\mathbf{r})$ is the electron envelope function.

In Si wires, the six anisotropic valleys are not equivalent, and to find the electron states, an effective mass equation must be solved for each pair of valleys [31, 32].

2.1.1.1.1. Eigenstates for the [0, 0, ±α] Valleys

The effective mass equation is

$$\left\{ -\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial y^2} + \frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial z} \mp a\right)^2 + V(X,Z) \right\} \varphi_i(X,Y,Z) = E_i \varphi_i(X,Y,Z) \tag{2}$$

where $V(X, Z)$ is the 2D confining potential for electrons in the wire and E_i is defined with respect to the bottom of the conduction band. By transforming the coordinate system from (x, y, z) to (X, Y, Z), Eq. (2) is transformed to

$$\left\{ -\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial Y^2} + \frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial Z} \mp a\right)^2 + V(X,Z) \right\} \varphi_i(X,Y,Z) = E_i \varphi_i(X,Y,Z) \tag{3}$$

To solve the above equation, we introduce:

$$\varphi_c(X, Y, Z) = \frac{1}{\sqrt{L_Y}} e^{ik_Y Y} e^{\mp i\alpha Z} \chi(X, Z) \tag{4}$$

where k_Y is the wave number in the Y direction and $\chi(X, Z)$ satisfies

$$\left\{ -\frac{\hbar^2}{2m_t} \frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m_l} \frac{\partial^2}{\partial Z^2} + V(X, Z) \right\} \chi(X, Z) = E^c_{XZ} \chi(X, Z) \tag{5}$$

Using the energy eigenvalue of the above equation and expressing the electron energy E_c with respect to the top of the valence band:

$$E_c = E_g + E^c_{XZ} + \frac{\hbar^2}{2m_t} k_Y^2 \tag{6}$$

where E_g is the bulk band gap energy.

Equation (5) is, in the general case, solved numerically for a given confining potential profile [HoNaTa]. For wires with rectangular cross section, it can be solved analytically:

$$\chi(X, Z) = \frac{2}{\sqrt{L_X L_Z}} \sin\left(\frac{n_X \pi X}{L_X}\right) \sin\left(\frac{n_Z \pi Z}{L_Z}\right) \tag{7}$$

where L_X, L_Z are the confinement dimensions in the X and Z directions, respectively, and n_X, $n_Z = 1, 2, 3, \ldots$.

2.1.1.1.2. Eigenstates for the [0, ±α, 0] Valleys The effective mass equation is

$$\left\{ -\frac{\hbar^2}{2m_t} \frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_t} \frac{\partial^2}{\partial z^2} + \frac{\hbar^2}{2m_l} \left(-i \frac{\partial}{\partial y} \mp a \right)^2 + V(X, Z) \right\} \varphi_c(X, Y, Z) = E_c \varphi_c(X, Y, Z) \tag{8}$$

By transforming the coordinate system from (x, y, z) to (X, Y, Z), this equation is transformed to

$$\left[-\frac{\hbar^2}{2} \left\{ \frac{1}{m_X} \frac{\partial^2}{\partial X^2} + \frac{1}{m_Y} \frac{\partial^2}{\partial Y^2} - 2\sin\theta\cos\theta \left(\frac{1}{m_t} - \frac{1}{m_l} \right) \frac{\partial^2}{\partial X \partial Y} \mp \frac{2ia}{m_l} \left(\sin\theta \frac{\partial}{\partial X} + \cos\theta \frac{\partial}{\partial Y} \right) - \frac{a^2}{m_l} + \frac{1}{m_l} \frac{\partial^2}{\partial Z^2} \right\} + V(X, Z) \right] \varphi_c(X, Y, Z) = E_c \varphi_c(X, Y, Z) \tag{9}$$

where

$$\begin{aligned} m_X &= \left(\frac{\cos^2\theta}{m_t} + \frac{\sin^2\theta}{m_l} \right)^{-1} \\ m_Y &= \left(\frac{\sin^2\theta}{m_t} + \frac{\cos^2\theta}{m_l} \right)^{-1} \end{aligned} \tag{10}$$

To solve this equation, we introduce:

$$\phi_c(X, Y, Z) = \frac{1}{\sqrt{L_Y}} e^{ik_Y Y} e^{i[\mp a \cdot \sin\theta - \Phi(k_Y)]X} \chi(X, Z) \tag{11}$$

where

$$\Phi(k_Y) = m_X \left(\frac{1}{m_t} - \frac{1}{m_l} \right) \sin\theta\cos\theta (k_Y \mp \alpha\cos\theta) \tag{12}$$

where k_Y is the wave number in the Y direction and $\chi(X, Z)$ satisfies

$$\left\{ -\frac{\hbar^2}{2m_X} \frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m_l} \frac{\partial^2}{\partial Z^2} + V(X, Z) \right\} \chi(X, Z) = E^c_{XZ} \chi(X, Z) \tag{13}$$

Using the energy eigenvalue of the above equation, we obtain

$$E_c = E_g + E^c_{XZ} + \frac{\hbar^2}{2m^*_Y}(k_Y \mp \alpha\cos\theta)^2 \tag{14}$$

where $m^*_Y = m_t\cos^2\theta + m_l\sin^2\theta$. For rectangular cross-section wires, the solution of (Eq. 13) is given by Eq. (7).

2.1.1.1.3. Eigenstates for the [±α, 0, 0] Valleys The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial y^2} - \frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial z^2} + \frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial x} \mp a\right)^2 + V(X,Z)\right\}\varphi_c(X,Y,Z) = E_c\varphi_c(X,Y,Z) \tag{15}$$

By transforming the coordinate system from (x, y, z) to (X, Y, Z), the above equation is transformed to

$$\left[-\frac{\hbar^2}{2}\left\{\frac{1}{m_X}\frac{\partial^2}{\partial X^2} + \frac{1}{m_Y}\frac{\partial^2}{\partial Y^2} + 2\sin\theta\cos\theta\left(\frac{1}{m_t} - \frac{1}{m_l}\right)\frac{\partial^2}{\partial X\partial Y} \pm \frac{2ia}{m_l}\left(\cos\theta\frac{\partial}{\partial X} - \sin\theta\frac{\partial}{\partial Y}\right)\right.\right.$$
$$\left.\left. - \frac{a^2}{m_l} + \frac{1}{m_t}\frac{\partial^2}{\partial Z^2}\right\} + V(X,Z)\right]\varphi_c(X,Y,Z) = E_c\varphi_c(X,Y,Z) \tag{16}$$

where

$$m_X = \left(\frac{\sin^2\theta}{m_t} + \frac{\cos^2\theta}{m_l}\right)^{-1}$$
$$m_Y = \left(\frac{\cos^2\theta}{m_t} + \frac{\sin^2\theta}{m_l}\right)^{-1} \tag{17}$$

To solve the above equation, we introduce:

$$\varphi_c(X,Y,Z) = e^{i(k_Y)}\, e^{i[\pm\alpha\cos\theta + \Phi(k_Y)]X}\chi(X,Z) \tag{18}$$

where

$$\Phi(k_Y) = -m_X\left(\frac{1}{m_t} - \frac{1}{m_l}\right)\sin\theta\cos\theta(k_Y \pm \alpha\sin\theta) \tag{19}$$

and k_Y is the wave number in the Y direction and $\chi(X,Z)$ satisfies Eq. (13):

$$\left\{-\frac{\hbar^2}{2m_X}\frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial Z^2} + V(X,Z)\right\}\chi(X,Z) = E^c_{XZ}\chi(X,Z)$$

Using the energy eigenvalue of the above equation, we obtain

$$E_c = E_g + E^c_{XZ} + \frac{\hbar^2}{2m^*_Y}(k_Y \pm \alpha\sin\theta)^2 \tag{20}$$

where $m^*_Y = m_t\sin^2\theta + m_l\cos^2\theta$. For rectangular cross-section wires, the solution of Eq. (13) is given by Eq. (7).

2.1.1.2. Quantum Dots States in a quantum dot are confined by the finite size, and electron and hole wave functions are spread in k-space. The electron state in the dot can be expanded in crystal basis states [33]:

$$\psi_c(\mathbf{r}) = \int d\mathbf{k}\,\varphi_c(\mathbf{k} - \mathbf{k}_0)\,u_{c\mathbf{k}}(\mathbf{r}) \tag{21}$$

where $u_{c\mathbf{k}}(\mathbf{r})$ is the Bloch function at a k-point of the bulk conduction band and the Fourier transform of the envelope function φ_c is developed around the band-edge position $\mathbf{k}_0$.

The electron envelope function $\varphi_c(\mathbf{r})$ and the energy states for dots are determined in a way similar to that described above for wires [34]. A 3D infinitely deep confining potential $V(X, Y, Z)$ is assumed in this case.

2.1.1.2.1. Eigenstates for the [0, 0, ±α] Valleys The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial y^2}+\frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial z}\mp a\right)^2+V(X,Y,Z)\right\}\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{22}$$

By transforming the coordinate system from (x, y, z) to (X, Y, Z), the above equation is transformed to

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial X^2}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial Y^2}+\frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial Z}\mp a\right)^2+V(X,Y,Z)\right\}\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{23}$$

where E_c is the electron energy with respect of the bottom of the conduction band. To solve this equation, the electron envelope function $\varphi_c(X, Y, Z)$ is written as

$$\varphi_c(X,Y,Z)=e^{\pm iaZ}\chi(X,Y,Z) \tag{24}$$

where $\chi(X, Y, Z)$ satisfies the following equation:

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial X^2}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial Y^2}-\frac{\hbar^2}{2m_l}\frac{\partial^2}{\partial Z^2}+V(X,Y,Z)\right\}\chi(X,Y,Z)=E_c\chi(X,Y,Z) \tag{25}$$

2.1.1.2.2. Eigenstates for the [0, ±α, 0] Valleys The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial z^2}+\frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial y}\mp a\right)^2+V(X,Y,Z)\right\}\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{26}$$

By transforming the coordinate system from (x, y, z) to (X, Y, Z), the above equation is transformed to

$$\left[-\frac{\hbar^2}{2}\left\{\frac{1}{m_X}\frac{\partial^2}{\partial X^2}+\frac{1}{m_Y}\frac{\partial^2}{\partial Y^2}-2\sin\theta\cos\theta\left(\frac{1}{m_t}-\frac{1}{m_l}\right)\frac{\partial^2}{\partial X\partial Y}\mp\frac{2ia}{m_l}\left(\sin\theta\frac{\partial}{\partial X}+\cos\theta\frac{\partial}{\partial Y}\right)\right.\right.$$
$$\left.\left.-\frac{a^2}{m_l}+\frac{1}{m_l}\frac{\partial^2}{\partial Z^2}\right\}+V(X,Y,Z)\right]\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{27}$$

where

$$m_X=\left(\frac{\cos^2\theta}{m_t}+\frac{\sin^2\theta}{m_l}\right)^{-1}$$
$$m_Y=\left(\frac{\sin^2\theta}{m_t}+\frac{\cos^2\theta}{m_l}\right)^{-1} \tag{28}$$

To solve this equation, the electron envelope function is written as

$$\varphi_c(X,Y,Z)=e^{id_X X}e^{id_Y Y}\chi(X,Y,Z) \tag{29}$$

where

$$d_X=\pm\frac{m_t}{m_Y}a\sin\theta\mp m_t\frac{c_{XY}}{2}a\cos\theta$$
$$d_Y=\pm\frac{m_t}{m_X}a\cos\theta\mp m_t\frac{c_{XY}}{2}a\sin\theta \tag{30}$$
$$c_{XY}=-2\sin\theta\cos\theta\left(\frac{1}{m_t}-\frac{1}{m_l}\right)$$

and $\chi(X, Y, Z)$ satisfies the following equation:

$$\left\{-\frac{\hbar^2}{2m_X}\frac{\partial^2}{\partial X^2}-\frac{\hbar^2}{2m_Y}\frac{\partial^2}{\partial Y^2}-\frac{\hbar^2}{2}c_{XY}\frac{\partial}{\partial X\partial Y}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial Z^2}\right.$$
$$\left.+V(X,Y,Z)\right\}\chi(X,Y,Z)=E^c_{XYZ}\chi(X,Y,Z) \tag{31}$$

2.1.1.2.3. Eigenstates for the [±α, 0, 0] Valleys The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial y^2}-\frac{\hbar^2}{2m_t}\frac{\partial^2}{\partial z^2}+\frac{\hbar^2}{2m_l}\left(-i\frac{\partial}{\partial x}\mp a\right)^2+V(X,Y,Z)\right\}\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{32}$$

By transforming the coordinate system from (x, y, z) to (X, Y, Z), the above equation is transformed to

$$\left[-\frac{\hbar^2}{2}\left\{\frac{1}{m_X}\frac{\partial^2}{\partial X^2}+\frac{1}{m_Y}\frac{\partial^2}{\partial Y^2}+2\sin\theta\cos\theta\left(\frac{1}{m_t}-\frac{1}{m_l}\right)\frac{\partial^2}{\partial X\partial Y}\mp\frac{2ia}{m_l}\left(\cos\theta\frac{\partial}{\partial X}-\sin\theta\frac{\partial}{\partial Y}\right)-\frac{a^2}{m_l}+\frac{1}{m_t}\frac{\partial^2}{\partial Z^2}\right\}+V(X,Y,Z)\right]\varphi_c(X,Y,Z)=E_c\varphi_c(X,Y,Z) \tag{33}$$

where

$$\begin{aligned} m_X &= \left(\frac{\sin^2\theta}{m_t}+\frac{\cos^2\theta}{m_l}\right)^{-1} \\ m_Y &= \left(\frac{\cos^2\theta}{m_t}+\frac{\sin^2\theta}{m_l}\right)^{-1} \end{aligned} \tag{34}$$

To solve this equation, the electron envelope function is written as in Eq. (29):

$$\varphi_c(X,Y,Z)=e^{id_X X}e^{id_Y Y}\chi(X,Y,Z)$$

where

$$\begin{aligned} d_X &= \pm\frac{m_t}{m_Y}a\cos\theta\pm m_t\frac{c_{XY}}{2}a\sin\theta \\ d_Y &= \mp\frac{m_t}{m_X}a\sin\theta\mp m_t\frac{c_{XY}}{2}a\cos\theta \\ c_{XY} &= 2\sin\theta\cos\theta\left(\frac{1}{m_t}-\frac{1}{m_l}\right) \end{aligned} \tag{35}$$

and $\chi(X, Y, Z)$ satisfies Eq. (31).

The electron envelope function $\phi_c(X, Y, Z)$ and the electron energy eigenvalues E_c for the three sets of valleys are obtained by solving numerically the equations for $\chi(X, Y, Z)$ for a given form of confining potential $V(X, Y, Z)$. The electron energy levels E_c, expressed with respect to the top of the valence band, are:

$$E_c = E_g + E^c_{XYZ} \tag{36}$$

where E_g is the bulk Si energy gap.

2.1.2. Holes

In the valence band (Γ_8), we have to take into account two different carrier types, which originate from the $k = 0$ degenerate heavy-hole (hh) and light-hole (lh) bands of bulk Si. The valence band wave function can be written as

$$\psi_v(\mathbf{r})=\sum_{m_j}\varphi_v^{m_j}(\mathbf{r})u_v^{m_j}(\mathbf{r}) \tag{37}$$

where $u_v^{m_j}(\mathbf{r})$ are the degenerate Bloch functions at the top of the bulk valence band. The sum extends over the four expectation values ($m_j = \pm\frac{3}{2}$ for the hh and $\pm\frac{1}{2}$ lh) of the $j = 3/2$ multiplet.

A very useful description of the band structure is given by the Kohn–Luttinger formulation [Kohn], according to which the calculation of the Γ_8 valence band starts from

the Hamiltonian:

$$H_{k_z} = \begin{bmatrix} H_{hh} & b & c & 0 \\ b^* & H_{lh} & 0 & c \\ c^* & 0 & H_{lh} & -b \\ 0 & c^* & -b^* & H_{hh} \end{bmatrix} \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{bmatrix} \tag{38}$$

where

$$\begin{aligned} H_{hh} &= \frac{\hbar^2}{2m_o}[(\gamma_1 + \gamma_2)(k_x^2 + k_y^2) + (\gamma_1 - 2\gamma_2)k_z^2] \\ H_{lh} &= \frac{\hbar^2}{2m_o}[(\gamma_1 - \gamma_2)(k_x^2 + k_y^2) + (\gamma_1 + 2\gamma_2)k_z^2] \\ b &= \frac{-\sqrt{3}i\hbar^2}{m_o}\gamma_3(k_x - ik_y)k_z \\ c &= \frac{\sqrt{3}\hbar^2}{2m_o}[\gamma_2(k_x^2 - k_y^2) - 2i\gamma_3 k_x k_y] \end{aligned} \tag{39}$$

and γ_1, γ_2, and γ_3 are the Kohn–Luttinger parameters, which can be obtained by fitting experimentally obtained hole masses, and (k_x, k_y, k_z) is the wave vector of the bulk electron. This formalism can be extended to lower-dimensionality structures by adding the confining potential term, $V(\mathbf{r})$, by replacing the k_α components in the confining dimension α by $-i\partial_\alpha$ for all confining dimensions, and by symmetrizing any product of noncommuting factors [36].

For the sake of simplicity, it is often assumed that the confining potential can be written as $V(X, Z) = V(X) + V(Z)$ for a quantum wire and $V(X, Y, Z) = V(X) + V(Y) + V(Z)$ for a quantum dot. Moreover, any dependence of the effective mass in a given direction on the motion perpendicular to it is neglected. Then, the envelope wave functions separate in the three spatial coordinates. By applying H_{k_z} to the wave function, a system of coupled eigenvalue equations for the envelope functions and the eigenenergy is obtained [36].

The degeneracy at the bottom of the valence band is removed by spin–orbit interaction. Because confinement is considerable in nanostructures, only the ground and first states are important in the optical properties. For this reason and for the sake of simplicity, hole states are in some cases calculated neglecting the mixing of hole bands. Therefore, holes are assumed to occupy parabolic subbands with minimum at the Γ-point ($k_v = 0$). Then:

2.1.2.1. Quantum Wires The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial Y^2} + \frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial Z^2} + V(X, Z)\right\}\varphi_v(X, Y, Z) = E_v\varphi_v(X, Y, Z) \tag{40}$$

The hole envelope functions $\phi_v(X, Y, Z)$ and the hole energy, E_v with respect to the top of the valence band, are

$$\varphi_v(X, Y, Z) = e^{(k_y)} \frac{2}{\sqrt{L_X L_Z}} \sin\left(\frac{n_X \pi X}{L_X}\right) \sin\left(\frac{n_Z \pi Z}{L_Z}\right) \tag{41a}$$

$$E_v = -E_{XZ}^v + \frac{\hbar^2}{2m_v}k_y^2 \tag{41b}$$

$$E_{X,Z}^v = \frac{\hbar^2}{2m_v}\left(\frac{n_X \pi}{L_X}\right)^2 - \frac{\hbar^2}{2m_v}\left(\frac{n_Z \pi}{L_Z}\right)^2, n_{X(Z)} = 1, 2, 3, \ldots \tag{42}$$

The index v runs over the hole bands (hh, lh), and m_v is the appropriate mass depending on v.

2.1.2.2. Quantum Dots The effective mass equation is

$$\left\{-\frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial X^2}-\frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial Y^2}+\frac{\hbar^2}{2m_v}\frac{\partial^2}{\partial Z^2}+V(X,Y,Z)\right\}\varphi_v(X,Y,Z)=E_v\varphi_v(X,Y,Z) \quad (43)$$

This equation is solved numerically in general. For a rectangular dot, it is solved analytically, and for a dot with confining dimensions L_X, L_Y, and L_Z, the hole envelope functions, φ_v (X, Y, Z), and the hole energy levels E_v, with respect to the top of the valence band, are

$$\varphi_v(X,Y,Z)=\frac{2}{\sqrt{L_X L_Y L_Z}}\sin\left(\frac{n_X\pi X}{L_X}\right)\sin\left(\frac{n_Y\pi Y}{L_Y}\right)\sin\left(\frac{n_Z\pi Z}{L_Z}\right) \quad (44)$$

$$E_v=-\varepsilon_v \quad (45)$$

$$\varepsilon_v=\frac{\hbar^2}{2m_v}\left(\frac{n_X\pi}{L_X}\right)^2+\frac{\hbar^2}{2m_v}\left(\frac{n_Y\pi}{L_Y}\right)^2+\frac{\hbar^2}{2m_v}\left(\frac{n_Z\pi}{L_Z}\right)^2,\ n_{X(Y,Z)}=1,2,3,\ldots \quad (46)$$

The index v runs over the hole bands (hh, lh), and m_v is the appropriate mass depending on v.

2.2. Atomistic Approaches

2.2.1. Tight Binding Method

The TB method is an atomistic description, and the structural details of the nanostructure are taken into account. In the TB method, all bands can be included and realistic band structures and reliable structural properties can be obtained. Its advantage is that it can provide quantitative predictions of electronic and optical properties of nanostructures at a computational cost that is comparable to that of continuum methods.

TB approaches range from empirical to *ab initio* formulations, depending on how the atom interactions are implemented. It involves a large number of parameters that are determined by fitting the bulk band structure, which is known by other methods, such as first principle calculations or experimental data. In some systems to be modeled, the parameters need to be tested on atom clusters. The transferability of the parameters to the physical system under investigation is an essential feature of the method.

In the TB method, the system wave function Ψ is expressed as a linear combination of atomic orbitals (LCAO):

$$\Psi=\sum_{a\mathbf{R}}c_{a\mathbf{R}}\varphi_{a\mathbf{R}} \quad (47)$$

where $\varphi_{a\mathbf{R}}$ is the α-th free atom orbital of atom at site $\mathbf{R}$. The Schrödinger equation is then written as

$$\sum_{a'\mathbf{R}'}c_{a'\mathbf{R}'}[H_{a\mathbf{R},a'\mathbf{R}'}-ES_{a\mathbf{R},a'\mathbf{R}'}]=0 \quad (48)$$

where E is the energy, $H_{a\mathbf{R},a'\mathbf{R}'}$ is the Hamiltonian matrix element:

$$H_{a\mathbf{R},a'\mathbf{R}'}=\langle a,\mathbf{R}|H|a',\mathbf{R}'\rangle \quad (49)$$

and $S_{a\mathbf{R},a'\mathbf{R}'}$ is the overlap matrix element between the atomic-like orbitals:

$$S_{a\mathbf{R},a'\mathbf{R}'}=\langle a,\mathbf{R}|a',\mathbf{R}'\rangle \quad (50)$$

The whole set of $\varphi_{a\mathbf{R}}$ is overcomplete [37], and one has to truncate the expansion on α in Eq. (47). This is done within the "minimal basis set" approximation. An analysis and discussion of this process can be found in Ref. [38]. The orthogonalization of the basis set can be done by using a Löwdin orthogonalization procedure [39] that preserves the symmetry of the basis function. When an orthogonal basis set is used, the overlap matrix S becomes the identity matrix. Although this facilitates the calculations, the use of a nonorthogonal basis set offers some other advantages such as the transferabilty of the TB parameters [40]. This is

because the nonorthogonal wave functions have a shorter range than the nonorthogonal wave functions that extends in the local environment of each atom [41].

The eigenvalues of the whole system are obtained by solving the secular problem of Eq. (48). For this purpose, the matrix elements of the Hamiltonian must be evaluated. The TB models can be classified according to the orthogonality of the basis set, the dimension of the basis set, the type of the matrix elements included (two- or three- center), and the range of interaction (the number of nearest neighbors and by the method of evaluation of the matrix elements (empirical, semiempirical, *ab initio*). ETB methods are often used to treat electronic and optical properties of nanostructured devices, whereas *ab initio* implementations of TB are mainly used in material science. Because we are interested here in the optical properties of Si nanostructures, we focus on ETB.

2.2.1.1. Empirical Tight Binding Method In ETB, the Hamiltonian matrix elements are treated in an approximate way without attempting to model the potential and the explicit form of the basis functions [39]. Both the Hamiltonian and the overlap matrix elements are fitting parameters of well-established quantities, such as the band structure or the total energy of the material obtained by first-principles calculations or extracted from experimental data.

To describe a nanostructure, one has first to describe bulk materials and establish the appropriate parameterization.

2.2.1.1.1. The Bulk Hamiltonian In the case of a crystalline bulk material, the symmetry of the system reduces the problem in the primitive unit cell. The one-electron wave function can be expressed as a 3D Bloch sum on the localized basis:

$$\Psi_{n\mathbf{k}} = \sum_{\alpha\beta} c_{\alpha\beta}(n, \mathbf{k}) \left\{ \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}_\beta} \varphi_{\alpha,\mathbf{R}_\beta} \right\} \tag{51}$$

where the subscripts refer to the base atom index and to the atomic orbital index, respectively; n is the band index and $\mathbf{k}$ is the wave vector. The coefficients $c_{\alpha\beta}(n, \mathbf{k})$ of the linear combination are determined by solving Schrödinger's equation:

$$H\Psi_{n,\mathbf{k}} = \varepsilon_{n,\mathbf{k}}\Psi_{n,\mathbf{k}} \tag{52}$$

where $\varepsilon_{n,\mathbf{k}}$ represents the energy band dispersion. In the localized basis, the mean-field one-electron TB Hamiltonian is written as [42]

$$\hat{H} = \sum_{\alpha\mathbf{R}_\beta} \varphi^*_{\alpha\mathbf{R}_\beta} \varepsilon_{\alpha\beta} \varphi_{\alpha\mathbf{R}_\beta} + \sum_{\substack{\alpha,\alpha' \\ \mathbf{R}'_{\beta'} \neq \mathbf{R}_\beta}} \varphi^*_{\alpha'\mathbf{R}'_{\beta'}} t_{\alpha\alpha'}(\mathbf{R}_\beta - \mathbf{R}'_{\beta'}) \varphi_{\alpha\mathbf{R}_\beta} \tag{53}$$

where

$$t_{\alpha\alpha'}(\mathbf{R}_\beta - \mathbf{R}'_{\beta'}) = \langle \alpha, \mathbf{R}_\beta | H | \alpha', \mathbf{R}'_{\beta'} \rangle \tag{54}$$

are the hopping matrix elements for $\mathbf{R}_\beta \neq \mathbf{R}'_{\beta'}$, and

$$\varepsilon_{\alpha\beta} = \langle \alpha, \mathbf{R}_\beta | H | \alpha, \mathbf{R}_\beta \rangle \tag{55}$$

are the on-site matrix elements.

2.2.1.1.2. Parameterization The system properties are sensitive to the parameters chosen, and parameterization is crucial for obtaining a reliable description of the system. The type of parameterization is closely related to the type of properties of interest. For example, to describe the electronic and optical properties, one needs to reproduce energy gaps, effective masses, and so forth, so a band structure parameterization would be appropriate.

ETB parameterizations in which the matrix elements were determined by comparison with experimental data are given by different authors [39, 43]. Harrison [44] developed a simple and universal parameterization. The diagonal matrix elements are proportional to atomic ionization energies, and the off-diagonal elements are universal functions of the interatomic distance. This has served to get simple and correct description of many physical properties

(e.g., Ref. [45]). However, a more accurate description of the materials is needed for the study of the electronic and optical properties of nanostructures.

The choice of the basis demands particular care. A sp^3 basis, for instance, succesfully reproduces the valence band of zinc-blende semiconductors, but fails to reproduce the conduction band. Vogt [42] added for this an additional s^* orbital representing high-energy orbitals. For silicon, this basis reproduces the lowest conduction bands more correctly than a first-nearest neighbor sp^3 mode [37]. The inclusion of more distant neighbors provides more accuracy. For instance, a sp^3 ETB model [46] including up to third-nearest-neighbor interactions and three center integrals has shown to provide an excellent fit to the silicon band structure, both for the valence and for the lowest conduction bands. However, this model is less good in describing the curvature of the bands near their extrema. In Ref. [47], the same method was used, but the parameters were obtained by minimizing the root-mean-square error on a weighted average of bulk band energies and the effective masses taken from *ab initio* GW calculation [48]. The same accuracy can be obtained by a first-nearest neighbor $sp^3d^5s^*$ model [49, 50]. This parameterization gives excellent agreement with pseudopotential calculations, up to 6 eV above the valence band maximum, even without the inclusion of interactions with more distant atoms and three-center integrals.

The advantage of the ETB is its simplicity and the fact that, in most cases, it provides qualitatively correct pictures. In some physical problems, fitting the bulk band structures does not lead to a unique answer (e.g., point defects [51]). To test the accuracy of ETB in intermediate situations, such as large size systems (e.g., dots), one must check the predictions in two limiting cases, bulk and the smallest dots, against those of the *ab initio* calculations.

The drawback of the ETB is the high number of parameters needed to parameterize all of the matrix element. For example, in the $sp^3d^5s^*$ model of Ref. [52] for silicon, 32 parameters are needed. Even though very sophisticated methods have been developed to define such parameters [49, 53], the parameterization and the scaling law of such matrix elements remain a critical issue in ETB methods.

2.2.1.2. Spin–Orbit Interaction Spin–orbit interaction was included in the ETB model by Chadi (see Ref. [40]). The spin–orbit matrix elements are

$$\langle a, \mathbf{R}_\beta, \sigma | H_{SO} | a', \mathbf{R}'_\beta, \sigma' \rangle = \left\langle a, \mathbf{R}_\beta \left| \frac{\hbar}{4m^2c^2} \frac{1}{r} \frac{dV_c(r)}{dr} \mathbf{L} \right| a', \mathbf{R}'_\beta \right\rangle \cdot \langle \sigma | \sigma | \sigma' \rangle \tag{56}$$

where σ are the Pauli matrices, V_c is the crystal potential and L is the angular momentum.

In the TB model, the spin–orbit interaction H_{SO} couples atomic orbitals on the same site. The spin–orbit matrix elements are evaluated between zero-order states [40]. Thus, the spin–orbit interaction is characterized by a single parameter λ. To get the correct spin–orbit splitting, (Δ) in bulk Si in [ossi-216] is taken $\lambda = \Delta/3$, where $\Delta = 0.044$ eV.

2.2.1.3. Quantum Wires and Dots In quantum wires and dots, the confinement is in 2D and 3D respectively. A proper account of the dimensionality and symmetry of the system must be used in the calculations. It is useful to introduce the idea of the "perpendicular" space, in which the translational symmetry of the system is broken, and of the "parallel" space, in which the full periodicity of the crystal is preserved. It is then possible to define a unit cell of the system under investigation in such a way that some basis vectors belong to the perpendicular space and the rest of the basis vectors belong to the parallel space [40]. A given Bravais vector can thus be decomposed into perpendicular $\mathbf{R}_\perp$ and parallel $\mathbf{R}_\parallel$ components: $\mathbf{R} = \mathbf{R}_\perp + \mathbf{R}_\parallel$. For $\mathbf{R}_\parallel$, a reciprocal parallel space can be defined, as can, consequently, the Brillouin zone for the $\mathbf{k}_\parallel$ vectors. In the case of dots, where the translational symmetry is lost in all three dimensions, the parallel space will be empty and the perpendicular space will coincide with the Bravais lattice ($\mathbf{R} = \mathbf{R}_\perp$).

In the directions, where the system is not confined, the translational symmetry is preserved and Bloch's theorem can be applied. This allows for a restriction of the problem to the unit cell and a consequent reduction of the number of atoms considered. The wave function of

the system $\Psi_{n\mathbf{k}}$ can be written as linear combination of planar Bloch sums in the parallel space $\Phi_{\alpha,\mathbf{R}_\parallel,\mathbf{k}_\parallel}$[54]:

$$\Psi_{n\mathbf{k}} = \sum_{\alpha,\mathbf{R}} C^n_{\alpha,\mathbf{R}}(\mathbf{k})\,\Phi_{\alpha,\mathbf{R}_\parallel,\mathbf{k}} \tag{57}$$

with

$$\Phi_{\alpha,\mathbf{R}_\parallel,\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_\parallel} e^{i\mathbf{k}_\parallel\cdot\mathbf{R}}\,\Phi_{\alpha,\mathbf{R}} \tag{58}$$

where $\mathbf{k}_\parallel$ is the parallel space vector and N is the number of unit cells in the parallel space. The subscript α refers both to the basis atom and to the atomic orbital.

For a given $\mathbf{k}_\parallel$, the eigenenergies E are calculated by solving the secular equation

$$H\Psi_{n\mathbf{k}} = (H^S + V^H)\Psi_{n\mathbf{k}} = E\Psi_{n\mathbf{k}} \tag{59}$$

where H^S is the system TB Hamiltonian and V^H is the Hartree potential, which can represent an externally applied potential or the internal potential resulting from charge interactions. Considering orthonormalized basis functions in the Bloch sum expansion, the secular equation reduces to

$$\sum_{\mathbf{R}',\alpha'} H_{\mathbf{R}_\parallel,\alpha;\mathbf{R}'_\parallel,\alpha'} C_{\mathbf{R}'_\parallel,\alpha'} = EC_{\mathbf{R}_\parallel,\alpha} \tag{60}$$

which needs to be solved with appropriate boundary conditions.

The boundary conditions depend on the physical problem investigated. When the nanostructure consists of an active part in a confined region, the system can be described, including the atoms belonging to this region and a few more atoms of the confining layers. In this case, cluster boundary conditions can be applied, and only a limited number of atoms or atomic planes in an empty space are considered. Outside the cluster of atoms, the wave function vanishes. The dangling bonds at the surface of the nanostructure introduce spurious surface states. These can be removed by hydrogenating the dangling bonds or by saturating them with atoms, depending on the system investigated (e.g., with oxygen in surface oxidized nanostructures).

2.2.2. The Pseudopotential Method

The pseudopotential method is an atomistic approach that represents a higher level of sophistication than TB. The central idea of this method is the pseudopotential, which is defined with the use of some parameters. The number of the needed parameters is small, which is an advantage of the method compared with TB. The empirical pseudopotential method (EPM) has been proved very successful in describing the properties of Si nanostructures, and we focus on this method here.

The electronic structure of nanostructures is described in terms of the solutions of an effective single-particle Schrödinger's equation [55]:

$$\hat{H}\Psi_i = \varepsilon_i\Psi_i \tag{61}$$

The mean-filed potential $V(\mathbf{r})$ is constructed as a superposition of atomic pseudopotentials:

$$V(\mathbf{r}) = \sum_{atom} v_{atom}(\mathbf{r} - \mathbf{R}_{atom}) \tag{62}$$

and the wave functions are expanded in plane waves

$$\Psi_i(\mathbf{r}) = \sum_{\mathbf{q}} a(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{63}$$

where $\{\mathbf{R}_{atom}\}$ are atomic position vectors and $a(\mathbf{q})$ are variationally determined expansion coefficients at the reciprocal lattice vector $\mathbf{q}$ of the supercell. The above expansion of the

wave function ensures a microscopic description and allows for multiband and intervalley coupling [56].

To solve (Eq. 61), one needs [55], to determine the position of the atoms, to determine the atomic pseudopotentials, and to diagonalize the Hamiltonian.

To determine the positions of the atoms, the positions of the atoms and the composition of the nanostructure are supplied as an input. Inside the nanostructure, the atomic positions and the interatomic distances are assumed the same as in the bulk structure, except in cases of other experimental evidence. At the surface of the nanostructure, the atoms are relaxed using first-principle calculations [56] and experimental data.

To determine the atomic pseudopotentials, they are expressed in terms of parameters that are fit to the measured band structure, anisotropic effective masses of the underlying bulk materials as well as bulk wave functions, deformation potentials, and band offsets that are provided by first-principles calculations [57–60].

The pseudopotential Hamiltonian is diagonalized with the Folded Spectrum method [61, 62]. This method scales as O(N), and therefore fast and enables us to study clusters with more than 1000 atoms. The inspiration of the folded spectrum method has been the fact that in physical properties such as optical properties, only the part of the electronic structure that is near the band gap is of interest. Thus, this method "folds" the lowest levels to very high energies, leaving the highest occupied or the lowest unoccupied states as the lowest solution of the modified Hamiltonian [61, 62]. Another modification of the method is Linear Expansion in Bloch Bands [63], which is even faster and allows the study of even larger nanostructures, such as million-atom dots.

The outcome of the above procedure is then used in the calculation of the physical properties of interest. For instance, the density of states and the optical absorption spectra can be calculated with the generalized moments method [64].

2.3. Quantum Confinement

A number of calculations on Si quantum wires [52, 65–77, 78–86] and quantum dots [87–104] have been reported using EMA, ETB, EPM, and *ab initio* density-functional theory, usually within the local density approximation, (LDA). In most calculations, single, ideal wires and dots, often of rectangular cross section, are considered. When the Si atoms at the surfaces are passivated by H, all calculations conclude that quantum confinement causes an opening of the fundamental energy gap, quantization of the energy states, and new selection rules for optical dipole transitions across the fundamental gap. These effects increase with decreasing confinement dimensions.

Figure 3 [37] shows results of ETB [88–89], EPM [61], and LDA [105] on the dependence of the energy gap from the confinement dimension of Si dots and wires. It is apparent that the band gap opening increases with confinement relative to the bulk, and it is therefore bigger in dots than in wires. The agreement between these methods gives confidence that they are reliable. The reliability of ETB originates from the good fit to the bulk band structure and the transferability of the obtained parameters to the nanostructures [37, 46, 88, 89, 106]. ETB descriptions that do not provide a good description of the bulk conduction band of Si [87, 97, 107, 108] predict lower values for the gap than *ab initio* calculations and more reliable ETB methods. The reliability of EPM originates by the good choice of the pseudopotentials. By comparing the results of the atomistic calculations with those of EMA [55, 106], it is shown that EMA overestimates the band-gap opening in quantum dots and wires. The energy gap does not follow the $1/d^2$ rule predicted by quantum confinement within EMA. Atomistic calculations showed the following behavior for wires and dots.

2.3.1. Quantum Wires

The great majority of calculations on Si quantum wires [65, 70, 72, 78–80, 109, 110], ranging from *ab initio* calculations to empirical or semiempirical methods, show a remarkable agreement. Si wires whose dimensions are of the order of 2–3 nm have energy gaps in the range of the experimental PL energies. In addition to that, the calculated dimensions compare well with the experimentally determined ones. Results for the calculated band gaps for

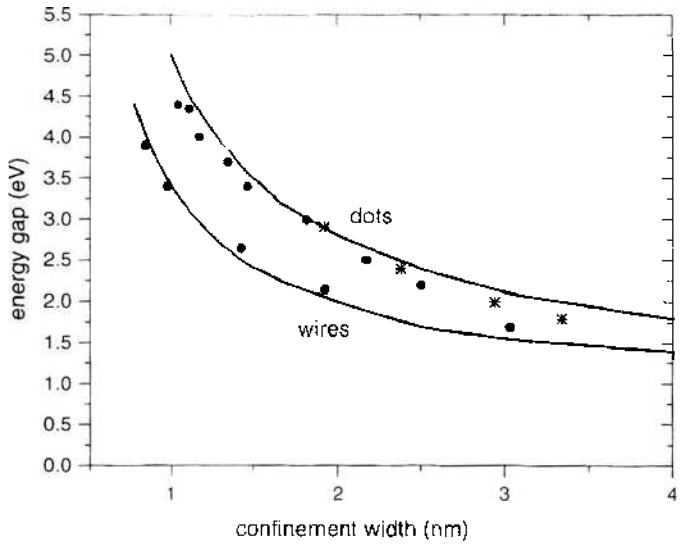


Figure 3. Energy gap versus confinement width for hydrogenated Si dots and wires (ETB: solid lines, EPM: stars, LDA: dots). The LDA results have been corrected by a rigid shift of 0.6 eV, as LDA underestimates the bulk bandgap by this amount.

various wire orientations and widths are shown in Figure 4 [80, 111]. Excitonic corrections to the band gaps, which can be as large as ~200 (100) meV for wires ~1 (2) nm wide [65, 112], do not alter the previous conclusions. The calculated band gap reduces with increasing size, but it does not follow the d^{-2} rule predicted by EMA. Figure 2 collects results for the calculated band gaps of Si quantum wires for various orientations and widths of empirical or semiempirical calculations [52, 76, 81, 82] and of local density-functional calculations [65, 70, 72, 78—80, 109]. These calculations can be fitted [80] by

$$E_g(d) = E_g(\infty) + \frac{c_1}{d} + \frac{c_2}{d^2}(eV) \tag{64}$$

where c_1 and c_2 are constants.

For wire orientations in the [100] direction of bulk Si, the calculated band gap appears at $k = 0$ [80, 112]. The near band-gap states are linear combinations of off-Γ bulk states [112]. The valence band maximum originates mainly from the coupling of the two highest-bulk

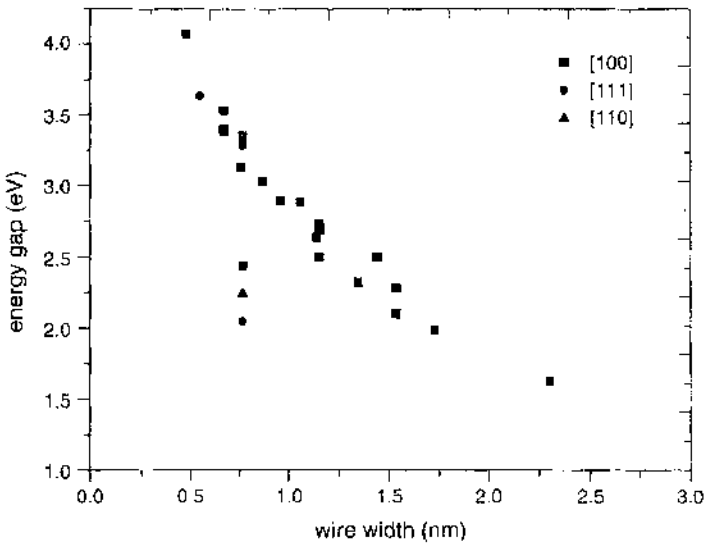


Figure 4. Collection of results for the calculated band gaps of Si quantum wires for various orientations and widths of empirical or semiempirical calculations and of local density-functional calculations as referenced in the main text.

valence bands at an off-Γ **k**-point. It is not a surface state, despite the fact that the wave function is localized near the surface. The conduction band minimum comes mainly from the lowest two bulk conduction bands at a different off-Γ **k**-points. The Si quantum wire band gap is called pseudo-direct, because of the fact that the valence band maximum and the conduction band minimum states project into bulk states of different **k** wave vectors.

The changeover from indirect conduction band in bulk Si to direct conduction band in a quantum wire along a main crystallographic direction can be qualitatively explained as follows [31, 32, 52]. In bulk Si, the indirect conduction band consists of six equivalent X valleys with minima at $\pm 0.85\frac{2\pi}{a}$ along ⟨100⟩ directions. These valleys are anisotropic ellipsoids with two light transverse masses of m_t and a heavy longitudinal mass of m_l. In a quantum wire along the [010] direction, for instance, the projections of four of these valley minima (oriented along [100] and [001] directions) onto the wire axis are at the zone center, and their energies are determined by the effective masses along the confinement directions of the wire boundaries. When projected onto the wire axis, these states give rise to the four closely spaced direct conduction subbands (also see equations of Section 2.1 for $\theta = 0$ deg). The quantum wire subbands derived from the two X valleys along [010] are indirect because the projections of their valley minima onto the wire axis are at $\pm 0.85\frac{2\pi}{a}$. The states derived from the two X valleys along [010] have higher energies than the direct minimum because the [010] valleys have light transverse masses along both confinement directions. Similar arguments explain the direct or indirect character of the band gap in quantum wires depending on the crystallographic orientation and the confinement dimension (also see discussion in Section 4.1.1). Similar to the bulk X-like states, the quantum wire conduction bands are spin-split. The ordering of the states varies with the wire width in a complicated way because of the intervalley mixing effect [52].

Both first principles and EMA results predict that the band-gap upshift is divided roughly equally between the valence and conduction edges [113], in qualitative agreement with the photoemission and X-ray absorption measurements [114]. In other atomistic calculations [72, 78], it is found that the gap opening is in fact not symmetric: Roughly 1/3 of the widening is in the valence band, while 2/3 in the conduction band.

2.3.2. Quantum Dots

In dots, the energy spectrum consists of discrete atomic-like levels. Thus, the energy gap is defined as the energy difference between the highest occupied level and the lowest unoccupied level. The energy gap of Si dots as it is calculated by EPM can be fitted by a "universal" curve [55]:

$$E_g(d) = 1.167 + 88.34/d^{1.37}(eV) \tag{65}$$

where the effective size d is defined for a cluster with N_{Si} atoms, as

$$d(N_{Si}) = \left(\frac{3}{4\pi\rho}\right)^{1/3} N_{Si}^{1/3} \tag{66}$$

where ρ is the mass density of bulk Si. These results are in agreement with third-nearest-neighbor TB results [88]. The discrepancy found between this result [55] and with EMA [90] and a model calculation [115] is another strong indication that to give reliable predictions for the energy gap of Si quantum dots one needs to use a method able to provide a good description of the Si bulk band structure.

2.4. Optical Properties

Experiments have shown that the optical properties of Si nanostructures are much different than those of bulk Si. The optical properties are closely related to the electronic structure of these systems. We discuss the results of calculations on the electronic structure and the optical properties, starting with EMA calculations. EMA is useful for obtaining the physical evidences—it is easy to handle, and it has important application in properties of Si wires, such as the effects of fluctuations in the wire thickness along its length. We proceed then with the results of atomistic calculations that provide a more accurate description.

2.4.1. Quantum Wires

2.4.1.1. Effective Mass Approximation We discuss results of a single band EMA calculation, in which the coupling of hole subbands is neglected and they are approximated by parabolic functions. The calculations are based on the formalism given in Section 2.1. It is found that the anisotropy of the conduction band of bulk Si, together with the anisotropy related to the geometry of a quantum wire, give, for Si wires, spectra for conduction subband energies very sensitive to structural parameters of the wire, such as the shape of the cross section and the crystallographic direction.

In Si quantum wires, each set of bulk Si ellipsoid valleys gives a sequence of 2D energy subbands [Eq. (5) and (13)]. This is shown in Figure 5, where the electron subband energies are plotted as a function of the wire direction for two quantum wires grown in the X–Y plane, which have the same cross section and different ratios $\lambda \equiv L_Z/L_X$: wire A ($\lambda > 1$), and wire B ($\lambda < 1$) [116]. The energy subbands from the [001] valleys (D-subbands) do not depend on the wire direction, as the change in the direction is in the X–Y plane, and their minimum is at the Γ-point [Eq. (6)]. Energy subbands originating from [010] and [100] valleys show directional dependence (I-subbands). Their minima are at the Γ-point (direct subbands) or away from the Γ-point (indirect subbands), depending on the wire direction and the valley of origin [Eqs. (14) and (20)]. The order and the separation of the energy subbands of the wires depend strongly not only on their confinement dimensions (i.e., L_X and L_Y for a wire in the Y-direction) but also on their ratio L_X/L_Z [31].

In Si quantum wires in the nanometric regime, because of considerable confinement, the number of excited e–h pair is small, and the system consists of independent e–h pairs. The lifetime, τ, is defined as the inverse of the recombination rate of a single e–h pair. The expressions for direct and for indirect (phonon assisted) recombination are, respectively [117],

$$R_{sp}(\hbar\omega) = \frac{2\pi}{\hbar} P_{cv} G(\hbar\omega) \times I_{cv}(k_y, \vartheta)\delta(E_c(k_y) - E_v(k_y) - \hbar\omega) \tag{67}$$

$$R_{sp}(\hbar\omega) = \frac{2\pi}{\hbar} P_{cv} G(\hbar\omega)|S_i^v|^2 \left(N_i + \frac{1}{2} \mp \frac{1}{2}\right)\delta(E_c - E_v \mp \hbar\omega_i - \hbar\omega) \tag{68}$$

$$|S_i^v|^2 = \frac{I_{cv}(k_y^v, \theta) \sum_{q_c, q_v} |Z_{c,q_c}^{i}(k_y^c, k_y^v)|^2}{[E_c(k_y^c) - E_v(k_y^v) - \hbar\omega]^2} \tag{69}$$

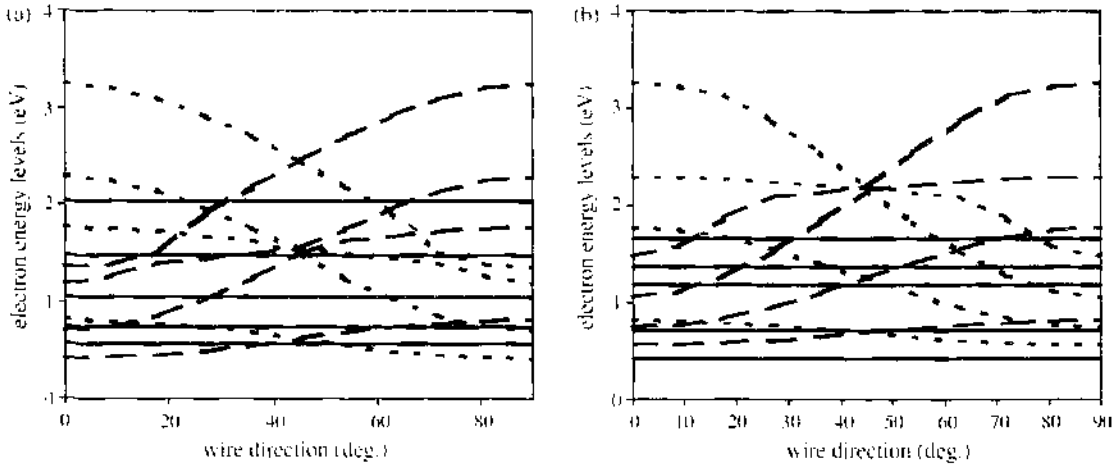


Figure 5. Electron energy levels versus wire direction for two quantum wires of confinement dimensions (a) wire A ($L_X = 2.0$ nm, $L_Y = 2.5$ nm, $L_Z/L_X > 1$) and (b) wire B ($L_X = 2.5$ nm, $L_Y = 2.0$ nm, $L_Z/L_X < 1$). Electron states are shown from the three pairs of valleys in [001] (solid lines), [010] (dotted lines), and [100] (dashed lines) directions. Energy levels are expressed relative to the bottom of the bulk Si conduction band.

In the above expressions, $\hbar\omega$ is the energy of the emitted photon, and $\hbar\omega_\ell$ is the energy of a phonon of branch ℓ. Phonons are considered dispersionless. For illustration purposes, the results shown here involve TO phonons and conduction band phonon-assisted transitions. The parameters are those for bulk Si [118]. P_{cv} is related to the bulk momentum matrix element, and $G(\hbar\omega)$ is the optical density of states. $I_{cv}(k_y, \theta)$ is the overlap integral for transition between the conduction and the valence band, and it is given in terms of the electron and hole envelope functions [116]. The overlap integrals for intersubband transitions depend on the phase of the electron envelope function, and they are consequently strongly dependent on the wire direction and the wire dimensions. The variable $Z_c^{\ell'}$ is the overlap integral for intrasubband transitions in the conduction that are induced by phonon absorption or emission.

For the sake of discussion, we choose a quantum wire with $L_X = 2.0$ nm, $L_Z = 2.2$ nm, and $\lambda = 1.1$. Because $\lambda > 1$, the ground subband is the I-subband for small angles θ, and the D-subband for bigger angles (Fig. 5). The lifetime for the D-subband does not depend on the wire direction and is on the order of microseconds (Fig. 6). This order of magnitude is determined by the overlap integral that is 10^{-4} in the present case and, in general is well smaller than unity in the case of Si wires in the nanometric regime [117]. The overap integral is determined by the confinement dimensions and by the phase of the wave function. The phase of the wave function [Eqs. (4), (11), and (18)] is related to the bulk Si conducton band minimum and effective masses and to the direction of the quantum wire. The lifetime for the I-subband depends on the direction of the wire, following the directional dependence of the overlap integral. It can vary from being on the order of milliseconds to being on the order of microseconds, with the wire direction (Fig. 6). These features are smoothed out by the thermal averaging over the possible occupied carrier states (Fig. 7). Moreover, at finite temperatures, thermally activated direct transitions are possible for the I-subband for small angles θ, because the subband minimum is close to the Γ-point (Fig. 7). These transitions become less probable as θ increases, and which is why the PL lifetime increases. At bigger angles, recombination is a result of indirect transitions, and the magnitude of lifetime depends on the wire direction in the same way as in the low-temperature regime.

The distinct features of the PL lifetimes when electrons occupy the I-subband or the D-subband, discussed in the previous section, show that the PL lifetime strongly depends on the electronic structure of the quantum wire. Thermal activation from one subband to the other is also important, as is shown in the following analysis. At low temperatures, only the ground subband is occupied and the dependence of the lifetime on the wire direction

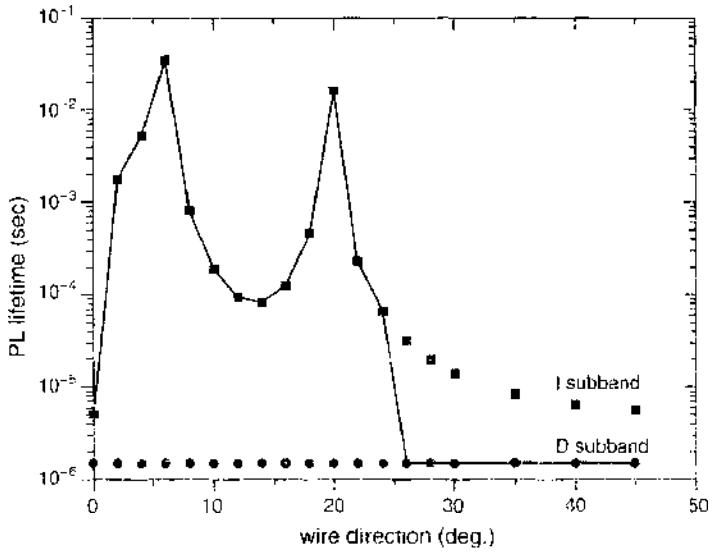


Figure 6. The calculated PL lifetime as a function of the wire direction ($L_X = 2$ nm, $L_Z = 2.2$ nm) at very low temperatures, for the I-subband, (squares) and the D-subband (dots). The solid line is for the total PL lifetime of the quantum wire.

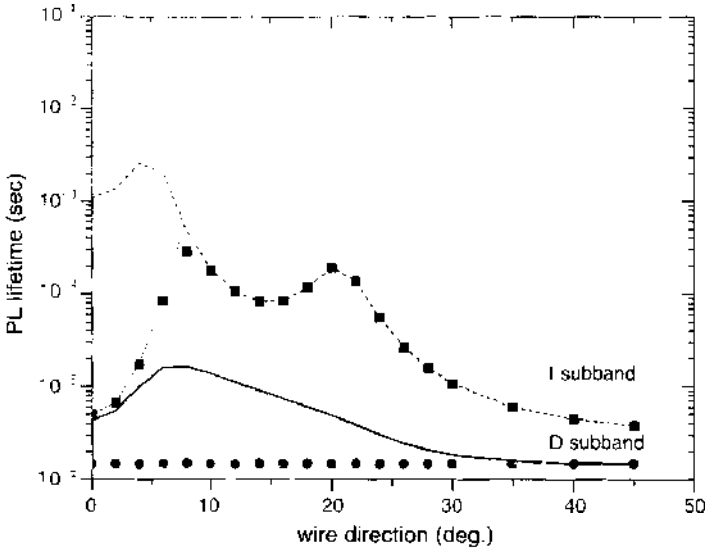


Figure 7. The calculated PL lifetime versus the wire direction ($L_x = 2$ nm, $L_z = 2.2$ nm), at room temperature, for the I-subband (squares) and the D-subband (dots). The dashed line is for the direct transitions of the I-subband, and the dotted line is for the phonon-assisted transitions of the I-subband. The solid line is for the total PL lifetime of the quantum wire.

is shown in Figure 6. At room temperature, occupation of the first excited subband may be activated, and the resulting lifetime is then given by the following expression

$$\tau^{-1} = \frac{\tau_F^{-1} e^{-\Delta E/kT} + \tau_S^{-1}}{1 + e^{-\Delta E/kT}} \tag{70}$$

where τ stands for the total PL lifetime and $\Delta E = E_F - E_S$ for the energy separation of the two subbands. The subscripts F and S are for "fast" and "slow," respectively, and refer to the D-subband or the I-subband, depending on which is faster or slower than the other. The lifetime shows a variation of one order of magnitude at room temperature as the wire orientation changes (Fig. 7).

The behavior presented in the previous paragraph is typical for wires with similar ratios of confinement dimensions λ [117]. As the wire's size increases, the lifetime depends more weakly on the crystallographic direction in wires with small degree of asymmetry ($\lambda \sim 1$). This is because the subband's energy separation becomes smaller for bigger sizes, the thermal activation is more efficient, and the effect of the D-subband is more important. In contrast, in quantum wires with $\lambda < 1$, the D-subband is the ground subband, and the lifetime is weakly dependent on the crystallographic direction of the quantum wire for all sizes. It is evident that quantum wires with similar crystallographic orientations and confinement dimensions can have PL lifetimes of different order of magnitude.

For direct gap wires, the PL lifetime increases with the diameter of the quantum wires (Fig. 8). As the size of the quantum wires changes from 2 to 4 nm, the magnitude of the calculated lifetimes varies from 0.1 μsec to 1 msec. These values are in agreement with first-principle calculations [52]. The lifetimes are higher for indirect transitions. For some crystallographic directions (Fig. 8), although the tendency is an increase of the lifetime with the size of the wires, nonmonotonic behavior has been found. This behavior is caused by the nonmonotonic size dependence of the overlap integral I_{cv} in Si wires in the nanometric regime. The sizes of the quantum wires in highly luminescent porous Si samples are below 3 nm. The magnitudes of the calculated lifetimes are in agreement with those measured in porous Si samples [119–121].

2.4.1.2. Atomistic Calculations Silicon wires [001]-oriented with a square cross section with diameters up to about 1.5 nm have been studied within LDA [70, 72]. These calculations showed that, unlike bulk c-Si, the gap is direct, at the center of the Brillouin zone, and involves states of predominately bulk Si character. The lowest transitions are dipole

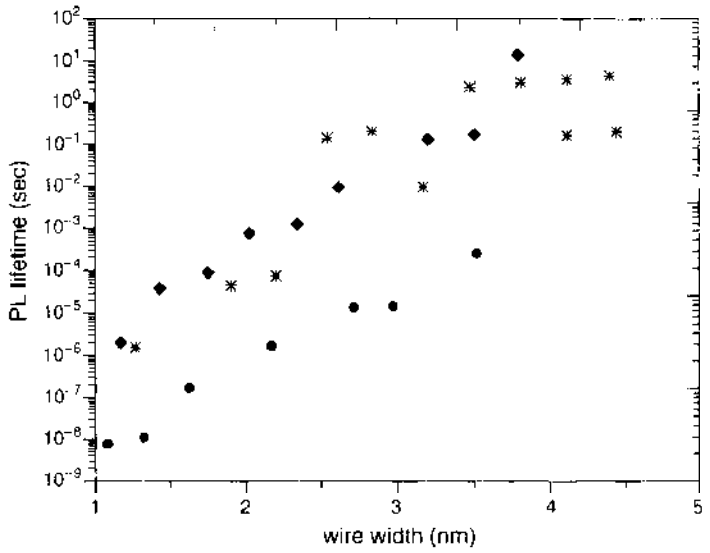


Figure 8. Size dependence of the calculated PL lifetimes, at very low temperature, for three wire directions: 0 (dots), 10 (triangles), and 30 degrees (stars).

allowed with matrix elements induced because of confinement. The LDA calculations support a more general analysis based on envelope function considerations [66, 90, 116, 117, 122] and on a TB approach [52, 88]. First-principles pseudopotential plane-wave calculations have been performed [67, 113] for rectangular wires along the [001] direction, with wire surfaces corresponding to (110)-type surfaces of bulk Si, with each surface dangling bond passivated by a H atom. Wire thicknesses up to 1.6 nm have been considered. The study of silicon wires [001]-oriented with a square cross section with diameters up to about 3. nm has been done, using a second-neighbor $sp^3d^2s^*$ ETB model that includes the spin–orbit interaction [52]. The results are in agreement with empirical pseudopotential calculations [23] on the electronic states and corresponding lifetimes of optical transitions for Si thin quantum wires.

Even the ideal wires considered have a rich electronic structure: The detailed valley-orbit splitting and ordering depends on the symmetry of the wire, as well as the structure in the surface region. The dipole matrix elements vary by more than one order of magnitude among the various low-energy interband transitions in a given wire, with interesting consequences for the optical properties and implications relevant to light-emitting porous Si.

The wires have been found to be direct with an X-like conduction-band minimum and a Γ-like valence band maximum both occuring at the zone center. The obtained valence band maximum wave functions can relate to those for a particle in a 2D box [52]:

$$f_{n_1 n_2}(x, y) = \frac{2}{L} \sin(n_1 \pi x / L) \sin(n_2 \pi y / L) \quad 0 < x, y < L \tag{71}$$

where n_1 and n_2 are the two principal quantum numbers. Detailed band structures and effective masses for the conduction and the valence band are given in Ref. [52].

The ordering of the levels at the valence band maximum agree with other first-principles results [65, 72] and are in disagreement with EMA [70, 113]. The calculated effective masses can be used in approximate EMA calculations.

The band structure of a wire about 1 nm wide is shown in Figure 9 [72]. The quantum wire bands are clearly pulled away from the bulk band edges by the kinetic energy of confinement. The bands with the most Si-H character are several eV from the band edges, with a Si-H bonding to antibonding separation of 8–10 eV. The main feature of the valence band states is the splitting at the zone center. The lowest four conduction bands derive from the $\Delta_{x,y}$ valleys, whereas the next two come from the Δ_z valleys. The latter are higher because the lighter (transverse) mass of the Δ_z valleys controls the kinetic energy of confinement. These valleys are further split by valley–orbit interaction because of the potential at the surface. The level ordering and spacing are sensitive to the details of the surface potential.

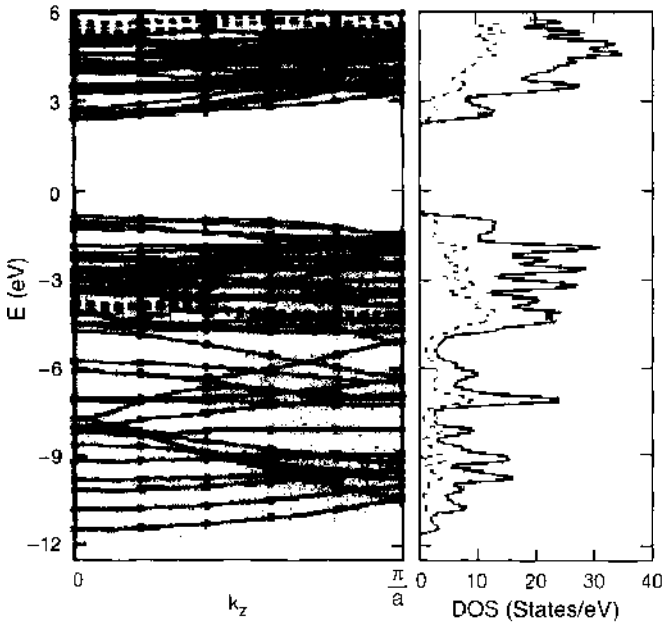


Figure 9. The band structure of the 5 × 5 wire together with the projected bulk energy bands of Si (gray) plotted versus wave vector along the wire and the associated density of states (solid curve at right with Gaussian broadening of 0.1 eV). The dashed lines indicate bands with more than 50% Si-H character. The Si-H projected density of states (dashed curve at right) is defined so as to include two electrons per Si-H bond. Reprinted with permission from [66], M. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993). © 1993, American Physical Society.

The imaginary part of the dielectric function resulting from band-to-band transitions in a quantum wire is given by Fermi's golden rule [124]:

$$\varepsilon_2(\hbar\omega) = \frac{4\pi^2 e^2}{m_e^2 V \omega^2} \sum_{n,n'} \sum_{\mathbf{k}} |\hat{e} \cdot P_{nn'}(\mathbf{k})|^2 \delta(E_{n'}(\mathbf{k}) - E_n(\mathbf{k}) - \hbar\omega) \tag{72}$$

where V is the volume of the quantum wire, $\hat{e}$ is the polarization vector, and $P_{nn'}(\mathbf{k})$ is the momentum matrix element between valence band n and conduction band n' at wave number $\boldsymbol{k}$.

The main contribution in the $\varepsilon_2(\omega)$ is a result of bulk-like excitations, but a side peak appears if the light is polarized in the direction of the wire (Fig. 10) [70]. This side peak

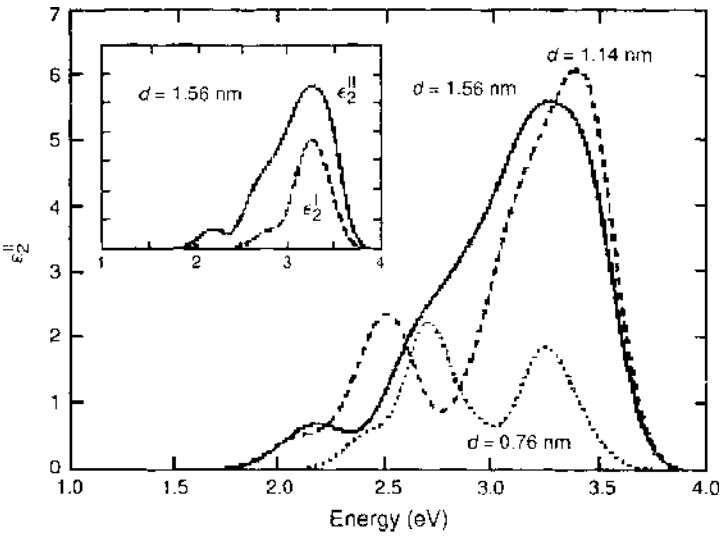


Figure 10. Imaginary part of the dielectric function polarized in the direction of the wire for three different values of the side length of the quantum wire. Inset: the same for the largest wire (solid line) compared to the component polarized in the orthogonal plane (dashed line). Reprinted with permission from [70], Buda et al., *Phys. Rev. Lett.* 69, 1272 (1992). © 1992, American Physical Society.

is caused by wire-related features and is absent in bulk c-Si. The size dependence of this optical feature can be closely related to the behavior of the luminescence peak observed in porous Si. Moreover, a high degree of polarization has been found in the calculations [52, 70] in agreement with experimental observations [125–127]. Near the band edge, the dielectric function is larger for polarization along the wire length. In contrast, the dielectric function of bulk Si is independent of polarization. Optical transitions for polarized light parallel to the wire axis are allowed between valence and conduction states of the same symmetry, whereas for light polarized normal to the wire's axis, allowed optical transitions are between states of opposite parity with respect to the planes normal to the polarization axis [52].

The absorption coefficient is

$$\alpha(\hbar\omega) = (\omega/n_v c)\varepsilon_2(\hbar\omega) \tag{73}$$

where n_v is the refraction index. Remarkable polarization effects have been found in the calculated absorption coefficient [72]. The absorption cross section and threshold are reduced for the wire of larger cross-sectional area in agreement with effective-mass models [66, 90, 122]. The confinement in these narrow wires (~1 nm) causes mixing of the bulk-band-edge-derived states with other states across the Brillouin zone. As a consequence, the low-energy transitions have direct, nonzero dipoles, in contrast to bulk Si, where phonons or defects mediate the absorption or emission of light [72]. The exact set of transitions is sensitive to the structure of the wire.

Of direct importance to understanding the luminescence of the ideal wires is the radiative rate. It is interesting to note that both empirical [52] and first-principle [72] calculations show that within a given wire, several low-energy transitions (only 0.1–0.2 eV above the lowest ones) exist with widely different radiative lifetimes that depend on the symmetry of the states involved in the transition. The calculated radiative times span several orders of magnitude, with the fastest times being on the order of tens of nanoseconds for these small wires [72]. These data are consistent with other LDA calculations [128] and with effective mass theory, which predicts strongly size-dependent induced dipoles [52, 66, 116, 117]. The decrease of the lifetime with the dimension of the wire agrees with experimental data on porous silicon [70] and this agreement is another strong indication in favor of the quantum confinement model.

2.4.1.2.1 Excitonic Effects Studies of excitons in silicon quantum wires using a two-band effective mass model are presented [52]. The electron and hole bands are assumed parabolic, and they are obtained from fits to computed quantum wire band structure. It is assumed that the exciton wave function may be written as a product of electron and wave functions describing the motion in the confining dimensions (x, y) and an exciton envelope function $G(z)$ describing the relative motion of electron and hole along the wire axis (z). The exciton wave function is given by

$$\psi(r_e, r_h) = f_{n_1 n_2}(x_e, y_e) f_{n_1' n_2'}(x_h', y_h') G(z_e - z_h) \tag{74}$$

where $f_{n_1 n_2}(x, y)$ is the wave function of a particle in a rigid two-dimensional box of area L^2 with principal quantum numbers (n_1, n_2). The envelope function $G(z)$ satisfies an effective one-dimensional Scrödinger's equation:

$$\left(\frac{-\hbar^2}{2\mu}\frac{d^2}{dz^2} + V(z)\right)G(z) = -E_B G(z) \tag{75}$$

where E_B is the binding energy. The reduced mass μ is obtained from the zone-center effective masses and the effective Coulomb potential is given by

$$V(z) = \frac{-e^2}{\varepsilon_v} \int \frac{|f_{n_1 n_2}(x_e, y_e)|^2 |f_{n_1 n_2}(x_h, y_h)|^2}{\sqrt{(x_e - x_h)^2 + (y_e - y_h)^2 + z^2}} dx_e dx_h dy_e dy_h \tag{76}$$

where ε_0 is the static dielectric constant. It has been found that $V(z)$ can be fit fairly well by a simple analytic expression with single adjustable parameter β [52],

$$V(z) \approx -\frac{e^2}{\varepsilon_0 z}(1 - e^{-\beta z}) \tag{77}$$

The lowest exciton energy is smaller than the quantum wire band gap by 0.1–0.2 eV [52]. Both decrease in the same way with increasing wire width as expected by the quantum confinement model. Excitonic transitions in narrow wires give emission energies far above the band gap of bulk silicon.

The total dielectric function, including the exciton states is given by

$$\tilde{\varepsilon}_2(\hbar\omega) = \varepsilon_2(\hbar\omega) + \frac{4\hbar\pi^2 e^2}{m_0 \omega \Omega} \sum_{n,n'} f_{nn'} \frac{\gamma/\pi}{(\hbar\omega - \hbar\omega_{nn'})^2 + \gamma^2} \tag{78}$$

where $\varepsilon_2(\hbar\omega)$ is the band-to-band dielectric function. In one dimension (or in quasi-one dimension), the spectra is singular near the threshold, and the excitonic effect does not change the spectrum appreciably [52], as is the case in three [124] and two dimensions [129, 130]. For this reason, the excitonic effect on the band-to-band transition can be ignored. The oscillator strength $f_{nn'}$ for an exciton derived from valence band n and conduction band n' is given by

$$f_{nn'} = \frac{2}{m_0 \hbar \omega_{nn'}} |\hat{e} \cdot P_{nn'}|^2 |G(0)|^2 L^{-2} \Omega \tag{79}$$

where $\hbar\omega_{nn'}$ is the exciton transition energy, $G(0)$ is the exciton envelope function at $z = 0$, L is the width of the square wire cross section, and Ω is the Si bulk unit cell. The averaged exciton oscillator strength decreases slightly faster than $1/L^5$ [52]. This is because $|G(0)|^2$ is approximately proportional to $1/L$, whereas the optical matrix element $|P_{nn'}|^2$ decreases slightly faster than $1/L^2$.

The relation between the radiative lifetime of free excitons in quantum wires and the exciton oscillator strength is given by [131]

$$\tau = \frac{m_0 c^3}{2\pi_0 e^2 \omega f_{ex}} \tag{80}$$

where f_{ex} refers to the oscillator strength per exciton instead of the oscillator strength per molecule as defined in Eq. (79). It holds [132]

$$f_{ex} = f \frac{L^2}{\Omega} \int |z| |G(z)|^2 dz \tag{81}$$

where $G(z)$ is the exciton envelope function.

The absorption spectrum near the fundamental gap is dominated by excitonic effects [52]. The anisotropy of the absorption coefficient is enhanced as the quantum wire becomes narrower. A unique feature of Si quantum-wire absorption is that higher-lying excitons can absorb more strongly than lower-lying excitons. This is attributed to intervalley mixing in the quantum-wire states.

The luminescence of the quantum wires resulting from direct recombination of excitons is proportional to the exciton oscillator strengths averaged over the occupied levels. The thermally averaged exciton oscillator strength is the physical quantity to be compared with experiment. This is given by

$$\langle f \rangle = \frac{\sum_{n,n'} f_{nn'} e^{-\hbar\omega_{nn'}/k_B T}}{\sum_{n,n'} e^{-\hbar\omega_{nn'}/k_B T}} \tag{82}$$

where k_B is the Boltzmann constant and T is the temperature. As the temperature increases, the oscillator strength varies as the population of exciton states with different oscillator strengths is modified. For quantum wires with higher-lying exciton states with much larger oscillator strengths than that of the lowest-lying exciton state, $\langle f \rangle$ increases quickly when the temperature increases [52].

2.4.1.2.2. Photoluminescence Decay Several experiments report nonexponential decay of the PL intensity in porous Si [133–136]. Both the effective mass and atomistic calculations exhibit strong size and symmetry-dependent radiative times. The nonexponential decay could be modeled by an ensemble of active nanostructures. The PL decay characteristics of porous Si are observed to be strongly temperature dependent [134–136]. The lifetime decreases as the temperature increases. The PL spectrum is inhomogeneous, which is related to the dependence of the lifetime on the wavelength. The PL decay times increase monotonically with decreasing detection energy (few microseconds for 2.2 eV and 100 μsec for 1.6 eV). It is remarkable that long decay times were systematically found in porous Si compared with other direct band-gap materials. The PL decay in porous Si has been modeled by several models. The distribution in the lifetimes has been attributed to the varying carrier localization in the porous Si nanocrystals [137]; to localized excitons that thermalize in the two excitonic states [52, 128, 138]; to the distribution of crystal shapes for a given porous Si volume [139]; and to transitions in a system of quantum wires or dots with dispersion in size, in shape, and in crystallographic direction [116, 117]. Detailed lineshape fitting of the luminescence decays with stretched exponentials as a function of T, energy, and porosity are reported [120, 137, 140, 141]. Recent papers question the presence of any carrier hopping [143, 144]. The different approaches conclude with similar results for the values of the mean lifetime.

2.4.1.2.3. Surface Passivation The surface passivation of the quantum wires strongly affects their optical properties. The presence of dangling bond states gives rise to drastic changes in the optical properties [76, 80, 112, 145]. Dangling bond-to-band state transitions are found in the energy gap, whereas band-to-band transitions are found at higher energies. Dangling bond–to–dangling bond transitions are intense and dominate the spectrum [146] of the imaginary part of the dielectric function. H chemisorption passivates the dangling bonds at the surface of the quantum wires. It results in valence band maximum and conduction band minimum wave function concentrated toward the interior of the wire, and so the band-edge states are expected to reflect the properties of the Si skeleton. H chemisorption changes the order of the states and gives denser level spacings.

The presence of oxygen reduces the band gap. The oxygen-related transitions are located in the energy gap, and the intensity of the lower-energy side peaks, in comparison with the H case, is reduced [145, 146]. Experimental findings [147] on the PL of porous silicon verify the theoretical findings. During oxidation treatment, a low-energy peak attributed to oxygen transitions appears. This peak is independent of the oxide formation. Another peak is observed that is blueshifted as the oxidation proceeds. This is related to the reduction of the confinement dimension and can be explained by quantum confinement.

2.4.1.2.4. Undulating and Interacting Wires Porous Si has been often modeled as a system of undulating quantum wires, and the theoretical results have been reviewed in Ref. [148]. Model calculations [149], LDA calculations [150, 151], and first-principles pseudopotential calculations [113] have shown that width fluctuations in a quantum wire may result in localized states.

The interaction between Si quantum wires causes reduction of the band gap resulting from interwire bonded states [80]. These states are strongly localized at the surface of the Si wires, and their tails extend in the vacuum region. As expected, the energy separation between Si core crystalline states remains unchanged. Furthermore, the energy localization of the interwire states reflects mostly the distance between wires and is practically insensitive to their dimensions. When Si quantum wires are forced to get closer, the H atoms repel each other, the Si atoms at the interface move by 0.2–0.3 a.u., while the subsurface and inner Si atoms remain practically in the same position. The results on ε_2 are [80]: the reduction of the energy threshold, the increase of the absolute value of ε_2 caused by the reduced porosity of the cell, and the presence of new transitions in the low energy tail related to the interface localized interwire states. The presence of states localized at the interfaces seems to be important for the enhancement of the transition probabilities.

2.4.2. Quantum Dots

2.4.2.1. Effective Mass Approximation In Si quantum dots, the three pairs of bulk Si valleys give distinct sequences of energy levels [Eqs. (22), (26), and (32)]. The effective mass equations are solved numerically [34], even for simple shapes such as cubic dots, because the anisotropy of bulk Si valleys causes a deviation of the electron states from the simple sinusoidal functions. The order and the separation of the energy levels depend on the size, on the crystallographic direction, and on the shape of the dot (Fig. 11). The energy levels do not depend on the direction of the dot for the energy levels from the valleys [001] (the [001]-levels), whereas they exhibit dependence for the other two sets of valleys.

Because of the indirect gap, band-edge optical transitions in bulk Si are only possible with the assistance of phonons to supply the momentum in a second-order process. In Si quantum dots, the strong confinement of the electron and hole wave functions in real space leads to a spread of the wave functions in momentum space. Thus radiative recombination can proceed by zero phonon transitions.

For single electron-hole pair recombination, the radiative lifetime, τ, is given by the following expressions [131].

2.4.2.1.1. For Zero-Phonon Transitions

$$\frac{1}{\tau} = \frac{4\pi}{3}\eta\frac{e^2}{\hbar^2 c^3 \varepsilon_0 m_e^2}(\hbar\omega)|p_{eh}^{(1)}|^2\delta(E_g + E_e + E_h - \hbar\omega) \tag{83}$$

where η is the Si refractive index, m_e is the electron mass, $p_{eh}^{(1)}$ is the momentum matrix element between the final and the initial states, and $\hbar\omega$ is the emitted photon energy.

2.4.2.1.2. For Phonon-Assisted Transitions

$$\frac{1}{\tau} = \frac{4\pi}{3}\eta\frac{e^2}{\hbar^2 c^3 \varepsilon_0 m_e^2}(\hbar\omega)\sum_{\lambda\mathbf{q}}|p_{eh,\lambda\mathbf{q}}^{(2\pm)}|^2\left(n_{\lambda\mathbf{q}} + \frac{1}{2} \mp \frac{1}{2}\right)\delta(E_g + E_e + E_h \mp \hbar\omega_{\lambda\mathbf{q}} - \hbar\omega) \tag{84}$$

where the symbol $\pm$ is to denote that the upper (lower) sign is for phonon absorption (emission). Summation over all phonon modes λq is included.

In the EMA [33],

$$\begin{aligned} p_{eh}^{(1)} &= p_{CV} I_{eh}(\mathbf{k}_0) \\ p_{eh}^{(2\pm)} &= p_{CV} R_\lambda I_{eh}(\mathbf{q} - \mathbf{k}_0)/N_a^{1/2} \\ I_{eh}(\mathbf{q}) &= \int d\mathbf{r}\phi_e^*(r)\phi_h(r)e^{-i\mathbf{q}\cdot\mathbf{r}} \end{aligned} \tag{85}$$

where $\mathbf{k}_0$ is bulk Si conduction band minimum wave vector and ϕ_e, ϕ_h are the electron and hole envelope functions, respectively The bulk value $|p_{CV}|^2/2m_e = 4$ eV is used. The number

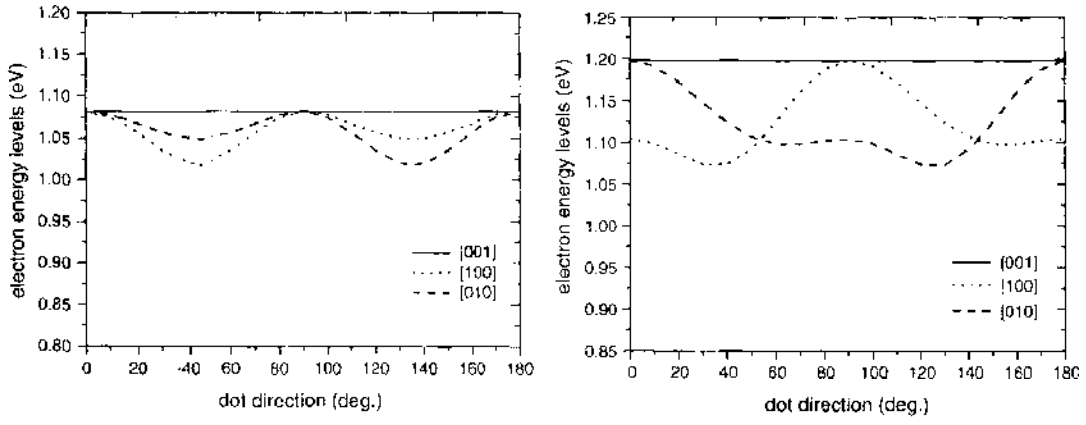


Figure 11. The electron energy levels versus the dot direction from the [001] valleys (solid line), the [100] valleys (dotted lines) and the [010] valleys (dashed line) for a cubic dot of width 2 nm (9a) and for an orthogonal dot with $L_y = L_z = 2$ nm and $L_x = 0.9L_y$ (9b).

of atoms per crystallite N_a enters explicitly, and the electron- and hole-phonon matrix elements are combined with the energy denominator into the net ratio R_λ for each phonon branch. The dependence on crystallite geometry is carried by the overlap integral I_{eh}. Because the important phonon wave vectors $\mathbf{q}$ come from a restricted region, the phonon frequencies can be taken to be approximately constant, and the sum on $\mathbf{q}$ can be done explicitly:

$$\sum_{\mathbf{q}} |I_{eh}(\mathbf{q})|^2 = \frac{V}{(2\pi)^3} \int dr |\phi_e(r)|^2 |\phi_h(r)|^2 \tag{86}$$

where V is the volume of the dot.

Calculated lifetimes for a 2-nm-wide cubic dot are shown in Figure 12 [34]. The lifetime for the [001]-level does not depend on the dot direction and is on the order of microseconds. Its magnitude depends on the overlap integral that is a function of the size of the dot and of the phase of the wave function. The phase of the electron wave function in Si dots, as in Si quantum wires, explicitly depends on the minimum of the bulk Si valleys, the effective masses, and the crystallographic orientation of the dot [Eqs. (24), (29), (30), and (35)]. The overlap integral in Si quantum dots, even in the nanometric regime, is much smaller than unity. The lifetime for the [100]-level shows a strong directional dependence, which varies in the range from micro- to milliseconds (Fig. 12).

At room temperature, although thermal averaging [Eq. (70)] makes the features of the lifetime smoother, in dots that are so small, a considerable dispersion in the magnitudes of the lifetime is found as a function of the crystallographic direction (Fig. 12). This dispersion is less remarkable in bigger dots.

2.4.2.2. Atomistic Calculations The electronic states of small spherical Si dots of sizes 1–4 nm calculated within TB [47] and EPM [55] are in agreement. These methods have been compared with EMA [47, 55–63]. It is concluded that EMA produces a rough sketch of the dot's microscopic wave functions. The limited accuracy originating from the neglect of multiband coupling restricts the accuracy of the calculated wave-function expectation values such as energy levels, exchange energies, and interelectronic Coulomb repulsion, and overestimates the effect of confinement. EMA is still useful, however, to fit experimental results and for qualitative understanding of the optical behavior of systems of dots that simulate the experimental conditions.

In Figure 13 [88, 89] results of the PL recombination rate on hydrogen terminated dots with spherical shape are presented, and they are compared with experimental data [110]. In quantum dots of size lower than 2.5 nm (the optical gap is larger than 2.3 eV), the mixing of different $\mathbf{k}$-bulk states is important, and the dots get optical properties intermediate between

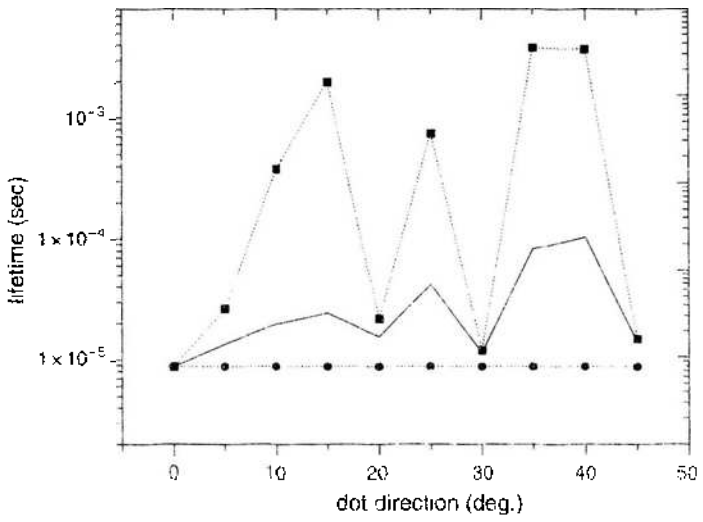


Figure 12. The lifetime versus the dot direction for a cubic dot of width 2 nm for the [001]-level (dots), the [100]-level (squares), and the total lifetime (solid line). The dashed curves are guide for the eye.

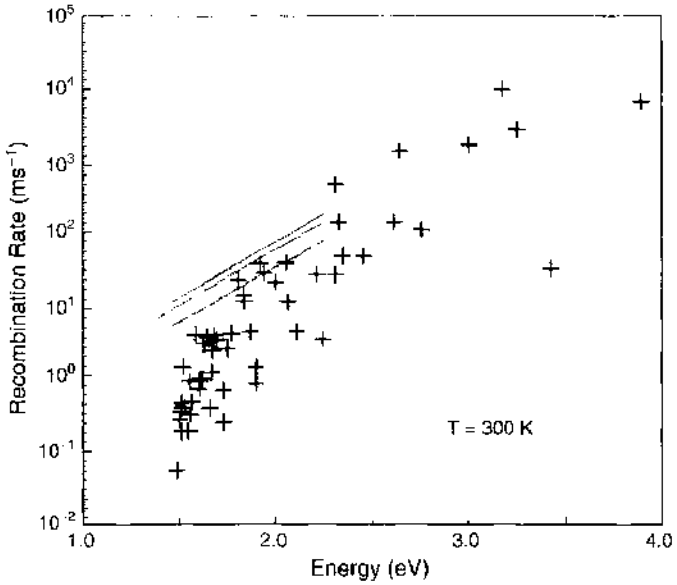


Figure 13. Calculated [89] recombination rate (ms^{-1}) of an excited electron-hole pair in silicon crystallites (crosses) with respect to the photon energy at 300 K. The spin degeneracy is not included: Its inclusion would divide the calculated recombination rates by a factor 2. Continuous lines plot the experimental dependence [110] of decay rates on photon energy for the three 65% porosity layers that differ by oxidation level. Reprinted with permission from [89], C. Delerue et al., *Phys. Rev. B* 48, 11024 (1993). © 1993, American Physical Society.

an indirect gap and direct gap material [52, 152]. The radiative recombination is quite efficient with a characteristic time lower than 10 μs. In quantum dots of bigger size (lower photon energy lower than 2 eV), the recombination rate presents an abrupt strong decrease. This is a result of the indirect nature of bulk Si, which gives a radiative recombination rate equal to zero in the limit of bulk silicon. The calculated radiative recombination rates are on average about one order of magnitude lower than the experimental decay rate, particularly for photon energy lower than 2 eV. This means that even if the confinement brings some momentum mixing to allow first-order optical transitions, the radiative recombination rates resulting from dipolar transitions are not efficient enough to account for the observed values. For large clusters, one must then take into account the contribution of phonon-assisted transitions, which are the dominant mechanism in bulk Si.

An EMA calculation on the strength of phonon-assisted transitions in Si quantum dots showed that the phonon-assisted transitions indeed dominate in large clusters, but the no-phonon transitions are likely to become comparable for dots sizes 1.5–2 nm [33]. Another more elaborated calculation showed different evidence, however [153]. In this model, the electronic structure is calculated in ETB, whereas the phonon matrix modes are obtained from a valence force field model. The electron–phonon matrix elements are evaluated numerically for each individual phonon mode by using a d^{-2} dependence of the TB matrix elements on distance, following Ref. [44]. The relative strengths of no-phonon and phonon-assisted processes have been determined, using the Fermi golden rule with Born–Oppenheimer products. It has been concluded that phonon-assisted processes should dominate whatever the size of the dot. The discrepancy between the two calculations may be explained by the fact that EMA is not accurate enough for the smaller dots. Size effects on the phonon modes in Si nanocrystals have been calculated in Refs. [154, 155], and a frequency-dependent behavior is obtained.

The polarization of the luminescence of Si quantum dots has been calculated within TB and interpreted using EM theory [156]. It is shown that simple selection rules cannot be established because the degree of linear polarization presents large oscillations with respect to the size of the clusters. This effect is a result of the indirect nature of the Si band gap, which leads to a dependence of the optical matrix elements on the oscillatory overlaps between electron and hole states in momentum space. However, in a statistical ensemble of crystallites elongated in a given direction, and with size larger than 2–3 nm, it has been

determined that the light is on average polarized along this direction, in agreement with experiments.

2.4.2.2.1 Surface Passivation Surface passivation effects have been recently investigated. Pseudopotential (PP) calculations of Si dots embedded in crystalline matrix have been performed [157], and the validity of effective-medium theory for optical properties has been explored [158]. The minimum crystallite diameter treatable by effective-medium theory is found to be approximately 1 nm. The influence of synthesis conditions on the structural and optical properties of passivated Si quantum dots has been investigated with density-functional and quantum Monte Carlo calculations [159]. The results demonstrate that kinetically limited nanostructures form different core structures than bulk-derived crystalline clusters, and the type of core structure that forms depends on how the cluster is passivated during synthesis. It has been found that during the formation process of 1-nm dots, relaxation of the high strain induced by the curvature and dangling bond states at the surface is in direct competition with the preference of the interior atoms to be tetrahedrally coordinated. The resulting structures consistently have a noncrystalline, double tetrahedral core, with optical gaps in good agreement with experimental measurements, illustrating that in some cases noncrystalline clusters may in fact be responsible for experimentally observed luminescence. Reconstructions unique to highly curved nanostructures that give evidence for larger optical gaps than previously predicted in reconstructed 1-nm crystalline nanoclusters have been found.

The effects of surface reconstruction on the absorption spectra and optical gaps of hydrogen-terminated Si clusters has been studied within the time-dependent LDA [160]. The calculations showed that the structural reconstruction of cluster surfaces produces new absorption bands in the lower part of the spectrum and causes significant reduction of the optical gap. The optical gaps in surface-reconstructed silicon clusters appear to be similar in size to those in clusters with partially oxidized surfaces. The stability and optical properties of Si clusters (up to 148 atoms, ~1.5–2 nm) with reconstructed surfaces have been studied with density-functional and quantum Monte Carlo calculations [161]. It is predicted that Si nanostructures with reconstructed surfaces are more stable than those with unreconstructed surfaces, and it is shown that surface step geometries unique to highly curved surfaces dramatically (2–1 eV for 1–1.8 nm) reduce the optical gaps and decrease (from one to two orders of magnitude in $d < 1.6$ nm) the excitonic lifetimes. These predictions provide an explanation of both the variations in the PL spectra of colloidally synthesized nanoparticles and observed deep-gap levels in porous silicon.

The comparison between calculations and observations revealed that experimental data on PL energies are systematically lower than calculated, whereas the optical absorption observations are in agreement with theoretical calculations (Fig. 14) [162]. This behavior has been first attributed by Koch [163] and Kanemitsu [164] to the existence of deep luminescence centers such as surface states. Theoretical calculations have indeed verified this assumption. ETB total energy calculations and LDA calculations have shown that in small crystallites (diameter <1 nm), the exciton can indeed be trapped at surface atoms and recombine from this state of the self-trapped exciton, giving PL a smaller energy than if it would give with a direct recombination [165]. Moreover, it has been shown that Si dots in their excited state relax to highly distorted equilibrium configurations that correspond to transitions with lower PL energies [166, 167]. ETB calculations [168] showed that the strong blue double peak (417 and 437 nm) PL obtained from fresh and thermally annealed porous Si with crystallite sizes smaller than 3 nm can be explained by the electronic states caused by the vacancy defects in the gap of small Si nanocrystallites.

TB calculations [37] have shown that in hydrogen-passivated quantum dots, recombination is via free exciton states for all sizes. For oxygen-passivated dots, three recombination mechanisms are present depending on the size of the cluster. In dots with diameter d bigger than about 3 nm, recombination occurs via free excitons, and the PL energy increases as the size decreases in agreement with the quantum confinement model. For smaller dots (< 1.5 nm $< d < 3$ nm), recombination occurs via a trapped electron and a free hole. Now the PL energy increases with decreasing size less than predicted by the quantum confinement

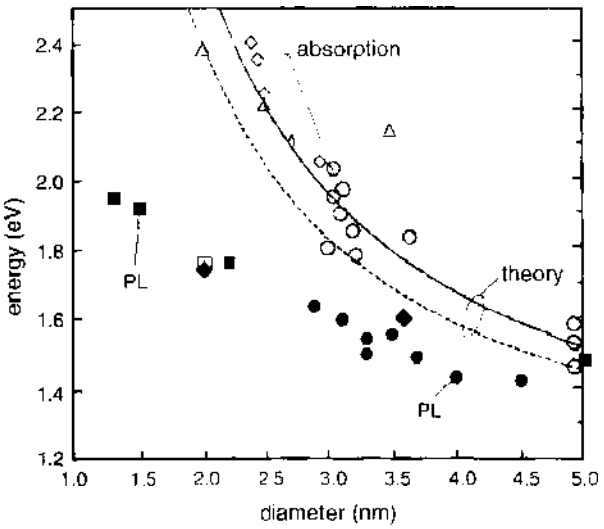


Figure 14. Compilation [200] of optical bandgaps of silicon crystallites and porous silicon samples obtained from optical absorption (empty symbols) and luminescence (full symbols). Straight line: one-electron bandgap calculated for spherical silicon crystallites [88, 89]; dashed line: the same but including the excitonic binding energy. Reprinted with permission from [106], C. Delerue et al. *Phys. Stat. Sol.* (b) 227, 115 (2001). © 2001, Wiley.

because the trapped electron state is size independent. For even smaller dots ($d < 1.5$ nm), recombination occurs via self-trapped excitons, and the PL energy is now independent of the size.

The oxygen role on the structural and optoelectronic properties of Si nanodots has been systematically investigated in Refs. [169–172]. First-principle calculations based on the density-functional theory in the LDA [169–174] and quantum Monte Carlo [173, 174] demonstrated that the structural, electronic, and optical properties of Si nanocrystals strongly depend on the passivation regimes. Starting from H-covered nanocrystals, it has been shown that single-bonded O atoms originate small variations in the electronic properties and huge changes in the structural properties. In contrast, double-bond O atoms make a small contribution to geometry variations, but a strong reduction in the energy gap. Thus, the presence of double-bond O atoms seems more appropriate to explain the huge red shift in PL observed after O exposure in high porosity porous Si samples. In the case of Si nanocrystals embedded in SiO_2 matrix, the results show that the matrix is only slightly deformed by the presence of the nanocrystals, that new electronic states are originated within the SiO_2 band gap, and that both the Si atoms in the dot and the O atoms at the interface play a role in the optical properties.

2.4.2.2.2. Excitons Excitons play an important role in the optical properties of quantum dots. They exhibit different physical behavior in quantum dots than do excitons in bulk semiconductors. EMA has been widely used to estimate the electron-hole Coulomb and exchange energies in nanostructures [175–178]. Assuming an infinite potential barrier at the boundaries of the quantum dot and using the envelope functions of a noninteracting electron-hole pair, simple, analytical expressions can be obtained for E_{Coul} and ΔE_{exch} [176, 177]:

$$E_{Coul}^{EMA} = C_{Coul}\frac{e^2}{\varepsilon R} \tag{87}$$

$$E_{exch}^{EMA} = C_{exch}\left(\frac{\alpha_x}{R}\right)E_x \tag{88}$$

where R is the dimension of the quantum dot; E_x and α_x are the bulk exciton exchange splitting and exciton radius, respectively; and C_{Coul}, C_{exch} are dimensionless constants that depend on the shape of the dot. The above equations show that the Coulomb energy E_{Coul} scales as $1/R$. Considering that EMA predicts that the single-particle energy gap shift $\Delta\varepsilon = \varepsilon_g - \varepsilon_g^{bulk}$ scales as $1/R^2$, with the size of the dot, in small crystallites it is predicted that

$E_{Coul} \ll \Delta\varepsilon$. As becomes apparent below, however, atomistic calculations have shown that the electron-hole Coulomb energy is underestimated in EMA.

Excitonic transitions and exchange splitting in Si quantum dots have been calculated within atomistic models for spherical Si dots, with all surface dangling bonds passivated with hydrogen, and very good agreement with experimental data has been obtained. The single-particle wave functions are first calculated (e.g., within EPM [179], within TB [180, 181]). From the solutions of this single-particle problem, a set of single-substitutions Slater determinants $\{\Phi_{e,h}\}$ are constructed. These are obtained from the ground-state Slater determinant by promoting an electron from the occupied valence state ψ_h of energy ε_h to the unoccupied conduction state ψ_e of energy ε_e. The exciton wave functions $\Psi^{(a)}$ are expanded in terms of this determinental basis set:

$$\Psi^{(a)} = \sum_{e=1}^{N_e} \sum_{h=1}^{N_h} C_{h,e}^{(a)} \Phi_{h,e} \tag{89}$$

where N_h and N_e denote the number of hole and electron states included in the expansion of the exciton wave functions. The matrix elements of the many-particle Hamiltonian H in the basis set $\{\Phi_{e,h}\}$ are calculated as

$$\begin{aligned} H_{he,h'e'} &\equiv \langle \Phi_{he} | H | \Phi_{h'e'} \rangle \\ &= (\varepsilon_e - \varepsilon_h)\delta_{e,e'}\delta_{h,h'} - J_{he,h'e'} + K_{he,h'e'} \end{aligned} \tag{90}$$

where J and K are the electron-hole Coulomb and exchange integrals, respectively:

$$J = e^2 \sum_{\sigma_1,\sigma_2} \iint \frac{\psi_{h'}^*(\mathbf{r}_1,\sigma_1)\psi_e^*(\mathbf{r}_2,\sigma_2)\psi_h(\mathbf{r}_1,\sigma_1)\psi_{e'}(\mathbf{r}_2,\sigma_2)}{\bar{\varepsilon}(|\mathbf{r}_1 - \mathbf{r}_2|, R)|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \tag{91}$$

$$K = e^2 \sum_{\sigma_1,\sigma_2} \iint \frac{\psi_{h'}^*(\mathbf{r}_1,\sigma_1)\psi_e^*(\mathbf{r}_2,\sigma_2)\psi_{e'}(\mathbf{r}_1,\sigma_1)\psi_h(\mathbf{r}_2,\sigma_2)}{\bar{\varepsilon}(|\mathbf{r}_1 - \mathbf{r}_2|, R)|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \tag{92}$$

The electron-hole Coulomb and exchange integrals of Eqs. (91) and (92) use a screening function that depends on the inter-particle distance and on the quantum dot radius R [182]. Correlation effects introduced via mixing of configurations of Eq. (87) shift the exciton energies downward [179, 180].

It has been found that the single-particle band-gap scales with size $\varepsilon_g \sim R^{-1.2}$ and Coulomb energy $E_{Coul} \sim R^{-1.5}$ [179]. This scaling comes from two factors: the fact that the pseudopotential wave functions are different than the envelope functions, and that the dielectric screening entering in the Coulomb energy calculation depends on the size R. For Si dots, even EMA using size-dependent screening gives $E_{Coul} \sim R^{-1.5}$. Thus, although simple theory indicates that Coulomb effects tend to become less important as size decreases, a more accurate calculation shows that Coulomb effects are more important than confinement effects at small sizes. Other atomistic calculations also support these results [62, 115, 183–185]. An ETB [181] calculation has shown that the electron-hole Coulomb interaction is insensitive to different ways of optimizing the TB parameters. However, it is sensitive to the choice of the atomic orbitals; this sensitivity decreases with increasing dot size. Quantitatively, TB treatments of Coulomb interactions are reliable for dots with radii larger than 1.5–2 nm, even for simple models for the basis orbitals. More detailed calculations of basis orbitals are required for smaller dots. The calculated excitonic gaps range from 2.5 to 1.5 eV for dots 1–2.5 nm and are in very good agreement with experimental data [186]. The exchange interaction is found to be oscillatory and to have a range of about 1.5 nm.

The calculations show that the electron-hole Coulomb interactions split the energies of excitons, which are degenerate in the single-particle approximation. When spin–orbit coupling is neglected, the lowest exciton level is split into two levels, the lowest triplet state and the singlet state. The calculated dark–bright excitonic splitting agrees well with experimental data. It ranges between 0 and 30 meV ($R > \sim 0.7$ nm) and has a size scaling of $R^{-2.5}$ [179]. The exchange interaction is found to increase with confinement. This can be understood by

a simple EMA argument. The exchange splitting is written as $\Delta = J \int |\phi(r, r')|^2 d^3r$, where $\phi(r_e, r_h)$ is the envelope function for the exciton and J is defined as twice the exchange integral per unit inverse volume for the conduction-band-minimum and valence band-maximum states in bulk Si. On introducing the spin–orbit coupling, the singlet and the triplet state are completely mixed together, because this coupling is of the same amplitude as the spacing between the levels [46, 187]. As a consequence, all the levels have, on average, similar lifetimes, and the two-level model is not valid. For ellipsoid crystallites, the degeneracy of the highest state in the valence band is lifted by the anisotropy. This behavior is also found in calculations on wires [69]. The excitonic spectrum remains quite complex, with several low-lying states having strongly varying lifetimes, because of the remaining degeneracy of the conduction states. For undulating ellipsoids, for maximum amplitude of the fluctuation equal to 25% of the average value of the ellipsoids, the orbital degeneracies of the valence and the conduction bands are lifted [180]. The excitonic spectrum becomes much simpler, with the lowest state having systematically a much longer lifetime than the first higher state. The two-level model then becomes valid to describe such anisotropic crystallites.

The calculated lifetimes [180] associated with the two levels of the undulating ellipsoids exhibit an important scattering that is consistent with the highly nonexponential decay of the luminescence in porous Si [188]. The difference of two or three orders of magnitude between the lifetimes of the upper and lower levels is in good agreement with experiments. However, the calculated lifetimes are between one and three orders of magnitude larger than the experimental ones. It has been found that this discrepancy can be partially explained by the Frank–Condon shift [180, 189]. In Ref. [180], it is found that $E_{abs} = E_l + 2d_{FC}$, where E_l is the luminescence energy, and d_{FC} is the Frank-Condon shift equal to the energy gain resulting from lattice relaxation after capture. It turns out that the onset in the selectively excited luminescence would be equal to $2d_{FC}$. The obtained values for d_{FC} are small but substantial; for example, on the order of 25 meV for a gap of 2.6 eV. As a consequence, the Frank–Condon shift could explain only partially the difference between the onset in the selectively excited PL and the calculated exchange splitting. It has become apparent that another mechanism, namely the existence of self-trapped excitons, is the main origin of this discrepancy.

2.4.2.23. Dielectric Constant The static dielectric constant ε_s is given by the integral of the absorption spectra $\varepsilon_2(E)$:

$$\varepsilon_s = 1 + \frac{2}{\pi} \int_0^\infty \varepsilon_2(E)/E \, dE \tag{93}$$

In bulk semiconductors, the calculated static dielectric constants are generally in good agreement with experiment [190–192]. In nanostructures, the size dependence of the dielectric constant is a key matter of interest. However, no systematic experimental or theoretical knowledge is available so far for the size dependence of ε_s. A generalization [193] of Penn's model [194] to an isotropic, spherical semiconductor of radius R gives

$$\varepsilon_s(R) = 1 + \frac{\varepsilon_b - 1}{1 + (\alpha/R)^j} \tag{94}$$

where $j = 2$, $\alpha = 1.093$ nm for Si, and $\varepsilon_b = 11.4$ is the bulk dielectric constant. The free exciton radius α_{eh} is proportional to ε_s. The evidence so far that ε_s is reduced as the size R diminishes implies that strong reduction of ε_s from the bulk value can change the ratio $\alpha_{eh}/2R$ from greater than 1 ("strong confinement" [195]) to less than 1 ("weak confinement" [195]). The generalized Penn's model (GPM) predicts that $\alpha_{eh}/2R$ is very close to unity for all $R < 2$ nm (i.e., "weak confinement").

GPM turned to be of questionable validity because an atomistic empirical pseudopotential calculation has predicted strong confinement for $R < 2$ nm [196]. The calculated $\varepsilon_s(R)$ is shown in Figure 15. This measures the total polarization P of a system of volume Ω, responding to a constant electric field F, (i.e., $\varepsilon_s = 1 + P/F\Omega$). This total polarization P consists of contributions from the interior Si atoms, as well as from the surface Si–H bonds.

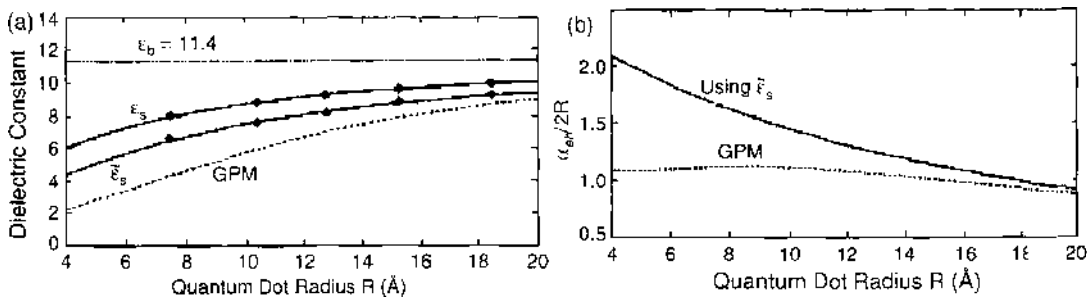


Figure 15. Dielectric constants (a) and the ratio between the free hydrogenic exciton radius and the quantum dot diameter (b) as a function of quantum dot radius R. Here, ε_s is for total polarization and $\tilde{\varepsilon}_s$ is for exciton screening. The diamond symbols in (a) denote the calculated results, while the solid lines are the fitted curves. The dashed curve corresponds to the GPM model of Equation (94). Reprinted with permission from [184], L. W. Wang and A. Zunger, *Phys. Rev. Lett.* 73, 1039 (1994). © 1994, American Physical Society.

It is found [196] that even though the surface-to-volume ration increases as the dot gets smaller, most of ε_s comes from the dot interior Si atoms, not the surface. The large reduction of ε_s with size (Fig. 15) cannot be explained in terms of surface effect and must reflect quantum confinement effects.

Another interesting quantity is the screening dielectric constant $\tilde{\varepsilon}_s(R)$, which measures the reduction in the Coulomb potential of an electron or hole resulting from screening by the medium. Because of the wave vector (q) dependence of the bulk dielectric function $\varepsilon_b(q)$ and the small dimension of the dot, it is expected that $\tilde{\varepsilon}_s(R) < \varepsilon_s(R)$. For a spherical quantum dot with radius R, the charge density of an uncorrelated electron or hole is [197]

$$\rho(r) = \left[\left(\frac{\pi}{2R^3} \right)^{1/2} \frac{\sin\left(\frac{\pi}{R} r\right)}{\left(\frac{\pi}{R} r\right)} \right]^2 \qquad r \le R \tag{95}$$

and zero elsewhere. This produces an unscreened external potential $v_{ext}(r)$ obtained from $\rho(r)$ by solving Poisson's equation with boundary condition $v_{ext}(R) = 0$. The $\tilde{\varepsilon}_s$ is defined as the ratio between the unscreened and screened Coulomb energies:

$$\tilde{\varepsilon}_s = \int v'_{ext}(r)\rho(r)d^3r \Big/ \int v'_{scr}(r)\rho(r)d^3r \tag{96}$$

where $v_{scr}(r)$ is the screened potential. With this definition, using perturbation theory, it is found that

$$\tilde{\varepsilon}_s = 1 + \frac{2}{\pi} \int_0^\infty \tilde{\varepsilon}_2(E)/E \, dE \tag{97}$$

and

$$\tilde{\varepsilon}_2(E) = \frac{2\pi}{\beta} \sum_{f,i} |\langle f | v_{ext}(r) | i \rangle|^2 \delta(E - E_{fi}) \tag{98}$$

where $\beta \equiv \int v_{ext}(r)\rho(r)d^3r$. Thus, $\tilde{\varepsilon}_2$ can be calculated in the same way as the spectrum operator $\hat{\mathbf{p}}$, replacing the momentum operator by the external potential $v_{ext}(r)$.

The calculated screening dielectric constants $\tilde{\varepsilon}_s$ for spherical dots [196] show that, as expected, $\tilde{\varepsilon}_s(R)$ is smaller than $\varepsilon_s(R)$. The ratio $\alpha_{eh}/2R$ between the bulklike (free) hydrogenic exciton radius and the system's size is predicted to be far larger than unity for $R < 2$ nm ("strong confinement"). An analytical expression for the size dependence of the dielectric constant can be obtained. Fitting the results to the analytic form of Eq. (94) gives $\alpha = 4.25$ Å, $\lambda = 1.25$ Å for the total polarization dielectric constant ε_s and $\alpha = 6.9$ Å, $l = 1.37$ Å for the screening dielectric constant $\tilde{\varepsilon}_s$. The GPM gives $\alpha = 10.93$ Å, $l = 2$ Å.

The exciton binding energy E_{BX}, with a carrier distribution close to $\sin(kr/r)$, is given by [198]

$$E_{BX}(R) = 1.79e^2/\varepsilon_{in}R - 2\delta\Sigma \tag{99}$$

where $\delta\Sigma$ is the shift in energy if an extra electron is put in the lowest conduction state. When $\varepsilon_{in} + \varepsilon_{out} \gg 1$, it holds [198]

$$\delta\Sigma = 0.47 \cdot \frac{e^2}{\varepsilon_{in} R} \left(\frac{\varepsilon_{in} - \varepsilon_{out}}{\varepsilon_{in} + \varepsilon_{out}} \right) \tag{100}$$

where the dot is considered a continuous medium of dielectric constant ε_{in} embedded in another medium of dielectric constant ε_{out}.

The results of a semiempirical LCAO calculation on the dielectric constant lie between those of Refs. [196] and [199], and their average value can be well approximated by $\tilde{\varepsilon}_{in} - 1 = (11.4 - 1)/[1 + (0.92/R)^{1.18}]$, with R in nanometers.

3. CONCLUSION

In this chapter, the different methods used in the study of Si quantum wires and dots were presented. The different models used in electronic structure calculations were discussed, including both continuum approaches (EMA) and atomistic approaches. The optical properties, which are closely related with the electronic structure of these systems, were also discussed. The results of calculations based on effective mass approximation were analyzed, and it was shown that EMA is useful for obtaining the physical evidence in an easy way and for incorporating into the calculations different effects, such as fluctuations in the wire thickness along its length, or the shape and crystallographic orientation of quantum wires and dots. Atomistic calculations were also presented, which provide a more accurate description of the system. The calculated PL lifetimes were discussed in comparison with experimental results from the literature.

REFERENCES

1. L. T. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
2. A. G. Cullis, in "Properties of Porous Silicon" (L. T. Canham, Ed.) (INSPEC, IEE, 1997). Emis Datareviews Series No. 18, p. 99 (1997).
3. A. G. Cullis, L. T. Canham, and P. D. J. Calcott, *J. Appl. Phys.* 82, 909 (1997).
4. A. G. Nassiopoulou, S. Grigoropoulos, E. Gogolides, and D. Papadimitriou, *Appl. Phys. Lett.* 66, 1114 (1995).
5. S. Grigoropoulos, A. G. Nassiopoulou, A. Travlos, D. Papadimitriou, S. Kennou, and S. Ladas, *Appl. Surf. Sci.* 102, 377 (1996).
6. D. Papadimitriou, Y. S. Raptis, and A. G. Nassiopoulou, *Phys. Rev. B* 58, 14089 (1998).
7. D. Papadimitriou and A. G. Nassiopoulou, *J. Appl. Phys.* 84, 1059 (1998).
8. A. G. Nassiopoulou, S. Grigoropoulos, and D. Papdimitriou, *Appl. Phys. Lett.* 69, 2267 (1996).
9. A. M. Morales and C. M. Lieber, *Science* 279, 208 (1998).
10. N. Wang, Y. H. Tang, Y. F. Zhang, C. S. Lee, and S. T. Lee, *Phys. Rev. B* 58, 16024 (1998).
11. P. Mutti, G. Ghislotti, L. Bonddi, G. F. Cerofolini, L. Mede, E. Grelli, and M. Guzzi, *Appl. Phys. Lett.* 66, 851 (1995).
12. H. Z. Song and X. M. Bao, *Phys. Rev. B* 55, 6988 (1997).
13. S. Guha, *J. Appl. Phys.* 84, 5210 (1998).
14. P. Normand, K. Beltsios, E. Kapetanakis, D. Tsoukalas, A. Travlos, J. Stoemenos, J. Van Den Berg, S. Zhang, C. Vieu, H. Launois, J. Gautier, F. Jourdan, and L. Palem, *Nucl. Instr. Meth. Phys. Res. B* 178, 74 (2001).
15. P. Photopoulos, A. G. Nassiopoulou, D. N. Kouvatsos, and A. Travlos, *Appl. Phys. Lett.* 76, 3588 (2000).
16. A. G. Nassiopoulou, V. Ioannou-Sougleridis, P. Photopoulos, A. Travlos, V. Tsakiri, and D. Papadimitriou, *Phys. Stat. Sol. A* 165, 79 (1998).
17. A. G. Nassiopoulou, in "Encyclopedia of Nanoscience & Nanotechnology" (H. S. Nalwa, Ed.), Vol. 9, p. 793 (2004).
18. A. Salonidou, A. G. Nassiopoulou, K. Giannakopoulos, A. Travlos, V. Ioannou-Sougleridis, and E. Tsoi, *Nanotechnology* 15, 1233 (2004).
19. W. L. Ellson, P. F. Szajowski, and L. E. Brus, *Science* 262, 1242 (1993).
20. V. Vinciguerra, G. Franzo, F. Priolo, F. Iacona, and C. Spinella, *J. Appl. Phys.* 87, 8165 (2000).
21. L. Tsybeskov, G. F. Grom, P. M. Fauchet, J. P. McCaffrey, J. M. Baribeau, G. I. Spoule, and D. J. Lockweed, *Appl. Phys. Lett.* 75, 2265 (1999).
22. B. V. Kamenev and A. G. Nassiopoulou, *J. Appl. Phys.* 90, 5735 (2001).
23. L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzo, and F. Priolo, *Nature* 408, 440 (2000).
24. L. Khriachtchen, M. Rasanen, S. Navikov, and J. Sukkorn, *Appl. Phys. Lett.* 79, 1249 (2001).
25. T. Baron, F. Mazen, C. Busseret, A. Souifi, P. Mur, F. Fournel, M. N. Séméria, H. Moriceau, B. Aspard, P. Gentile, and N. Magnea, *Microelect. Engin.* 61–62, 511 (2002).

26. D. N. Kouvatsos, V. Ioannou-Sougleridis, and A. G. Nassiopoulou, *Appl. Phys. Lett.* 82, 397 (2003).
27. V. Ioannou-Sougleridis, A. G. Nassiopoulou, and A. Travlos, *Nanotechnology* 14, 1 (2003).
28. M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allan, and C. Delerue, *Phys. Rev. Lett.* 82, 197 (1999).
29. H. Koyama, T. Ozaki, and N. Koshida, *Phys. Rev. B* 52, 11562 (1995).
30. I. Mihalcescu, J. C. Vial, A. Briesy, F. Muller, R. Romestain, E. Martin, C. Delerue, M. Lanno, and C. Allan, *Phys. Rev. B* 51, 17605 (1995).
31. S. Horiguchi, Y. Nakajima, Y. Takahashi, and M. Tabe, *Jpn. J. Appl. Phys.* 34, 5489 (1995).
32. S. Horiguchi, *Superlattices Microstruct.* 23, 355 (1998).
33. M. S. Hybertsen, *Phys. Rev. Lett.* 72, 1514 (1994).
34. X. Zianni and A. G. Nassiopoulou, unpublished data.
35. J. M. Luttinger and W. Kohn, *Phys. Rev.* 97, 869 (1955).
36. U. Blockelmann and G. Bastard, *Phys. Rev. B* 45, 1688 (1992).
37. C. Delerue, M. Lannoo, and G. Allan, *Phys. Stat. Sol. B* 227, 115 (2001).
38. S. G. Louie, *Phys. Rev. B* 22, 1933 (1980).
39. J. C. Slater and G. F. Koster, *Phys. Rev. B* 94, 1498 (1954).
40. A. Di Carlo, *Semicond. Sci. Technol.* 18, R1 (2003).
41. M. J. Mehl and D. A. Papaconstantopoulos, in "Computational Material Science" (C. Fong, Ed., World Scientific, Singapore, 1998.
42. P. Vogl, H. P. Hjalmarson, and J. D. Dow, *J. Chem. Solids* 44, 365 (1983).
43. G. Dresselhaus and M. S. Dresselhaus, *Phys. Rev.* 160, 649 (1967).
44. W. A. Harrison, in "The Physics of the Chemical Bond," Freeman, San Francisco, 1980.
45. J. A. Majewski and P. Vogl, in "Structure of Binary Compounds" (F. R. de Boer and D. G. Pettifor, Eds.), Elsevier, Amsterdam, 1989.
46. C. Tserbak, H. M. Polatoglou, and G. Theodorou, *Phys. Rev. B* 47, 7104 (1993).
47. Y. M. Niquet, C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 62, 5109 (2000).
48. L. Hedin and S. Lundqvist, in "Solid State Physics," Vol. 23, Academic Press, London, 1969.
49. J. M. Jancu, R. Scholz, F. Beltram, and F. Bassani, *Phys. Rev B* 57, 6493 (1998).
50. J. M. Jancu, F. Bassani, F. Della Sala, and R. Scholz, *Appl. Phys. Lett.* 81, 4838 (2002).
51. M. Lannoo, *J. Phys. C* 17, 3137 (1984).
52. G. D. Sanders and Yia-Chung Chang, *Phys. Rev. B* 45, 9202 (1992).
53. G. Klimeck, R. C. Bowen, T. B. Boykin, C. Salazar-Lazaro, T. A. Cwik, and A. Stoica, *Superlattices Microstruct.* 27, 77 (2000).
54. D. H. Lee and J. D. Joannopoulos, *Phys. Rev. B* 23, 4988 (1981).
55. A. Zunger and L.-W. Wang, *Appl. Surface Sci.* 102, 350 (1996).
56. A. Zunger, in "23rd International Conference Physics of Semiconductors," Vol. 2, World Scientific, Singapore, 1996.
57. K. Mader, L. W. Wang and A. Zunger, *Phys. Rev. Lett.* 74, 2555 (1995).
58. L. W. Wang and A. Zunger, *Phys. Rev. B* 51, 17398 (1995).
59. H. Fu and A. Zunger, *Phys. Rev. B* 55, 1642 (1997).
60. H. Fu and A. Zunger, *Phys. Rev. B* 56, 1496 (1997).
61. L. W. Wang and A. Zunger, *J. Phys. Chem.* 98, 2158 (1994).
62. L. W. Wang and A. Zunger, *J. Chem. Phys.* 100, 2394 (1994).
63. L. W. Wang and A. Zunger, *Phys. Rev. B* 59, 15806 (1999).
64. L. W. Wang, *Phys. Rev. B* 49, 10154 (1994).
65. T. Ohno, K. Shiraishi, and T. Ogawa, *Phys. Rev. Lett.* 69, 2400 (1992).
66. M. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993); M. S. Hybertsen, in "Light Emission from Silicon" (S. S. Iyer, L. T. Canham, and R. T. Collins, Eds.), Materials Research Society, Pittsburgh, 1992.
67. A. J. Read, R. J. Needs, K. J. Nash, L. T. Canham, P. D. J. Calcott, and A. Qteish, *Phys. Rev. Lett* 69, 1232 (1992).
68. A. J. Read, R. J. Needs, K. J. Nash, L. T. Canham, P. D. J. Calcott, and A. Qteish, *Phys. Rev. Lett* 70, 2050 (1993).
69. M. Voos, P. Uzan, C. Delalande, G. Bastard, and A. Halimaoui, *Appl. Phys. Lett.* 61, 1213 (1992).
70. F. Buda, J. Kohanoff, and M. Parrinello, *Phys. Rev. Lett.* 69, 1272 (1992).
71. J. W. Mintmire, *J. Vac. Sci. Technol. A* 11, 1733 (1993).
72. M. S. Hybertsen and M. Needels, *Phys. Rev. B* 48, 4608 (1993).
73. A. Shik, *J. Appl. Phys.* 74, 2951 (1993).
74. H. M. Polatoglou, *J. Lumin.* 57, 117 (1993).
75. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Appl. Phys. Lett.* 64, 3545 (1994).
76. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Phys. Rev. B* 50, 14405 (1994).
77. S. G. Lee, B.-H. Cheong, K.-H. Lee, and K. J. Chang, *Phys. Rev. B* 51, 1762 (1995).
78. L. Dorigoni, O. Bisi, F. Bernardini, and S. Ossicini, *Phys. Rev. B* 53, 4557 (1996).
79. A. M. Saitta, F. Buda, G. Fiumara, and P. V. Giaquinta, *Phys. Rev. B* 53, 1446 (1996).
80. S. Ossicini, C. M. Bertoni, M. Biagini, A. Lugli, G. Roma, and O. Bisi, *Thin Solid Films* 297, 154 (1997).
81. H. M. Polatoglou, *J. Lumin.* 57, 117 (1993).
82. A. B. Filonov, G. V. Potrov, V. A. Novikov, and V. E. Borisenko, *Appl. Phys. Lett.* 67, 1091 (1995).
83. K. Nishio, J. Koga, H. Ohtani, T. Yamaguchi, and F. Yonezawa, *J. Non-Crystalline Solids* 293, 705 (2001).

84. J. S. deSousa, V. N. Freire, G. A. Farias, R. R. Landim, and E. F. da Silva, *J. Korean Phys. Soc.* 39, 473 (2001).
85. A. G. Alexanian, A. S. Yeremyan, and V. M. Aroutiounian, *Phys. Stat. Sol. A* 197, 425 (2003).
86. M. G. Lisachenko, E. A. Konstantinova, P. K. Kashkarov, and V. Yu. Timoshenko, *Phys. Stat. Sol. A* 182, 297 (2000).
87. S. Y. Ren and J. D. Dow, *Phys. Rev. B* 45, 6492 (1992).
88. J. P. Proot, C. Delerue, and G. Allan, *Appl. Phys. Lett.* 61, 1948 (1992).
89. C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 48, 11024 (1993).
90. T. Takagahara and K. Takeda, *Phys. Rev. B* 46, 15578 (1992).
91. F. Huaxiang, Y. Ling, and X. Xide, *Phys. Rev. B* 48, 10978 (1993).
92. G. Fishman, I. Mihalcescu, and R. Romestain, *Phys. Rev. B* 48, 1464 (1993).
93. B. Delley and E. F. Steigmeier, *Phys. Rev. B* 47, 1397 (1993).
94. M. Hirao, in "Microcrystalline and Nanocrystalline Semiconductors" (R. W. Collins, C. C. Tsai, M. Hirose, F. Koch, and L. Brus, Eds.), Materials Research Society Symposium Proceedings, Vol. 358, Materials Research Society, Pittsburg, 1995.
95. M. Hirao, T. Uda, and Y. Murayama, *Mater. Res. Soc. Symp. Proc.* 283, 425 (1993).
96. J. L. Gavartin, C. C. Matthai, and I. Morrison, Thin Solid Films 255, 39 (1995).
97. N. A. Hill and K. B. Whaley, *Phys. Rev. Lett.* 75, 1130 (1995).
98. M. Nishida, *Phys. Rev. B* 66, 125313 (2002).
99. S. Ogut, R. Burdick, Y. Saad, and J. R. Chelikowsky, *Phys. Rev. Lett.* 90, 127401 (2003).
100. J. R. Chelikowsky, L. Kronik, and I. Vasiliev, *J. Phys. Condens. Matter* 15, R1517 (2003).
101. I. Vasiliev, S. Ogut and J. R. Chelikowsky, *Phys. Rev. B* 65, 115416 (2002).
102. C. S. Garoufalis, A. D. Zdetsis, and S. Grimme, *Phys. Rev. Lett.* 87, 276402 (2001).
103. Z. Zhou, L. Brus, and R. Friesner, *Nano Lett.* 3, 163 (2003).
104. V. A. Burdov, *Semiconductors* 36, 1154 (2002).
105. B. Delley and E. F. Steigmeier, *Appl. Phys. Lett.* 67, 2370 (1995).
106. C. Delerue, G. Allan, and M. Lannoo, in "Light Emission in Silicon: From Physics to Devices, Semiconductors and Semimetals" (D. J. Lockwood, Ed.), Vol. 49, Academic Press, New York, 1998; C. Delerue, M. Lannoo, and G. Allan, *Phys. Stat. Sol. B* 227, 115 (2001).
107. J. Van Der Rest and P. Pecheur, *J. Chem. Solids* 45, 563 (1984).
108. P. Vogl, H. P. Hjalmarson and J. D. Dow, *J. Phys. Chem. Solids* 44, 365 (1983).
109. S.-G. Lee, B.-H. Cheong, K.-H. Lo, and K. J. Chang, *Phys. Rev. B* 51, 1762 (1995).
110. J. C. Vial, A. Bsiesy, F. Gaspard, R. Herino, M. Ligeon, F. Muller, R. Romestain, and R. M. Macfarlane, *Phys. Rev. B* 45, 14171 (1992).
111. S. Ossicini, in "Properties of Porous Silicon" (L. T. Canham, Eds.), IEE INSPEC, Institution of Electrical Engineers, London, 1997.
112. C.-Y. Yeh, S. B. Zhang, and A. Zunger, *Appl. Phys. Lett.* 63, 3455 (1993).
113. R. J. needs, S. Bhattacharjee, K. J. Nash, A. Qteish, A. J. Read, and L. T. Canham, *Phys. Rev. B* 50, 14223 (1994).
114. T. van Buuren, T. Tiedje, J. R. Dahn, and B. M. Way, *Appl. Phys. Lett.* 63, 2911 (1993).
115. M. V. Rama Krishna and R. A. Friesner, *Phys. Rev. Lett.* 67, 629 (1991).
116. X. Zianni and A. G. Nassiopoulou, *Phys. Rev. B* 65, 035326 (2002).
117. X. Zianni and A. G. Nassiopoulou, *Phys. Rev. B* 66, 205323 (2002).
118. O. J. Grembocki and F. H. Pollak, *Phys. Rev. Lett.* 48, 413 (1982).
119. "Properties of Porous Silicon" (L. T. Canham, Ed.), Vol. 18, INSPEC, London, 1997.
120. Boris Kamenev, private communication; P. K. Kashkarov, B. V. Kamenev, E. A. Konstantinova, A. I. Efimova, A. V. Pavlikov, and V. Yu Timoshenko, *Usp. Fiz. Nauk* 168, 577 (1998).
121. J. Linnros, N. Lalic, A. Galeckas and V. Grivickas, *J. Appl. Phys.* 86, 6128 (1999).
122. M. Yamamoto, R. Hayashi, K. Tsunetomo, K. Kohno, and Y. Osaka, *Jpn. J. Appl. Phys.* 30, 136 (1991).
123. J.-B. Xia and K. W. Cheah, *Phys. Rev. B* 55, 15688 (1997).
124. F. Bassani and C. P. Parravacini, "Electronic States and Optical Properties in Solids." Pergamon, New York, 1975.
125. Z. Wu, T. Nakayama, S. Qiao, and M. Aono, *Appl. Phys. Lett.* 74, 3842 (1999).
126. A. N. Starukhin, A. A. Lebedev, B. S. Razbirin, and I. M. Kapitonova, *Sov. Technol. Phys. Lett.* 18, 535 (1992).
127. H. Kayama and N. Koshida, *Phys. Rev. B* 52, 2649 (1995).
128. A. J. Read, R. J. Needs, K. J. Nash, L. T. Canham, P. D. J. Calcott, and A. Qtiesh, *Phys. Rev. Lett.* 69, 1232 (1992); F. Buda, J. Kohanoff, and M. Parinello, *Phys. Rev. Lett.* 69, 1272 (1992); T. Ohno, K. Shiraishi, and T. Ogawa, *Phys. Rev. Lett.* 69, 2400 (1992).
129. P. Vogel, H. P. Hjalmarson, and J. D. Dow, *J. Phys. Chem. Solids* 44, 365 (1983).
130. G. D. Sanders and Y. C. Chang, *Phys. Rev. B* 35, 1300 (1987).
131. D. L. Dexter, in "Solid State Physics, Advances in Research and Applications" (F. Seitz and D. Turnbull, Eds.), Vol. 6, Academic Press, New York, 1958.
132. G. W't Hooft, W. A. J. A. van der Poel, L. W. Molenkamp, and C. T. Foxon, *Phys. Rev. B* 35, 8281 (1987).
133. J. C. Vial, A. Bsiesy, F. Gaspard, R. Herino, M. Ligeon, F. Muller, R. Romestain, and R. M. Macfarlane, *Phys. Rev. B* 45, 14171 (1992).
134. Y. H. Xie, W. L. Wilson, F. M. Ross, J. A. Mucha, E. A. Fitzgerald, J. M. Macaulay, and T. D. Harris, *J. Appl. Phys.* 71, 2403 (1992).

135. X. Chen, B. Henderson, and K. P. O'Donnell, *Appl. Phys. Lett.* 60, 2672 (1992).
136. P. D. J. Calcott, K. J. Nash, L. T. Canham, M. J. Kane, and M. D. Brumhead, *J. Phys. Condens. Matter* 5, L91 (1993).
137. I. Mihalcescu, J. C. Vial, and R. Romestain, *J. Appl. Phys.* 80, 2404 (1996).
138. P. D. J. Calcott, K. J. Nash, L. T. Canham and M. J. Kane, in "Microcystalline and Nanocrystalline Semiconductors" (R. W. Collins, C. C. Tsai, M. Hirosi, F. Koch, and L. Brus, Eds.). Materials Research Society, Pittsburgh, 1994.
139. E. Martin, C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 50, 18 258 (1994).
140. L. Pavesi and M. Ceschini, *Phys. Rev. B* 48, 17625 (1993).
141. J. Linnros, N. Lalic, A. Galeckas, and V. Grivickas, *J. Appl. Phys.* 86, 6128 (1999).
142. P. K. Kashkarov, B. V. Kamenev, E. A. Konstantinova, A. I. Efimova, A. V. Pavlikov, and V. Yu. Timoshenko, *Physics-Uspekhi* 41, 511 (1998).
143. S. Tanaka, H. Koyama, and N. Koshida, *Appl. Phys. Lett.* 73, 2334 (1998).
144. I. Mihalcescu, J. C. Vial, and R. Romestain, *Phys. Rev. Lett.* 80, 3392 (1998).
145. S. Ossicini and O. Bisi, *Solid State Phenom.* 54, 127 (1997).
146. S. Ossicini, *Phys. Stat. Sol. A* 170, 377 (1998).
147. S. Gardelis and B. Hamilton, *J. Appl. Phys.* 76, 5327 (1994).
148. K. J. Nash, in "Properties of Porous Silicon" (L. T. Canham, Ed.). IEE INSPEC, Institution of Electrical Engineers, London, 1997.
149. T. C. Tigelis, J. P. Xanthakis, and J. L. Vomvoridis, *Phys. Stat. Sol. A* 165, 125 (1998).
150. M. Luppi and S. Ossicini, *Phys. Rev. B*, 71, 035340 (2005).
151. E. Degoli, M. Luppi, and S. Ossicini, *Phys. Stat. Sol. A* 182, 301 (2000).
152. E. F. Steigmeier, B. Delley, and H. Auderset, *Phys. Scr. T* 45, 305 (1992).
153. C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 64, 193402 (2001).
154. X. Hu and J. Zi, *J. Phys. Condens. Matter* 14, 1671 (2002).
155. W. Cheng and S.-F. Ren, *Phys. Rev. B* 65, 205305 (2002).
156. G. Allan, C. Delerue, and Y. M. Niquet, *Phys. Rev. B* 63, 205301 (2001).
157. H.-Ch. Weissker, J. Furthmuller, and F. Bechstedt, *Phys. Rev. B* 65, 155327 (2002).
158. H.-Ch. Weissker, J. Furthmuller, and F. Bechstedt, *Phys. Rev. B* 67, 165322 (2003).
159. E. W. draeger, J. C. Grossman, A. J. Williamson, and G. Galli, *Phys. Rev. Lett.* 90, 167402 (2003).
160. I. Vasiliev and R. M. Martin, *phys. Stat. Sol. B* 233, 5 (2002).
161. A. Puzder, A. J. Williamson, F. A. Rebonero, and G. Galli, *Phys. Rev. Lett.* 91, 157405 (2003).
162. G. Allan, C. Delerue, and M. Lannoo, *Phys. Rev. B* 57, 6933 (1998).
163. W. Lang, *J. Lumin.* 57, 271 (1993).
164. Y. Kanemitsu, *Phys. Rev. B* 49, 16845 (1994).
165. G. Allan, C. Delerue, and M. Lannoo, *Phys. Rev. Lett.* 78, 3161 (1997).
166. R. J. Beierle, M. J. Caldas, E. Molinari, and S. Ossicini, *Solid State Commun.* 102, 545 (1997).
167. M. J. Caldas, R. J. Beierle, E. Molinari, and S. Ossicini, *Mater. Sci. Forum* 11, 258 (1997).
168. Z. Y. Zhang, X. L. Wu, S. J. Xiong, J. C. Shen, and X. M. Bao, *Phys. Lett. A* 313, 467 (2003).
169. M. Luppi and S. Ossicini, *Phys. Stat. Sol. A* 197, 251 (2003).
170. M. Luppi and S. Ossicini, *Mater. Sci. Eng. B* 101 34, (2003).
171. M. Luppi and S. Ossicini, *J. Appl. Phys.* 94, 2130 (2003).
172. I. Vasiliev, J. R. Chelikowsky, and R. M. Martin, *Phys. Rev. B* 65, 121302 (2002).
173. A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, *Phys. Rev. Lett.* 88, 097401 (2002).
174. A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, *Phys. Stat. Sol. B* 233, 39 (2002).
175. Y. Kayanuma, *Phys. Rev. B* 44, 13085 (1991).
176. L. E. Brus, *J. Chem. Phys.* 80, 4403 (1984).
177. T. Takagahara, *Phys. Rev. B* 47, 4569 (1993).
178. T. Ogawa and T. Takagahara, *Phys. Rev. B* 44, 8138 (1991).
179. F. A. Reboredo, A. Franceschetti, and A. Zunger, *Appl. Phys. Lett.* 75, 2972 (1999).
180. E. Martin, C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 50, 18258 (1994).
181. S. Lee, L. Joensson, J. Wilkins, G. W. Bryant, and G. Limeck, *Phys. Rev. B* 63, 195318 (2001).
182. A. Franceschetti, L. W. Wang, H. Fu, and A. Zunger, *Phys. Rev. B* 58, R13367 (1998).
183. P. E. Lippens and M. Lannoo, *Phys. Rev. B* 39, 10935 (1989); J. P. Proot, C. Delerue, and G. Allan, *Appl. Phys. Lett.* 61, 1948 (1992).
184. L. W. Wang and A. Zunger, *Phys. Rev. Lett.* 73, 1039 (1994).
185. M. Lannoo, C. Delerue, and G. Allan, *Phys. Rev. Lett.* 74, 3415 (1995); *Phys. Rev. B* 52, 11982 (1995).
186. M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allan, and C. Delerue, *Phys. Rev. Lett.* 82, 197 (1999).
187. K. J. Nash, P. D. J. Calcott, and R. J. Needs, *Phys. Rev. B* 51, 17698 (1995).
188. J. C. Vial, A. Bsiesy, F. Gaspard, R. Herino, M. Ligeon, F. Muller, R. Romestain, and R. M. Macfarlane, *Phys Rev. B* 45, 14171 (1992).
189. A. Franceschetti and S. T. Pantelides, *Phys. Rev. B* 68, 033313 (2003).
190. J. P. Walter and M. L. Cohen, *Phys. Rev. B* 2, 1821 (1970).
191. S. Baroni and R. Resta, *Phys. Rev. B* 33, 7017 (1986).
192. M. S. Hybertsen and S. G. Louie, *Comments Condens. Matter Phys.* 13, 273 (1987).
193. R. Tsu, L. Ioriatti, J. F. Harvey, H. Shen, and R. A. Lux, *Mater. Res. Soc. Symp. Proc.* 283, 437 (1993).
194. D. R. Penn, *Phys. Rev.* 128, 2093 (1962).

195. A. D. Yoffe, *Adv. Phys.* 42, 173 (1993).
196. L. W. Wang and A. Zunger, *Phys. Rev. Lett.* 73, 1039 (1994).
197. L. E. Brus, *J. Phys. Chem.* 90, 2555 (1986).
198. M. Lannoo, C. Delerue, and G. Allan, *Phys. Rev. Lett.* 74, 3415 (1995).
199. R. Tsu and D. Babic, in "Optical Properties of Low Dimensional Silicon Structures" (D. C. Bensahel, L. T. Canham, and S. Ossicini, Eds.), NATO ASI Series, Kluwer Academic, Dordrecht, 1993; R. Tsu and D. Babic, *Appl. Phys. Lett.* 64, 1806 (1994).
200. D. J. Lockwood, *Solid State Commun.* 92, 101 (1994).

CHAPTER 16

Real-Space Electronic-Property Calculations for Nanoscale Structures

T. Torsti,[1,2] V. Lindberg,[3] I. Makkonen,[1] E. Ogando,[4] E. Räsänen,[1] H. Saarikoski,[1] M. J. Puska,[1] R. M. Nieminen[1]

[1]*Laboratory of Physics, Helsinki University of Technology, Espoo, Finland*
[2]*CSC—Scientific Computing Ltd, Espoo, Finland*
[3]*Department of Physics, Växjö University, Växjö, Sweden*
[4]*Donostia International Physics Center, Donostia, Spain*

CONTENTS

1. Introduction 772
2. Equations of Density-Functional Theory 774
3. Real-Space Multigrid Methods 775
   3.1. Previous Real-Space Approaches 775
   3.2. RQMG Method 776
4. Technical Enhancements 778
   4.1. Double Grid Technique 778
   4.2. Traditional Mixing Schemes 780
   4.3. Response-Iteration Methods 781
   4.4. Higher-Order Compact Discretizations 782
   4.5. RQMG with Galerkin Conditions 783
   4.6. Alternative Eigenproblem Solvers 783
5. Status of the General-Purpose Real-Space Tool 784
6. Two-Dimensional Quantum Dots 785
   6.1. Introduction and the Model 785
   6.2. Computational Aspects 785
   6.3. Zero-Field Results 786
   6.4. Magnetic Fields and the Vortex Clusters 786
   6.5. Impurities in Quantum Dots 788
7. Nanophysics in Axial Symmetry 788

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (771–795)

7.1. Ultimate Jellium Model for a Breaking Nanowire 789
7.2. Adsorbed Na Quantum Dots on Cu(111) 790
8. Positron States at Nanoscale Defects 791
9. Summary and Outlook 792
References 793

1. INTRODUCTION

The physical and chemical properties of nanoscale objects are driven by their electronic structure. Ever since the visionary words of Dirac in the late 1920s there has been a tremendous effort to develop computationally effective, yet accurate methods for electronic-structure calculations. For ground-state properties, density-functional theory (DFT) in a multitude of numerical implementations provides the most popular route, as its computational scaling with increasing degrees of freedom is much more favorable than for wave function–based theories, such as configuration-interaction (CI) expansion of Hartree-Fock theory. For small systems, the CI, coupled-cluster, and similar "exact diagonalisation" methods provide important benchmarks, as do quantum Monte Carlo calculations with their explicit inclusion of electron–electron correlations.

Within ground-state DFT, considerable attention has been paid to the construction of increasingly accurate exchange-correlation functionals, local, semilocal, or nonlocal. Their ranges of accurate applicability obviously depend on the system at hand, that is, the strength and nature of the local correlations between electrons. For example, it appears that for semiconductors and insulators a certain amount of the long-range nonlocal character of the exchange interaction must be restored (e.g., via the exact-exchange or screened-exchange methods to obtain the correct value for the energy gap between occupied and unoccupied states.

For excited-state properties, such as those probed by various kinds of spectroscopies, density-functional theory can be generalised to its time-dependent form (TDDFT). It provides a powerful alternative to methods based on many-body perturbation theory, such as Hedin's GW (one-particle Green's function with screened Coulomb interaction) expansion and its generalizations.

The usual numerical implementation of any of the methods mentioned above is based on an expansion in terms of mathematically convenient or physically motivated basis functions. The popular choices include plane waves (Fourier techniques), Gaussians, localized atomic-like orbitals, or the combinations of these. The choice of the basis set and its necessary truncation to a finite size is based on accuracy analysis and accrued experience. Systematic control of basis set convergence is strictly possible only for plane waves, which basis set is in principle also free from Pulay corrections for interatomic forces (as the basis set accuracy is independent of atomic positions).

The plane wave basis set, however, usually requires that periodic boundary conditions are applied to the computational unit cell. While this is usually not a serious limitation for bulk systems (even with disorder or defects), it is somewhat inconvenient for low-symmetry nanoscale objects, such as point contacts, constrictions, adsorbate islands on surfaces, and so on. There artificial periodicity has often to be avoided, with the cost of the unit cells becoming exceedingly large.

For such systems, an alternative and attractive route is provided by discretising the equations directly in real space, that is, without resort to any basis set whatsoever. The advantage is that boundary conditions can be chosen much more flexibly and no periodicity is required (although it *can* be required). The accuracy is controlled simply by the discretizing distance (the grid spacing). The drawback is obviously the large amount of data points that need to be stored during the calculation, and the convergence of iterative solutions for the eigenstates (and thus the efficiency of the scheme) is an important issue. Real-space finite-difference (FD) methods have won increasing popularity in recent years, both due to the advent of powerful large-memory computers and the introduction of new algorithms which can speed up the calculations considerably. Finite-element (FE) techniques provide another alternative where boundary conditions can be efficiently implemented and where accuracy can be

controlled via the element distribution in the regions critical for accuracy and convergence. The basis functions in the FE method are piecewise polynomial functions with a support of only a few neighboring elements. Many of the results described can be transferred to the FE case as well. The choice between the FD and FE methods for DFT problems is not obvious, as both methods have their advantages.

This article provides an overview of the real-space FD-based program package called MIKA. It consists of several different modules for solving the Kohn-Sham equations of the density-functional theory in different geometries. The program package is visualized in Fig. 1. It is emphasized in this figure that all the four application codes `rspace`, `cyl2`, `doppler`, and `rs2dot` share common numerical multigrid routines implemented in the subroutine library `MGLIB`. This, however, is in practice not quite true. There is a common ancestor to the multigrid subroutines, but in practice each code comes with its own version of the subroutine library. Current versions of each of the codes are available from the MIKA Web page [1]. The full source codes are also available, and the source codes are licensed with the GNU general public license (GPL). This seems to be a general trend in the electronic-structure community, advocated by the `fsatom` initiative [2], and followed, for example, by the `abinit` [3] and `octopus` [4] projects. Therefore, it is fully acceptable, and even recommended, for other researchers to take a piece of software that he/she considers useful, make further improvements, and start distributing the derived product on his/her own Web page. We believe that such a distributed mode of development, based on voluntary work of researchers, should in the long run be more efficient than a centralized mode. Of course, some form of centralized effort is necessary to coordinate such activity.

The outline of this article is as follows. In Section 2, we give, for the sake of completeness, the Kohn-Sham equations, the numerical solution of which is the subject of the rest of this article. In Section 3, we give first some motivation for using grid-based real-space methods as opposed to plane-wave schemes and techniques based on the use of atom-centered basis functions. We use the standard and well-known multigrid solver for the Poisson equation to illustrate the basic ideas of the multigrid methodology. We then give a classification of three different types of multigrid techniques appearing in the literature for the eigenvalue problem. Without making any conclusions about the relative efficiency of the three approaches, we proceed to describe in detail our generalization of the Rayleigh-quotient minimization multigrid (RQMG) method [5]. In Section 4, we outline various technical improvements related to the RQMG method, to efficient methods for reaching self-consistency, and to the computationally inexpensive use of auxiliary finer grids to improve the numerical accuracy of real-space grid-based calculations in general. Sections 5–8 contain brief introductions to the applications of our MIKA package in general three-dimensional pseudopotential calculations for both finite and periodic systems, quantum dots in two-dimensional electron gas, axially symmetric jellium models for nanostructures, and for the calculation of positron states in solids, respectively. In addition to these, a modification of the RQMG method for one-dimensional problems has been applied by Ogando et al. to treat thin metallic films on solid surfaces [6, 7].

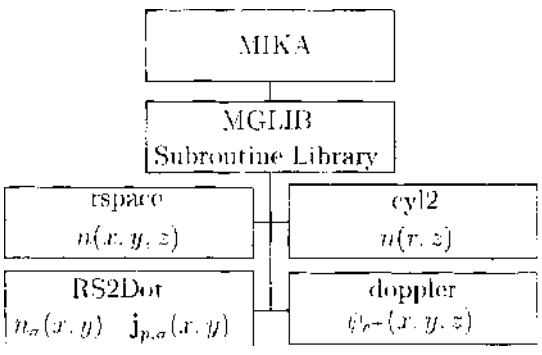


Figure 1. Schematic illustration of the software comprising the MIKA package. Each of the four applications comes with a version of the numerical multigrid subroutine library `MGLIB`. Examples of applications of each of these codes are given in the text.

2. EQUATIONS OF DENSITY-FUNCTIONAL THEORY

In the Kohn-Sham DFT approach to electronic structure calculations one solves for a set of equations self-consistently [8]. In the following, for the sake of completeness and simplicity, we present the equations in the spin-compensated form. The generalization to the spin-dependent formalism is straightforward and has been implemented in the actual computer programs. The set of equations reads as (atomic units with $\hbar = m_e = e = 1$ are used):

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right)\Psi_i = \epsilon_i \Psi_i \tag{1}$$

$$n(\mathbf{r}) = \sum_i^N |\Psi_i(\mathbf{r})|^2 \tag{2}$$

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ion}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \tag{3}$$

$$V_{\text{H}}(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{4}$$

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{5}$$

The first Eq. (1) is a Schrödinger equation for noninteracting particles in an effective potential $V_{\text{eff}}(\mathbf{r})$. For finite systems the wave functions are required to vanish at the boundaries of the computation volume. In the case of infinite periodic systems, the complex wave functions have to obey the Bloch theorem at the cell boundaries. The electron density $n(\mathbf{r})$ is obtained from a sum over the N occupied states. The effective potential consists of an external potential $V_{\text{ion}}(\mathbf{r})$ due to ions (or nuclei in all-electron calculations), the Hartree potential $V_{\text{H}}(\mathbf{r})$ calculated from the electron density distribution, and the exchange-correlation potential $V_{\text{xc}}(\mathbf{r})$. In the examples of the present article we use the local-density approximation (LDA) for the exchange-correlation energy

$$E_{\text{xc}}[n(\mathbf{r})] = \int \epsilon_{\text{xc}}(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r} \tag{6}$$

and for the exchange-correlation potential

$$V_{\text{xc}}(\mathbf{r}) = \epsilon_{\text{xc}}(n(\mathbf{r})) + n(\mathbf{r})\frac{d\epsilon_{\text{xc}}}{dn}\bigg|_{n=n(\mathbf{r})} \tag{7}$$

The Hartree potential is solved from the Poisson equation

$$\nabla^2 V_{\text{H}}(\mathbf{r}) = -4\pi n(\mathbf{r}) \tag{8}$$

In practice, the electron density $n(\mathbf{r})$ is substituted by the total charge density $\rho(\mathbf{r})$, which includes a positive charge neutralizing the system. This positive charge is composed of Gaussian charge (with charge Z_I) distributions centered at the ions I. Once Eq. (8) is numerically solved, the analytically known potential caused by the Gaussians is replaced by the local part of the pseudopotential. Obviously, this correction has finite range, since both potentials have the asumptotic behavior $Z_I/|\mathbf{r} - \mathbf{r}_I|$. In the case of finite systems, Dirichlet boundary conditions are used with the Coulomb potential values calculated using a multipole expansion. For periodic systems, we fix the average Coulomb potential to zero—this is equivalent to setting $V_C(G = 0) = 0$ in the plane-wave approach—and allow the periodic boundary conditions to result in the corresponding converged potential.

The self-consistent solution of the above Kohn-Sham equations leads to the ground state electronic structure minimizing the total energy

$$E_{\text{tot}} = \sum_i^N \Psi_i^*(\mathbf{r})\left(-\frac{1}{2}\nabla^2\right)\Psi_i(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int V_{\text{H}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}$$

$$+ \int V_{\text{ion}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{\text{xc}} + E_{\text{ion-ion}} \tag{9}$$

where $E_{\text{ion ion}}$ is the repulsive interaction between the ions (nuclei) of the system. Instead of the self-consistency iterations the solution of the Kohn-Sham problem can be found by minimizing directly the total energy with respect to the wave function parameters, e.g., plane-wave coefficients [9]. However, Kresse and Furthmüller [10, 11] have found this scheme less efficient than the self-consistency iterations.

3. REAL-SPACE MULTIGRID METHODS

The plane-wave pseudopotential method has proved to be an excellent computational tool for solving large-scale electronic structure problems in condensed matter physics [9, 12]. Notable strengths of the method are the ability to use the fast Fourier transform for updating the Kohn-Sham single-particle wave functions, lack of dependence of the basis on atom positions, and the clear control of convergence with the cutoff energy determined by the shortest-wavelength mode. However, the method encounters difficulties when treating widely varying length scales. This issue is relevant for core electron calculations and the hard pseudopotentials of first-row elements or transition metals. The periodic boundary conditions, also known as the supercell method, are often cumbersome for low-symmetry, nonperiodic systems such as isolated clusters, molecules, or point contacts between two media. Unlike with plane-wave methods, it is not necessary to use the supercell approximation when treating these types of systems with real-space methods. However, it should be noted that techniques to avoid periodic boundary conditions have been developed also for plane-wave methods [13].

Approaches where the basis functions are atom-centered or floating Gaussians or atomic orbitals are very well established and are used by the majority of the quantum-chemistry community as well as by a large number of condensed-matter theorists. A wide selection of well-established codes based on atom-centered basis functions is available, including, for example, DMOL [14], ADF [15], TURBOMOLE [16], NWCHEM [17], and SIESTA [18], among others. The basis sets used in these methods are at least an order of magnitude smaller than in the plane-wave methods, but the magnitude of the related basis-set truncation error is often difficult to estimate.

3.1. Previous Real-Space Approaches

Considerable effort has recently been focused also on developing "fully numerical" real-space methods [19], which permit systematic studies of convergence in the spirit of the plane-wave methods. These methods are based on finite elements (FE) [20–24], finite-difference (FD) discretizations [25–31], or wavelets [32]. Advantages of these approaches include the free choice of boundary conditions, allowing, for example, the treatment of finite and periodic systems with equal effort. Near-locality of the kinetic energy operator in real-space representations leads to simplicity in developing domain-decomposition parallel algorithms. In addition, it is possible to implement adaptive grid-refinement strategies to focus effort in spatial regions with large variations in the computed functions, for example near the nuclei. In finite-difference methods, the available strategies for mesh refinement include composite grids [33–35] and adaptive coordinates [31, 36]. In FE methods, on the other hand, there is in principle complete freedom in the choice of the computational mesh. However, generating an optimal FE mesh (or a finite-difference composite grid) for a given problem is a nontrivial task [37–39], which either requires *a priori* knowledge of the spatial dependence of the required density of the mesh, or involves a repeated sequence of solving the problem in a given mesh, making an *a posteriori* error estimation and then remeshing. Representations on real space grids allow also the use of multigrid (MG) algorithms with their excellent convergence characteristics and scaling properties [40, 41]. A real-space formulation is also often used in efficient implementations of $O(N)$ electronic structure methods, in which the computational work required scales linearly with the number of atoms [42, 43].

Among the pioneers of real-space methods for molecular systems were Becke [44, 45] and Pyykkö et al. [46–48], who made highly accurate fully numerical all-electron real-space calculations for diatomic molecules, employing the prolate spheroidal coordinate system. In the axial symmetry of diatomic molecules, the azimuthal dependence of the single-particle functions can be treated analytically and the ensuing numerical problem is two dimensional. Their

approach for diatomic molecules is very similar to our more general method for axially symmetric systems, described in Section 7. Besides standard density-functional theory, Pyykkö et al. applied their fully numerical approach to other quantum chemical approaches such as multiconfigurational Hartree-Fock methods and relativistic DFT [46].

Several approaches employing the multigrid acceleration idea within electronic structure calculations have appeared during the last decade [5, 27–30, 43, 49]. The main idea of multigrid methods is to avoid the critical slowing-down (CSD) phenomenon occuring when a partial differential equation discretized on a real space grid is solved with a simple relaxation method such as the Gauss-Seidel method. The discretized operators use information from a rather localized region of the grid at a time. Therefore, the high-frequency error of the length scale of the grid spacing is reduced very rapidly in the relaxation. However, once the high-frequency error has effectively been removed, the slow convergence of the low-frequency components dominates the overall error reduction rate [40], (i.e., CSD occurs). In multigrid methods one stops the relaxation on a given (fine) grid before CSD sets in and transfers the equation to a coarser grid (the so-called restriction operation) where the low-frequency components can be solved more efficiently. On the coarsest grid, the problem is solved exactly or as accurately as possible, after which one interpolates (the so-called prolongation operation) the correction to finer grids, performing simultaneously relaxations to remove the high-frequency errors introduced in the interpolation.

It is best to illustrate these ideas in the case of the simple Poisson problem $\nabla^2 V = f$. After a few relaxation sweeps on the fine grid f, the rapidly varying components of the residual $R_f = f - \nabla^2 V_f$ have been efficiently damped. A smooth correction V_c (approximately satisfying $V = V_f + I_c^f V_c$) is solved from another Poisson equation $\nabla^2 V_c = I_f^c R_f$. The transfer operators I_f^c and I_c^f are referred to as *restrictor* and *prolongator*, respectively.

References [50, 51] are classical textbooks on multigrid methods. Introductory material can be found in the recently appeared second edition of the *Multigrid Tutorial* by Briggs [52]. The full-approximation storage method [40] (FAS) is a standard recipe for nonlinear problems. Beck et al. [30, 53] have applied the FAS eigensolver of Brandt et al. [41] for electronic structure calculations of small molecules. Costiner and Ta'asan [54, 55] have made several technical improvements to overcome various obstacles related to the application of the FAS-method in electronic structure calculations. It has also been noted [55, 56] that the FAS-scheme, applicable to nonlinear systems of equations, can be directly applied to the nonlinear Kohn-Sham problem, bypassing the self-consistency iterations. However, according to the authors' knowledge, none of these methods are yet routinely applied in large-scale electronic structure calculations. When many eigenfunctions are solved simultaneously, the FAS methods may suffer from problems with representing the eigenfunctions accurately on the coarse levels, limiting the number of levels that can be used.

Briggs et al. [27, 49] apply a steepest descent method, with a special preconditioner. In the preconditioning step, the Hamiltonian is approximated by the kinetic energy term only—thus they end up solving a Poisson equation in the preconditioning step. The same preconditioner is applied also in the (almost) linear scaling method by Fattebert and Bernholc [43]. Mortensen et al. [57] apply this preconditioning technique in connection with the DIIS-method (direct inversion in the iterative subspace) in their real-space implementation of the projector augmented wave method.

There are efficient solvers for the matrix eigenproblems arising from the FD or FD discretizations of the Kohn-Sham equations involve that do not involve the multigrid concept. Chelikowsky et al. [25, 58–60] have succesfully applied iterative diagonalization schemes based on preconditioned Krylov techniques, such as the Lanczos method [61]. Their current implementation uses a preconditioned generalized Davidson algorithm [62]. Some form of a multigrid technique should probably play a role in the optimal preconditioning step of Krylov subspace techniques as well. There exists interesting recent development in the field of preconditioned eigensolvers by Knyazev and Neymeyr [63, 64].

3.2. RQMG Method

We have developed a generalization of the so-called Rayleigh quotient multigrid method first introduced by Mandel and McCormick [65]. Our generalization is presented in detail

in Ref. [5]. In this method, the coarse grid relaxation passes are performed so that the Rayleigh quotient calculated on the *fine* grid will be minimized. In this way, there is no requirement for the solution to be well represented on a coarse grid and thus the coarse grid representation problem is avoided. Mandel and McCormick [65] introduced the method for the solution of the eigenpair corresponding to the lowest eigenvalue. We have generalized it to the simultaneous solution of a desired number of lowest eigenenergy states by developing a scheme that keeps the eigenstates separated by the use of a penalty functional, Gram-Schmidt orthogonalization, and subspace rotations.

A basic ingredient of the scheme is a very simple relaxation method called *coordinate relaxation* [66]. Coordinate relaxation is a method of solving the discretized eigenproblem

$$Hu = \lambda Bu \tag{10}$$

by minimizing the Rayleigh quotient

$$\frac{\langle u|H|u\rangle}{\langle u|B|u\rangle} \tag{11}$$

Above, H and B are matrix operators chosen so that the Schrödinger equation discretized on a real-space point grid with spacing h is satisfied to a chosen order $O(h^n)$. In Eq. (11) u is a vector, containing the values of the Kohn-Sham orbitals at the grid points. In the relaxation method, the current estimate u is replaced by $u' = u + \alpha d$, where the search vector d is simply chosen to be unity in one grid point and to vanish in all other points, and α is chosen to minimize the Rayleigh quotient. This leads to a simple quadratic equation for α. For complex eigenfunctions, it is possible to either solve a remarkably complicated coupled pair of quadratic equations for the real and imaginary parts of α or to sequentially apply separate coordinate relaxation steps for the real and imaginary parts. A complete coordinate relaxation pass is then obtained by performing the minimization at each point in turn and these passes can be repeated until the lowest state is found with desired accuracy.

Naturally, also the coordinate relaxation suffers from CSD because of the use of local information only in updating u in a certain point. To avoid it one applies the multigrid idea. In the multigrid scheme by Mandel and McCormick [65], the crucial point is that *coarse* grid coordinate relaxation passes are performed so that the Rayleigh quotient calculated on the *fine* grid will be minimized. In this way, there is no requirement for the solution to be well represented on a coarse grid. In practice, a coarse grid search substitutes the fine grid solution by

$$u'_f = u_f + \alpha I_c^f e_c \tag{12}$$

where the subscripts f and c stand for the fine and coarse grids, respectively, and I_c^f a prolongation operator interpolating the coarse grid vector to the fine grid. The Rayleigh quotient to be minimized is then

$$\frac{\langle u_f + \alpha I_c^f d_c|H_f|u_f + \alpha I_c^f d_c\rangle}{\langle u_f + \alpha I_c^f d_c|B_f|u_f + \alpha I_c^f d_c\rangle} = \frac{\langle u_f|H_f u_f\rangle + 2\alpha\langle I_f^c H_f u_f|d_c\rangle + \alpha^2\langle d_c|H_c d_c\rangle}{\langle u_f|B_f u_f\rangle + 2\alpha\langle I_f^c B_f u_f|d_c\rangle + \alpha^2\langle d_c|B_c d_c\rangle} \tag{13}$$

The second form is obtained by relating the coarse grid operators, H_c and B_c, with the fine grid ones, H_f and B_f, by the Galerkin condition

$$H_c = I_f^c H_f I_c^f; \quad B_c = I_f^c B_f I_c^f; \quad I_f^c = (I_c^f)^T \tag{14}$$

Note, however, that the Galerkin condition was not satisfied in our original implementation instead, we discretized the original equation separately on each grid to obtain H_c and B_c (discretization coarse grid approximation DCA).

The key point to note is that when $H_f u_f$ and $B_f u_f$ are provided from the fine grid to the coarse grid; the remaining integrals can be calculated on the coarse grid itself. Thus one really applies coordinate relaxation on the coarse grids to minimize the *fine-level* Rayleigh quotient. This is a major departure from the earlier methods, which, to some extent, rely

on the ability to represent the solution of some coarse grid equation on the coarse grid itself. Here, however, one can calculate the *exact* change in the Rayleigh quotient due to *any* coarse grid change, no matter how coarse the grid itself is. There is no equation whose solution would have to be representable.

Next we consider the generalization of the RQMG method to the simultaneous solution of several (N) mutually orthogonal eigenpairs. The separation of the different states is divided into two or three subtasks. First, to make the coarse grid relaxations converge toward the desired state, we apply a penalty functional scheme. Given the current approximations for the k lowest eigenfunctions, the next lowest, $(k+1)$th state is updated by minimizing the functional

$$\frac{\langle u_{k+1}|H|u_{k+1}\rangle}{\langle u_{k+1}|B|u_{k+1}\rangle} + \sum_{i=1}^{k} q_i \frac{|\langle u_i|u_{k+1}\rangle|^2}{\langle u_i|u_i\rangle \cdot \langle u_{k+1}|u_{k+1}\rangle} \tag{15}$$

The modulus of the overlap integral in the penalty term is squared to make the penalty positive definite. The denominator is required to make the functional independent of the norms of u_i, $i = 1 \ldots k+1$. The minimization of this functional is equivalent to imposing the orthonormality constraints against the lower k states, when $q_i \to \infty$. By increasing the shifts q_i any desired accuracy can be obtained, but in order to obtain a computationally efficient algorithm a reasonable finite value should be used, for example,

$$q_i = (\lambda_{k+1} - \lambda_i) + Q \tag{16}$$

where Q is a sufficiently large positive constant. In our calculations, we have typically used the value $Q = 2$ Ha.

The substitution (12) is introduced in the functional (15) and the minimization with respect to α leads again to a quadratic equation. This time the coefficients contain terms due to the penalty part.

While the penalty functional keeps the states separated on the coarse levels, we apply a simple relaxation method (Gauss-Seidel) on the finest level. The Gauss-Seidel method converges to the nearest eigenvalue, so ideally no additional orthogonalizations would be needed. In practice, however, we use Gram-Schmidt orthogonalizations and subspace rotations. However, the number of fine-grid orthogonalizations remains small, for example, in comparison with the conjugate gradient search of eigenpairs employing only the finest grid [26].

4. TECHNICAL ENHANCEMENTS

4.1. Double Grid Technique

Representing functions with high-frequency components on coarse grids has problems, sometimes referred to as *aliasing* (the high-frequency components, not representable on the coarse grids, are aliased to lower-frequency components), if one simply takes the pointwise values of the continuous functions at each grid point. The most direct way to see this in electronic structure calculations is to monitor the calculated total energy of an atom or a molecule as the system is moved with respect to the grid. Figure 6.16 on page 254 of Ref. [8] illustrates this *egg-box effect* in the case of CH_4 molecule when calculated with the real-space code `octopus`. The total energy varies with an amplitude of 70 meV when the central atom is moved from one grid point to another (the molecule rigidly following the movement). The first cure suggested to this problem was to simply to Fourier filter away the high-frequency components of such a function [27, 49]. This, however, may not be the best solution to the problem.

A much more elegant solution was suggested by Ono and Hirose [67]. This scheme applies most directly to the evaluation of the inner products of the wave-function with the nonlocal projectors occuring in norm-conserving nonlocal pseudopotentials or the projector augmented wave (PAW)-method.

A careful and well-documented implementation of the Ono-Hirose scheme was recently made by Mortensen et al. in the context of the PAW method [57]. A key factor in this

implementation was to use the scheme not only for the projector functions, but as a general recipe to transfer any function defined on a radial atom-centered mesh to the coarse grid where the wave functions are computed.

In a grid presentation, integrals over space are turned into sums over grid points. In our implementation of the nonlocal norm-conserving pseudopotentials, we often need to calculate the integral of a localized function, centered on an atom, multiplied by the wave function extended over all space. The most obvious example is the projector function $\xi^a_{lm}(\mathbf{r} - \mathbf{R}^a)$ centered on atom a at position $\mathbf{R}^a$ multiplied by an extended wave function $\psi_{nc}(\mathbf{r})$ defined by its values at the grid points of the computational grid (as indicated by the subscript c).

Using G to index the grid points for the wave functions and transforming the integral to a sum over grid points with V_c being the volume per grid point, we obtain

$$P^a_{lm} = V_c \sum_G \xi^a_{lm,G} \psi_{nG} = \langle \xi^a_{lm,c} | \psi_{nc} \rangle \tag{17}$$

where $\psi_{nG} = \psi_n(\mathbf{r}_G)$, and $\mathbf{r}_G$ is the position of grid point G (only grid points in the localized region around atom a need to be summed over). For $\xi^a_{lm,G}$ we could use $\xi^a_{lm}(\mathbf{r}_G - \mathbf{R}_a)$. However, this is not accurate enough, unless we use a very fine grid that would compromise efficiency. Instead we use the elegant double grid technique of Ono and Hirose [67]. Here we interpolate the wave function to a finer grid f and evaluate the inner product there, using pointwise values of the projector function $\xi^a_{lm,f}$ on the finer grid. It is most convenient in the following discussion to use the prolongator (interpolation) operator I^f_c and its transpose, the restriction operator I^c_f.

$$P^a_{lm} = \langle \xi^a_{lm,f} | I^f_c \psi_{nc} \rangle = \langle I^c_f \xi^a_{lm,f} | \psi_{nc} \rangle \tag{18}$$

Here the relation $I^f_c = (I^c_f)^T$ (Eq. [14]) is used. We see, that the interpolation of the wave function needs not be done in practice. Comparing Eqs. (17) and (18) it is easy to see that

$$\xi^a_{lm,c} = I^c_f \xi^a_{lm,f} \tag{19}$$

that is, the projector is first evaluated on the auxiliary fine grid (this needs to be done only once in the beginning of the calculation) and then brought to the coarser grid used to represent the wave functions by the restriction operator. The fine grid can, for example, be a uniform grid with five times the density of the computation grid, as suggested in Ref. [67]. However, a more economic choice would be to select special integration points corresponding to a Gaussian quadrature for this finer grid and properly account for the integration weights.

The projector functions are not the only localized atom-centered functions in our implementation of the pseudopotential method. In addition, we have atom-centered Gaussian positive charge distributions, used to neutralize the charge density occuring in (Eq. [8]), and associated local potential corrections $\Delta V_{\text{loc}} = V_{\text{loc}}(r) - Z_{\text{eff}}(r/r_c)/r$. In practice, we have used the same scheme of evaluating these functions on a fine grid and restricting them to the coarse computational grid, and found two orders of magnitude reduction in the egg-box effect, as reported also by Mortensen et al. in the case of the PAW scheme [57].

Note, however, that strictly speaking a nonlocal operator on the coarse grid should replace the local potential of the fine grid:

$$\langle I^f_c \psi_{nc} | (\Delta V_f) I^f_c \psi_{nc} \rangle = \langle \psi_{nc} | I^c_f (\Delta V_f) I^f_c \psi_{nc} \rangle \tag{20}$$

The sparse matrix $I^c_f(\Delta V_f) I^f_c$ can be constructed in the same way as the Galerkin form H_c in Section 4.5. A similar accuracy—and similar nonlocal operator replacing the local potential—is obtained if the wave functions are expanded in the nonorthogonal basis of piecewise polynomial functions associated with the grid points (the finite-element method), and the matrix-elements of the potential in this basis are properly evaluated.

In practice, the simple restriction may be justified, besides, by the fact that it seems to work well in practice, also by the fact that we evaluate the density directly from the pointwise

values of wave functions on the coarse grid, and the related correction to the total energy satisfies (note that we use the notation $\langle f|g\rangle$ as a shorthand for $\int dr f^* g$):

$$\langle I_c^f n_c | \Delta V_f \rangle = \langle n_c | I_f^c (\Delta V_f) \rangle \tag{21}$$

These relations seem to suggest that in a numerically correct implementation a local potential (obtained by simple restriction from the auxiliary fine grid in the double grid scheme) should operate on the density, for example, in expressions for the total energy, while the nonlocal form $I_f^c V I_c^f$ should be used in the eigenvalue problem when operating on the wave functions. It could appear that treating the local potential numerically accurately and also computing the density, the exchange-correlation potential and the electrostatic Hartree potential on a grid finer than that used for the wave functions would allow the use of even coarser grids for the wave functions. This would result in substantial memory and CPU savings. After all, plane-wave and grid presentations of the wave functions should in principle be equivalent. Yet the conclusion made in, for example, in Ref. [57] is that an order of magnitude more memory is needed to represent the wave functions on a real-space grid than by storing the plane-wave coefficients.

A possible improvement to the existing methodology would indeed be to evaluate the density on a finer grid (with grid spacing halved):

$$n_f = \sum_i |I_c^f \psi_{i,c}|^2 \tag{22}$$

The Hartree and exchange-correlation potentials would then be evaluated on this finer grid and returned to the coarse, wave function grid through either the above described method yielding a short-range nonlocal potential operator or by the simple restriction, which may be accurate enough in practice (or, as suggested, both of these operators would be needed for different purposes). Note that the use of Eq. (22) requires also a modification for the normalization condition for the wavefunctions, $\langle \psi, I_f^c I_c^f \psi \rangle = 1$. This would be analogous to the plane-wave scheme, where the cutoff wave-vector for the density is twice the cutoff wave vector for the wave functions. Such a scheme is not yet implemented in `rspace` (or at least not properly tested), and not in any other real-space method either according to the authors knowledge. Note, however, that in the PAW-scheme by Mortensen et al., such a finer mesh is used for the solution of the Poisson equation. However, the relation (22) is not used; instead the density is first evaluated on the coarse grid and then interpolated to the fine grid.

Finally, it would be interesting to see, if the idea can be applied also to the kinetic energy stencil T_c (or $A_{\text{mehr,c}}$ and $B_{\text{mehr,c}}$ of section 4.4), by replacing it with $I_f^c T_f I_c^f$. Again this could be done using the same algorithm as explained for the B-stencil in Section 4.5.

We encourage the interested reader to obtain the source code of `rspace` from the Web page [1] (or use his/her own favorite implementation of real-space grid methods) and implement these ideas. There is no guarantee that we will have time to do so in the near future.

4.2. Traditional Mixing Schemes

Reaching self-consistency in the solution of the Kohn-Sham equations (1)–(5) can be a tricky task for systems where the Hartree and exchange-correlation potentials dominate over the external potential (e.g., quantum dots described in Section 6), and/or when the system size is large (e.g., thousands of electrons as in the axially symmetric jellium model calculations described in Section 7). Defining the output density n_{out} as the density obtained from Eq. (2), when the orbitals are solved from Eq. (1) in the potential generated through Eqs. (3)–(5) by the input density n_{in}, the simplest stable iteration scheme results by choosing for the input density of the next iteration the linear combination

$$n_{\text{in}}^{\text{next}} = (1 - \kappa) n_{\text{in}} + \kappa n_{\text{out}} \tag{23}$$

Instead of the density n, the potential V_{en} can equally well be mixed. In large systems, or quantum dots in low confinement, the mixing parameter κ may have to be chosen smaller than 1%. Clearly more sophisticated mixing schemes should be used.

We have recently implemented the standard Pulay mixing scheme [68, 69] and its several variants. An interesting new variant is the guaranteed reduction Pulay (GRPulay) method [70], where no mixing parameter needs to be given. The key idea of the GRPulay method is that at the first iteration one computes also the residual of the output density. Then, in the two-dimensional space spanned by the input and output densities, the residual is minimized (assuming a linear dependence between input density and the residual), and the input and output densities for the next iteration are predicted, and the new residual of the output density computed.

A preconditioning to the Pulay scheme that damps the long wavelength changes of the density was proposed by Kerker [71]. The Kerker scheme was originally motivated by a real-space scheme [72–74], where, instead of the Poisson equation, a modified Helmholz equation is solved. We have implemented this idea as well and found it very useful in some calculations. Also the use of a special metric that weights long wavelength errors stronger than short wavelength errors was proposed by Kresse and Furthmüller [10, 11]. In these schemes, a Fourier decomposition of the density to components of different wavelengths is required. We have used the following multigrid-based method to make this decomposition. Here N denotes the coarsest grid used, so that n_N is the density component of the longest wavelength. $I_1^N = I_1^2 \ldots I_{N-1}^N$ is the restriction operator from the finest grid 1 to the coarsest grid N:

$$n_N = I_N^1 I_1^N n \tag{24}$$

$$n_{k+1} = I_{k+1}^1 I_1^k \left(n - \sum_{i=k}^{N} n_i \right) \tag{25}$$

It can be easily seen that this decomposition satisfies $\sum_{i=1}^{N} n_i = n$, and obviously each component n_i contains features with a wavelength characteristic for level i. Furthermore, it can be checked that the norm of n_N is equal to the number of electrons in the charge distribution n, while the norm of other components is zero.

4.3. Response-Iteration Methods

While the mixing schemes referred to above use mainly mathematical tricks in accelerating the convergence towards self-consistency, Auer and Krotscheck [75–77] have suggested a "physical" method where the static dielectric function of the nonuniform electron gas is utilized to obtain a rapidly converging algorithm. They define the functional $F[n_{\text{in}}](\mathbf{r}) = n_{\text{out}}[n_{\text{in}}](\mathbf{r}) - n_{\text{in}}$ and obtain the following linear integral equation for the density correction $\delta n(\mathbf{r})$ to be added to n_{in}:

$$F[n](\mathbf{r}) = \int d^3r' \epsilon(\mathbf{r}, \mathbf{r}'; 0) \delta n(\mathbf{r}') \tag{26}$$

where

$$\epsilon(\mathbf{r}, \mathbf{r}'; 0) = \delta(\mathbf{r} - \mathbf{r}') - \int d^3\mathbf{r}'' \chi_0(\mathbf{r}, \mathbf{r}''; 0) V_{p-h}(\mathbf{r}'', \mathbf{r}') \tag{27}$$

is the static dielectric function of non-uniform electron gas, $\chi_0(\mathbf{r}, \mathbf{r}'; 0)$ is the zero-frequency Lindhard function of the noninteracting system (and can be expressed in terms of the occupied and unoccupied orbitals), and the particle-hole interaction (or Hartree-exchange-correlation kernel) is defined as

$$V_{p-h}(\mathbf{r}, \mathbf{r}') = \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta V_{\text{xc}}(\mathbf{r})}{\delta n(\mathbf{r}')} = K_{\text{hxc}}(\mathbf{r}, \mathbf{r}') \tag{28}$$

In a state-space formulation [76], they obtain the following form for this equation:

$$u(\mathbf{r}) = F[n](\mathbf{r}) - \delta n(\mathbf{r}) = 2 \sum_{p,h} \frac{\phi_p(\mathbf{r}) \phi_h(\mathbf{r})}{\varepsilon_p - \varepsilon_h} \int d^3\mathbf{r}' d^3\mathbf{r}'' \phi_p(\mathbf{r}') \phi_h(\mathbf{r}') V_{p-h}(\mathbf{r}', \mathbf{r}'') \delta n(\mathbf{r}'')$$

$$\equiv \sum_{h,p} u_{p,h} \phi_p(\mathbf{r}) \phi_h(\mathbf{r}) \tag{29}$$

where the subscript p refers to "particle-states" (unoccupied states) and h refers to "hole-states" (occupied states). In practice, the number of unoccupied states needed for a calculation to converge is often no more than the number of occupied states. Nevertheless, a scheme were only occupied states are required is desirable, and was indeed derived [76] by making the *collective approximation* that the coefficients $u_{p,h}$ are matrix elements of a *local* function $\omega(\mathbf{r})$, that is,

$$u_{p,h} = \int d^3\mathbf{r}\phi_p(\mathbf{r})\phi_h(\mathbf{r})\omega(\mathbf{r}) \tag{30}$$

After some manipulation one arrives at the following equation for $\tilde{\omega}(\mathbf{r}) = \sqrt{n(\mathbf{r})}\omega(\mathbf{r})$:

$$[H_0 + 2S_F * \tilde{V}_{p-h} * S_F] * \tilde{\omega} = 2S_F * \tilde{V}_{p-h} * \frac{\Delta n}{\sqrt{n}} \tag{31}$$

where

$$H_0(\mathbf{r}) = \frac{1}{2}\left[-\nabla^2 + \frac{\nabla^2\sqrt{n(\mathbf{r})}}{\sqrt{n(\mathbf{r})}}\right] \tag{32}$$

$$\tilde{V}_{p,h} = \sqrt{n(\mathbf{r})}V_{p,h}\sqrt{n(\mathbf{r})} \tag{33}$$

and

$$S_F(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') - \frac{1}{\nu}\frac{1}{\sqrt{n(\mathbf{r})n(\mathbf{r}')}}\sum_{h,h'}\phi_h(\mathbf{r})\phi_h(\mathbf{r}')\phi'_h(\mathbf{r})\phi'_h(\mathbf{r}') \tag{34}$$

where $\nu = 2$ for a spin-restricted calculation, and $\nu = 1$ otherwise. Finally, one obtains the density correction from

$$u(\mathbf{r}) = \sqrt{n(\mathbf{r})}[S_F * \tilde{\omega}](\mathbf{r}) \tag{35}$$

Above $[A * B](\mathbf{r},\mathbf{r}')$ denotes the ordinary matrix product defined as $\int dr'' A(r,r'')B(r'',r')$, referred to as a *convolution* in Ref. [76]. In collaboration with Aichinger, we have implemented the response iteration schemes utilizing both the state-space formulation of the full-response method (Eq. [29]) and the collective approximation (Eq. [31]) in connection with the `rs2dot` and `cyl2`-programs. To solve the integral Eqs. (29) and (31), we have used either the conjugate gradient or the generalized mimimum residual (GMRES)-method [78]. In the case of very large systems, a more efficient solver, maybe a multigrid scheme, for Eq. (31) could be useful.

4.4. Higher-Order Compact Discretizations

The Mehrstellen-discretized Schrödinger equation

$$H_{\text{Mehr}}\psi_i = A_{\text{Mehr}}\psi_i + B_{\text{Mehr}}(V\psi_i) = \varepsilon_i\psi_i \tag{36}$$

was first introduced to electronic structure calculation by Briggs et al. [27, 49]. The matrices of this fourth-order discretization only have elements corresponding to the nearest and second nearest-neighbor grid points in three-dimensional space. This is in contrast to the traditional central finite difference (CDS) fourth-order discretization, which is more nonlocal, and involves 13 points in a starlike-type constellation consisting of three orthogonal line segments of five grid points each. Going to higher order, the CDS stencils become more and more nonlocal [25]. It is argued in Ref. [49] that the fourth-order Mehrstellen discretization has the accuracy comparable to the sixth-order CDS stencil. In addition to the Mehrstellen discretization, we have derived and implemented a set of higher-order compact, Mehrstellen-type stencils for the Schrödinger and Poisson equations [79].

One immediately notes that although $B^{-1}_{\text{Mehr}}H_{\text{Mehr}}$ is hermitian [27], H_{Mehr} itself is not. In the original implementation of the RQMG method [5], the hermiticity of H and B is assumed, and this can degrade the performance of the RQMG method when used with Mehrstellen-type stencils. A generalization of the RQMG method to nonhermitian discretizations is under construction.

4.5. RQMG with Galerkin Conditions

As explained in Section 3.2, our original implementation of the RQMG method does not respect the Galerkin conditions of Eq. (14) but replaces the H and B matrices on the coarse levels by a rediscretization of the original problem (DCA). This can in some cases result in a limitation of the coarsest level that can be used during the multigrid V-cycle, and hinder the convergence of the higher, unoccupied levels, that are needed in, for example, the linear-response calculations based on time-dependent DFT (TDDFT) or the full-response formulation of the response-iteration scheme (see Section 4.3). In practice, convergence for the required number of states can be obtained by selecting a grid dense enough and tuning the coarsest grid size, but in large three-dimensional calculations, it is not desirable to have to use dense grids simply because of an improperly implemented numerical scheme.

We have recently implemented also the full Galerkin version of the RQMG method. Note that even for CDS-stencils, for which $B_f = I$ on the fine level, $B_c \neq I$ on the coarse level. We note that multiplying a coordinate vector e_G (which has the value 1 at a single grid point G and 0 in all other points) by a matrix A produces the column vector of A corresponding to the point G. Then, the column of H_c^I (the required row of H_c) corresponding to G is given by

$$H_c^I e_{c,I}^c = I_f^c H_f^T I_c^f e_G^c \tag{37}$$

Here, e_G is first multiplied by I_c^f, and the (well-known) column vector of the prolongator is obtained. The result is multiplied by the transpose of H_f (given in stencil representation on the finest level, and compressed row storage (CRS)-format [80] on coarse levels). Finally, multiplication with the restriction operator I_f^c gives the row of the Galerkin H_c. This vector will be nonzero only in the immediate vicinity of G, and thus we have obtained a row of the sparse matrix H_c, to be stored in CRS format. This scheme is much faster than the more obvious alternative of representing each of the three matrices appearing above in Eq. (37) in CRS format and computing the two-matrix products by standard methods. The matrix B_c is obtained in exactly the same way, but since B_f is independent of $\mathbf{r}$, so will B_c be also, and this can be stored in the simple stencil format, and is very fast to compute. The current implementation of this scheme is a bit slower than using the stencil representation, but results in guaranteed convergence even on very coarse grids. More speed may be obtained by using the Galerkin matrices only on the very coarse grids, while keeping the CDA on, for example, the two finest levels. Also, more convenient formats than CRS could result in additional speed—in fact on each row of the matrix the nonzero elements have the same pattern, so that simple array could be a more convenient storage format. Note that the kinetic energy operator still allows a simple stencil representation, as does B; it is the potential energy which requires the nonlocal nonstencil form in the Galerkin formulation.

4.6. Alternative Eigenproblem Solvers

It would be desirable to implement a few alternative eigenvalue solvers, in addition to RQMG, in connection with MIKA. For example, the approach chosen by Mortensen et al., to follow as accurately as possible the plane-wave scheme of Kresse et al. [10, 11], based on the DIIS-method, only replacing the preconditioning operator by a multigrid V-cycle, seems promising. On the other hand, Krylov subspace methods such as Lanczos or block Davidson are well-known efficient schemes, and careful comparisons between them and the RQMG-method would be interesting. A parallel implementation of the generalized Davidson algorithm is used by Chelikowsky et al. [62] to treat three-dimensional systems of up to a thousand electrons. Recently, a new preconditioned, Krylov-space technique has been introduced by Knyazev [63, 64]—this method is claimed to be more efficient than its precursors. We have made some comparisons, according to which it seems not to be competitive with RQMG, but more work needs to be done before making a definite conclusion. Auer, Krotcheck, and Aichinger [81, 77] have developed a surprisingly efficient scheme based on time evolution in imaginary time with fourth order factorization of the operator $\exp(-\epsilon H)$.

5. STATUS OF THE GENERAL-PURPOSE REAL-SPACE TOOL

The development of `rspace` has been driven largely by thesis projects and student assignments for technical improvements as described in Section 4. It offers a versatile alternative for large-scale static DFT calculations especially for geometries awkward for periodic boundary conditions. Among the most promising general directions where real-space grid methods are actively used, is in the applications of the real-time [82] as well as in the linear-response [83] formulations of time-dependent density-functional theory. This direction, including also the quantitative calculation of interatomic forces in excited systems, carries also the highest priority in the future developments in the MIKA project.

The `rspace` code allows both periodic and cluster boundary conditions (Fig. 2). In the cluster case, the boundary values for the Coulomb potential in Eq. (8) are computed by a multipole expansion including terms up to the quadrupole term. In the periodic case, the average potential is set to zero. Boundary conditions for general k-points have been implemented. A simple generalization to a combination of cluster boundary conditions in one or two directions and periodic in the other directions would allow computations for surfaces or polymers, respectively, avoiding the periodic images problematic in plane-wave calculations. Even a special boundary condition for long polymers where a unit cell is invariant with respect to a combination of a translation and a rotation about the axis of translation can be implemented in the real-space grid context.

The code has been parallelized through domain decomposition in real-space, and also over the k-points. Forces and structural optimization have been implemented; in fact we have two implementations of structural optimization, one written in `Fortran90` and the other in `Python`. The `Fortran90` implementation has been tested by relaxing the structures of various defects in silicon [84], the results being in agreement with plane-wave calculations.

Recently, the Perdew-Burke-Ernzerhof (PBE) [85] generalized gradient approximation (GGA) has been included, adopted more or less directly from the other open-source real-space package `octopus` [4]. As a sequel to this project, we have observed that the numerical accuracy of this implementation can be improved by following the advice given by Mortensen et al. [57], and computing the potential $V_{xc}(\mathbf{r}_G)$ exactly as the numerical derivative of the discretized E_{xc} (where the gradient of the density is evaluated via finite differences) with respect to the density at the grid point at $\mathbf{r}_G$, a trick similar to that of White and Bird [86] used in the plane-wave context.

Next we move to describe a few application areas where the real-space codes from the MIKA project have been applied. The chosen examples highlight the usage of real-space methods in nanoscale systems in particular.

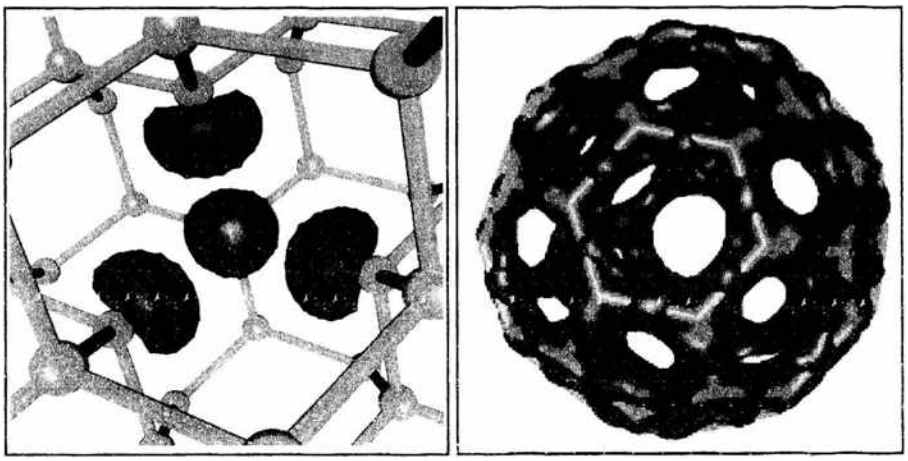


Figure 2. Typical applications of electronic structure calculations with periodic and cluster boundary conditions. (left) Electron-density isosurface corresponding to the deep states localized at the neutral, ideal (no ion relaxation) silicon vacancy in Si. (right) Isosurface of the Kohn-Sham orbital corresponding to the lowest eigenvalue in the C_{60}-molecule.

6. TWO-DIMENSIONAL QUANTUM DOTS

6.1. Introduction and the Model

In the rapidly expanding field of nanotechnology, semiconductor quantum dots (QD) represent basic elements of novel nanoelectronic components. They have dimensions from nanometers to a few microns and contain a controlled number of electrons, typically from one to several thousands. Semiconductor QDs are fabricated with several different methods [87]. The common objective between the techniques is to produce a lateral confinement of the two-dimensional electron gas (2DEG) at the interface of a semiconductor heterostructure, for example, GaAs/AlGaAs, so that the transverse dimensions are considerably larger than the thickness of the dot. Hence, the corresponding model system is usually two dimensional, and the shape of the lateral confining potential may be varied at will. The most common approximation is a parabolic confinement that has been shown to model the conventionally fabricated QDs with a reasonable accuracy [88]. In this section, however, we review some recent results for various geometries with direct relevance to experimental studies.

We define the quantum dot to be located on the xy plane and use the effective-mass approximation with the material parameters for GaAs, that is, the effective mass $m^* = 0.067\ m_e$ and the dielectric constant $\epsilon = 12.4$. The many-body Hamiltonian for this system in the presence of an external magnetic field can be written in SI units as

$$H = \frac{1}{2m^*}\sum_{i=1}^{N}[-i\hbar\nabla_i + e\mathbf{A}(\mathbf{r}_i)]^2 + \sum_{i<j}^{N}\frac{e^2}{4\pi\epsilon_0\epsilon|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N}\left[V_{\text{ext}}(\mathbf{r}_i) + g^*\mu_B B s_{z,i}\right] \qquad (38)$$

where the vector potential is chosen in the symmetric gauge, $\mathbf{A} = \frac{B}{2}(-y, x, 0)$. This determines the magnetic field perpendicular to the dot plane, that is, $\mathbf{B} = \nabla \times \mathbf{A} = B\hat{z}$. The last term is the Zeeman energy, where g^* is the effective gyromagnetic ratio for GaAs (typically, -0.44), μ_B is the Bohr magneton, and $s_z = \pm\frac{1}{2}$ for the electron spin $\sigma = \uparrow, \downarrow$, respectively. The spin-orbit interaction is excluded in the Hamiltonian, since it is supposed to be relatively small in a wide gap material like GaAs.

6.2. Computational Aspects

In the calculations, we apply mostly the spin-density-functional theory (SDFT) in the conventional self-consistent Kohn-Sham formulation. In high magnetic fields, we have also employed the computationally more demanding current spin-density functional theory (CSDFT), which does not, however, represent a major improvement over the SDFT. A detailed comparison between these two methods for a six-electron quantum dot can be found in Ref. [89].

We have also tested different parametrizations for the exchange-correlation functionals in the local-spin-density approximation (LSDA) [89]. Quantum Monte Carlo energies for a six-electron QD in zero and finite magnetic fields were taken as benchmark results. According to our calculations, the functional by Attaccalite and co-workers [90] generally gives more accurate results than the form by Tanatar and Ceperley [91]. However, the CSDFT suffers from the lack of accurate interpolation forms between the zero and high magnetic field limits for a given spin polarization.

In the real-space QD program `rs2dot`, the RQMG-method is used for solving the effective single-electron Schrödinger equation on a two-dimensional point grid. In practical calculations, the number of grid points is set between 80 and 128 in one direction. This gives $\sim$1 nm for a typical grid spacing, which is sufficient for describing electrons in GaAs, when the discretizations are fourth order. The accuracy of the calculations has been checked with the Richardson extrapolation, leading to a typical error of less than $\sim$1% in the total energy ($\lesssim$3% in the low-density limit). A converged solution typically takes 100...500 self-consistency iterations, but that number can be remarkably reduced by using the density-response functions [75–77] which are currently implemented into the computational code.

6.3. Zero-Field Results

As a symmetry-unrestricted method the real-space approach is suitable for treating QDs defined by a noncircular confining potential. In zero magnetic fields, we have studied Wigner crystallization in polygonal systems [92], namely, how the electrons localize into a regular lattice at sufficiently low densities as the dot size is increased. The phenomenon is due to the different scalings of the potential and kinetic parts of the total energy. The former part becomes gradually dominant over the latter as the density decreases, and finally the kinetic energy remains in the zero-point motion of the vibrational modes.

For two-electron polygonal QDs, we find the Wigner-molecule formation around the density parameter value $r_s = \sqrt{A/(N\pi)} \approx 3$, where A is the area of the polygonal potential well. This agrees well with the exact diagonalization (ED) results by Creffield et al. [93]. The qualitative behavior in the electron density is similar in both DFT and ED, leading to the localization to the corners of the QD as A increases (see Fig. 3). For $N > 2$, we find the formation of extra density peaks along the sides of the QD. In the case of a double number of electrons with respect to the number of corners in the dot, the enlargement of the dot area leads to N density peaks at $r_s \simeq 4.0$ in all polygonal QD's studied. This value is defined as the critical density parameter for the Wigner crystallization in those systems.

In Ref. [94], we have presented a detailed study on the electronic structure of rectangular QDs with a hard-wall confinement potential similar to the above presented polygonal system, that is,

$$V_{\text{ext}}(x, y) = \begin{cases} 0, & 0 \le x \le \beta L,\ 0 \le y \le L \\ \infty, & \text{elsewhere} \end{cases} \tag{39}$$

The deformation parameter β thus determines the ratio between the side lengths of the rectangle, and the area of the rectangle is set to $\beta L^2 = \pi^2 a_B^{*2} \approx 1{,}000\,\text{nm}^2$.

The chemical potentials and the addition energies are calculated as a function of β with both the SDFT and the variational quantum Monte Carlo method. The agreement between the two methods is very precise. In addition, the comparison of our results with the experiments and previous simulations [95] shows that the hard-wall approximation is slightly more realistic than the elliptic one. However, more experimental data over a wider range of β would be needed.

In agreement with Hund's rule, we find several partially spin-polarized states with $S = 1$ as β and N are varied. In the SDFT, those states are bracketed by spin-density waves where the spin-up and spin-down densities are symmetrically coupled with each other. In Ref. [96], we have proven explicitly that those states represent a wrong mixing of different spin states. The underlying problem is the fact that the SDFT can not properly describe ensemble-v-representable densities, that is, systems with more than one major configuration in the ground-state wave function.

6.4. Magnetic Fields and the Vortex Clusters

The solutions in high magnetic fields predict the existence of a completely spin-polarized finite structure called the maximum-density droplet (MDD). The MDD is related to the quantum Hall effect with one completely filled Landau level, that is, the filling factor $\nu = 1$, and its existence has also been verified experimentally [97]. The MDD state can be found in various QD geometries. In the MDD state of a circular QD, electrons occupy successive

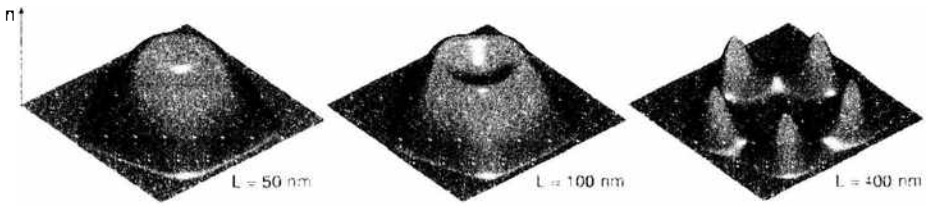


Figure 3. Electron densities for three different sizes of a pentagonal, two-electron quantum dot. As the size of the dot increases, the electrons localize in the corners and form a Wigner molecule.

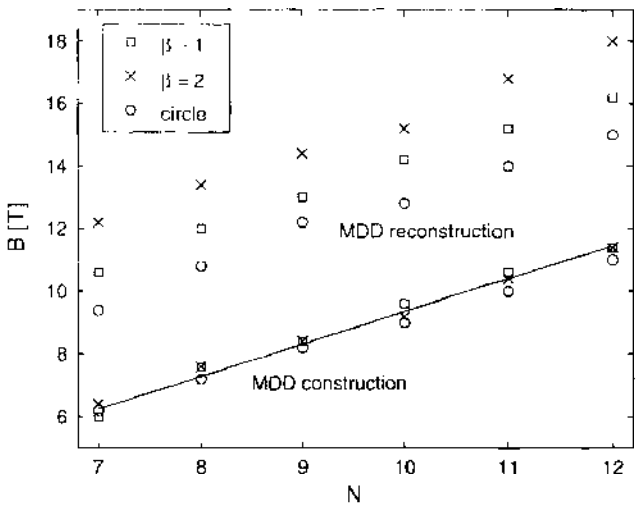


Figure 4. MDD-window limits obtained from the kinks in $\mu(N)$ as a function of N in different QD geometries. The line shows the prediction for the MDD formation based on the number of flux quanta penetrating the dot.

angular momentum levels from $l = 0$ to $l = -N + 1$, where N is the number of electrons in the dot. The MDD state is stable for a rather wide range of the magnetic field, but at a certain field strength it reconstructs into a lower-density droplet (see the end of this section).

In Ref. [98], we study the MDD formation in noncircular hard-wall QDs as previously defined. We identify the MDD window in the calculated chemical potentials $\mu(N)$. In addition, we predict the onset of the MDD from the number of flux quanta N_Φ penetrating through the QD and find a good agreement with the kinks in the chemical potentials (see Fig. 4). Because of the Coulomb interactions, the MDD electron density in a hard-wall dot is pronouncedly localized in the corners and on the edges, in contrast with the parabolic case that exhibits a smooth density distribution [99].

When the magnetic field is further increased the MDD state breaks down to a lower density droplet. The mechanism of this breakdown has been a focus of much theoretical and experimental work. Recently, we have calculated beyond MDD states of different QDs. SDFT predicts formation of vortex structures, that is, holes in the charge density associated with rotating currents around them [100]. They can be seen directly in the total electron density obtained by our symmetry-unrestricted approach. However, these symmetry breaking solutions do not give the physical particle density in the laboratory frame of reference (since it must remain rotationally symmetric) but it may reveal electron–electron correlations in the true many-body wave function which is inaccessible in the density-functional approach.

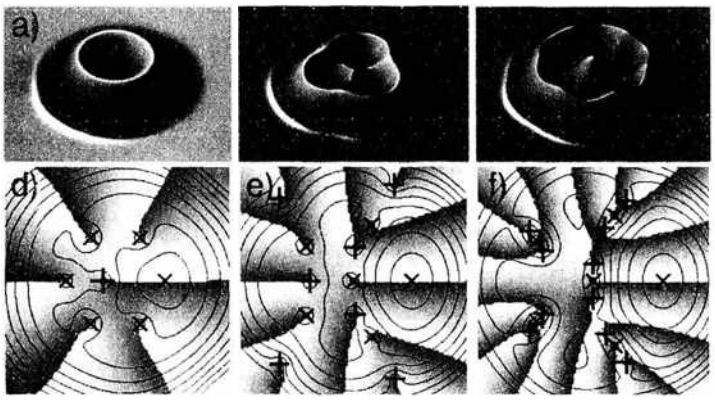


Figure 5. (a–c) Vortex holes in the density-functional electron density of six-electron quantum dots and (d–f) the corresponding conditional wave functions from exact diagonalization. The fixed electrons are marked with crosses. The shading shows the wave function phase that changes by 2π in a path around a vortex. There are vortices on top of each electron and additional vortices moving between the electrons (+ signs). The right-most solution is related to the the $\nu = 1/3$ quantum Hall state with three vortices near each electron.

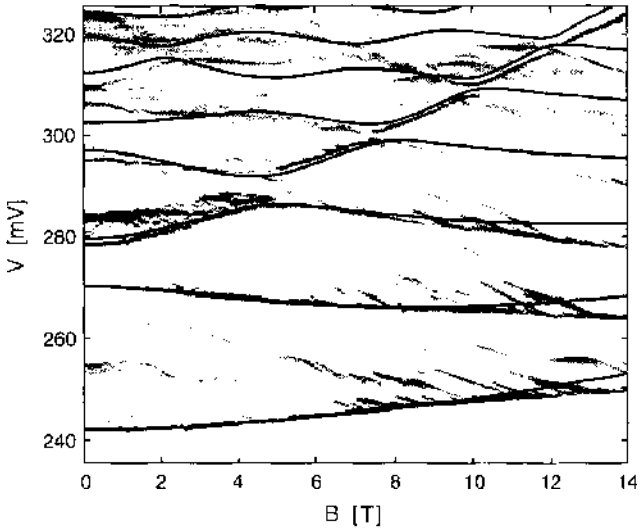


Figure 6. Measured transport spectrum (gray scale) of a GaAs/AlGaAs QD and the calculated single-electron energies (red lines) corresponding to the model potential given in Ref. [98].

Detailed calculations using exact many-body methods lead to similar vortex structures giving credence to the interpretation of the SDFT results [100].

Using different symmetry-breaking QD geometries it was found that the vortices are stable in high magnetic fields and they correspond to density minima also in the ED results [101]. The vortex formation is a considerable energetic effect and it could be observed in transport experiments similar to those of Oosterkamp et al. [97].

Vortex solutions were analyzed further using conditional wave functions in both the ED and the SDFT [102]. The results show that there are two types of vortices: vortices that are on top of an electron and additional vortices which are not bound to a particular electron (see Fig. 5). For the correct particle statistics (fermion antisymmetry), the number of vortices on top of each electron must be odd. The off-electron vortices were found to give rise to charge minima associated with rotating currents around them. The vortex formation reduces the interaction energy and cause strong correlations between the electrons. Some of the solutions have much in common with the fractional quantum Hall states. For instance, the solution with three vortices near each electron was identified as a finite size precursor of the $\nu = 1/3$ fractional quantum Hall state [see Fig. 5(e) and 5(f)]. Moreover, there appear many similarities between vortex formation in bosonic and fermionic case, suggesting that the vortex formation is a universal phenomenon in 2D quantum systems [103].

6.5. Impurities in Quantum Dots

Theoretical modeling of quantum dots is usually based on the approximation of clean samples, although in real semiconductor devices the effects due to impurities or donor-scattering centers may be remarkable. In Ref. [104], a measured transport spectrum of a vertical QD is shown to have clear deviations from the Fock-Darwin energies. We model the system with an external potential consisting of a parabolic confinement and a negatively charged Coulombic impurity placed in the vicinity of the QD. As demonstrated in Fig. 6, the model leads to a good agreement between the calculated single-electron eigenenergies and the experimental spectrum. We also show with the SDFT that in the high magnetic field regime the increasing electron number reduces the distortion induced by the impurity.

7. NANOPHYSICS IN AXIAL SYMMETRY

In Refs. [105–107], we have applied the RQMG-method for studies of various nanostructures. We have found it convenient in all these applications to use axially symmetric model systems instead of fully atomistic models. This approximation reduces the computational demands and allows one to study rather large systems encompassing hundreds (Refs. [105, 107]) and

even thousands (Ref. [106]) of electrons. In addition, by restricting the geometry to the axial symmetry and resorting To uniform-background ("jellium-like") models, many random effects related to the detailed and sometimes unimportant atomic structure disappear, and the relevant physics is easier to extract from the simulations.

In axial symmetry, Eq. (1) for the Kohn-Sham orbital

$$\psi_{mkn}(\mathbf{r}) = e^{im\varphi} U_{mkn}(r, z) \tag{40}$$

can be replaced by the following equation:

$$-\frac{1}{2}\left(\frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2} - \frac{m^2}{r^2} + \frac{\partial^2}{\partial z^2} - 2V_{\text{eff}}\right) U_{mkn}(r, z) = \varepsilon_{mkn} U_{mkn}(r, z) \tag{41}$$

We denote the components of the **k**-vector by k_z and $k_\parallel$. The z-component k_z of the **k**-vector only has relevance in periodic systems, such as the nanowires studied in Ref. [107]. In the periodic case, the following Bloch boundary condition

$$U_{mkn}(r, z + L_{\text{cell}}) = e^{ik_z L_{\text{cell}}} U_{mkn}(r, z) \tag{42}$$

is satisfied. The radial component $k_\parallel$ enters in Ref. [106], where we approximate a planar system by a hexagonal lattice of circles. We see that the numerical problem is reduced to a two-dimensional one. Furthermore, the problem is conveniently split into a number of independent subproblems—a property that can be exploited in a massively parallel computer environment. The Kohn-Sham orbitals with different $(m, \mathbf{k})$ (or $(m, \mathbf{k}, s)$, should we treat spin-polarized systems) are automatically orthogonal, and can be solved simultaneously.

7.1. Ultimate Jellium Model for a Breaking Nanowire

In Ref. [107], we have studied the stability of nanowires and the nanowire breaking process performing self-consistent calculations within the so-called ultimate jellium model (Fig. 7). In this model, both electron and the positive background charge densities acquire the optimal profile minimizing the total energy. The model enables thus studies of shape-dependent properties of nanoscopic systems such as quantum dots or, as in the present work, quantum wires. The model advocates the idea the geometry and ionic structure, also in a partially confined system, are determined by the quantum-mechanical electronic shell structure.

First, we have analyzed the stability of infinite periodic quantum wires pointing out the ability of the electronic band structure to stabilize the nanowires at magic radii, that is, any small deformation of the nanowire along the z-axis always increases the energy. At the unstable radii corresponding to maximum values in the energy oscillations, the wire is uniform up

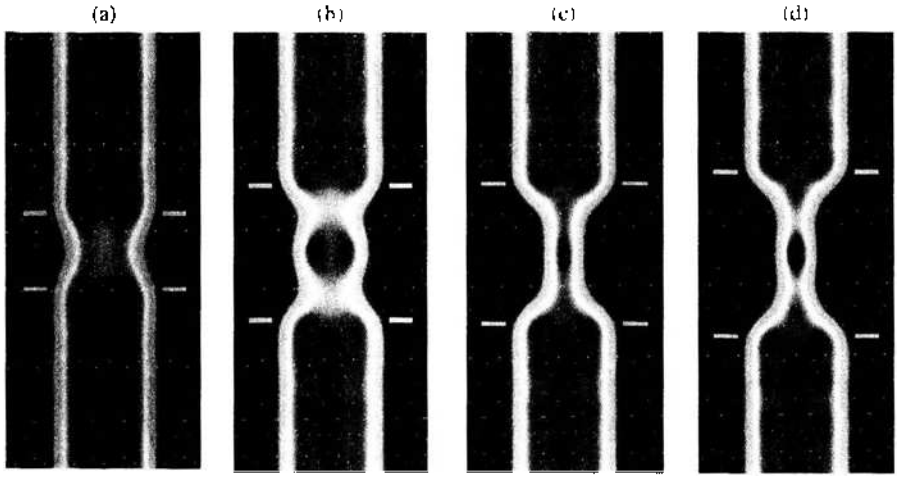


Figure 7. Snapshots from a simulation of nanowire breaking by the ultimate jellium model. A catenoid surface (a), cluster-derived structures (b and d), and uniform cylindrical shape (c) can be seen. Green rectangles mark the lead-constriction boundary.

to a critical value of the unit cell length. The critical values found are close to $L_{cell}/R = 4.5$. Above this limit the local energy minimum disappears and a deformation of the wire lowers the total energy. This length is shorter than the classically expected 2π value, thus the wire electronic structure also has a destabilizing effect.

Then we have investigated the elongation process of finite nanowires between two leads. The elongation force, conductance and effective radius of the constriction have been calculated simultaneously. The importance of the charge relaxation in order to obtain results in agreement with experiments has been shown, for example, in the case of the elongation force. The ability of the ultimate jellium (electron density) to acquire the optimal shape allows the selection of magic radii wires that stabilize the nanocontact, as well as the formation of cluster derived structures (CDS) showing the importance of electron states in the formation of these structures. The related resonance states and their origin is also shown. We have found CDS's that can be linked with the eight- and two-electron freestanding clusters.

In summary, three different types of nanocontact stabilization mechanisms have been found during the breaking process: catenoid-like shape stabilized by classical *surface tension*, straight magic wires stabilized by the *wire electronic shell structure* and CDS's stabilized by *cluster electronic shell structure*.

7.2. Adsorbed Na Quantum Dots on Cu(111)

In Ref. [106], we model electronic properties of the second monolayer Na adatom islands (quantum dots) on the Cu(111) surface covered homogeneously by the wetting layer of one monolayer of Na (Fig. 8). An axially-symmetric three-dimensional jellium model, taking into account the effects due to the first Na monolayer and the Cu substrate, has been developed. The model enables the study of systems consisting of thousands of Na-atoms.

We have modeled quantum dots as small cylindrical jellium islands, and the underlying Na monolayer and Cu substrate as a two-density jellium slab. The two parameters of the model have been chosen to fit experimental spectroscopic data and calculated first-principles band structures for one and two completed monolayers of Na on the Cu(111) surface.

The calculated results are compared with experimental findings in scanning tunneling microscopy and photoemission experiments. The model gives local densities of states which are in a quantitative agreement with constant current topographs and dI/dV spectra and maps. Thereby the idea of surface states which are localized as resonances at the quantum dots is supported. The future applications of the model will include studies of the adsorption and dissociation of molecules in the vicinity of alkali metal quantum dots.

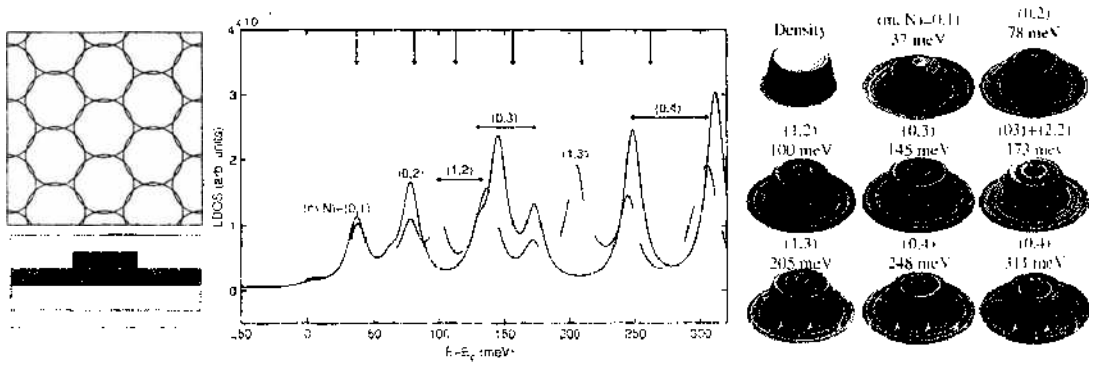


Figure 8. (top left) Hexagonal lattice of area-covering circles. (bottom left) Schematic view of the background charge density in a plane containing the z-axis in our two-density model for a quantum dot on top of a full monolayer of Na on Cu(111). (middle) Local density of states on top of a cylindrical QD of 550 electrons on two-jellium substrate calculated at a height of $i8\,a_{ll}$ above the jellium edge at the axis (solid line) and at $r = 20a_{ll}$ (dashed line). The (shifted) experimental peak positions are given by vertical arrows pointing downward. The peaks are identified with (m, N) resonance states having two horizontal node planes in the QD. (right) Calculated isosurfaces of the electron density (upper-left corner) and LDOS at the energies corresponding to the dominant peaks of the middle panel.

8. POSITRON STATES AT NANOSCALE DEFECTS

The use of positron annihilation in defect studies is based on the trapping of positrons from a delocalized bulk state to a localized state at the defect (see Fig. 9). The trapping is due to the reduced nuclear repulsion at the open-volume defects. Because the electronic structure seen by the positron at the defect differs from that in the perfect bulk crystal the annihilation characteristics change. The positron lifetime increases because the average electron density decreases. For the same reason, the momentum distribution of annihilating electron-positron pairs becomes more peaked at low momenta (see Fig. 10[a]) [108, 109]. However, the positron density may sample the different atomic species of a compound material with different relative probabilities in the bulk and at a defect. The defect may also be surrounded by impurity atoms. In these cases the high-momentum region of the distribution, which is mainly due to annihilations with core electrons, reflects the chemical structure of the defect (see Fig. 10[b]) [110]. The changes in the bond structure between the atoms neighboring the defect may also affect the low-momentum part of the distribution. To understand these changes and fully benefit from them in defect identification, theoretical calculations with high-predictive power are indispensable.

The description of the electron-positron system can be formulated as a two-component density-functional theory [111]. In the measurements there is only one positron in the solid sample at the time. Therefore the density-functional scheme has to be properly purified from positron self-interaction effects. Comparisons with the two-component and experimental results have shown that the following scheme is adequate. First, the electron density $n(\mathbf{r})$ of the system is solved without the effect of the positron. This can be done using different (all-electron) electronic structure calculation methods. A surprisingly good approximation for the positron lifetime and core-electron momentum calculations is to simply superimpose free atom charges. Then the potential $V_+(\mathbf{r})$ felt by positron is constructed as a sum of the Coulomb potential $\phi(\mathbf{r})$ and the so-called correlation potential $V_{\text{corr}}(\mathbf{r})$, which is treated in a local density approximation, that is

$$V_+(\mathbf{r}) = \phi(\mathbf{r}) + V_{\text{corr}}(n\ (\mathbf{r})) \tag{43}$$

The ensuing single-particle Schrödinger equation can be solved using similar techniques as the electron states. For example, we use the three-dimensional real-space Schrödinger equation solver of the MIKA package.

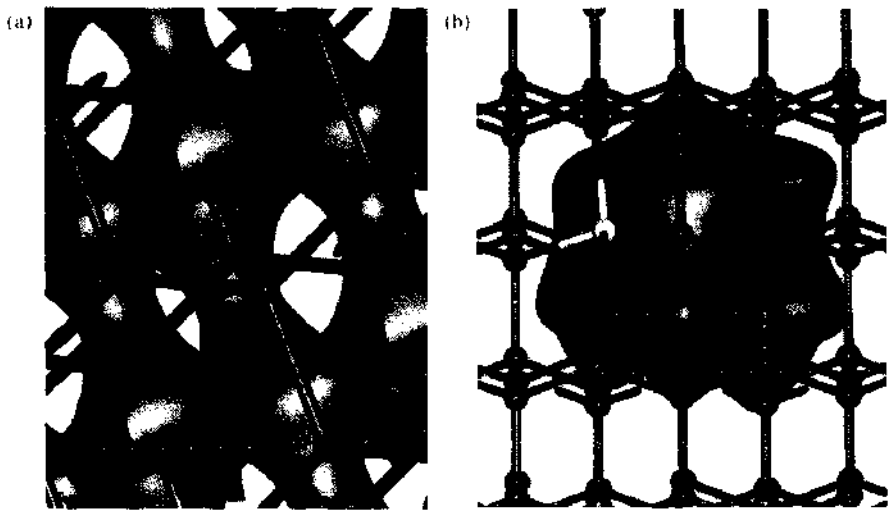


Figure 9. (a) An isosurface of the positron wave function in a perfect Si lattice. The positions of the Si atoms are denoted by blue spheres, and the electronic interatomic bonds as blue sticks. The positron lifetime in this state is according to experiments and theory about 220 ps. (b) An isosurface of the positron wave function at a vacancy surrounded by one Sb impurity. The Sb atom is denoted by a yellow sphere. The positron lifetime in this state is according to theory about 230 ps.

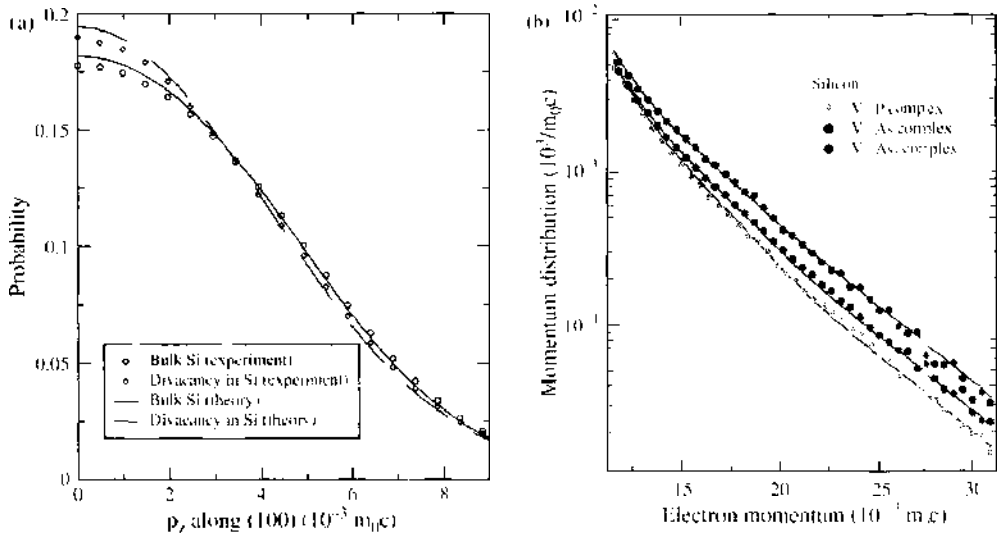


Figure 10. Momentum distributions of annihilating electron-positron pairs in Si. (a) Low momentum parts. The theoretical predictions [108] (lines) are compared with the spectra measured by the Doppler broadening technique (markers) [109]. (b) High-momentum parts (K. Saarinen et al. [110]). The theoretical predictions (solid lines) are compared with the spectra measured by the Doppler broadening techniques (markers). The comparison identifies vacancy-P complexes in electron-irradiated P-doped Si (green circles), vacancy-As complexes in electron-irradiated As-doped Si (blue circles), and vacancy-As_4 complexes in as-grown highly As-doped Si (red circles). The annihilation with As 3d electrons raises the intensity. The study concludes that the saturation of the free electron density in highly As-doped Si is mainly caused by the formation of vacancy-As_4 complexes.

When the electron density $n(\mathbf{r})$ and the positron density $n_+(\mathbf{r}) = |\psi^+(\mathbf{r})|^2$ are known, the positron annihilation rate is calculated within the LDA as an overlap integral

$$\lambda = \pi r_0^2 c \int \mathrm{d}\mathbf{r}\, n_+(\mathbf{r}) n_-(\mathbf{r}) \gamma(n_-(\mathbf{r})) \tag{44}$$

where r_0 is the classical electron radius, c the speed of light, and γ the enhancement factor taking into account the pile up of electron density at the positron (a correlation effect). The inverse of the annihilation rate is the positron lifetime.

The momentum distribution of the annihilating electron-positron pairs is calculated as

$$\rho(\mathbf{p}) = \pi r_0^2 c \sum_j \left| \int \mathrm{d}\mathbf{r}\, e^{-\mathrm{i}\mathbf{p}\cdot\mathbf{r}} \psi^+(\mathbf{r}) \psi_j(\mathbf{r}) \sqrt{\gamma(n_-(\mathbf{r}))} \right|^2 \tag{45}$$

Good results, especially for high-momentum part due to the core electrons, are obtained using a state-dependent constant enhancement factor by replacing $\gamma(n_-(\mathbf{r}))$ above with a constant γ_j, which is determined from the annihilation rate of the state j [112]. It is this state-dependent form, which we use in practice.

The `doppler`-program delivered within the MIKA package uses the atomic superposition method. The scheme cannot be used for the low-momentum part due to valence electrons. For that purpose, self-consistent all-electron wave functions have to be constructed. For example, we have used the projector augmented-wave (PAW) method implemented in the plane-wave code VASP [10, 11, 113, 114].

9. SUMMARY AND OUTLOOK

We have given an overview of the real-space, multigrid-based program package called `MIKA`, and several examples of its applications in research of quantum dots, nanostructures, and positron physics. We hope that the work invested in developing these codes will be useful to a wide group of researchers. Therefore, following the model given, for example, by the `octopus`-project [4], and advocated by the `fsatom` -project [2], we have decided to license the code with the GNU general public license (GPL), and distribute the software

on a Web page [1]. This does not imply that the codes are particularly easy to use or of commercial quality, neither does it include any promise of user support. On the contrary, we hope that other researchers will take parts of the code, inspect them critically, modify them for their purposes, and distribute the derived product further. Such a distributed mode of development should accelerate the development and adoption of real-space methods in the computational nanoscience community.

ACKNOWLEDGMENTS

We are grateful to J. J. Mortensen, M. Marques, A. Castro, and J. Wang for useful discussions on real-space methods; to M. Aichinger and E. Krotscheck for collaboration in the implementation of the response iteration technique to the MIKA-package; and to J. J. Mortensen and M. Aichinger for sharing their manuscripts [57, 77] before publication. B. Hellsing and N. Zabala are acknowledged for collaboration in the research reported in Section 7. M. Hakala, A. Harju, V. Lehtola, V. Ranki, S. Riikonen, K. Ritvanen, P. Valminen and T. Hakala are acknowledged for their contibutions to the MIKA package. S. F. McCormick, M. Lyly, J. Ruokolainen, T. Eirola, M. Huhtanen, and A. Knyazev are acknowledged for discussions on numerical methods. A. Harju deserves a special acknowledgement for providing us results from accurate many-particle methods and also for constructive criticism on density-functional theory in general and the MIKA-package in particular. Special thanks to M. Heiskanen for his original development of the multigrid subroutines.

REFERENCES

1. `http://www.csc.fi/physics/mika`.
2. `http://fsatom.org; http://www.tddft.org/fsatom`.
3. `http://www.abinit.org`.
4. `http://www.tddft.org/programs/octopus`.
5. M. Heiskanen, T. Torsti, M. Puska, and R. Nieminen, *Phys. Rev. B* 63, 245106 (2001).
6. E. Ogando, N. Zabala, E. V. Chulkov, and M. J. Puska, *Phys. Rev. B* 69, 153410 (2004).
7. E. Ogando, N. Zabala, E. V. Chulkov, and M. J. Puska, *Phys. Rev. B* 71, 205401 (2005).
8. F. Nogueira, A. Castro, and M. Marques, in "A Primer in Density Functional Theory" (C. Fiolhais, F. Nogueira, and M. Marques, Eds.), p. 218, Heidelberg, 2003.
9. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* 64, 1045 (1992).
10. G. Kresse and J. Furthmüller, *Phys. Rev. B* 54, 11169 (1996).
11. G. Kresse and J. Furthmüller, *Comp. Mater. Sci.* 6, 15 (1996).
12. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
13. R. N. Barnett and U. Landman, *Phys. Rev. B* 48, 2081 (1993).
14. B. Delley, *J. Chem. Phys.* 113, 7756 (2000).
15. G. te Velde, F. Bickelhaupt, E. Baerends, C. F. Guerra, S. van Gisbergen, J. Snijders, and T. Ziegler, *J. Comput. Chem.* 22, 931 (2001).
16. R. Ahlrichs and M. V. Arnim, in "Methods and Techniques in Computational Chemistry: METECC-95" (E. Clementi and G. Corongiu), pp. 509–504, STEF, Cagliari, 1995.
17. High Performance Computational Chemistry Group, NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.1, Pacific Northwest National Laboratory, Richland, Washington 99352, 2002.
18. J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *J. Phys.* 14, 2745 (2002).
19. T. L. Beck, *Rev. Mod. Phys.* 72, 1041 (2000).
20. S. R. White, J. W. Wilkins, and M. P. Teter, *Phys. Rev. B* 39, 5819 (1989).
21. J. Pask, B. Klein, C. Fong, and P. Sterne, *Phys. Rev. B* 135, 1 (1999).
22. E. Tsuchida and M. Tsukada, *Phys. Rev. B* 52, 5573 (1995).
23. E. Tsuchida and M. Tsukada, *J. Phys. Soc. Jpn.* 67, 3844 (1998).
24. K. Tagami, E. Tsuchida, and M. Tsukada, *Surf. Sci.* 446, L108 (2000).
25. J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, *Phys. Rev. B* 50, 11355 (1994).
26. A. P. Seitsonen, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 51, 14057 (1995).
27. E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Phys. Rev. B* 54, 14362 (1996).
28. F. Ancilotto, P. Blandin, and F. Toigo, *Phys. Rev. B* 59, 7868 (1999).
29. Y.-G. Jin, J.-W. Jeong, and K. Chang, *Physica B* 274, 1003 (1999).
30. J. Wang and T. L. Beck, *J. Chem. Phys.* 112, 9223 (2000).
31. U. V. Waghmare, H. Kim, I. J. Park, N. Modine, P. Maragakis, and E. Kaxiras, *Comput. Phys. Commun.* 137, 341 (2001).
32. T. A. Arias, *Rev. Mod. Phys.* 71, 267 (1999).

33. D. Bai and A. Brandt, *SIAM J. Sci. Stat. Comput.* 8, 109 (1987).
34. E. J. Bylaska, S. R. Kohn, S. B. Baden, A. Edelman, R. Kawai, M. E. G. Ong, and J. H. Weare, in "Proceedings of the 7th SIAM Conference on Parallel Processing for Scientific Computing" (D. H. Bailey, P. E. Bjørstad, J. R. Gilbert, M. Mascagni, R. S. Schreiber, H. D. Simon, V. J. Torczon, and L. T. Watwson, Eds.), p. 219, 1995.
35. J.-L. Fattebert, *J. Comput. Phys.* 149, 75 (1999).
36. F. Gygi and G. Galli, *Phys. Rev. B* 52, R2229 (1995).
37. I. Babuska and Rheinboldt, *Int. J. Num. Meth. Eng.* 12, 1579 (1978).
38. I. Babuska and B. Szabo, *Int. J. Num. Meth. Eng.* 18, 323 (1982).
39. I. Babuska and M. Suri, *Comp. Meth. Appl. Mech. Eng.* 80, 5 (1990).
40. A. Brandt, *Math. Comput.* 31, 333 (1977).
41. A. Brandt, S. F. McCormick, and J. W. Ruge, *SIAM J. Sci. Comput. (USA)* 4, 244 (1983).
42. S. Goedecker, *Rev. Mod. Phys.* 71, 1085 (1999).
43. J.-L. Fattebert and J. Bernholc, *Phys. Rev. B* 62, 1713 (2000).
44. A. D. Becke, *J. Chem. Phys.* 76, 6037 (1982).
45. A. D. Becke, *Phys. Rev. A* 33, 2786 (1986).
46. L. Laaksonen, P. Pyykkö, and D. Sundholm, *Comp. Phys. Rep.* 4, 313 (1986).
47. D. Sundholm, *Comp. Phys. Commun.* 49, 409 (1988).
48. J. Kobus, L. Laaksonen, and D. Sundholm, *Comp. Phys. Commun.* 98, 346 (1996).
49. E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Phys. Rev. B* 52, R5471 (1995).
50. W. Hackbush, "Multi-Grid Methods and Applications," Springer, Berlin, 1985.
51. P. Wesseling, "An Introduction to Multigrid Methods," Wiley, New York, 1992.
52. W. L. Briggs, V. E. Henson, and S. F. McCormick, "A Multigrid Tutorial," 2nd Edn., SIAM, Philadelphia, 2000.
53. T. L. Beck, K. A. Iyer, and M. P. Merrick, *Int. J. Quantum Chem.* 61, 341 (1997).
54. S. Costiner and S. Ta'asan, *Phys. Rev. E* 51, 3704 (1995).
55. S. Costiner and S. Ta'asan, *Phys. Rev. E* 52, 1181 (1995).
56. J. Wang and K. Marthinsen (unpublished, 2003).
57. J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, *Phys. Rev. B* 71, 035109 (2005).
58. J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* 72, 1240 (1994).
59. X. Jing, N. Troullier, D. Dean, J. R. Chelikowsky, K. Wu, and Y. Saad, *Phys. Rev. B* 50, 12234 (1994).
60. J. R. Chelikowsky, X. Jing, K. Wu, and Y. Saad, *Phys. Rev. B* 53, 12071 (1996).
61. G. H. Golub and C. F. V. Loan, "Matrix Computations," 2nd Edn., Johns Hopkins University Press, London, 1989.
62. A. Stathopoulos, S. Ögüt, Y. Saad, J. Chelikowsky, and H. Kim, *Comput. Sci. Eng.* 2, 19 (2000).
63. A. Knyazev, *Siam. J. Sci. Comput.* 23, 517 (2001).
64. A. Knyazev and K. Neymeyr, *El. Trans. Numer Anal.* 15, 38 (2003).
65. J. Mandel and S. McCormick, *J. Comput. Phys.* 80, 442 (1989).
66. J. L. B. Cooper, *Q. Appl. Math.* 6, 179 (1948).
67. T. Ono and K. Hirose, *Phys. Rev. Lett.* 82, 5016 (1999).
68. P. Pulay, *Chem. Phys. Lett.* 73, 393 (1980).
69. P. Pulay, *J. Comp. Chem.* 3, 556 (1982).
70. D. R. Bowler and M. J. Gillan, *Chem. Phys. Lett.* 325, 473 (2000).
71. G. P. Kerker, *Phys. Rev. B* 23, 3082 (1981).
72. M. Manninen, R. M. Nieminen, P. Hautojärvi, and J. Arponen, *Phys. Rev. B* 12, 4012 (1975).
73. J. Arponen, P. Hautojärvi, R. M. Nieminen, and E. Pajanne, *J. Phys. F* 3, 2092 (1973).
74. R. M. Nieminen, *J. Phys. F* 7, 375 (1977).
75. J. Auer and E. Krotscheck, *Comput. Phys. Commun.* 118, 139 (1999).
76. J. Auer and E. Krotscheck, *Comput. Phys. Commun.* 151, 139 (2003).
77. M. Aichinger and E. Krotscheck, *Comp. Mater. Sci.* 34, 188 (2005).
78. H. A. van der Vorst, "Iterative Krylov Methods for Large Linear Systems," Cambridge University, Cambridge, 2003.
79. T. Torsti (unpublished, 1997); M. Heiskanen (unpublished, 1998).
80. J. Dongarra, in "Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide" (Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, Eds.), p. 372, SIAM, Philadelphia, 2000.
81. J. Auer, E. Krotscheck, and S. A. Chin, *J. Chem. Phys.* 115, 6841 (2001).
82. M. A. L. Marques, A. Castro, G. F. Bertsch, and A. Rubio, *Comput. Phys. Commun.* 151, 60 (2003).
83. J. R. Chelikowsky, L. Kronik, and I. Vasiliev, *J. Phys.* 15, R1517 (2003).
84. S. Riikonen, T. Torsti, M. J. Puska, and R. M. Nieminen (unpublished, 2002).
85. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996).
86. J. A. White and D. M. Bird, *Phys. Rev. B* 50, 4954 (1994).
87. For an overview, see L. Jacak, P. Hawrylak, and A. Wójs, "Quantum Dots," Springer, Berlin, 1998.
88. See, e.g., L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, *Rep. Prog. Phys.* 64, 701 (2001); S. M. Reimann and M. Manninen, *Rev. Mod. Phys.* 74, 1283 (2002).
89. H. Saarikoski, E. Räsänen, S. Siljamäki, A. Harju, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 67, 205327 (2003).
90. C. Attaccalite, S. Moroni, P. Gori-Giorgi, and G. B. Bachelet, *Phys. Rev. Lett.* 88, 256601 (2002).
91. B. Tanatar and D. M. Ceperley, *Phys. Rev. B* 39, 5005 (1989).

92. E. Räsänen, H. Saarikoski, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 67, 035326 (2003).
93. C. E. Creffield, W. Häusler, J. H. Jefferson, and S. Sarkar, *Phys. Rev. B* 59, 10719 (1999).
94. E. Räsänen, H. Saarikoski, V. N. Stavrou, A. Harju, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 67, 235307 (2003).
95. D. G. Austing, S. Sasaki, S. Tarucha, S. M. Reimann, M. Koskinen, and M. Manninen, *Phys. Rev. B* 60, 11514 (1999).
96. A. Harju, E. Räsänen, H. Saarikoski, M. J. Puska, R. M. Nieminen, and K. Niemelä, *Phys. Rev. B* 69, 153101 (2004).
97. T. H. Oosterkamp, J. W. Janssen, L. P. Kouwenhoven, D. G. Austing, T. Honda, and S. Tarucha, *Phys. Rev. Lett.* 82, 2931 (1999).
98. E. Räsänen, A. Harju, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 69, 165309 (2004).
99. S. M. Reimann, M. Koskinen, M. Manninen, and B. R. Mottelson, *Phys. Rev. Lett.* 83, 3270 (1999).
100. H. Saarikoski, A. Harju, M. J. Puska, and R. M. Nieminen, *Phys. Rev. Lett.* 93, 116802 (2004).
101. H. Saarikoski, S. M. Reimann, E. Räsänen, A. Harju, and M. J. Puska, *Phys. Rev. B.* 71, 035421 (2005).
102. H. Saarikoski, A. Harju, M. J. Puska, and R. M. Nieminen, *Physica E* 26, 317 (2005).
103. M. Toreblad, M. Borgh, M. Koskinen, M. Manninen, and S. M. Reimann, *Phys. Rev. Lett.* 93, 090407 (2004).
104. E. Räsänen, J. Könemann, R. J. Haug, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 70, 115308 (2004).
105. P. Havu, T. Torsti, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 66, 075401 (2002).
106. T. Torsti, V. Lindberg, M. J. Puska, and B. Hellsing, *Phys. Rev. B* 66, 235420 (2002).
107. E. Ogando, T. Torsti, M. J. Puska, and N. Zabala, *Phys. Rev. B* 67, 075417 (2003).
108. M. Hakala, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B* 57, 7621 (1998).
109. K. Saarinen and V. Ranki, *J. Phys.* 15, S2791 (2003).
110. K. Saarinen, J. Nissilä, H. Kauppinen, M. Hakala, M. J. Puska, P. Hautojärvi, and C. Corbel, *Phys. Rev. Lett.* 82, 1883 (1999).
111. E. Boroński and R. M. Nieminen, *Phys. Rev. B* 34, 3820 (1986).
112. M. Allatalo, B. Barbiellini, M. Hakala, H. Kauppinen, T. Korhonen, M. J. Puska, K. Saarinen, P. Hautojärvi, and R. M. Nieminen, *Phys. Rev. B* 54, 2397 (1996).
113. G. Kresse and D. Joubert, *Phys. Rev. B* 59, 1758 (1999).
114. I. Makkonen, M. Hakala, and M. J. Puska, *Phys. Rev. B* in press.

CHAPTER 17

Electronic Structure of Clusters and Nanocrystals

James R. Chelikowsky

Department of Chemical Engineering and Materials Science, Institute for the Theory of Advanced Materials in Information Technology, Digital Technology Center, University of Minnesota, Minneapolis, Minnesota, USA

Yousef Saad

Department of Computer Science, Institute for the Theory of Advanced Materials in Information Technology, Digital Technology Center, University of Minnesota, Minneapolis, Minnesota, USA

CONTENTS

1. Introduction 798
2. Quantum Descriptions of Matter 798
   2.1. The Hartree Approximation 800
   2.2. The Hartree–Fock Approximation 801
3. Density-Functional Approaches 803
   3.1. Free-Electron Gas 803
   3.2. Hartree–Fock Exchange in a Free-Electron Gas 804
   3.3. Density-Functional Theory 805
   3.4. Time-Dependent Density Functional Theory 807
4. Pseudopotentials 812
5. Solving the Eigenvalue Problem 818
6. Properties of Confined Systems: Clusters 822
   6.1. Structure 822
   6.2. Photoemission Spectra 824
   6.3. Vibrational Modes 825
   6.4. Polarizabilities 828
   6.5. Optical Spectra 830

ISBN: 1-58883-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8. Pages (797–852)

7. Quantum Confinement in Nanocrystals 833
7.1. Role of Oxygen in Silicon Quantum Dots 840
7.2. Doping Quantum Dots 845
References 849

1. INTRODUCTION

Frequently one categorizes material properties at the macroscopic scale by terms such as "intensive" and "extensive." An intensive property is not dependent on the sample size or mass. For example, if one specifies the temperature of a sample as that of room temperature, the description is complete. One does not need to specify the size of the sample. Other intensive properties include pressure and density. An extensive property does depend on the size of the sample. For example, the volume of a sample depends on the size: Two moles of a gas occupies twice the volume as one mole. Therefore, one often specifies volume in terms of the volume per mole, or the specific volume. Other extensive properties include thermodynamic properties such as the heat capacity, enthalpy, entropy, and free energy of the system.

Matter at the nanoscale is different. Properties that are intensive at the macroscopic scale may not be intensive at the "nano-" or "subnano-" scale. In fact, such properties may be hard to define at very small length scales. Consider a small cluster of atoms (e.g., a dozen silicon atoms). Such an ensemble contains so few atoms that it is difficult to define a property such as temperature or volume. Although extensive properties such as the heat capacity or free energy of the system remain, in the sense of changing with the size of the system, such properties may no longer scale linearly with size. For example, the free energy of two dozen atoms of silicon may not be twice the free energy of one dozen atoms of silicon.

In this latter example, one can make some simple arguments to explain this behavior. Suppose we consider the scaling of the free energy for a spherical sample of matter whose radius is R. The volume energy term would scale as R^3; the surface term as R^2. As R tends toward the nanoscale, the surface terms can become the dominant term in the free energy and do not scale linearly with the volume or mass of the sample. This different scaling between volume and surface terms is well known in nucleation theory. A manifestation of this effect is that particles must exceed a certain size before they are stable.

One definition of the nanoscale is the size at which deviations from intrinsic intensive and extensive properties at the macroscopic scale occur. Consider an intensive electronic property such as the band gap of silicon. For specificity, let us consider a spherical sample of silicon whose radius is R. If R changes from 10 cm to 1 cm, the optical gap will remain unchanged, as expected for an intensive property. However, if one considers values of R changing from 10 nm to 1 nm, the optical gap will be strongly altered. In this size regime, the optical gap is no longer an intensive property; the properties of such systems are said to characterize nanoscale. One of the first manifestations of this effect was observed in porous silicon, which exhibits remarkable room temperature luminescence [1]. This is in strong contrast to crystalline silicon, which is optically inactive. It is widely accepted that the localization of optical excitations at the nanoscale results in the luminescence of porous silicon.

In this chapter, the cornerstones of theoretical methods for understanding the electronic and structural properties of matter at the nanoscale will be reviewed.

2. QUANTUM DESCRIPTIONS OF MATTER

Quantum mechanical laws that describe the behavior of matter at the nanoscale were discovered in the early part of the twentieth century. These laws mark one of the greatest scientific achievements of humankind. Using these laws, it is possible to predict the electronic properties of matter from the nanoscale to the macroscale, at least in principle.

Although it is relatively easy to write down the Hamiltonian for interacting fermions, obtaining a solution to the problem that is sufficiently accurate to make predictions is another matter.

Consider N nucleons of charge Z_n at positions $\{\vec{R}_n\}$ for $n = 1, \ldots, N$ and I electrons at positions $\{\vec{r}_i\}$ for $i = 1, \ldots, M$. This is shown schematically in Fig. 1. The Hamiltonian for this system in its simplest form can be written as

$$\mathcal{H}(\vec{R}_1, \vec{R}_2, \vec{R}_3, \ldots; \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots) = \sum_{n=1}^{N} \frac{-\hbar^2 \nabla_n^2}{2M_n} + \frac{1}{2} \sum_{n,m=1, n\neq m}^{N} \frac{Z_n Z_m e^2}{|\vec{R}_n - \vec{R}_m|} + \sum_{i=1}^{M} \frac{-\hbar^2 \nabla_i^2}{2m} - \sum_{n=1}^{N} \sum_{i=1}^{M} \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} + \frac{1}{2} \sum_{i,j=1, i\neq j}^{M} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (1)$$

M_n is the mass of the nucleon, $\hbar$ is Planck's constant divided by 2π, and m is the mass of the electron. This expression omits some terms such as those involving relativistic interactions, but it captures the essential features for nanoscale matter.

Using the Hamiltonian in Eq. (1), the quantum mechanical equation known as the Schrödinger equation for the electronic structure of the system can be written as

$$\mathcal{H}(\vec{R}_1, \vec{R}_2, \vec{R}_3, \ldots; \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots)\Psi(\vec{R}_1, \vec{R}_2, \vec{R}_3, \ldots; \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots) = E\Psi(\vec{R}_1, \vec{R}_2, \vec{R}_3, \ldots; \vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots) \quad (2)$$

where E is the total electronic energy of the system and Ψ is the many-body wave function.

Soon after the discovery of the Schrödinger equation, it was recognized that this equation provided the means of solving for the electronic and nuclear degrees of freedom. Using the variational principle, which states that an approximate wave function will always have a less favorable energy than the true ground-state energy, one has an equation and a method to test the solution. One can estimate the energy from

$$E = \frac{\int \Psi^* \mathcal{H} \Psi d^3R_1 d^3R_2 d^3R_3 \cdots d^3r_1 d^3r_2 d^3r_3 \cdots}{\int \Psi^* \Psi d^3R_1 d^3R_2 d^3R_3 \cdots d^3r_1 d^3r_2 d^3r_3 \cdots} \quad (3)$$

However, a solution of Eq. (2) for anything more complex than a few particles becomes problematic even with the most powerful computers. Obtaining an approximate solution for systems with many atoms is difficult, but considerable progress has been made since the advent of reliable digital computers.

A number of highly successful approximations have been made to solve for both the ground-state and excited-state energies. For the most part, these approximations are used

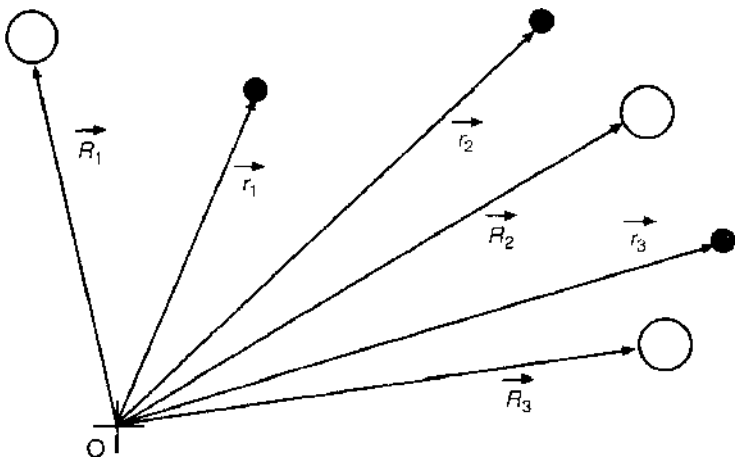


Figure 1. Atomic and electronic coordinates. The electrons are illustrated by filled circles; the nuclei by open circles.

to remove as many "irrelevant" degrees of freedom from the system as possible. One common approximation is to separate the nuclear and electronic degrees of freedom. Because the nuclei are considerably more massive than the electrons, it can be assumed that the electrons will respond "instantaneously" to the nuclear coordinates. This approximation is called the Born–Oppenheimer or adiabatic approximation. It allows one to treat the nuclear coordinates as classical parameters. For most condensed matter systems, this assumption is highly accurate [2, 3].

2.1. The Hartree Approximation

Another common approximation is to construct a specific form for the many-body wave function. If one can obtain an accurate wave function, then via the variational principle, an accurate estimate for the energy will emerge. The most difficult part of this exercise is using physical intuition to define a trial wave function that is close to the true wave function.

One can use some simple limiting cases to illustrate the construction of many-body wave functions. Suppose one considers a solution for noninteracting electrons; that is, in Eq. (1), the last term in the Hamiltonian is ignored. In this limit, it is possible to write the many-body wave function as a sum of independent Hamiltonians. Using the adiabatic approximation, the electronic part of the Hamiltonian becomes

$$\mathcal{H}_{\text{el}}(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots) = \sum_{i=1}^{M} \frac{-\hbar^2 \nabla_i^2}{2m} - \sum_{n=1}^{N} \sum_{i=1}^{M} \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} \tag{4}$$

Let us define a nuclear potential, V_N, which the ith electron sees as

$$V_N(\vec{r}_i) = -\sum_{n=1}^{N} \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} \tag{5}$$

One can now rewrite our simplified Schrödinger equation as

$$\mathcal{H}_{\text{el}}(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots)\psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots) = \sum_{i=1}^{M} H^i \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots) \tag{6}$$

where the Hamiltonian for the ith electron is

$$H^i = \frac{-\hbar^2 \nabla_i^2}{2m} + V_N(\vec{r}_i) \tag{7}$$

For this simple Hamiltonian, let us write the many-body wave function as

$$\psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)\phi_3(\vec{r}_3)\cdots \tag{8}$$

The $\phi_i(\vec{r})$ orbitals can be determined from a "one-electron" Hamiltonian

$$H^i \phi_i(\vec{r}) = \left[\frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r})\right]\phi(\vec{r}) = E_i \phi_i(\vec{r}) \tag{9}$$

The index i for the orbital $\phi_i(\vec{r})$ can be taken to include the spin of the electron plus any other relevant quantum numbers. This type of Schrödinger equation can be easily solved for fairly complex condensed matter systems. The many-body wave function in Eq. (8) is known as the Hartree wave function. If one uses this form of the wave function as an approximation to solve the Hamiltonian including the electron–electron interactions, this is known as the Hartree Approximation. By ignoring the electron–electron terms, the Hartree approximation treats the electrons moving independently in the nuclear potential. The total energy of the system is given by the sum of the eigenvalues, E_i.

To obtain a realistic Hamiltonian, the electron–electron interactions must be reinstated in Eq. (6).

$$\mathcal{H}_{el}(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots)\psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots) = \sum_{i=1}^{M}\left(H^i + \frac{1}{2}\sum_{j=1, j\neq i}^{M}\frac{e^2}{|\vec{r}_i - \vec{r}_j|}\right)\psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots) \tag{10}$$

In this case, the individual orbitals $\phi_i(\vec{r})$ can be determined by minimizing the total energy as per Eq. (3), with the constraint that the wave function be normalized. This minimization procedure results in the following Hartree equation.

$$H^i\phi_i(\vec{r}) = \left[\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1, j\neq i}^{M}\int\frac{e^2|\phi_j(\vec{r}')|^2}{|\vec{r} - \vec{r}'|}d^3r'\right]\phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{11}$$

Using the orbitals $\phi(\vec{r})$ from a solution of Eq. (11), the Hartree many-body wave function can be constructed and the total energy determined from Eq. (3).

The Hartree approximation is useful as an illustrative tool, but it is not a very accurate approximation. A significant failing of the Hartree wave function is that it does not reflect the antisymmetric nature of the electrons, as required by the Pauli principle [4]. Moreover, the Hartree equation is difficult to solve. The Hamiltonian is orbital dependent because the summation in Eq. (11) does not include the ith orbital. This means that if there are M electrons, then M Hamiltonians must be considered and Eq. (11) must be solved for each orbital.

2.2. The Hartree–Fock Approximation

It is possible to write down a many-body wave function that reflects the antisymmetric nature of the wave function. In this approach, the spin coordinate of each electron is explicitly treated. The coordinates of an electron may be specified by $\vec{r}_i s_i$, where s_i represents the spin coordinate. Starting with one-electron orbitals, $\phi_i(\vec{r}s)$, the following form can be invoked.

$$\Psi(\vec{r}_1 s_1, \vec{r}_2 s_2, \vec{r}_3 s_3, \ldots) = \begin{vmatrix} \phi_1(\vec{r}_1 s_1) & \phi_1(\vec{r}_2 s_2) & \cdots & \cdots & \phi_1(\vec{r}_M s_M) \\ \phi_2(\vec{r}_1 s_1) & \phi_2(\vec{r}_2 s_2) & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \phi_M(\vec{r}_1 s_1) & \cdots & \cdots & \cdots & \phi_M(\vec{r}_M s_M) \end{vmatrix} \tag{12}$$

This form of the wave function is called a Slater determinant. It reflects the proper symmetry of the wave function and the Pauli principle. If two electrons occupy the same orbit, two rows of the determinant will be identical, and the many-body wave function will have zero amplitude. Likewise, the determinant will vanish if two electrons occupy the same point in generalized space (i.e., $\vec{r}_i s_i = \vec{r}_j s_j$), as two columns of the determinant will be identical. If two particles are exchanged, this corresponds to a sign change in the determinant. The Slater determinant is a convenient representation, but it is an *ansatz*. It is probably the simplest many-body-waved function that incorporates the required symmetry properties for fermions, or particles with noninteger spins.

If one uses a Slater determinant to evaluate the total electronic energy and maintains wave function normalization, the orbitals can be obtained from the following Hartree–Fock equations:

$$H^i\phi_i(\vec{r}) = \left[\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1}^{M}\int\frac{e^2|\phi_j(\vec{r}')|^2}{|\vec{r} - \vec{r}'|}d^3r'\right]\phi_i(\vec{r})$$
$$- \sum_{j=1}^{M}\int\frac{e^2}{|\vec{r} - \vec{r}'|}\phi_j^*(\vec{r}')\phi_i(\vec{r}')d^3r'\delta_{s_i, s_j}\phi_j(\vec{r}) = E_i\phi_i(\vec{r}) \tag{13}$$

It is customary to simplify this expression by defining an electronic charge density, ρ:

$$\rho(\vec{r}) = \sum_{j=1}^{M} |\phi_j(\vec{r})|^2 \tag{14}$$

and an orbital dependent exchange-charge density ρ_i^{HF} for the ith orbital

$$\rho_i^{HF}(\vec{r}, \vec{r}') = \sum_{j=1}^{M} \frac{\phi_j^*(\vec{r}')\phi_i(\vec{r}')\phi_i^*(\vec{r})\phi_j(\vec{r})}{\phi_i^*(\vec{r})\phi_i(\vec{r})} \delta_{s_i s_j} \tag{15}$$

This "density" involves a spin-dependent factor that couples only states (i, j) with the same spin coordinates (s_i, s_j).

With these charge densities defined, it is possible to define corresponding potentials. The Coulomb or Hartree potential V_H is defined by

$$V_H(\vec{r}) = \int \rho(\vec{r}') \frac{e^2}{|\vec{r} - \vec{r}'|} d^3r' \tag{16}$$

and an exchange potential can be defined by

$$V_x^i(\vec{r}) = -\int \rho_i^{HF}(\vec{r}, \vec{r}') \frac{e^2}{|\vec{r} - \vec{r}'|} d^3r' \tag{17}$$

This combination results in the following Hartree–Fock equation

$$\left[\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + V_H(\vec{r}) + V_x^i(\vec{r})\right]\phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{18}$$

Once the Hartree–Fock orbitals have been obtained, the total Hartree–Fock electronic energy of the system, E_{HF}, can be obtained from

$$E_{HF} = \sum_i^M E_i - \frac{1}{2}\int \rho(\vec{r})V_H(\vec{r})d^3r - \frac{1}{2}\sum_i^M \int \phi_i^*(\vec{r})\phi_i(\vec{r})V_x^i(\vec{r})\, d^3r \tag{19}$$

E_{HF} is not a sum of the Hartree–Fock orbital energies E_i. The factor of $\frac{1}{2}$ in the electron–electron terms arises because the electron–electron interactions have been double counted in the Coulomb and exchange potentials. The Hartree–Fock Schrödinger equation is only slightly more complex than the Hartree equation. Again, the equations are difficult to solve because the exchange potential is orbital dependent.

There is one notable difference in the Hartree–Fock summations compared to the Hartree summation. The Hartree–Fock sums include the $i = j$ terms found in Eq. (13). This difference arises because the exchange term corresponding to $i = j$ cancels an equivalent term in the Coulomb summation. The $i = j$ term in both the Coulomb and exchange term is interpreted as a "self-screening" of the electron. Without a cancellation between Coulomb and exchange terms, a "self-energy" contribution to the total energy would occur. Approximate forms of the exchange potential often do not have this property. The total energy then contains a self-energy contribution that one needs to remove to obtain a correct Hartree–Fock energy.

The Hartree–Fock equation is an approximate solution to the true ground-state many-body wave functions. Terms not included in the Hartree–Fock energy are referred to as correlation contributions. One definition for the correlation energy E_{corr} is to write it as the difference between the exact total energy of the system E_{exact} and the Hartree–Fock energies: $E_{corr} = E_{exact} - E_{HF}$. Correlation energies may be included by considering Slater determinants composed of orbitals that represent excited-state contributions. This method of including unoccupied orbitals in the many body wave function is referred to as configuration interactions or CIs [5].

Applying Hartree–Fock wave functions to systems with many atoms is not routine. The resulting Hartree–Fock equations are often too complex to be solved for extended systems.

except in special cases. The number of electronic degrees of freedom grows rapidly, with the number of atoms often prohibiting an accurate solution, or even one's ability to store the resulting wave function. As such, it has been argued that a "wave function" approach to systems with many atoms does not offer a satisfactory approach to the electronic structure problem. An alternate approach is based on density-functional theory.

3. DENSITY-FUNCTIONAL APPROACHES

Descriptions of quantum states based on a knowledge of the electronic charge density [Eq. (14)] have existed since the 1920s. For example, the Thomas–Fermi description, based on a knowledge of $\rho(\vec{r})$, was one of the first attempts at quantitative theory for the electronic structure of atoms [6–8]. However, most treatments of density-functional theory begin by considering a free-electron gas of uniform charge density. The justification for this starting point comes from the observation that simple metals like aluminum and sodium have properties that appear to resemble those of a free-electron gas. A "free-electron" model cannot be applied to systems with localized electrons, such as highly covalent materials like carbon or highly ionic materials like sodium chloride.

3.1. Free-Electron Gas

Perhaps the simplest description of a condensed matter system is to imagine noninteracting electrons contained within a box of volume Ω. The Schrödinger equation for this system is similar to Eq. (9) with the potential set to zero.

$$\frac{-\hbar^2\nabla^2}{2m}\phi(\vec{r}) = E\phi(\vec{r}) \tag{20}$$

Ignoring spin for the moment, the solution of Eq. (20) is

$$\phi(\vec{r}) = \frac{1}{\sqrt{\Omega}}\exp(i\vec{k}\cdot\vec{r}) \tag{21}$$

The variable $\vec{k}$ is called a wave vector. The energy is given by $E(k) = \hbar^2k^2/2m$, and the charge density by $\rho = 1/\Omega$.

A key issue in describing systems with a very large number of atoms is to account properly for the number of states. In the limit of systems corresponding to crystalline states, the eigenvalues are closely spaced and essentially "infinite" in number. For example, if one has a mole of atoms, then one can expect to have $\sim 10^{23}$ occupied states. In such systems, the number of states per unit energy is a natural measure to describe the energy distribution of states.

It is easy to do this with periodic boundary conditions. Suppose one considers a one-dimensional specimen of length L. In this case the wave functions obey the rule $\phi(x + L) = \phi(x)$, as $x + L$ corresponds in all physical properties to x. For a free-electron wave function, this requirement can be expressed as $\exp[ik(x + L)] = \exp(ikx)$ or as $\exp(ikL) = 1$ or $k = 2\pi n/L$, where n is an integer.

Periodic boundary conditions force k to be a discrete variable, with allowed values occurring at intervals of $2\pi/L$. For very large systems, one can describe the system as continuous in the limit of $L \to \infty$. Electron states can be defined by a density of states, defined as follows

$$D(E) = \lim_{\Delta E \to 0}\frac{(N[E + \Delta E] - N[E])}{\Delta E} = \frac{dN}{dE} \tag{22}$$

where $N(E)$ is the number of states whose energy resides below E. For the one-dimensional case, $N(k) = 2k/(2\pi/L)$ (the factor of two coming from spin) and $dN/dE = (dN/dk)\cdot(dk/dE)$. Using $E(k) = \hbar^2k^2/2m$, we have $k = \sqrt{2mE}/\hbar$ and $dk/dE = \frac{1}{2}\sqrt{2m/E}/\hbar$. This result for this one-dimensional density of states is

$$D(E) = \frac{L}{2\pi\hbar}\sqrt{2m/E} \tag{23}$$

The density of states for a one-dimensional system diverges as $E \rightarrow 0$. This divergence of $D(E)$ is not a serious issue, as the integral of the density of states remains finite. In three dimensions, it is straightforward to show

$$D(E) = \frac{\Omega}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E} \tag{24}$$

The singularity is removed, although a discontinuity in the derivative exists as $E \rightarrow 0$.

One can determine the total number of electrons in the system by integrating the density of states up to the highest occupied energy level. The energy of the highest occupied states is called the Fermi level or Fermi Energy, E_f

$$N = \frac{\Omega}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\int_0^{E_f}\sqrt{E}\,dE \tag{25}$$

and

$$E_f = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N}{\Omega}\right)^{2/3} \tag{26}$$

By defining a Fermi wave vector as $k_f = (3\pi^2 n_{el})^{1/3}$, where n_{el} is the electron density $n_{el} = N/\Omega$ of the system, one can write

$$E_f = \frac{\hbar^2 k_f^2}{2m} \tag{27}$$

It should be noted that typical values for E_f for simple metals like sodium or potassium are on the order of several eVs. If one defines a temperature T_f, where $T_f = E_f/k_b$, and k_b is the Boltzmann constant, typical values for T_f might be 10^4 K. At ambient temperatures, one can often neglect the role of temperature in determining the Fermi Energy.

3.2. Hartree–Fock Exchange in a Free-Electron Gas

For a free-electron gas, it is possible to evaluate the Hartree–Fock exchange energy directly [9, 10]. The Slater determinant is constructed using free-electrons orbitals. Each orbital is labeled by a $\vec{k}$ and a spin index. The Coulomb potential for an infinite free-electron gas diverges, but this divergence can be removed by imposing a compensating uniform positive charge. The resulting Hartree–Fock eigenvalues can be written as

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{1}{\Omega}\sum_{k'<k_f}\frac{4\pi e^2}{|\vec{k}-\vec{k}'|^2} \tag{28}$$

where the summation is over occupied $\vec{k}$-states. It is possible to evaluate the summation by transposing the summation to an integration. This transposition is often done for solid-state systems, as the state density is so high that the system can be treated as a continuum

$$\frac{1}{\Omega}\sum_{k'<k_f}\frac{4\pi e^2}{|\vec{k}-\vec{k}'|^2} = \frac{1}{(2\pi)^3}\int_{k'<k_f}\frac{4\pi e^2}{|\vec{k}-\vec{k}'|^2}d^3k \tag{29}$$

This integral can be solved analytically [10]. The resulting eigenvalues are given by

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_f}{\pi}\left[1 + \frac{1-(k/k_f)^2}{2(k/k_f)}\ln\left|\frac{k+k_f}{k-k_f}\right|\right] \tag{30}$$

Using the above expression and Eq. (19), the total electron energy E_{HF}^{FEG} for a free-electron gas within the Hartree–Fock approximation is given by

$$E_{HF}^{FEG} = 2\sum_{k<k_f}\frac{\hbar^2 k^2}{2m} - \frac{e^2 k_f}{\pi}\sum_{k<k_f}\left[1 + \frac{1-(k/k_f)^2}{2(k/k_f)}\ln\left|\frac{k+k_f}{k-k_f}\right|\right] \tag{31}$$

The factor of 2 in the first term comes from spin. In the exchange term, there is no extra factor of 2 because one can subtract off a "double counting term" [see Eq. (19)]. The summations can be executed as per Eq. (29) to yield

$$E_{HF}^{UEG}/N = \frac{3}{5}E_f - \frac{3e^2}{4\pi}k_f \tag{32}$$

The first term corresponds to the average energy per electron in a free-electron gas. The second term corresponds to the exchange energy per electron. The exchange energy is attractive and scales with the cube root of the average density. This form provides a clue as to what form the exchange energy might take in an interacting electron gas or a nonuniform electron gas.

Slater was one of the first to propose replacing V_x^i in Eq. (18) by a term that depends only on the cube root of the charge density [11–13]. By analogy to Eq. (32), he suggested that V_x^i be replaced by

$$V_x^{\text{Slater}}[\rho(\vec{r})] = -\frac{3e^2}{2\pi}[3\pi\rho(\vec{r})]^{1/3} \tag{33}$$

This expression is not orbital dependent. As such, a solution of the Hartree–Fock equation [Eq. (18)] using V_x^{Slater} is much easier to implement. Although Slater exchange was not rigorously justified for nonuniform electron gases, it has been quite successful in replicating the essential features of atomic and molecular systems as determined by Hartree–Fock calculations [11–13].

3.3. Density-Functional Theory

In a number of classic papers, Hohenberg, Kohn, and Sham established a theoretical basis for justifying the replacement of the many-body wave function by one-electron orbitals [8, 14, 15]. In particular, they proposed that the charge density played a central role in describing the electronic structure of matter. A key aspect of their work is the local density approximation or LDA. Within this approximation, one can express the exchange energy as

$$E_x[\rho(\vec{r})] = \int \rho(\vec{r})\mathscr{E}_x[\rho(\vec{r})]d^3r \tag{34}$$

where $\mathscr{E}_x[\rho]$ is the exchange energy per particle of uniform gas at a density of ρ. Within this framework, the exchange potential in Eq. (18) is replaced by a potential determined from the functional derivative of $E_x[\rho]$

$$V_x[\rho] = \frac{\delta E_x[\rho]}{\delta\rho} \tag{35}$$

One serious issue is the determination of the exchange energy per particle $\mathscr{E}_x$ or the corresponding exchange potential V_x. The exact expression for either of these quantities is unknown, save for special cases. If one assumes the exchange energy is given by Eq. (32) (i.e., the Hartree–Fock expression for the exchange energy of the free-electron gas), then one can write

$$E_x[\rho] = -\frac{3e^2}{4\pi}(3\pi^2)^{1/3}\int[\rho(\vec{r})]^{4/3}d^3r \tag{36}$$

and, taking the functional derivative, one obtains

$$V_x[\rho] = -\frac{e^2}{\pi}[3\pi^2\rho(\vec{r})]^{1/3} \tag{37}$$

Comparing this equation to the form chosen by Slater, this form, known as Kohn–Sham exchange V_x^{KS}, differs by a factor of 2/3 (i.e., $V_x^{KS} = 2V_x^{\text{Slater}}/3$). For a number of years, some controversy existed as to whether the Kohn–Sham or Slater exchange was more accurate for realistic systems [8]. Slater suggested that a parameter be introduced that would allow

one to vary the exchange between the Slater and Kohn–Sham values [13]. The parameter, α, was often placed in front of the Slater exchange: $V_{x\alpha} = \alpha V_x^{\text{Slater}}$. The variable α was often chosen to replicate some known feature of a exact Hartree–Fock calculation, such as the total energy of an atom or ion. Acceptable values of α were viewed to range from $\alpha = 2/3$ to $\alpha = 1$. Slater's so-called "X_α" method was very successful in describing molecular systems [13]. Notable drawbacks of the X_α method center on its *ad hoc* nature through the α parameter and the omission of an explicit treatment of correlation energies.

In contemporary theories, α is taken to be 2/3, and correlation energies are explicitly included in the energy functionals [8]. Numerical studies have been performed on uniform electron gases, resulting in local density expressions of the form $V_{xc}[\rho(\vec{r})] = V_x[\rho(\vec{r})] + V_c[\rho(\vec{r})]$, where V_c represents contributions to the total energy beyond the Hartree–Fock limit [16]. It is also possible to describe the role of spin explicitly by considering the charge density for up signs and down spins: $\rho = \rho_\uparrow + \rho_\downarrow$. This approximation is called the local spin density approximation [8].

The Kohn–Sham equation[15] for the electronic structure of matter is given by

$$\left\{\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + V_H(\vec{r}) + V_{xc}[\rho(\vec{r})]\right\}\phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{38}$$

This equation is usually solved "self-consistently." An approximate charge is assumed to estimate the exchange-correlation potential, and this charge is used to determine the Hartree potential from Eq. (16). These approximate potentials are inserted in the Kohn–Sham equation, and the total charge density is determined as in Eq. (14). The "output" charge density is used to construct new exchange-correlation and Hartree potentials. The process is repeated until the input and output charge densities or potentials are identical to within some prescribed tolerance.

Once a solution of the Kohn–Sham equation is obtained, the total energy can be computed from

$$E_{KS} = \sum_i^M E_i - 1/2\int \rho(\vec{r})V_H(\vec{r})d^3r + \int \rho(\vec{r})\{E_{xc}[\rho(\vec{r})] - V_{xc}[\rho(\vec{r})]\}d^3r \tag{39}$$

where E_{xc} is a generalization of Eq. (34), (i.e., the correlation energy density is included). The electronic energy as determined from E_{KS} must be added to the ion–ion interactions to obtain the structural energies. This is a straightforward calculation for confined systems. For extended systems such as crystals, the calculations can be done using Madelung summation techniques [17].

Owing to its ease of implementation and overall accuracy, the LDA is a popular choice for describing the electronic structure of matter. It is relatively easy to implement and surprisingly accurate. Recent developments have included so-called gradient corrections to the local density approximation. In this approach, the exchange-correlation energy depends on the gradient of the density. This approach is called the generalized gradient approximation, or GGA [18].

When first proposed, density-functional theory was not widely accepted within the chemistry community. The theory is not "rigorous" in the sense that it is not clear how to improve the estimates for the ground-state energies. For wave function–based methods, one can include more Slater determinants, as in a configuration interaction approach. As the wave functions improve via the variational theorem, the energy is lowered. The Kohn–Sham equation is variational, but because of its approximate form, it need not approach the true ground-state energy. This is not a problem, provided that one is interested in relative energies and that any inherent density-functional errors cancel in the difference. For example, if the Kohn–Sham energy of an atom is 10% too high and the corresponding energy of the atom in a crystal is also 10% too high, the cohesive energies that involve the difference of the two energies can be better than the nominal 10% error of the absolute energies. An outstanding fundamental issue of using density-functional theory is obtaining an *ab initio* estimate of the cancellation errors.

In some sense, density-functional theory is an *a posteri* theory. Given the transference of the exchange-correlation energies from an electron gas, it is not surprising that errors would arise in its implementation to highly nonuniform electron gas systems, as found in realistic systems. However, the degree of error cancellations is rarely known *a priori*. The reliability of density-functional theory has been established by numerous calculations for a wide variety of condensed matter systems. For example, the cohesive energies, compressibility, structural parameters, and vibrational spectra of elemental solids have been calculated within the density-functional theory [19]. The accuracy of the method is best for systems in which the cancellation of errors is expected to be complete. Because cohesive energies involve the difference in energies between atoms in solids and in free space, error cancellations are expected to be significant. This is reflected by the fact that historically, cohesive energies have presented greater challenges for density-functional theory; the errors between theory and experiment are typically ~5–10%, depending on the nature of the density function. In contrast, vibrational frequencies that involve small structural changes within a given crystalline environment are easily reproduced to within 1–2%.

3.4. Time-Dependent Density Functional Theory

One of the most significant limitations of "conventional" density-functional formalism is its inability to deal with electronic excitations. Within time-independent, or static, density-functional theory, a quantum mechanical system is described through the ground-state electronic charge density. Although this approach can be accurate for the ground-state of a many-electron system, unoccupied electronic states cannot be identified as those belonging to electronic, or quasi-particle, excitations [20, 21]. The inability of "conventional" density-functional theory to describe excitations severely restricts its range of applications, as many important physical properties such as optical absorption and emission are associated with excited states.

Explicit calculations for excited states present enormous challenges for theoretical methods. Accurate calculations for excitation energies and absorption spectra typically require computationally intensive techniques, such as the configuration interaction method [22, 23], quantum Monte Carlo simulations [24–26], or Green's function methods [27–29]. Although these methods describe electronic excitations properly, they are usually limited to very small systems because of high computational demands. An alternative approach is to consider methods based on time-dependent density-functional theory such as those using the time-dependent local density approximation (TDLDA) [20, 21, 30–38].

The TDLDA technique can be viewed as a natural extension of the ground-state density-functional LDA formalism, designed to include the proper representation of excited states. TDLDA excitation energies of a many-electron system are usually computed from conventional, time-independent Kohn–Sham transition energies and wave functions. Compared to other theoretical methods for excited states, the TDLDA technique requires considerably less computational effort. Despite its relative simplicity, the TDLDA method incorporates screening and relevant correlation effects for electronic excitations [20, 21, 30, 31]. In this sense, TDLDA represents a fully *ab initio* formalism for excited states.

As in the case of time-independent density functional theory, the time-dependent formalism reduces the many-electron problem to a set of self-consistent single-particle equations [39, 40]

$$\left\{-\frac{\nabla^2}{2} + v_{\text{eff}}[\rho(\mathbf{r}, t)]\right\}\psi_i(\mathbf{r}, t) = i\frac{\partial}{\partial t}\psi_i(\mathbf{r}, t). \tag{40}$$

In this case, the single-particle wave functions $\psi_i(\mathbf{r}, t)$ and the effective potential $v_{\text{eff}}[\rho(\mathbf{r}, t)]$ explicitly depend on time. The effective potential is given by

$$v_{\text{eff}}[\rho(\mathbf{r}, t)] = \sum_a v_{\text{ion}}(\mathbf{r} - \mathbf{r}_a) + v_H[\rho(\mathbf{r}, t)] + v_{\text{xc}}[\rho(\mathbf{r}, t)] \tag{41}$$

The three terms on the right side of Eq. (41) describe the external ionic potential, the potential Hartree, and the exchange-correlation potential, respectively. In the adiabatic approximation, which is local in time, the exchange-correlation potential and its first derivative can be expressed in terms of the time-independent exchange-correlation energy $E_{xc}[\rho]$

$$v_{xc}[\rho(\mathbf{r}, t)] \cong \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \qquad \frac{\delta v_{xc}[\rho(\mathbf{r}, t)]}{\delta \rho(\mathbf{r}', t')} \cong \delta(t - t') \frac{\delta^2 E_{xc}[\rho]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \tag{42}$$

The LDA makes a separate local approximation (i.e., within the LDA, the exchange-correlation energy density is local in space).

Although LDA in time-dependent density-function theory has proven itself for molecules, clusters, and small quantum dots and small clusters, several questions remain as areas of active research. The application of TDLDA to large, extended systems remains problematic. It is widely accepted that TDLDA as outlined here will approach the LDA results for extended systems and, consequently, suffer the flaws of LDA such as exhibiting band gaps much smaller than experiment [41–43].

Most implementations of time-dependent density-functional theory are based on the LDA or the GGA [44, 45]. However, these approximations are known to have the wrong asymptotic behavior (e.g., the potential does not scale as $1/r$ for large distances). It is widely believed that more accurate TDLDA methods will necessitate other forms of the density function. Examples of such an approach are the asymptotically corrected LDAs introduced by Casida and Salahub [46], and by van Leeuwen and Baerends [47]. These potentials have recently been investigated using the current formalism [48].

The linear response formalism within time-dependent density functional theory (TDDFT) provides a theoretical basis for the TDLDA method. In this section, we illustrate how TDLDA excitation energies and oscillator strengths are derived from single-electron Kohn–Sham eigenvalues and eigen wave functions. A comprehensive analysis of time-dependent density functional response theory can be found elsewhere [20, 21, 30, 31]. The notation in the work by Casida [30, 31] is implemented within this section.

The response of the Kohn–Sham density matrix within TDDFT is obtained by introducing a time-dependent perturbation $\delta v_{appl}(\mathbf{r}, t)$. Because of the self-consistent nature of the Kohn–Sham Hamiltonian, the effective perturbation includes the response of the self-consistent field, $\delta v_{SCF}[\rho(\mathbf{r}, t)]$

$$\delta v_{eff}[\rho(\mathbf{r}, t)] = \delta v_{appl}(\mathbf{r}, t) + \delta v_{SCF}[\rho(\mathbf{r}, t)] \tag{43}$$

where the self-consistent field is given by the last two terms in Eq. (41)

$$v_{SCF}[\rho(\mathbf{r}, t)] = \int \frac{\rho_n(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}[\rho(\mathbf{r}, t)] \tag{44}$$

With the frequency domain, the response of the Kohn–Sham density matrix, $\delta \mathbf{P}(\omega)$ to the perturbation can be derived using a generalized susceptibility $\chi(\omega)$. For quasi-independent Kohn–Sham particles, the sum-over-states representation of the generalized susceptibility is given by

$$\chi_{ij\sigma, kl\tau}(\omega) = \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \frac{\lambda_{lk\tau}}{\omega - \omega_{lk\tau}} \tag{45}$$

where $\lambda_{lk\tau} = n_{l\tau} - n_{k\tau}$ is the difference between the occupation numbers, and $\omega_{lk\tau} = \epsilon_{k\tau} - \epsilon_{l\tau}$ is the difference between the eigenvalues of the lth and kth single-particle states. The susceptibility in Eq. (45) is expressed in the basis of the unperturbed Kohn–Sham orbitals $\{\psi_{i\sigma}\}$, and the indices refer to space (i, j, or k, l) and spin (σ or τ) wave components. The linear response of the density matrix is

$$\delta P_{ij\sigma}(\omega) = \sum_{kl\tau} \chi_{ij\sigma, kl\tau}(\omega) \delta v^{eff}_{kl\tau}(\omega) = \frac{\lambda_{ji\sigma}}{\omega - \omega_{ji\sigma}} \left[\delta v^{appl}_{ij\sigma}(\omega) + \delta v^{SCF}_{ij\sigma}(\omega) \right] \tag{46}$$

Equation (46) is, however, complicated by the fact that $\delta v_{\mathrm{SCF}}(\omega)$ depends on the response of the density matrix

$$\delta v_{ij\sigma}^{\mathrm{SCF}}(\omega) = \sum_{kl\tau} K_{ij\sigma,kl\tau}\delta P_{kl\tau}(\omega) \tag{47}$$

where the coupling matrix $\mathbf{K}$ describes the response of the self-consistent field to changes in the charge density. Within the adiabatic approximation, this matrix is frequency independent. The analytical expression for the adiabatic coupling matrix $K_{ij\sigma,kl\tau} = \partial v_{ij\sigma}^{\mathrm{SCF}}/\partial P_{kl\tau}$ can be derived from Eq. (44) by making use of the functional chain rule

$$K_{ij\sigma,kl\tau} = \iint \psi_{i\sigma}^*(\mathbf{r})\psi_{j\sigma}(\mathbf{r})\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho_\sigma(\mathbf{r})\delta\rho_\tau(\mathbf{r}')}\right)\psi_{k\tau}(\mathbf{r}')\psi_{l\tau}^*(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{48}$$

The functional derivative in Eq. (48) is evaluated with respect to the unperturbed charge densities. By using the coupling matrix, Eq. (46) can be rewritten as

$$\sum_{kl\tau}^{\lambda_{kl\tau}\neq 0}\left(\delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}\frac{\omega-\omega_{lk\tau}}{\lambda_{lk\tau}} - K_{ij\sigma,kl\tau}\right)\delta P_{kl\tau}(\omega) = \delta v_{ij\sigma}^{\mathrm{appl}}(\omega) \tag{49}$$

Because the summation in Eq. (49) is performed over all occupied and unoccupied orbitals, it contains both the particle-hole and hole-particle contributions. These contributions can be written as two separate equations: the particle-hole part of $v_{\mathrm{appl}}(\omega)$ is given by

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0}\left(\delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}\frac{\omega-\omega_{lk\tau}}{\lambda_{lk\tau}} - K_{ij\sigma,kl\tau}\right)\delta P_{kl\tau}(\omega) - \sum_{kl\tau}^{\lambda_{kl\tau}<0} K_{ij\sigma,lk\tau}\delta P_{lk\tau}(\omega) = \delta v_{ij\sigma}^{\mathrm{appl}}(\omega) \tag{50}$$

and the hole-particle part of $v_{\mathrm{appl}}(\omega)$ is

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0}\left(\delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}\frac{\omega-\omega_{kl\tau}}{\lambda_{kl\tau}} - K_{ji\sigma,lk\tau}\right)\delta P_{lk\tau}(\omega) - \sum_{kl\tau}^{\lambda_{kl\tau}<0} K_{ji\sigma,kl\tau}\delta P_{kl\tau}(\omega) = \delta v_{ji\sigma}^{\mathrm{appl}}(\omega) \tag{51}$$

Combining Eqs. (50) and (51), one can separate the real and imaginary parts of the density matrix response $\delta\mathbf{P}(\omega)$. If the basis functions $\{\psi_{i\sigma}\}$ in Eq. (48) are real, the coupling matrix $\mathbf{K}$ is also real and symmetric with respect to the interchange of space indices $i \leftrightarrow j$ and $k \leftrightarrow l$. Because $\delta\mathbf{P}(\omega)$ is hermitian (i.e., $\delta P_{ji\sigma} = \delta P_{ij\sigma}^*$), the real part of $\delta\mathbf{P}(\omega)$ for a real perturbation $v_{\mathrm{appl}}(\omega)$ is given by

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0}\left[\frac{\delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}}{\lambda_{kl\tau}\omega_{kl\tau}}(\omega^2-\omega_{kl\tau}^2) - 2K_{ij\sigma,kl\tau}\right]\Re(\delta P_{kl\tau})(\omega) = \delta v_{ij\sigma}^{\mathrm{appl}}(\omega) \tag{52}$$

where $\Re(\delta P_{ij\sigma})(\omega)$ denotes the Fourier transform of the real part of $\delta P_{ij\sigma}(t)$.

Equation (52) can be used to obtain the density-functional expression for the dynamic polarizability. This is accomplished by introducing a perturbation $\delta\hat{v}_{\mathrm{appl}}(t) = \hat{\gamma}\mathscr{E}_\gamma(t)$, where $\mathscr{E}_\gamma$ is an external electric field applied along the γ-axis, $\gamma = \{x, y, z\}$. The linear response of the dipole moment $\delta\mu(\omega)$ is expressed through of the real part of $\delta\mathbf{P}(\omega)$ as

$$\delta\mu_\beta(\omega) = -2\sum_{ij\sigma}^{\lambda_{ij\sigma}>0}\beta_{ji\sigma}\Re(\delta P_{ij\sigma})(\omega) \qquad \beta = \{x, y, z\} \tag{53}$$

The components of the dynamic polarizability tensor are given by

$$\alpha_{\beta\gamma}(\omega) = \frac{\delta\mu_\beta(\omega)}{\mathscr{E}_\gamma(\omega)} = -\frac{2}{\mathscr{E}_\gamma(\omega)}\sum_{ij\sigma}^{\lambda_{ij\sigma}>0}\beta_{ji\sigma}\Re(\delta P_{ij\sigma})(\omega) \qquad \beta,\gamma = \{x, y, z\} \tag{54}$$

Solving Eq. (52) with respect to $\Re(\delta P_{ij\sigma})(\omega)$ and substituting the result into Eq. (54), one obtains the following matrix equation for the polarizability components

$$\alpha_{\beta\gamma}(\omega) = 2\hat{\boldsymbol{\beta}}\mathbf{R}^{1/2}(\mathbf{Q} - \omega^2\mathbf{1})^{-1}\mathbf{R}^{1/2}\hat{\boldsymbol{\gamma}} \qquad \beta, \gamma = \{x, y, z\} \tag{55}$$

where the matrices $\mathbf{R}$ and $\mathbf{Q}$ are given by

$$R_{ij\sigma,kl\tau} = \delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}\lambda_{kl\tau}\omega_{kl\tau} \tag{56}$$

$$Q_{ij\sigma,kl\tau} = \delta_{i,k}\delta_{j,l}\delta_{\sigma,\tau}\omega_{kl\tau}^2 + 2\sqrt{\lambda_{ij\sigma}\omega_{ij\sigma}}\,K_{ij\sigma,kl\tau}\sqrt{\lambda_{kl\tau}\omega_{kl\tau}} \tag{57}$$

The TDLDA expressions for excitation energies and oscillator strengths can be derived by comparing Eq. (55) with the general sum-over-states formula for the average dynamic polarizability, $\langle\alpha(\omega)\rangle = \mathrm{tr}(\alpha_{\beta\gamma}(\omega))/3 = \sum_l f_l/(\Omega_l^2 - \omega^2)$. The true excitation energies Ω_l, which correspond to the poles of the dynamic polarizability, are obtained from the solution of the eigenvalue problem

$$\mathbf{Q}\,\mathbf{F}_l = \Omega_l^2\,\mathbf{F}_l \tag{58}$$

The oscillator strengths f_l, which correspond to the residues of the dynamic polarizability, are given by

$$f_l = \frac{2}{3}\sum_{\beta=1}^{3}|\mathbf{B}_\beta^l\,\mathbf{R}^{1/2}\,\mathbf{F}_l|^2 \tag{59}$$

where $\mathbf{F}_l$ are the eigenvectors of Eq. (58), $(B_\beta)_{ij} = \int \psi_i^* r_\beta \psi_j d\mathbf{r}$, and $\{r_1, r_2, r_3\} = \{x, y, z\}$.

The adiabatic TDLDA calculations for optical spectra require only the knowledge of the time-independent single-electron Kohn–Sham transition energies and wave functions. The most computationally demanding part in such calculations is the evaluation of the coupling matrix given by Eq. (48). This equation can be split into two parts: $\mathbf{K} = \mathbf{K}^{(I)} + \mathbf{K}^{(II)}$. The first term represents a double integral over $1/|\mathbf{r} - \mathbf{r}'|$. Instead of performing the costly double integration by direct summation, we calculate this term by solving the Poisson equation within the boundary domain. The conjugate-gradient method is employed to solve

$$\nabla^2\Phi_{ij\sigma}(\mathbf{r}) = -4\pi\psi_{i\sigma}(\mathbf{r})\psi_{j\sigma}(\mathbf{r}) \tag{60}$$

The first term in Eq. (48) is calculated as

$$K^{(I)}_{ij\sigma,kl\tau} = \int \Phi_{ij\sigma}(\mathbf{r})\psi_{k\tau}(\mathbf{r})\psi_{l\tau}(\mathbf{r})\,d\mathbf{r} \tag{61}$$

The Poisson equation method provides a considerable speed-up as compared to the direct summation. The second term in Eq. (48) represents a double integral over the functional derivative of the exchange-correlation energy $\delta^2 E_{xc}[\rho]\,/\,\delta\rho_\sigma(\mathbf{r})\delta\rho_\tau(\mathbf{r}')$. Within the local approximation of the exchange-correlation potential, this term is reduced to a single integral

$$K^{(II)}_{ij\sigma,kl\tau} = \int \psi_{i\sigma}(\mathbf{r})\psi_{j\sigma}(\mathbf{r})\frac{\delta^2 E_{xc}[\rho]}{\delta\rho_\sigma(\mathbf{r})\delta\rho_\tau(\mathbf{r})}\psi_{k\tau}(\mathbf{r})\psi_{l\tau}(\mathbf{r})\,d\mathbf{r} \tag{62}$$

where the LDA exchange-correlation energy $E_{xc}[\rho]$ is analogous to Eq. (34).

Eq. (62) requires the evaluation of the second derivatives for the LDA exchange-correlation energy with respect to spin-up and spin-down charge densities. The LDA exchange energy per particle is normally approximated by that of the homogeneous electron gas [49]

$$\epsilon_x[\rho_\sigma(\mathbf{r})] = -\frac{3}{4\pi}[6\pi^2\rho_\sigma(\mathbf{r})]^{1/3} \qquad \sigma = \{\uparrow, \downarrow\} \tag{63}$$

The first derivative of the total exchange energy determines the LDA exchange potential

$$\frac{\delta E_x[\rho]}{\delta\rho_\sigma} = v_x[\rho_\sigma] = -\frac{1}{\pi}\left(6\pi^2\rho_\sigma\right)^{1/3} \qquad \sigma = \{\uparrow, \downarrow\} \tag{64}$$

The second derivatives are

$$\frac{\delta^2 E_x[\rho]}{\delta\rho_\uparrow \delta\rho_\uparrow} = -\left(\frac{2}{9\pi}\right)^{1/3} \rho_\uparrow^{-2/3} \qquad \frac{\delta^2 E_x[\rho]}{\delta\rho_\uparrow \delta\rho_\downarrow} = 0 \tag{65}$$

A parameterized form of the Ceperley–Alder functional [16, 50, 51] can be used for the LDA correlation energy. This functional is based on two different analytical expressions for $r_s < 1$ and $r_s \geq 1$, where $r_s = (3/4\pi\rho)^{1/3}$ is the local Seitz radius and $\rho = \rho_\uparrow + \rho_\downarrow$. One can adjust the parametrization for $r_s \ll 1$ to guarantee a continuous second derivative of the correlation energy. The adjusted interpolation formula for the correlation energy per particle is given by [52]

$$\epsilon_c^{U,P} = \begin{cases} A \ln r_s + B + Cr_s \ln r_s + Dr_s + Xr_s^2 \ln r_s & r_s < 1 \\ \gamma/(1 + \beta_1\sqrt{r_s} + \beta_2 r_s) & r_s \geq 1, \end{cases} \tag{66}$$

where two separate sets of coefficients are used for the polarized spin (P) and unpolarized spin (U) cases. The numerical values of all fitting parameters appearing in Eq. (66) can be found in Ref. [53]. The adjusted interpolation formula for the correlation energy is continuous up to its second derivative, whereas the original Perdew-Zunger parametrization is not [51].

Equations (63) through (66) describe only the cases of the completely polarized and unpolarized spin. For intermediate spin polarizations, the correlation energy can be obtained with a simple interpolation formula

$$\epsilon_c = \epsilon_c^U + \xi(\rho)(\epsilon_c^P - \epsilon_c^U) \tag{67}$$

where

$$\xi(\rho) = \frac{1}{1 - 2^{1/3}}\left(x_\uparrow^{4/3} + x_\downarrow^{4/3} - 2^{-1/3}\right) \qquad x_\uparrow = \frac{\rho_\uparrow}{\rho} \quad x_\downarrow = \frac{\rho_\downarrow}{\rho} \tag{68}$$

The expression for the second derivative of the correlation energy in case of an arbitrary spin polarization can be written as

$$\begin{aligned} \frac{\delta^2 E_c[\rho]}{\delta\rho_\sigma \delta\rho_\tau} = {} & \frac{\delta^2 E_c^U}{\delta\rho^2} + \xi(\rho)\left(\frac{\delta^2 E_c^P}{\delta\rho^2} - \frac{\delta^2 E_c^U}{\delta\rho^2}\right) + \left(\frac{\partial\xi(\rho)}{\partial\rho_\sigma} + \frac{\partial\xi(\rho)}{\partial\rho_\tau}\right)\left(\frac{\delta E_c^P}{\delta\rho} - \frac{\delta E_c^U}{\delta\rho}\right) \\ & + \frac{\partial^2\xi(\rho)}{\partial\rho_\sigma \partial\rho_\tau}\rho(\epsilon_c^P - \epsilon_c^U) \qquad \sigma, \tau = \{\uparrow, \downarrow\} \end{aligned} \tag{69}$$

where the spin polarization function $\xi(\rho)$ and its derivatives are given by

$$\frac{\partial\xi(\rho)}{\partial\rho_\uparrow} = \frac{4}{3\rho(1 - 2^{1/3})}\left(x_\uparrow^{1/3} - x_\uparrow^{4/3} - x_\downarrow^{4/3}\right) \tag{70}$$

$$\frac{\partial^2\xi(\rho)}{\partial\rho_\uparrow \partial\rho_\uparrow} = \frac{4}{9\rho^2(1 - 2^{1/3})}\left[x_\uparrow^{-2/3} - 8x_\uparrow^{1/3} + 7(x_\uparrow^{4/3} + x_\downarrow^{4/3})\right] \tag{71}$$

$$\frac{\partial^2\xi(\rho)}{\partial\rho_\uparrow \partial\rho_\downarrow} = \frac{4}{9\rho^2(1 - 2^{1/3})}\left[7(x_\uparrow^{4/3} + x_\downarrow^{4/3}) - 4(x_\uparrow^{1/3} + x_\downarrow^{1/3})\right] \tag{72}$$

The TDLDA formalism presented in previous sections can be further simplified for systems with the unpolarized spin. In this case, the spin-up and spin-down charge densities are equal, $\rho_\uparrow = \rho_\downarrow$, and Eqs. (68) and (70)–(72) yield

$$\begin{aligned} \xi(\rho) &= 0 \qquad & \frac{\partial^2\xi(\rho)}{\partial\rho_\uparrow \partial\rho_\uparrow} &= \frac{4}{9\rho^2(2^{1/3} - 1)} \\ \frac{\partial\xi(\rho)}{\partial\rho_\uparrow} &= 0 \qquad & \frac{\partial^2\xi(\rho)}{\partial\rho_\uparrow \partial\rho_\downarrow} &= -\frac{4}{9\rho^2(2^{1/3} - 1)} \end{aligned} \tag{73}$$

Because the coordinate parts of spin-up and spin-down Kohn–Sham wave functions for systems with the unpolarized spin are identical, $\psi_{i\uparrow} = \psi_{i\downarrow}$, it follows that $Q_{ij\uparrow,kl'} = Q_{ij\downarrow,kl\downarrow}$ and $Q_{ij\uparrow,kl\downarrow} = Q_{ij\downarrow,kl\uparrow}$. This allows us to separate "singlet" and "triplet" transitions by representing Eq. (58) in the basis set of $\{\mathbf{F}_+, \mathbf{F}_-\}$, chosen as

$$F_{ij}^{\{+,-\}} = \frac{1}{\sqrt{2}}\left(F_{ij\uparrow} \pm F_{ij\downarrow}\right) \tag{74}$$

In this basis, the matrix $\mathbf{Q}$ becomes

$$Q_{ij,kl}^{\{+,-\}} = \delta_{i,k}\delta_{j,l}\omega_{kl}^2 + 2\sqrt{\lambda_{ij}\omega_{ij}}\,K_{ij,kl}^{\{+,-\}}\sqrt{\lambda_{kl}\omega_{kl}} \tag{75}$$

where $K_{ij,kl}^{\{+,-\}} = K_{ij\uparrow,kl\uparrow} \pm K_{ij\uparrow,kl\downarrow}$. The components of $\mathbf{K}^{\{+,-\}}$ in their explicit form are given by

$$K_{ij,kl}^{+} = 2\int\int \frac{\psi_i(\mathbf{r})\psi_j(\mathbf{r})\psi_k(\mathbf{r}')\psi_l(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}' + 2\int \psi_i(\mathbf{r})\psi_j(\mathbf{r})\left(\frac{\delta^2 E_c^U}{\delta\rho^2} - \frac{1}{(9\pi)^{1/3}\rho^{2/3}}\right)\psi_k(\mathbf{r})\psi_l(\mathbf{r})\,d\mathbf{r} \tag{76}$$

$$K_{ij,kl}^{-} = 2\int \psi_i(\mathbf{r})\psi_j(\mathbf{r})\left(\frac{4(\epsilon_c^P - \epsilon_c^U)}{9(2^{1/3}-1)} - \frac{1}{(9\pi)^{1/3}\rho^{2/3}}\right)\psi_k(\mathbf{r})\psi_l(\mathbf{r})\,d\mathbf{r} \tag{77}$$

For most practical applications, only "singlet" transitions represented by the $\mathbf{F}_+$ basis vectors are of interest. Triplet transitions described by the $\mathbf{F}_-$ vectors have zero dipole oscillator strength and do not contribute to optical absorption. By solving Eq. (58) for the $\mathbf{F}_+$ vectors only, one can reduce the dimension of the eigenvalue problem by a factor of two. Eqs. (75) through (77), however, can only be applied to systems with unpolarized spin. In case of an arbitrary spin polarization, the general form of the matrix $\mathbf{Q}$ presented by Eq. (57), with the coupling matrix given by Eq. (48) and the functional derivatives given by Eqs. (64) through (72), must be used.

Other than the adiabatic local density approximation, no other approximations have been made. The exact solution of the matrix equation (58) incorporates all relevant correlations among single-particle transitions.

The frequency-domain approach presented here lends itself naturally to a massively parallel solution within the real-space grid. A practical solution begins with solving the Kohn–Sham equation (Eq. 38). Once the Kohn–Sham equation is solved, the wave functions and eigenvalues can be used to set up the TDLDA equation for the excited states. Forming the coupling matrix [Eq. (48)] is the most computationally intensive step. Fortunately, each $ij\sigma$ combination defines a row in the coupling matrix. As such, it is easy to parallelize a solution to Eq. (60). Once this step is accomplished, one may evaluate the integrals of Eqs. (61) and (62) as simple sums for each element within a matrix row. The evaluation of each matrix row is completely independent of the evaluation of another row, leading to embarrassingly simple parallelization, with no need for communication between processors working on different matrix rows. The matrix is diagonalized once, using any off-the-shelf diagonalization approach. Further implementation details can be found elsewhere [54].

4. PSEUDOPOTENTIALS

The pseudopotential model of a solid has led the way in providing a workable model for computing the electronic properties of materials [55]. For example, it is now possible to predict accurately the properties of complex systems such as quantum dots or semiconductor liquids with hundreds, if not thousands, of atoms.

The pseudopotential model treats matter as a sea of valence electrons moving in a background of ion cores (Fig. 2). The cores are composed of nuclei and inert inner electrons.

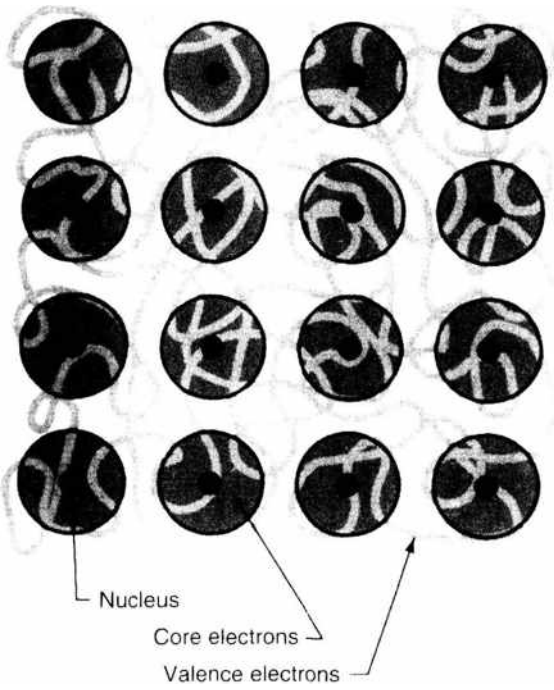


Figure 2. Standard pseudopotential model of a solid. The ion cores composed of the nuclei and tightly bound core electrons are treated as chemically inert. The pseudopotential model describes only the outer, chemically active, valence electrons.

Within this model, many of the complexities of an all-electron calculation are avoided. A group IV solid such as C, with six electrons, is treated in a similar fashion to Pb, with 82 electrons, because both elements have four valence electrons.

Pseudopotential calculations center on the accuracy of the valence electron wave function in the spatial region away from the core (i.e., within the "chemically active" bonding region). The smoothly varying pseudo–wave function is taken to be identical to the appropriate all-electron wave function in the bonding regions. A similar construction was introduced by Fermi in 1934 [56] to account for the shift in the wave functions of high-lying states of alkali atoms subject to perturbations from foreign atoms. In his remarkable paper, Fermi introduced the conceptual basis for both the pseudopotential and the scattering length. In his analysis, Fermi noted that it was not necessary to know the details of the scattering potential. Any number of potentials that reproduced the phase shifts of interest would yield similar scattering events.

A significant advance in the construction of pseudopotentials occurred with the development of density-functional theory [8, 14, 15]. Within density-functional theory, the many-body problem is mapped onto a one-electron Hamiltonian. The effects of exchange and correlation are subsumed into a one-electron potential that depends only on the charge density. This procedure allows for a great simplification of the electronic structure problem. Without this approach, most electronic structure methods would not be feasible for systems of more than a few atoms. The chief limitation of density-functional methods is that they are appropriate only for the ground-state structure and cannot be used to describe excited states without other approximations.

In this section, the procedure for constructing an *ab initio* pseudopotential within density-functional theory will be illustrated. Using the approach of Kohn and Sham [15], one can write down a Hamiltonian corresponding to a one-electron Schrödinger equation. It is not difficult to solve the Kohn–Sham equation [Eq. (87)] for an atom, as the atomic charge density is taken to be spherically symmetric. The problem reduces to solving a one-dimensional equation. The Hartree and exchange-correlation potentials can be iterated to form a self-consistent field. Usually the process is so quick that it can be done on a desktop or laptop computer in a matter of seconds. In three dimensions, as for a complex atomic cluster,

the problem is highly nontrivial. One major difficulty is the range of length scales involved. For example, in the case of a multielectron atom, the most tightly bound core electrons can be confined to within ~0.01 Å, whereas the outer valence electrons may extend over ~1–5 Å. In addition, the nodal structure of the atomic wave functions are difficult to replicate with a simple basis—especially the wave function cusp at the origin where the Coulomb potential diverges. The pseudopotential approximation eliminates this problem and is quite efficacious when combined with density-functional theory. However, it should be noted that the pseudopotential approximation is not dependent on the density-functional theory: Pseudopotentials can be created without resorting to density-functional theory (e.g., pseudopotentials can be created within Hartree–Fock theory).

To illustrate the construction of an *ab initio* pseudopotential, let us consider a sodium atom. Extensions to more complex atoms are straightforward. The starting point of any pseudopotential construction is to solve for the atom within density-functional theory (i.e., the electronic structure problem for the Na atom is solved, including the core and valence electrons). In particular, one extracts the eigenvalue ϵ_{3s} and the corresponding wave function $\psi_{3s}(r)$ for the single-valence electron. Several conditions for the Na pseudopotential are employed: the potential binds only the valence electron: the 3*s*-electron: the eigenvalue of the corresponding valence electron is identical to the full-potential eigenvalue (the "full potential" is also called the "all-electron" potential); and the wave function must be nodeless and identical to the all-electron wave function outside the core region. For example, we construct a pseudo–wave function $\phi_{3s}(r)$ such that $\phi_{3s}(r) = \psi_{3s}(r)$ for $r > r_c$, where r_c defines the size spanned by the ion core (i.e., the nucleus and core electrons). For Na, this means knowing the spatial extent of the size of the ion core, which includes the $1s^2 2s^2 2p^6$ states and the nucleus. Typically, the size of the ion core is taken to be less than the distance corresponding to the maximum of the valence wave function as measured from the nucleus, but greater than the distance of the outermost node from the nucleus. These valence wave functions are depicted in Fig. 3.

The pseudo–wave function $\phi_p(r)$ must be identical to the all-electron wave function $\psi_{AE}(r)$ outside the core: $\phi_p(r) = \psi_{AE}(r)$ for $r > r_c$. This condition will guarantee that the pseudo–wave function possesses identical properties to the all-electron wave function. For $r < r_c$, one may alter the all-electron wave function as desired within certain limitations (e.g., the wave function in this region must be smooth and nodeless).

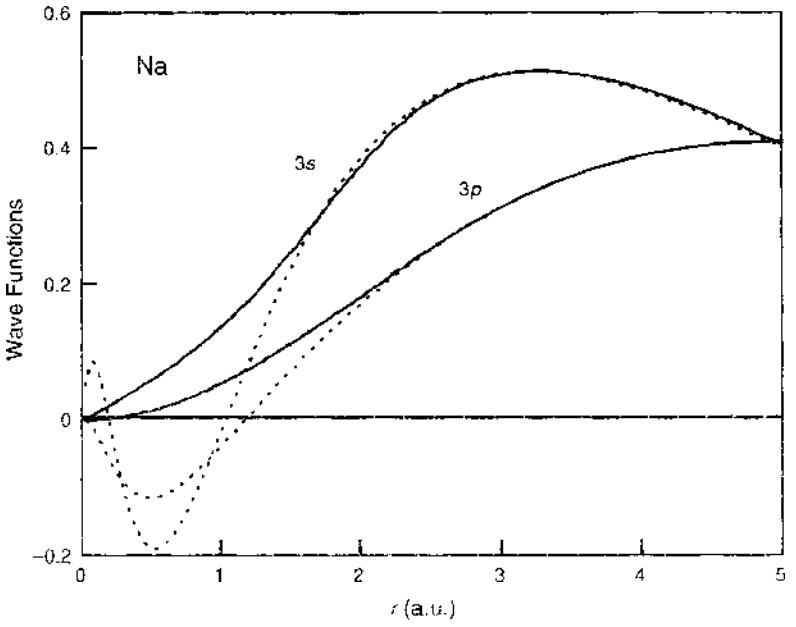


Figure 3. Pseudopotential wave functions compared with all-electron wave functions for the sodium atom. The all-electron wave functions are indicated by the dashed lines.

Another very important criterion is mandated; namely, the integral of the pseudocharge density (i.e., square of the wave function $|\phi_p(r)^2|$) within the core should be equal to the integral of the all-electron charge density. Without this condition, the pseudo–wave function can differ by a scaling factor from the all-electron wave function; that is, $\phi_p(r) = C \times \psi_{AE}(r)$ for $r > r_c$, where the constant C may differ from unity. Because we expect the chemical bonding of an atom to be highly dependent on the tails of the valence wave functions, it is imperative that the normalized pseudo–wave function be identical to the all-electron wave functions. The criterion by which one ensures $C = 1$ is called norm conserving [57]. Some of the earliest *ab initio* potentials did not incorporate this constraint [58]. These potentials are not used for accurate computations. The chemical properties resulting from these calculations using these non–norm conserving pseudopotentials are quite poor when compared to experiments or to the more accurate norm-conserving pseudopotentials [55].

In 1980, Kerker [59] proposed a straightforward method for constructing local density pseudopotentials that retained the norm-conserving criterion. He suggested that the pseudo-wave function have the following form

$$\phi_p(r) = r^l \exp[p(r)] \quad \text{for } r < r_c \tag{78}$$

where $p(r)$ is a simple polynomial $p(r) = -a_0 r^4 - a_1 r^3 - a_2 r^2 - a_3$ and

$$\phi_p(r) = \psi_{AE}(r) \qquad \text{for } r > r_c \tag{79}$$

This form of the pseudo–wave function for ϕ_p assures us that the function will be nodeless and have the correct behavior at large r. Kerker proposed criteria for fixing the parameters (a_0, a_1, a_2, and a_3). One criterion is that the wave function be norm conserving. Other criteria include that the all-electron and pseudo–wave functions have the same valence eigenvalue, that the pseudo–wave function be nodeless and be identical to the all-electron wave function for $r > r_c$; and that the pseudo–wave function must be continuous as well as the first and second derivatives of the wave function at r_c.

Other local density pseudopotentials include those proposed by Hamann, Schlüter, and Chiang [57]; Bachelet, Hamann, and Schlüter [60]; and Greenside and Schlüter [61]. These pseudopotentials were constructed from a different perspective. The all-electron potential was calculated for the free atom. This potential was multiplied by a smooth, short-range cutoff function that removes the strongly attractive and singular part of the potential. The cutoff function is adjusted numerically to yield eigenvalues equal to the all-electron valence eigenvalues and to yield nodeless wave functions converged to the all-electron wave functions outside the core region. Again, the pseudocharge within the core is constrained to be equal to the all-electron value.

As indicated, there is some flexibility in constructing pseudopotentials. Although all local density pseudopotentials impose the condition that

$$\phi_p(r) = \psi_{AE}(r)$$

for $r > r_c$, the construction for $r < r_c$ is not unique. The nonuniqueness of the pseudo–wave function was recognized early in its inception [55]. This attribute can be exploited to optimize the convergence of the pseudopotentials for the basis of interest. Much effort has been made to construct "soft" pseudopotentials. By soft, one means a rapidly convergent calculation using a simple basis such as plane waves. Typically, soft potentials are characterized by a large core radius r_c. As the core radius becomes larger, the convergence between the all-electron and pseudo–wave functions is postponed to larger distances. The quality of the pseudo–wave functions starts to deteriorate, and the transferability of the pseudopotential between the atom and complex environments such as a cluster of atoms becomes limited.

Several schemes have been developed to generate soft pseudopotentials for species that extend effectively the core radius while preserving transferability. The primary motivation for such schemes is to reduce the size of the basis. One of the earliest discussions of such issues is from Vanderbilt [62]. A common measure of pseudopotential softness is to examine the behavior of the potential in reciprocal space. For example, a hard-core pseudopotential

(i.e., one that scales as $1/r^2$ for small r) will decay only as $1/q$ in reciprocal space. This rate of decay is worse than using the bare coulomb potential, which scales as $1/q^2$.

The Kerker pseudopotential [59] does no better than the coulomb potential, as the Kerker pseudopotential has a discontinuity in its third derivative at the origin and at the cutoff radius. This gives rise to a slow $1/q^2$ decay of the potential, although one should examine each case, as the error introduced by truncation of such a potential in reciprocal space may still be acceptable in terms of yielding accurate wave functions and energies. Hamann-Schluter–Chiang [57] potentials often converge better than the Kerker potentials [59], in that they contain no such discontinuities.

An outstanding issue that remains unresolved is the "best" criterion to use in constructing an "optimal" pseudopotential. An optimal pseudopotential is one that minimizes the number of basis functions required to achieve the desired goal; it yields a converged total energy yet does not sacrifice transferability.

One straightforward approach to optimizing a pseudopotential is to build additional constraints into the polynomial given in Eq. (78). For example, suppose we write

$$p(r) = c_0 + \sum_{n=1}^{N} c_n r^n \tag{80}$$

In Kerker's scheme, $N = 4$. However, there is no compelling reason for demanding that the series terminate at this particular point. If we extend the expansion, we may impose additional constraints. For example, we might try to constrain the reciprocal space expansion of the pseudo–wave function so that beyond some momentum cutoff, the function vanishes. A different approach has been suggested by Troullier and Martins [63]. They write Eq. (80) as

$$p(r) = c_0 + \sum_{n=1}^{6} c_{2n} r^{2n} \tag{81}$$

As usual, they constrained the coefficients to be norm conserving. In addition, they demanded continuity of the pseudo–wave functions and the first four derivatives at r_c. The final constraint was to demand zero curvature of the pseudopotential at the origin. These potentials tend to be quite smooth and converge very rapidly in reciprocal space.

Once the pseudo–wave function is defined, as in Eqs. (78) and (79), one can invert the Kohn–Sham equation and solve for the ion core pseudopotential $V_{\text{ion},p}$:

$$V^n_{\text{ion},p}(\vec{r}) = E_n - V_H(\vec{r}) - V_{xc}[\vec{r}, \rho(\vec{r})] + \frac{\hbar^2 \nabla^2 \phi_{p,n}}{2m\phi_{p,n}} \tag{82}$$

This potential, when self-consistently screened by the pseudocharge density

$$\rho(\vec{r}) = -e \sum_{n,\,\text{occup}} |\phi_{p,n}(\vec{r})|^2 \tag{83}$$

will yield an eigenvalue of E_n and a pseudo-wave function $\phi_{p,n}$. The pseudo–wave function, by construction, will agree with the all-electron wave function away from the core.

There are some important issues to consider about the details of this construction. First, the potential is state dependent, as written in Eq. (82) (i.e., the pseudopotential is dependent on the quantum state n). This issue can be handled by recognizing the nonlocality of the pseudopotential. The potential is different for an s-, p-, or d-electron. The nonlocality appears in the angular dependence of the potential, but not in the radial coordinate.

A related issue is whether the potential is highly dependent on the state energy (e.g., if the potential is fixed to replicate the $3s$ state in Na, will it also do well for the $4s$, $5s$, and $6s$ states, etc.? Of course, one could also question how dependent the pseudopotential is on the atomic state used for its construction. For example, would a Na potential be very different for a $3s^1 3p^0$ versus a $3s^{1.5} 3p^{1.5}$ configuration? Finally, how important are loosely bound core states in defining the potential? For example, can one treat the $3d$ states in copper as part of the core or part of the valence shell?

Each of this issues has been carefully addressed in the literature. In most cases, the separation between the core states and the valence states is clear. For example, in Si there is no issue that the core is composed of the $1s^2 2s^2 p^6$ states. However, the core in Cu could be considered to be the $1s^2 2s^2 p^6 3s^2 3p^6 3d^{10}$ configuration, with the valence shell consisting of the $4s^1$ state. Alternatively, one could well consider the core to be the $1s^2 2s^2 p^6 3s^2 3p^6$ configuration, with the valence shell being composed of the $3d^{10} 4s^1$ states. It is quite apparent that solely treating the valence state in Cu as a 4*s* state cannot be correct: otherwise, K and Cu would be chemically similar, as K has the same electron core as Cu, save the 3*d* shell. It is the outer 3*d* shell that distinguishes Cu from K. Such issues are traditionally considered on a case-by-case basis. It is always possible to construct different pseudopotentials, one for each core–valence dichotomy. One can examine the resulting electronic structure for each potential and verify the role of including a questionable state as a valence or core state.

Another aspect of this problem centers on "core–valence" exchange-correlation. In the all electron exchange-correlation potential, the charge density is composed of the core and valence states; in the pseudopotential treatment, only the valence electrons are included. This separation neglects terms that may arise between the overlap of the valence and core states. There are well-defined procedures for including these overlap terms. It is possible to include a fixed-charge density from the core and to allow the valence overlap to be explicitly included. This procedure is referred to as a partial core correction [64]. This correction is especially important for elements such as Zn, Cd, and Hg, in which the outermost filled *d*-shell can contribute to the chemical bonding. Again, the importance of this correction can be tested by performing calculations with and without the partial core. Of course, one might argue that the most accurate approach would be to include any "loosely bound" core states as valence states. This approach is often not computationally feasible or desirable. For example, the Zn core without the 3*d* states results in dealing with an ion core pseudopotential for Zn^{+12}. This results in a very strong pseudopotential, which is required to bind 12 valence electrons. The basis must contain highly localized functions to replicate the *d*-states plus extended states to replicate the *s*-states. Moreover, the number of occupied eigenstates increases by a factor of six. Because most "standard" algorithms to solve for eigenvalues scale superlinearly in time with the number of eigenvalues (such as N_{eig}^2 where N_{eig} is the number of require eigenvalues), this is a serious issue.

With respect to the state dependence of the pseudopotential, these problems can be overcome with little computational effort. Because the core electrons are tightly bound, the ion core potential is highly localized and is not highly sensitive to the ground-state configuration used to compute the pseudopotential. There are well-defined tests for assessing the accuracy of the pseudopotential, especially in terms of the phase shifts [55]. In addition, it should be noted that higher excited states sample the tail of the pseudopotential. This pseudopotential should converge to the all-electron potential outside of the core. A significant source of error here is the LDA. The LDA yields a potential that scales exponentially at large distances, and not as one would expect for an image charge [i.e., intuitively, the true potential should incorporate an image potential such that $V_{xc}(r \to \infty) \to -e^2/r$].

Nonlocality in the pseudopotential is often treated in Fourier space, but it may also be expressed in real space. The interactions between valence electrons and pseudoionic cores may be separated into a local potential and a Kleinman and Bylander [65] form of a nonlocal pseudopotential in real space [63]

$$V_{\text{ion}}^{p}(\vec{r})\phi_n(\vec{r}) = \sum_a V_{\text{loc}}(|\vec{r}_a|)\phi_n(\vec{r}) + \sum_{a,n,lm} G_{n,lm}^{a} u_{lm}(\vec{r}_a)\Delta V_l(r_a) \tag{84}$$

$$K_{n,lm}^{a} = \frac{1}{\langle \Delta V_{lm}^{a} \rangle} \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)\psi_n(\vec{r})\, d^3r \tag{85}$$

and $\langle \Delta V_{lm}^{a} \rangle$ is the normalization factor,

$$\langle \Delta V_{lm}^{a} \rangle = \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)u_{lm}(\vec{r}_a)\, d^3r \tag{86}$$

where $\vec{r}_a = \vec{r} - \vec{R}_a$, and the u_{lm} are the atomic pseudopotential wave functions of angular momentum quantum numbers (l, m), from which the l-dependent ionic pseudopotential $V_l(r)$ is generated. $\Delta V_l(r) = V_l(r) - V_{\text{loc}}(r)$ is the difference between the l component of the ionic pseudopotential and the local ionic potential.

In the case of Na, one might choose the local part of the potential to replicate only the $l = 0$ component as defined by the $3s$ state. The nonlocal parts of the potential would then contain only the $l = 1$ and $l = 2$ components. For simple metals like Na or electronic materials such as Si and GaAs, the angular momentum components for $l = 3$ (or higher) are not significant in the ground-state wave functions. In these systems, one can treat the summation over $l = 0, 1, 2$ to be complete.

The choice of the angular component for the local part of the potential is somewhat arbitrary. It is often convenient to chose the highest l-component present. This avoids the complex projections with the highest l. These issues can be tested by choosing different components for the local potential.

In Fig. 4, the ion core pseudopotential for Na is presented using the Troullier–Martins formalism for creating pseudopotentials. The nonlocality of the potential is evident by the existence of the three potentials corresponding to the s-, p-, and d-states.

5. SOLVING THE EIGENVALUE PROBLEM

Once the pseudopotential has been determined, the resulting eigenvalue problem needs to be solved for the system of interest

$$\left\{\frac{-\hbar^2\nabla^2}{2m} + V^p_{\text{ion}}(\vec{r}) + V_H(\vec{r}) + V_{xc}[\vec{r}, \rho(\vec{r})]\right\}\phi_n(\vec{r}) = E_n\phi_n(\vec{r}) \tag{87}$$

where V^p_{ion} is the ionic pseudopotential for the system. Because the ion cores can be treated as chemically inert and highly localized, it is a simple matter to write

$$V^p_{\text{ion}}(\vec{r}) = \sum_{\vec{R}_a} V^p_{\text{ion},a}(\vec{r} - \vec{R}_a) \tag{88}$$

where $V^p_{\text{ion},a}$ is the ion core pseudopotential associated with the atom a at a position $\vec{R}_a$.

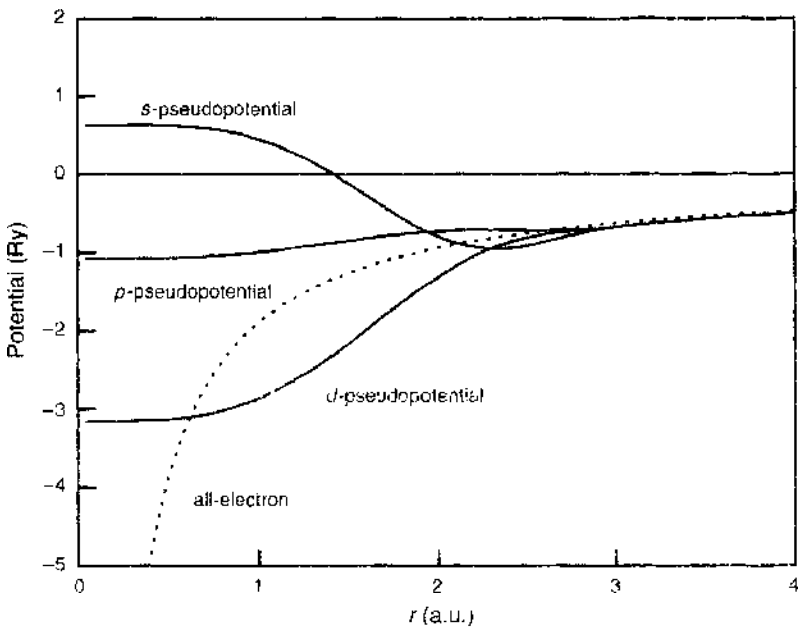


Figure 4. Pseudopotential compared to the all-electron potential for the sodium atom.

A major difficulty in solving the eigenvalue problem in Eq. (87) is the length and energy scales involved. The inner (core) electrons are highly localized and tightly bound compared to the outer (valence electrons). A simple basis function approach is frequently ineffectual. For example, a plane wave basis might require 10^6 waves to represent converged wave functions for a core electron, whereas only 10^2 waves are required for a valence electron [55]. The pseudopotential overcomes this problem by removing the core states from the problem and replacing the all-electron potential by one that replicates only the chemically active, valence electron states [55]. By construction, the pseudopotential reproduces the valence state properties such as the eigenvalue spectrum and the charge density outside the ion core.

Because the pseudopotential is weak, simple basis sets such as a plane wave basis can be quite effective for crystalline matter. For example, in the case of crystalline silicon, only 50–100 plane waves need to be used. The resulting matrix representation of the Schrödinger operator is dense in Fourier (plane wave) space, but it is not formed explicitly. Instead, matrix–vector product operations are performed with the help of fast Fourier transforms. This plane wave approach is akin to spectral techniques used in solving certain types of partial differential equations [66]. The plane wave method uses a basis of the form

$$\psi_{\vec{k}}(\vec{r}) = \sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp[i(\vec{k} + \vec{G}) \cdot \vec{r}] \tag{89}$$

where $\vec{k}$ is the wave vector, $\vec{G}$ is a reciprocal lattice vector, and $\alpha(\vec{k}, \vec{G})$ represent the coefficients of the basis. In a plane wave basis, the Laplacian term of the Hamiltonian is represented by a diagonal matrix. The potential term V_{ion}^{ps} gives rise to a dense matrix.

In real space, it is trivial to operate with the potential term that is represented by a diagonal matrix, and in Fourier space it is trivial to operate with the Laplacian term, which is also represented by a diagonal matrix. The use of plane wave bases also leads to natural preconditioning techniques that are obtained by simply employing a matrix obtained from a smaller plane wave basis, neglecting the effect of high-frequency terms on the potential. For periodic systems, where $\vec{k}$ is a good quantum number, the plane wave basis coupled with pseudopotentials is quite effective. However, for nonperiodic systems such as clusters, liquids, or glasses, the plane wave basis must be combined with a supercell method [55]. The supercell repeats the localized configuration to impose periodicity to the system. This preserves the validity of $\vec{k}$ and Bloch's theorem, which Eq. (89) obeys. There is a parallel to be made with spectral methods that are quite effective for simple periodic geometries, but that lose their superiority when more generality is required. In addition to these difficulties, the two fast Fourier transforms performed at each iteration can be costly, requiring $n \log n$ operations, where n is the number of plane waves, versus $O(N)$ for real space methods, where N is the number of grid points. Usually, the matrix size $N \times N$ is larger than $n \times n$ but only within a constant factor. This is exacerbated in high-performance environments where fast Fourier transforms require an excessive amount of communication and are particularly difficult to implement efficiently.

Another popular basis employed with pseudopotentials include Gaussian orbitals [67–70]. Gaussian bases have the advantage of yielding analytical matrix elements, provided the potentials are also expanded in Gaussians. However, the implementation of a Gaussian basis is not as straightforward as with plane waves. For example, numerous indices must be employed to label the state, the atomic site, and the Gaussian orbitals employed. On the positive side, a Gaussian basis yields much smaller matrices and requires less memory than plane wave methods. For this reason, Gaussians are especially useful for describing transition metal systems.

An alternative approach is to avoid the use of a explicit basis. For example, one can use a real space method that avoids the use of plane waves and fast Fourier transforms altogether. This approach has become popular, and different versions of this general approach have been implemented [66, 71–82].

A real space approach overcomes many of the complications involved with nonperiodic systems, and although the resulting matrices can be larger than with plane waves, they are

sparse, and the methods are easier to parallelize. Even on sequential machines, real space methods can be an order of magnitude faster than the traditional approach.

Real space algorithms avoid the use of fast Fourier transforms by performing all calculations in real physical space instead of Fourier space. Fast Fourier transforms require global communications; as such, they are not efficient for implementation on multiprocessor platforms. The only global operation remaining as real space approaches is that of the inner products. These inner products are required when forming the orthogonal basis used in the generalized Davidson procedure, as discussed below.

The simplest real space method uses finite difference discretization on a cubic grid. A key aspect to the success of the finite difference method is the availability of higher-order finite difference expansions for the kinetic energy operator (i.e., expansions of the Laplacian [83]. Higher-order finite-difference methods significantly improve convergence of the eigenvalue problem when compared with standard finite-difference methods. If one imposes a simple, uniform grid on our system where the points are described in a finite domain by (x_i, y_j, z_k), we approximate $\frac{\partial^2\psi}{\partial x^2}$ at (x_i, y_j, z_k) by

$$\frac{\partial^2\psi}{\partial x^2} = \sum_{n=-M}^{M} C_n \psi(x_i + nh, y_j, z_k) + O(h^{2M+2}) \tag{90}$$

where h is the grid spacing and M is a positive integer. This approximation is accurate to $O(h^{2M+2})$ on the assumption that ψ can be approximated accurately by a power series in x. Algorithms are available to compute the coefficients C_n for arbitrary order in h [83].

With the kinetic energy operator expanded, as in Eq. (90), one can set up a one-electron Schrödinger equation over a grid. One may assume a uniform grid, but this is not a necessary requirement. $\psi(x_i, y_j, z_k)$ is computed on the grid by solving the eigenvalue problem

$$\begin{aligned} -\frac{h^2}{2m}\left[\sum_{n_1=-M}^{M} C_{n_1}\psi_n(x_i + n_1 h, y_j, z_k) + \sum_{n_2=-M}^{M} C_{n_2}\psi_n(x_i, y_j + n_2 h, z_k) + \sum_{n_3=-M}^{M} C_{n_3}\psi_n(x_i, y_j, z_k + n_3 h)\right] \\ + \left[V_{\text{ion}}(x_i, y_j, z_k) + V_H(x_i, y_j, z_k) + V_{xc}(x_i, y_j, z_k)\right]\psi_n(x_i, y_j, z_k) \\ = E_n \psi_n(x_i, y_j, z_k) \end{aligned} \tag{91}$$

If we have L grid points, the size of the full matrix resulting from the above problem is $L \times L$.

The grid we use is based on points uniformly spaced in a three-dimensional cube, as shown in Fig. 5, with each grid point corresponding to a row in the matrix. However, many points in the cube are far from any atoms in the system, and the wave function on these points may be replaced by zero. Special data structures may be used to discard these points and to keep only those having a nonzero value for the wave function. The size of the Hamiltonian matrix is usually reduced by a factor of two to three with this strategy, which is quite important considering the large number of eigenvectors that must be saved. Further, because the Laplacian can be represented by a simple stencil, and as all local potentials sum up to a simple diagonal matrix, the Hamiltonian need not be stored. Handling the ionic pseudopotential is complex, as it consists of a local and a nonlocal term [Eqs. (84) and (85)]. In the discrete form, the nonlocal term becomes a sum over all atoms a and quantum numbers (l, m) of rank-one updates

$$V_{\text{ion}} = \sum_a V_{\text{loc},a} + \sum_{a,l,m} c_{a,l,m} U_{a,l,m} U^T_{a,l,m} \tag{92}$$

where $U_{a,l,m}$ are sparse vectors that are only nonzero in a localized region around each atom; $c_{a,l,m}$ is a normalization coefficient.

There are several difficulties in solving the eigen problems, in addition to the size of the matrices. First, the number of required eigenvectors is proportional to the atoms in the system and can grow up to thousands of vectors, if not more. In addition to storage, maintaining the orthogonality of these vectors can be a formidable task. Second, the relative separation of the eigenvalues becomes increasingly poor as the matrix size increases, and this

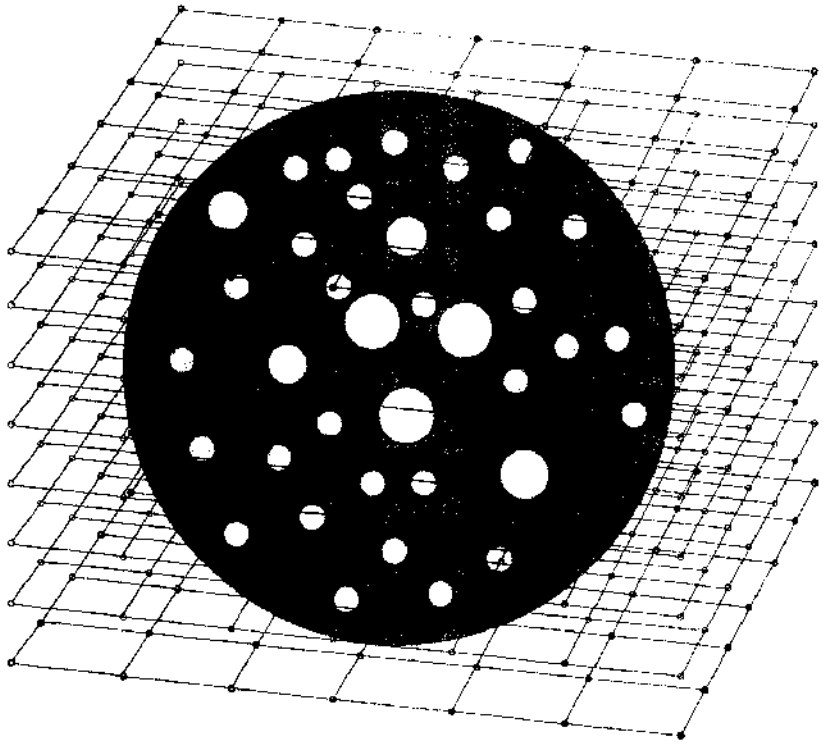


Figure 5. Uniform grid illustrating a typical configuration for examining the electronic structure of a localized system. The dark gray sphere represents the domain where the wave functions are allowed to be nonzero. The light spheres within the domain are atoms.

has an adverse effect on the rate of convergence of the eigenvalue solvers. Preconditioning techniques attempt to alleviate this problem. On the positive side, the matrix need not be stored, as was mentioned earlier, and this reduces storage requirement. In addition, good initial eigenvector estimates are available at each iteration from the previous self-consistency loop.

A popular form of extracting the eigenpairs is based on the generalized Davidson [84] method, in which the preconditioner is not restricted to be a diagonal matrix, as in the Davidson method. (A detailed description can be found in Ref. [85].) Preconditioning techniques in this approach are typically based on filtering ideas and on the fact that the Laplacian is an elliptic operator [86]. The eigenvectors corresponding to the few lowest eigenvalues of ∇^2 are smooth functions and so are the corresponding wave functions. When an approximate eigenvector is known at the points of the grid, a smoother eigenvector can be obtained by averaging the value at every point with the values of its neighboring points. Assuming a cartesian (x, y, z) coordinate system, the low-frequency filter acting on the value of the wave function at the point (i, j, k), which represents one element of the eigenvector, is described by:

$$\left(\frac{\psi_{i-1,j,k}+\psi_{i,j-1,k}+\psi_{i,j,k-1}+\psi_{i+1,j,k}+\psi_{i,j+1,k}+\psi_{i,j,k-1}}{12}\right)+\frac{\psi_{i,j,k}}{2}\rightarrow(\psi_{i,j,k})_{\text{Filtered}} \quad (93)$$

It is worth mentioning that other preconditioners that have been tried have resulted in mixed success. The use of shift-and-invert [87] involves solving linear systems with $A-\sigma I$, where A is the original matrix and the shift σ is close to the desired eigenvalue. These methods would be prohibitively expensive in most situations, given the size of the matrix and the number of times that $A-\sigma I$ must be factored. Alternatives based on an approximate factorization such as ILUT [88] are ineffective beyond the first few eigenvalues. Methods based on approximate inverse techniques have been somewhat more successful, performing better than filtering at additional preprocessing and storage cost. Preconditioning "interior" eigenvalues (i.e., eigenvalues located well inside the interval containing the spectrum) is still a very hard problem. Current solutions only attempt to dampen the effect of eigenvalues

that are far away from the ones being computed. This is, in effect, what is achieved by filtering and sparse approximate inverse preconditioning. These techniques do not reduce the number of steps required for convergence in the same way that shift-and-invert techniques do. However, filtering techniques are inexpensive to apply and result in fairly substantial savings in iterations.

6. PROPERTIES OF CONFINED SYSTEMS: CLUSTERS

The electronic and structural properties of atomic clusters stand as one of the outstanding problems in materials physics. Clusters possess properties that are characteristic of neither the atomic nor solid state. For example, the energy levels in atoms may be discrete and well separated in energy relative to kT. In contrast, solids have continuum of states (energy bands). Clusters may reside between these limits (i.e., the energy levels may be discrete, but with a separation much less than kT).

Real space methods are ideally suited for investigating these systems. In contrast to plane wave methods, real space methods can examine non-periodic systems without introducing artifacts such as supercells. In addition, one can easily examine charged clusters. In supercell configurations, unless a compensating background charge is added, the Coulomb energy diverges for charged clusters.

A closely related issue concerns electronic excitations. In periodic systems, it is nontrivial to consider localized excitation (e.g., with band theory, exciting an atom in one cell excites all atoms in all the equivalent cells). Density functional formalisms often avoid these issues by considering localized or nonperiodic systems.

6.1. Structure

Perhaps the most fundamental issue in dealing with clusters is determining their structure. Before any accurate theoretical calculations can be performed for a cluster, the atomic geometry of a system must be defined. However, this can be a formidable exercise. Serious problems arise from the existence of multiple local minima in the potential-energy-surface of these systems; many similar structures can exist with vanishingly small energy differences.

A convenient method for determining the structure of small or moderate sized clusters is simulated annealing [89]. Within this technique, atoms are randomly placed within a large cell and allowed to interact at a high (usually fictive) temperature. Within this temperature regime, atoms will sample a large number of configurations. As the system is cooled, the number of high-energy configurations sampled is restricted. If the anneal is done slowly enough, the procedure should quench out structural candidates for the ground-state structures.

Langevin molecular dynamics is well suited for simulated-annealing methods. In Langevin dynamics, the ionic positions $\mathbf{R}_j$ evolve according to $\ddot{\mathbf{R}}$

$$M_j\ddot{\mathbf{R}}_j = \mathbf{F}(\{\mathbf{R}_j\}) - \gamma M_j\dot{\mathbf{R}}_j + \mathbf{G}_j \tag{94}$$

where $\mathbf{F}(\{\mathbf{R}_j\})$ is the interatomic force on the jth particle, and $\{M_j\}$ are the ionic masses. The last two terms on the right-hand side of Eq. (94) are the dissipation and fluctuation forces, respectively. The dissipative forces are defined by the friction coefficient γ. The fluctuation forces are defined by random Gaussian variables $\{\mathbf{G}_j\}$ with a white noise spectrum

$$\langle G_j^\alpha(t)\rangle = 0 \quad \text{and} \quad \langle G_j^\alpha(t)G_j^\alpha(t')\rangle = 2\gamma M_j k_B T\delta_{\alpha\beta}\delta(t-t') \tag{95}$$

The angular brackets denote ensemble or time averages, and α stands for the Cartesian component. The coefficient of T on the right-hand side of Eq. (95) ensures that the fluctuation–dissipation theorem is obeyed [i.e., the work done on the system is dissipated by the viscous medium ([90, 91])]. The interatomic forces can be obtained from the Hellmann-Feynman theorem, using the pseudopotential wave functions.

Langevin simulations using quantum forces can be contrasted with other techniques such as the Car-Parrinello method [92, 93]. Langevin simulations as outlined above do not

employ fictitious electron dynamics; at each time step, the system is quenched to the Born–Oppenheimer surface and the quantum forces are determined. This approach requires a fully self-consistent treatment of the electronic structure problem; however, because the interatomic forces are true quantum forces, the resulting molecular dynamics simulation can be performed with much larger time steps. Typically, it is possible to use steps an order of magnitude larger than in the Car–Parrinello method [94]. It should be emphasized that neither of these methods is particularly efficacious without the pseudopotential approximation.

To illustrate the simulated annealing procedure, we consider a silicon cluster of seven atoms. With respect to the technical details for this example, the initial temperature of the simulation was taken to be about 3000 K; the final temperature was taken to be 300 K. The annealing schedule lowered the temperature 500 K each 50 time steps. The time step was taken to be 5 fs. The friction coefficient in the Langevin equation was taken to be 6×10^{-4} a.u. (a.u. is defined by $\hbar = m = e = 1$. The unit of energy is the hartree [1 hartree = 27.2 eV]; the unit of length is the bohr radius [1 bohr = 0.529 Å]). After the clusters reached a temperature of 300 K, they were quenched to 0 K. The ground-state structure was found through a direct minimization by a steepest descent procedure.

Choosing an initial atomic configuration for the simulation takes some care. If the atoms are too far apart, they will exhibit Brownian motion, which is appropriate for Langevin dynamics with the interatomic forces zeroed. In this case, the atoms may not form a stable cluster as the simulation proceeds. Conversely, if the atoms are too close together, they may form a metastable cluster from which the ground state may be kinetically inaccessible even at the initial high temperature. Often the initial cluster is formed by a random placement of the atoms, with a constraint that any given atom must reside within 1.05 and 1.3 times the bond length from at least one atom, where the bond length is defined by the crystalline environment. The cluster in question is placed in a spherical domain. Outside of this domain, the wave function is required to vanish. The radius of the sphere is such that the outmost atom is at least 6 a.u. from the boundary. Initially, the grid spacing was 0.8 a.u. For the final quench to a ground-state structure, the grid spacing was reduced to 0.5 a.u. As a rough estimate, one can compare this grid spacing with a plane wave cutoff of $(\pi/h)^2$ or about 40 Ry for $h = 0.5$ a.u.

In Fig. 6, we illustrate the simulated anneal for the Si_7 cluster. The initial cluster contains several incipient bonds, but the structure is far removed from the ground state by approximately 1 eV/atom. In this simulation, at about 100 time steps, a tetramer and a trimer form. These units come together and precipitate a large drop in the binding energy. After another ~100 time steps, the ground-state structure is essentially formed. The ground state of Si_7 is a bicapped pentagon, as is the corresponding structure for the Ge_7 cluster. The binding energy shown in Fig. 6 is relative to that of an isolated Si atom. Gradient corrections [18, 45]

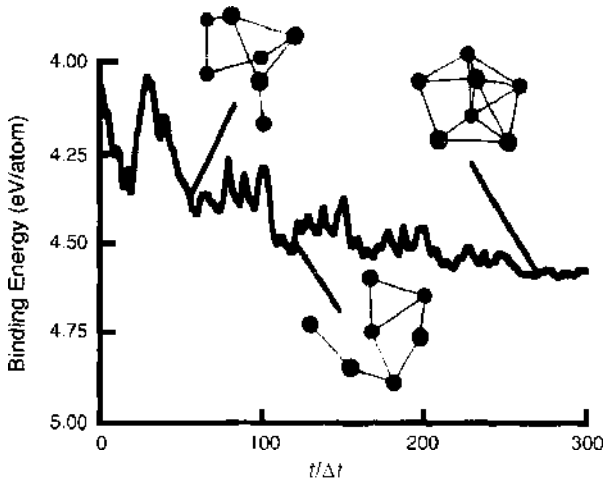


Figure 6. Binding energy of Si_7 during a Langevin simulation. The initial temperature is 3000 K; the final temperature is 300 K. Bonds are drawn for interatomic distances of less than 2.5 Å. The time step is 5 fs.

or spin polarization [95] have not been included in this example. Therefore, the binding energies indicated in the figure are likely to be overestimated by ~20% or so.

In Fig. 7, the ground-state structures for Si_n are presented for $n \leq 7$. The structures for Ge_n are very similar to those of Si_n. The primary difference resides in the bond lengths. The Si bond length in the crystal is 2.35 Å, whereas in Ge the bond length is 2.44 Å. This difference is reflected in the bond lengths for the corresponding clusters; Si_n bond lengths are typically a few percent shorter than the corresponding Ge_n clusters.

It should be emphasized that this annealing simulation is an optimization procedure. As such, other optimization procedures may be used to extract the minimum-energy structures. Recently, a genetic algorithm has been used to examine carbon clusters [99]. In this algorithm, an initial set of clusters is "mated" with the lowest-energy offspring "surviving." By examining several thousand generations, it is possible to extract a reasonable structure for the ground state. The genetic algorithm has some advantages over a simulated anneal, especially for clusters that contain more than ~20 atoms. One of these advantages is that kinetic barriers are more easily overcome. However, the implementation of the genetic algorithm is more involved than an annealing simulation (e.g., in some cases, "mutations," or *ad hoc* structural rearrangements, must be introduced to obtain the correct ground state [99]).

6.2. Photoemission Spectra

A very useful probe of condensed matter involves the photoemission process. Incident photons are used to eject electrons from a solid. If the energy and spatial distributions of the electrons are known, then information can be obtained about the electronic structure of the materials of interest. For crystalline matter, the photoemission spectra can be related to the electronic density of states. For confined systems, the interpretation is not as straightforward. One of the earliest experiments performed to examine the electronic structures of small semiconductor clusters examined negatively charged Si_n and Ge_n ($n \leq 12$) clusters [100]. The photoemission spectra obtained in this work were used to gauge the energy gap between the highest occupied state and the lowest unoccupied state. Large gaps were assigned to the "magic number" clusters, whereas other clusters appeared to have vanishing gaps. Unfortunately, the first theoretical estimates [101] for these gaps showed substantial disagreements with the measured values. It was proposed by Cheshnovsky et al. [10] that sophisticated calculations including transition cross sections and final states were necessary to identify the cluster geometry from the photoemission data. The data were first interpreted in terms of the gaps obtained for neutral clusters; it was later demonstrated that atomic relaxations within the charged cluster are important in analyzing the photoemission data

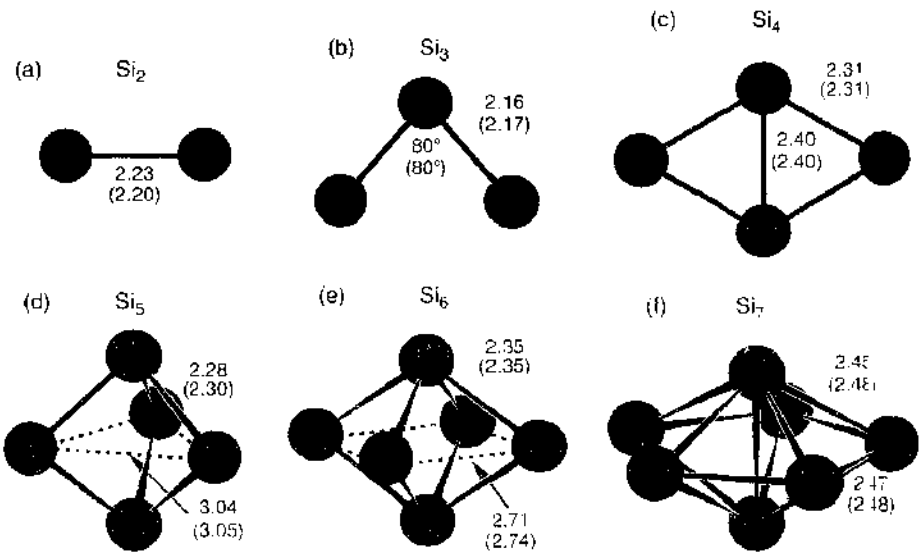


Figure 7. Ground-state geometries and some low-energy isomers of Si_n ($n \leq 7$) clusters. Interatomic distances are in Angstroms. The values in parentheses are from Ref. [96–98]

[102, 103]. In particular, atomic relaxations as a result of charging may change dramatically the electronic spectra of certain clusters. These charge-induced changes in the gap were found to yield very good agreement with the experiment.

The photoemission spectrum of Ge_{10} illustrates some of the key issues. Unlike Si_{10}^-, the experimental spectrum for Ge_{10} does not exhibit a gap. Cheshnovsky et al. interpreted this to mean that Ge_{10}^- does not exist in the same structure as Si_{10}. This is a strange result. Si and Ge are chemically similar, and the calculated structures for both neutral structures are similar. The lowest-energy structure for both 10-atom clusters is the tetracapped trigonal prism (labeled [*I*] in Fig. 8). The photoemission spectra for these clusters can be simulated by using Langevin dynamics. The clusters are immersed in a fictive heat bath and subjected to stochastic forces. If one maintains the temperature of the heat bath and averages over the eigenvalue spectra, a density of states for the cluster can be obtained. The heat bath resembles a buffer gas, as in the experimental setup, but the time intervals for collisions are not similar to the true collision processes in the atomic beam. The simulated photoemission spectrum for Si_{10}^- is in very good agreement with the experimental results, reproducing both the threshold peak and other features in the spectrum. If a simulation is repeated for Ge_{10}^-, using the tetracapped trigonal prism structure, the resulting photoemission spectrum is *not* in good agreement with the experiment. Moreover, the calculated electron affinity is 2.0 eV, in contrast to the experimental value of 2.6 eV. In addition, there is no reason to believe that the tetracapped trigonal prism structure is correct for Ge_{10} when charged. In fact, we find that the bicapped antiprism structure is lower in energy for Ge_{10}. The resulting spectra, (using both structures [I] and [II] in Fig. 8), are presented in Fig. 9 and are compared to the photoemission experiment. The calculated spectrum using the bicapped antiprism structure is in very good agreement with the photoemission. The presence of a gap is indicated by a small peak removed from the density of states (Fig. 9a). This feature is absent in the bicapped antiprism structure (Fig. 9b) and is consistent with experiment. For Ge_{10}, charging the structure reverses the relative stability of the two structures. This accounts for the major differences between the photoemission spectra.

It is difficult to assign a physical origin to a particular structure, owing to the smaller energy differences involved. However, the bicapped structure has a higher coordination. Most chemical theories of bonding indicate that Ge is more metallic than Si and, as such, would prefer a more highly coordinated structure. The addition of an extra electron may induce a structure reflecting the metallic characteristic of Ge relative to Si.

6.3. Vibrational Modes

Experiments on the vibrational spectra of clusters can provide us with very important information about their physical properties. Recently, Raman experiments have been performed on clusters that have been deposited on inert substrates [104]. Because different structural configurations of a given cluster can possess different vibrational spectra, it is possible to compare the vibrational modes calculated for a particular structure with the Raman experiment in a manner similar to the previous example with photoemission. Good agreement

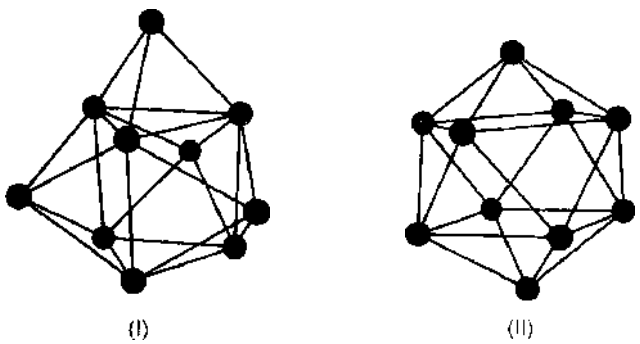


Figure 8. Two possible isomers for Si_{10} or Ge_{10} clusters. (I) is a tricapped trigonal prism cluster, and (II) is a bicapped antiprism cluster.

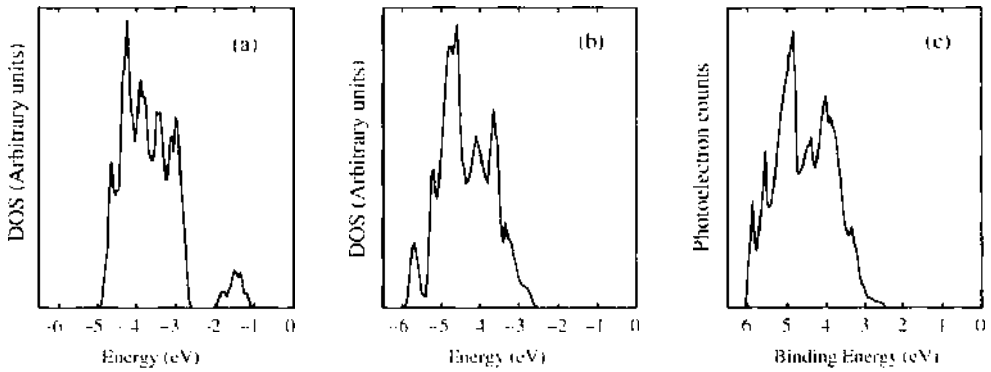


Figure 9. Calculated density of states for Ge_{10} in the tetracapped trigonal prism structure (a) and the bicapped antiprism structure (b). (c) Experimental photoemission spectra from Ref. [100].

between the predicted vibrational spectra and the measured spectra is a necessary condition for the validity of the predicted structure of the cluster.

There are two common approaches for determining the vibrational spectra of clusters. One approach is to calculate the dynamical matrix for the ground state structure of the cluster

$$M_{i\alpha,j\beta} = \frac{1}{m}\frac{\partial^2 E}{\partial R_i^\alpha \partial R_j^\beta} = -\frac{1}{m}\frac{\partial F_i^\beta}{\partial R_j^\alpha} \tag{96}$$

where m is the mass of the atom, E is the total energy of the system, F_i^α is the force on atom i in the direction α, and R_i^α is the α component of coordinate for atom i. One can calculate the dynamic matrix elements by numerically calculating the first-order derivative of force versus atom displacement. From the eigenvalues and eigenmodes of the dynamical matrix, one can obtain the vibrational frequencies and modes for the cluster of interest [105].

The other approach to determine the vibrational modes is to perform a molecular dynamics simulation. The cluster in question is excited by small, random displacements. By recording the kinetic (or binding) energy of the cluster as a function of the simulation time, it is possible to extract the power spectrum of the cluster and determine the vibrational modes. This approach has an advantage for large clusters in that one never explicitly has to do a mode analysis. Another advantage is that anharmonic mode couplings can be examined. It has the disadvantage of the simulation needing to be performed over a long time to extract accurate values for all the modes.

As an example, consider the vibrational modes for a small silicon cluster: Si_4. The starting geometry was taken to be a planar structure for this cluster, as established from a simulated annealing calculation [105].

Determining the dynamical matrix and eigenmodes for this cluster is straightforward process. In Fig. 10, the fundamental vibrational modes are illustrated, and in Table 1, the frequency of these modes are presented. One can also determine the modes via a simulation. To initiate the simulation, one can perform a Langevin simulation [102] with a fixed temperature at 300 K. After a few dozen time steps, the Langevin simulation is turned off and the simulation proceeds, following Newtonian dynamics with "quantum" forces. This procedure allows a stochastic element to be introduced and establishes initial conditions for the simulation without bias toward a particular mode. For this example, time step in the molecular dynamics simulation was taken to be 3.7 fs. The simulation was allowed to proceed for 1000 time steps, or roughly 4 ps. The variation of the kinetic and binding energies is given in Fig. 11 as a function of the simulation time. Although some fluctuations of the total energy occur, these fluctuations are relatively small (i.e., less than $\sim$1 meV) and there is no noticeable drift of the total energy. Such fluctuations arise, in part, because of discretization errors. As the grid size is reduced, such errors are minimized [105]. Similar errors can occur in plane wave descriptions using supercells (i.e., the artificial periodicity of the supercell can introduce erroneous forces on the cluster). By taking the power spectrum of either the kinetic energy (KE) or binding energy (BE) over this simulation time, the vibrational modes

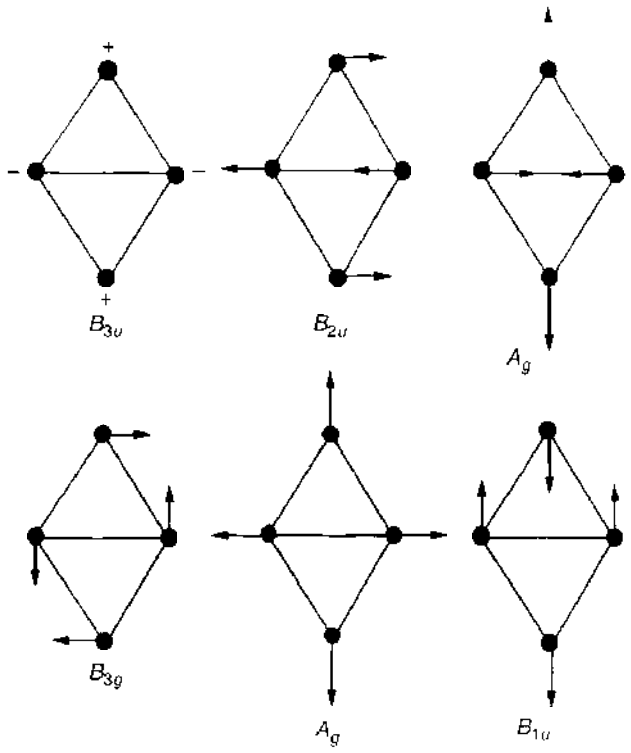


Figure 10. Normal modes for a Si_4 cluster. The + and − signs indicate motion in and out of the plane, respectively.

can be determined. These modes can be identified with the observed peaks in the power spectrum, as illustrated in Fig. 12.

A comparison of the calculated vibrational modes from the molecular dynamics simulation and from a dynamic matrix calculation is listed in Table 1. Overall, the agreement between the simulation and the dynamical matrix analysis is quite satisfactory. In particular, the softest mode (i.e., the B_{3u} mode, and the splitting between the (A_g, B_{1u}) modes are well replicated in the power spectrum. The splitting of the (A_g, B_{1u}) modes is less than 10 cm^{-1}, or about 1 meV, which is at the resolution limit of any *ab initio* method.

The theoretical values are also compared to experimental values. The predicted frequencies for the two A_g modes are surprisingly close to Raman experiments on silicon clusters [104]. The other allowed Raman line of mode B_{3g} is expected to have a lower intensity and has not been observed experimentally [104].

The theoretical modes using the formalism outlined here are in good accord (except the lowest mode) with other theoretical calculations given in Table 1: an LCAO calculation [106] and a Hartree–Fock calculation [96]. The calculated frequency of the lowest mode (i.e., the B_{3u} mode) is problematic. The general agreement of the B_{3u} mode as calculated by the simulation and from the dynamical matrix is reassuring. Moreover, the real space calculations agree with the Hartree–Fock value to within ~20–30 cm^{-1}. However, the LCAO method yields a value that is 50–70% smaller than either the real space or Hartree-Fock calculations. The origin of this difference is not apparent. For a poorly converged basis, vibrational

Table 1. Calculated and experimental vibrational frequencies in a Si_4 cluster.

	B_{3u}	B_{2u}	A_g	B_{3g}	A_g	B_{1u}
Experiment [104]			345		470	
Dynamical matrix [105]	160	280	340	460	480	500
Molecular dynamics simulation [105]	150	250	340	440	490	500
Hartree-Fock [96]	117	305	357	465	489	529
Linear combination of atomic orbitals [106]	55	248	348	436	464	495

Note: See Fig. 10 for an illustration of the normal modes. The frequencies are given in cm^{-1}.

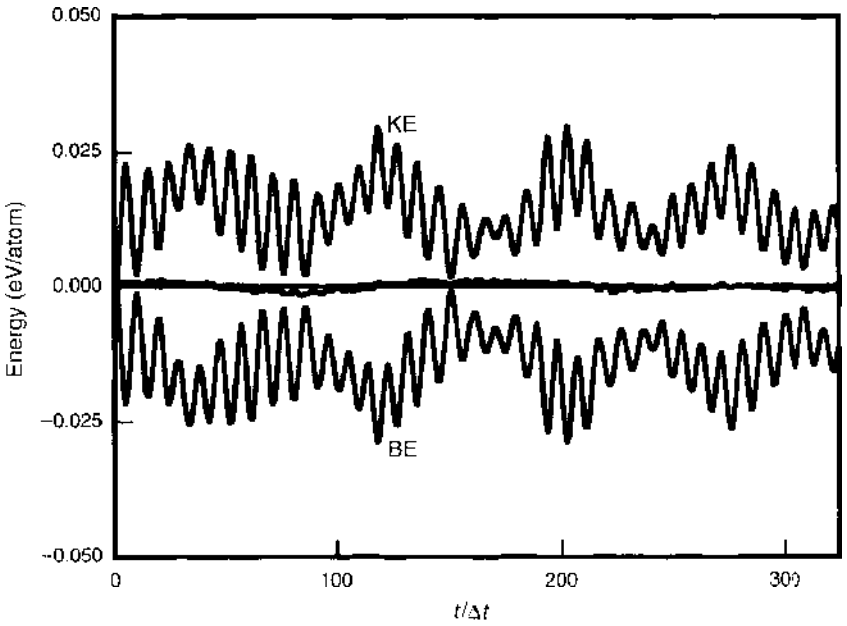


Figure 11. Simulation for a Si_4 cluster. The kinetic energy (KE) and binding energy (BE) are shown as a function of simulation time. The total energy (KE + BE) is also shown, with the zero of energy taken as the average of the total energy. The time step Δt is 3.7 fs.

frequencies are often overestimated, as opposed to the LCAO result, which underestimates the value, at least when compared to other theoretical techniques. Setting aside the issue of the B_{3u} mode, the agreement between the measured Raman modes and theory for Si_4 indicates that Raman spectroscopy can provide a key test for the structures predicted by theory.

6.4. Polarizabilities

Polarizability measurements [107] have been performed for small semiconductor clusters. The polarizability tensor α_{ij} is defined as the second derivative of the energy with respect to electric field components. For a noninteracting quantum mechanical system, the expression

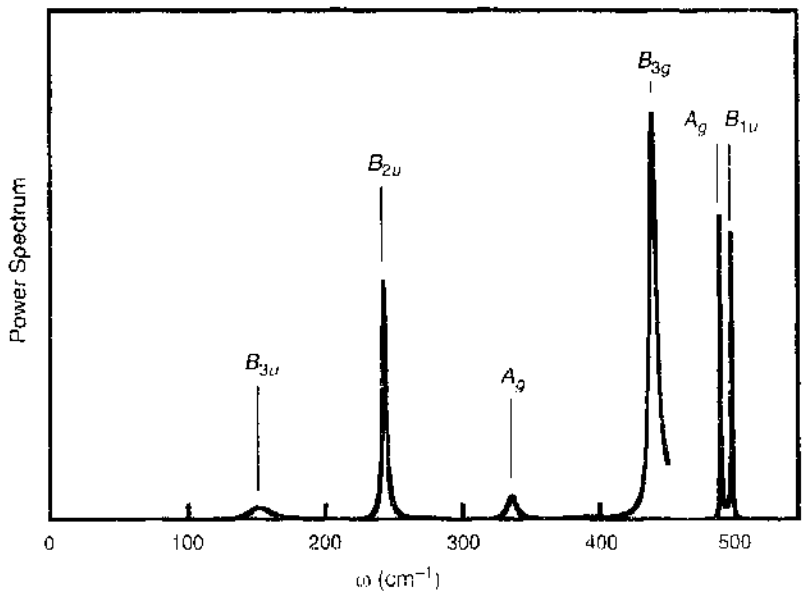


Figure 12. Power spectrum of the vibrational modes of the Si_5 cluster. The simulation time was taken to be 4 ps. The intensity of the B_{3g} and (A_g, B_{1u}) peaks has been scaled by 10^{-2}.

for the polarizability can be easily obtained by using second-order perturbation theory, where the external electric field $\mathcal{E}$ is treated as a weak perturbation.

Within the density-functional theory, as the total energy is not the sum of individual eigenvalues, the calculation of polarizability becomes a nontrivial task. One approach is to use the density-functional perturbation theory that has been developed recently in Green's function and variational formulations [108, 109].

Another approach, very convenient for handling the problem for confined systems, like clusters, is to solve the full problem exactly within the one-electron approximation. In this approach, the external ionic potential $V_{\text{ion}}(\mathbf{r})$ experienced by the electrons is modified to have an additional term given by $-e\mathcal{E}\cdot\mathbf{r}$. The Kohn–Sham equations are solved with the full external potential $V_{\text{ion}}(\mathbf{r}) - e\mathcal{E}\cdot\mathbf{r}$. For quantities like polarizability, which are derivatives of the total energy, one can compute the energy at a few field values, and differentiate numerically. Real space methods are suitable for such calculations on confined systems, because the position operator $\mathbf{r}$ is not ill-defined, as is the case for supercell geometries.

There is another point that should be emphasized. It is difficult to determine the polarizability of a cluster or molecule because of the need for a well-converged basis in the presence of an electric field. Often polarization functions are added to achieve an accurate basis, and the response of the system to the field can be sensitive to the details of basis. In both real space and plane wave methods, the lack of a "prejudice" with respect to the basis is a considerable asset. The real space method implemented with a uniform grid possesses a nearly "isotropic" environment with respect to the applied field. The response can be easily checked with respect to the grid size by varying the grid spacing. Typically, the calculated electronic response of a cluster is not sensitive to the magnitude of the field over several orders of magnitude. In Fig. 13, we illustrate the calculated polarizability as a function of the finite electric fields. For very small fields, the polarizability calculated by the change in dipole or energy is not reliable because of numerical inaccuracies such as roundoff errors. For very large fields, the cluster can be ionized by the field, and again the accuracy suffers. However, for a wide range of values of the electric field, the calculated values are stable.

In Table 2, we present some recent calculations for the polarizability of small Si and Ge clusters. It is interesting to note that some of these clusters have permanent dipoles. For example, Si_6 and Ge_6 both have nearly degenerate isomers. One of these isomers possesses a permanent dipole, the other does not. Hence, in principle, one might be able to separate the one isomer from the other via an inhomogeneous electric field.

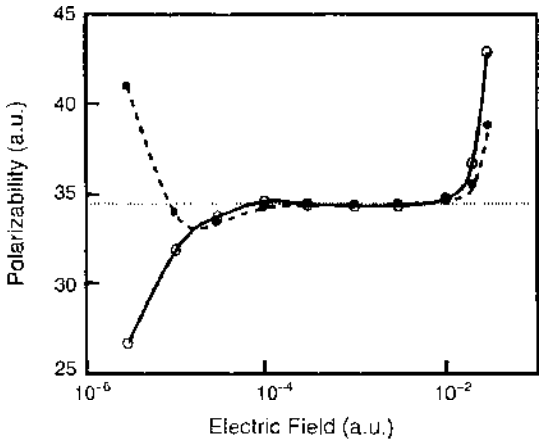


Figure 13. One component of the polarizability tensor of Si_5 as a function of the electric field. The dashed curve is the polarizability component from the second derivative of the energy with respect to the field; the solid curve is from the dipole derivative. For very small fields, the values are not accurate because of the strict convergence criteria required in the wave functions to get accurate values. For high fields, the cluster is ionized. The dashed value is the predicted value for the cluster using a small, but finite, field.

Table 2. Static dipole moments and average polarizabilities of small silicon and germanium clusters.

Silicon			Germanium		
Cluster	$\|\mu\|$ (D)	$\langle\alpha\rangle$ (Å^3/atom)	Cluster	$\|\mu$ (D)	$\langle\alpha\rangle$ (Å^3/atom)
Si_2	0	6.29	Ge_2	0	6.67
Si_3	0.33	5.22	Ge_3	0.43	5.89
Si_4	0	5.07	Ge_4	0	5.45
Si_5	0	4.81	Ge_5	0	5.15
Si_6 (I)	0	4.46	Ge_6 (I)	0	4.87
Si_6 (II)	0.19	4.48	Ge_6 (II)	0.14	4.88
Si_7	0	4.37	Ge_7	0	4.70

6.5. Optical Spectra

The time-dependent density-functional formalism (Section 3.4) is easy to implement in real space within the high-order finite difference pseudopotential method [72, 73, 81]. The TDLDA technique will be illustrated by considering the absorption spectra of sodium and hydrogenated silicon clusters. The ground-state structures of the clusters were determined by simulated annealing [102]. In all cases the obtained cluster geometries agreed well with the structures reported in other works [110, 111]. Because the wave functions for the unoccupied electron states are very sensitive to the boundary conditions, these calculations need to be performed within a relatively large boundary domain.

The calculated absorption spectrum for Na_4 is shown in Fig. 14 along with the measured spectrum. In addition, the spectrum generated by considering transitions between the LDA eigenvalues is shown. The agreement between TDLDA and experiment is remarkable, especially when contrasted with the LDA spectrum. TDLDA correctly reproduces the experimental spectral shape, and the calculated peak positions agree with experiment within 0.1–0.2 eV. The comparison with other theoretical work demonstrates that our TDLDA absorption spectrum is as accurate as the available CI spectra [112, 113]. Furthermore, the TDLDA spectrum for the Na_3 cluster appears to be in better agreement with experiment than the GW absorption spectrum calculated in Ref. [114].

The study of optical excitations in hydrogen-terminated silicon clusters is essential for understanding absorption and emission of visible light in porous silicon [1]. Over the last

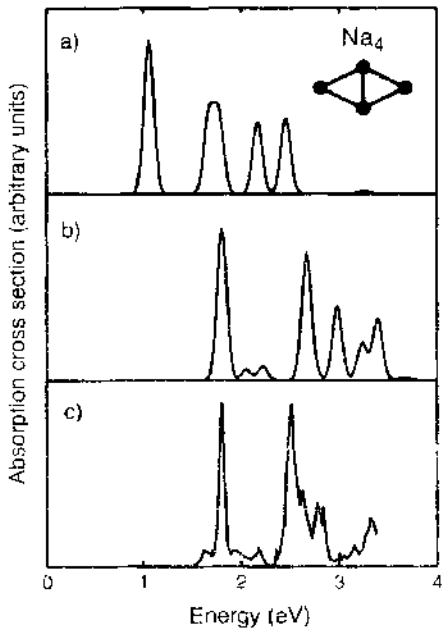


Figure 14. The calculated and experimental absorption spectrum for Na_4. (a) A local density approximation to the spectrum using Kohn–Sham eigenvalues. (b) A time-dependent local density approximation calculation. Technical details of the calculation can be found in [32]. (c) Panel is experiment from [115].

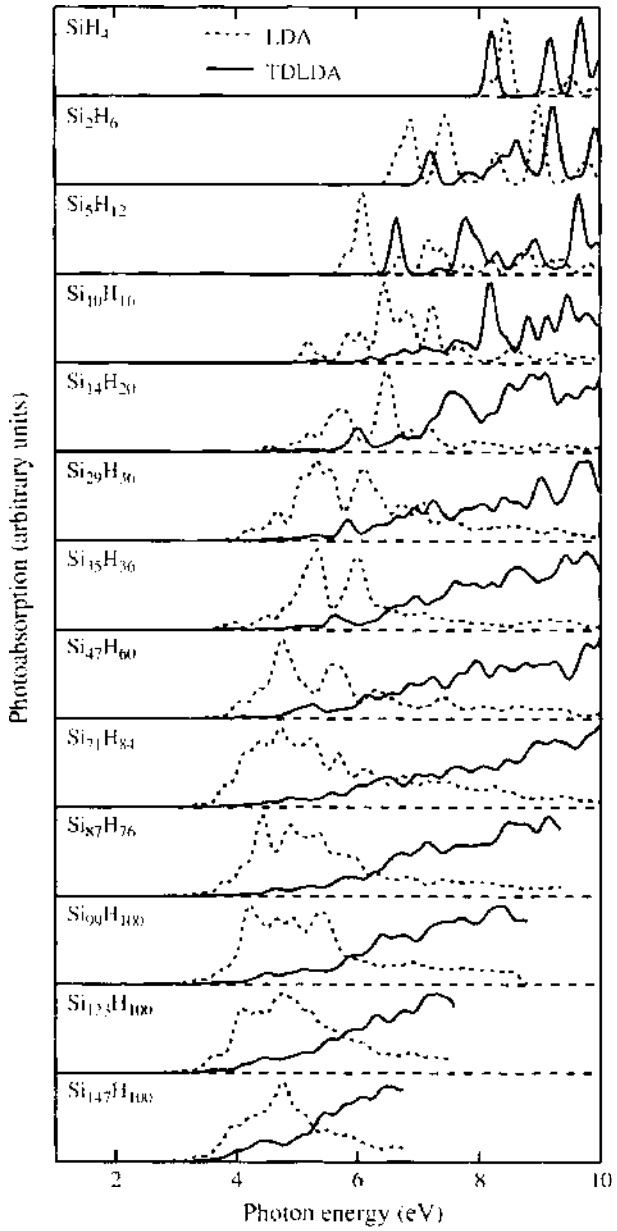


Figure 16. Calculated time-dependent local density approximation (TDLDA) absorption spectra of Si_nH_m clusters (solid lines). Spectra of time-independent Kohn–Sham LDA eigenvalues (dotted lines) are shown for comparison. All spectra are broadened by 0.1 eV using a Gaussian convolution.

Table 3. Excitation energies of hydrogenated silicon clusters.

Cluster	Transition	Experiment	BS	TDLDA	$-\epsilon_{HOMO}^{LDA}$
SiH_4	$4s$	8.8	9.0	8.2	8.6
	$4p$	9.7	10.2	9.2	
	$4d$	10.7	11.2	9.7	
Si_2H_6	$4s$	7.6	7.6	7.3	7.5
	$4p$	8.4	9.0	7.8	
Si_5H_{12}	—	6.5	7.2	6.6	7.3

Note: The experimental optical absorption energies are taken from Ref. [116] (silane and disilane) and Ref. [126] (neopentasilane). The assignment of electronic excitations for silane and disilane corresponds to the Rydberg transitions. The Bethe–Salpeter excitation energies are adapted from Ref. [121]. $-\epsilon_{HOMO}^{LDA}$ is the time-independent LDA "ionization" energy. All values are in eV.

decade, Si_nH_m clusters have been the subject of intensive experimental [116–120] and theoretical [121–127] research. However, disagreements among different theoretical models used for describing electronic excitations in these systems remain a subject of significant controversy. For the most part, the disagreements arise from the formulation of the optical gap in confined systems and the calculation of different components making up the optical gap [128–131].

A common approach to hydrogenated silicon clusters is to consider Si_nH_m clusters in which the arrangement of the silicon atoms corresponds to bulk silicon fragments. This is illustrated in Fig. 15. The calculated absorption spectra of Si_nH_m clusters are shown in Fig. 16. For the larger clusters shown, only electronic transitions below a chosen energy threshold are displayed as a result of computational constraints. The spectra of time-independent Kohn–Sham LDA eigenvalues will be illustrated here. As in the case of metallic and semiconductor clusters with free surfaces [32, 132, 133], the TDLDA spectra of Si_nH_m clusters are blue-shifted with respect to the Kohn–Sham eigenvalue spectra. Unlike optical spectra of "bare" semiconductor clusters considered in previous sections, the spectra of hydrogenated silicon clusters do not display the low-energy transitions associated with the surface states. Photoabsorption gaps for Si_nH_m clusters are much larger than those of Si_n clusters with open surfaces.

In Table 3, TDLDA values for the excitation energies of the first three Si_nH_m clusters are compared with experimental data [116, 126] as well as with the values calculated using the Bethe–Salpeter technique [121]. The last column in Table 3 shows the Kohn–Sham LDA "ionization" energies of the clusters $-\epsilon^{\mathrm{LDA}}_{\mathrm{HOMO}}$, given by the negative values of the energies for the highest occupied LDA electronic orbitals. Table 3 demonstrates that the calculated TDLDA excitation energies for the transitions below, or close to, $-\epsilon^{\mathrm{LDA}}_{\mathrm{HOMO}}$ agree well with the experimental data and the Bethe–Salpeter values. The agreement, however, deteriorates for higher excitations, which lie above $-\epsilon^{\mathrm{LDA}}_{\mathrm{HOMO}}$. As the size of clusters increases, the energy

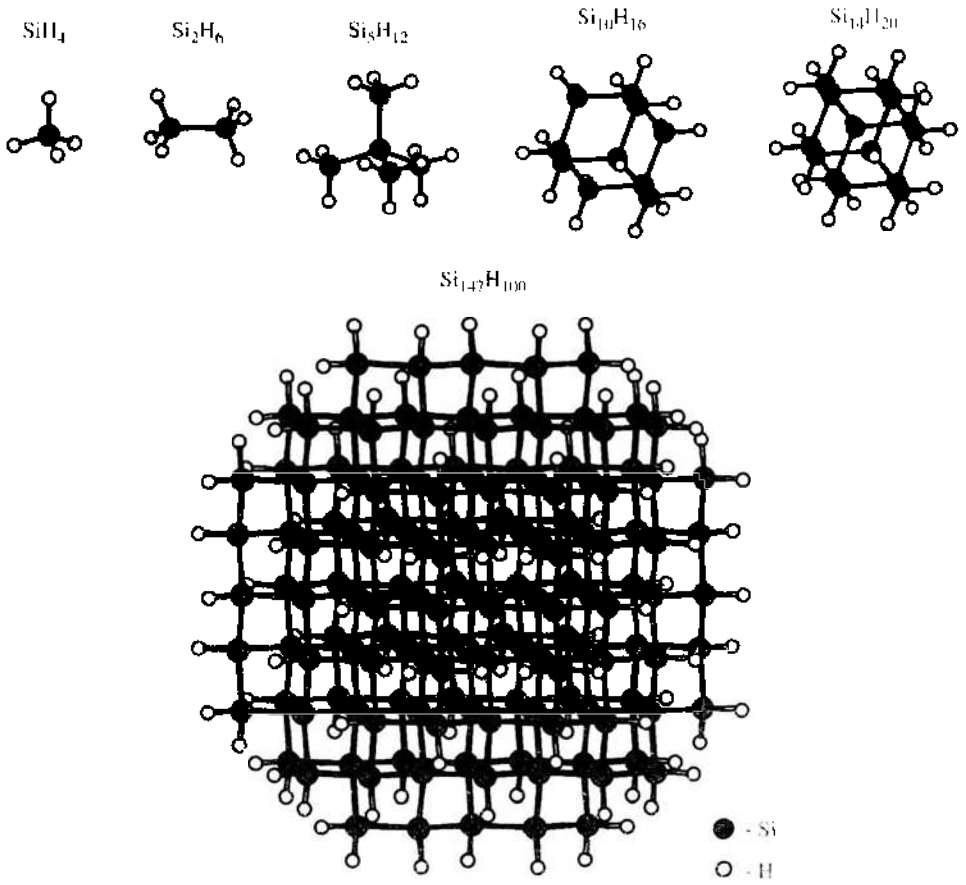


Figure 15. Ball-and-stick models for hydrogenated silicon clusters.

of the first allowed excitation moves further down from the LDA "ionization" energy, and agreement with experiment improves.

For large Si_nH_m clusters, the first allowed optical transitions are always located below $-\epsilon^{LDA}_{HOMO}$. On this basis, one can argue that TDLDA should provide an accurate description for the photoabsorption gaps and the low-energy optical transitions in larger Si_nH_m clusters.

7. QUANTUM CONFINEMENT IN NANOCRYSTALS

Nanocrystals are assemblages of atoms at the nanoscale where the atomic positions are characteristic of the crystalline state. Spherical nanocrystals are sometimes called "quantum dots." Understanding the role of quantum confinement in altering optical properties of nanocrystals made from semiconductor materials is a problem of both technological and fundamental interest. In particular, the discovery of visible luminescence from porous Si [1] has focused attention on optical properties of confined systems. Although there is still debate on the exact mechanism of photoluminescence in porous Si, there is a great deal of experimental and theoretical evidence that supports the important role played by quantum confinement in producing this phenomenon [119, 134].

Excitations in confined systems, such as porous Si, differ from those in extended systems because of quantum confinement. In particular, the components that make up the excitation energies, such as quasiparticle and exciton-binding energies, change significantly with the physical extent of the system. So far, most calculations that model semiconductor nanocrystals have been of an empirical nature because of the major challenges involved in simulating these systems from first principles [120, 122–124, 135–139]. Although empirical studies have shed some light on the physics of optical excitations in semiconductor nanocrystal, one often has to make assumptions and approximations that may not be justified. Efficient and accurate *ab initio* studies are necessary to achieve a better microscopic understanding of the size dependence of optical processes in semiconductor quantum dots. A major goal of our work is to develop such *ab initio* methods that handle systems from atoms to dots to crystals on an equal footing.

A problem using empirical approaches for semiconductor quantum dots centers on the transferability of the bulk interaction parameters to the nanocrystalline environment. The validity of this assumption, which postulates the use of fitted bulk parameters in a size regime of a few nanometers, is not clear and has been questioned in recent studies [138, 139]. More specifically, quantum confinement–induced changes in the self-energy corrections, which may affect the magnitude of the optical gaps significantly, are neglected in empirical approaches by implicitly assuming a "size-independent" correction that corresponds to that of the bulk. It follows that a reliable way to investigate optical properties of quantum dots would be to model them from first principles with no uncontrolled approximations or empirical data. However, there have been two major obstacles for the application of *ab initio* studies to these systems. First, because of large computational demand, accurate first-principles calculations have been limited to small system sizes, which do not correspond to the nanoparticle sizes for which experimental data are available. Second, *ab initio* calculations performed within the local density approximation suffer from the underestimate of the band gap [29].

Recent advances in electronic structure algorithms, have indicated that using pseudopotentials (as outlined earlier) [66, 71–82] and computational platforms, and alternative formulations of the optical gaps suitable for confined systems, the above-mentioned challenges for *ab initio* studies of quantum dots can be overcome. In particular, new electronic structure methods [66, 71–82], implemented on massively parallel computational platforms, allow one to model a cluster of more than 1000 atoms in a straightforward fashion [140]. Such approaches can be illustrated by focusing on silicon quantum dots such as $Si_{705}H_{300}$. This system corresponds to spherically bulk-terminated Si clusters passivated by hydrogens at the boundaries (Fig. 17). Computational details can be found in the literature [140].

For an n-electron system, the quasiparticle gap ε_g^{QP} can be expressed in terms of the ground-state total energies E of the $(n+1)$-, $(n-1)$-, and n-electron systems as

$$\varepsilon_g^{QP} = E(n+1) + E(n-1) - 2E(n) = \varepsilon_g^{HL} + \Sigma \tag{97}$$

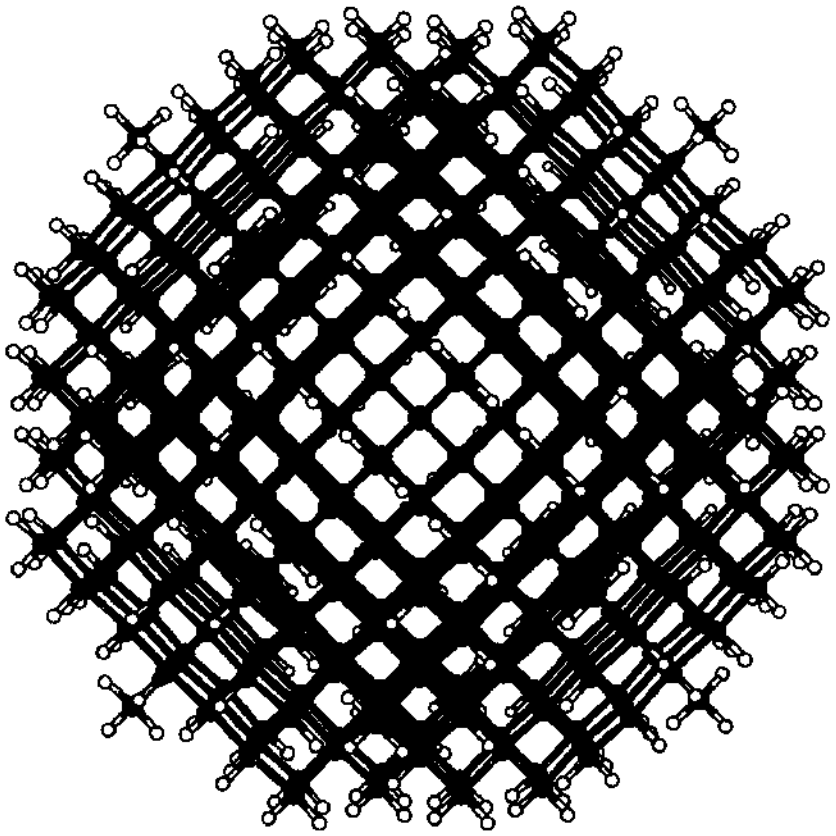


Figure 17. Atomic structure of a Si quantum dot with composition $Si_{525}H_{276}$. The gray and white balls represent Si and H atoms, respectively. This bulk-truncated Si quantum dot contains 25 shells of Si atoms and is 27.2 Å in diameter.

where Σ is the self-energy correction to the highest occupied molecular orbit–lowest unoccupied molecular orbit gap $\varepsilon_g^{\text{HL}}$ obtained within LDA. This definition is quite convenient for the calculation of the quasiparticle gap, as it is possible to excite individual electrons or holes from the ground-state electronic configuration of a confined system. The calculation of $\varepsilon_g^{\text{qp}}$ requires the self-consistent solutions of three different charge configurations of each quantum dot. The computational demand of this approach can be reduced significantly by using the wave functions of the neutral cluster to extract very good initial charge densities for the self-consistent solutions of the charged systems. Total energies for charged $(n+1)$- and $(n-1)$-electron systems can be calculated in a straightforward fashion [140].

Equation (97) yields the correct quasiparticle gap $\varepsilon_g^{\text{qp}}$, if the exact exchange-correlation functional is used. Within the local density approximation, in the limit of very large systems $(n \to \infty)$, the gaps calculated using Eq. (97) approach the highest occupied molecular orbit–lowest unoccupied molecular orbit gap $\varepsilon_g^{\text{HL}}$ [130]. However, for small systems, Eq. (97) captures the correction to the LDA highest occupied molecular orbit–lowest unoccupied molecular orbit gap quite accurately (Table 4) when compared with available quasi-particle calculations using the GW approach [121]. Small deviations appear as the system size reaches approximately 1000 atoms.

The size dependence of the quasiparticle and LDA highest occupied molecular orbit–lowest unoccupied molecular orbit gaps and self-energy corrections are shown in Fig. 18. Both gap values and self-energy corrections are enhanced substantially with respect to bulk values, and are inversely proportional to the quantum dot diameter d as a result of quantum confinement. Specifically, $\varepsilon_g^{\text{qp}}(d) - \varepsilon_{g,\text{bulk}}^{\text{qp}}$, $\varepsilon_g^{\text{band}}(d) - \varepsilon_{g,\text{bulk}}^{\text{band}}$, and $\Sigma(d) - \Sigma_{\text{bulk}}$ scale as $d^{-1.2}$, $d^{-1.1}$, and $d^{-1.5}$, respectively. The quasiparticle gaps shown in the figure are significantly higher compared to the gap values obtained in earlier semiempirical calculations, although it is problematic in terms of any comparisons. The empirical gaps are obtained from potentials obtained from crystalline environments and scaled to dot sizes. The nature of these gaps is problematic in that they do not strictly correspond to quasi-particle gaps owing to

Table 4. Highest occupied molecular orbit–lowest unoccupied molecular orbit and quasiparticle gaps ε_g^{qp} [Eq. (97)] calculated for hydrogenated Si clusters compared to quasiparticle gaps calculated within the GW approximation [121].

	ε_g^{HL}	ε_g^{qp}	GW
SiH_4	7.9	12.3	12.7
Si_2H_6	6.7	10.7	10.6
Si_5H_{12}	6.0	9.5	9.8
$Si_{14}H_{20}$	4.4	7.6	8.0

Note: All energies are in eV.

their empirical roots. The main reason for the size difference in the gaps is the significant enhancement of electron self-energies because of quantum confinement, which cannot be properly taken into account in semiempirical approaches.

Because the quasiparticle gap refers to the energy to create a noninteracting electron-hole $(e - h)$ pair, one cannot compare these gaps directly to measurements of the optical gap. This issue is especially important for quantum dots in which the exciton radius becomes comparable to the size of the dot. Quantum confinement in nanostructures enhances the bare exciton Coulomb interaction and also reduces the electronic screening so that the exciton Coulomb energy E_{Coul} becomes comparable to the quasiparticle gap. In order to extract the optical gaps

$$\varepsilon_g^{opt} = \varepsilon_g^{qp} - E_{Coul} \tag{98}$$

the exciton Coulomb energy needs to be calculated accurately. Compared to E_{Coul}, exciton exchange-correlation energies are much smaller for the quantum dots studied in this work, and will therefore be neglected. Equation (98) is a rigorous expression, provided the Coulomb energy can be properly computed.

A crude, yet commonly used, approximation to E_{Coul} comes from the effective mass approximation [141, 142]. Within the effective mass approximation, one assumes an infinite potential barrier at the boundary of the quantum dot, and envelope wave functions of the form $\psi(\mathbf{r}) \sim \frac{1}{r}\sin(2\pi r)/d$ for a noninteracting $e - h$ pair. This yields (in a.u.) $E_{Coul} = 3.572/\epsilon d$. The effective mass approximation, although commonly used, cannot be expected to yield accurate exciton Coulomb energies, as in this approximation the microscopic features

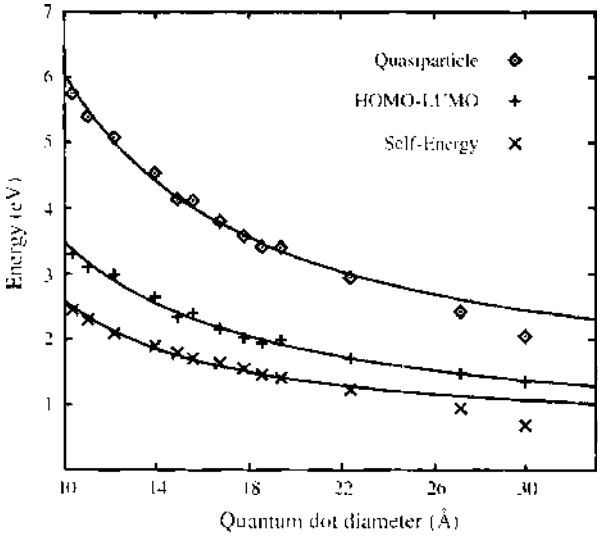


Figure 18. Calculated quasiparticle (◇) and highest occupied molecular orbit–lowest unoccupied molecular orbit (HOMO-LUMO) gaps (+) and self-energy corrections (×) as a function of the quantum dot diameter d (in Angstroms). The solid lines are power-law fits to the calculated data approaching the corresponding bulk limits. For small deviations from the fits for large system sizes, see the text.

of the $e - h$ wave functions inside the quantum dot are neglected, and the wave functions are constrained to vanish abruptly outside the quantum dots, instead of decaying relatively slowly into the vacuum.

Thus, E_{Coul} is better calculated directly using *ab initio* pseudo wave functions. The exciton Coulomb energy can be written as

$$\begin{aligned} E_{\text{Coul}} &= \int d\mathbf{r}_1 |\psi_e(\mathbf{r}_1)|^2 V^h_{\text{scr}}(\mathbf{r}_1) \\ &= \int d\mathbf{r}_1 |\psi_e(\mathbf{r}_1)|^2 \int d\mathbf{r} \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) V^h_{\text{unscr}}(\mathbf{r}) \\ &= \iiint \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) \frac{|\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r}\, d\mathbf{r}_1\, d\mathbf{r}_2. \end{aligned} \tag{99}$$

In this expression, V^h_{scr} and V^h_{unscr} are screened and unscreened potentials resulting from the hole, ψ_e and ψ_h are the electron and hole wave functions, and ϵ^{-1} is the inverse of the microscopic dielectric matrix. One can define $\tilde{\epsilon}^{-1}$ as

$$\int \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r} \equiv \tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{100}$$

then the exciton Coulomb energy can be written as

$$E_{\text{Coul}} = \iint \tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2) \frac{|\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1\, d\mathbf{r}_2 \tag{101}$$

If $\tilde{\epsilon}$ is taken to be unity, the unscreened E_{Coul} can be determined. The results are shown in Fig. 19, along with the predictions of the effective mass approximation (EMA) and recent empirical calculations [143]. The *ab initio* and empirical calculations for the unscreened Coulomb energy are in quite good agreement with each other, both predicting smaller Coulomb energies and a softer power-law decay compared to the EMA. In particular, fitting the calculated data to a power law of the diameter as $d^{-\beta}$, we find $\beta = 0.7$.

An accurate calculation of E_{Coul} requires the inverse dielectric matrix $\tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2)$ in Eq. (101). An *ab initio* calculation of $\tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2)$ is computationally very demanding, although recent progress has been made on this problem [144]. Earlier calculations used either the bulk dielectric constant or the reduced dielectric constant of the quantum dot for all $e - h$

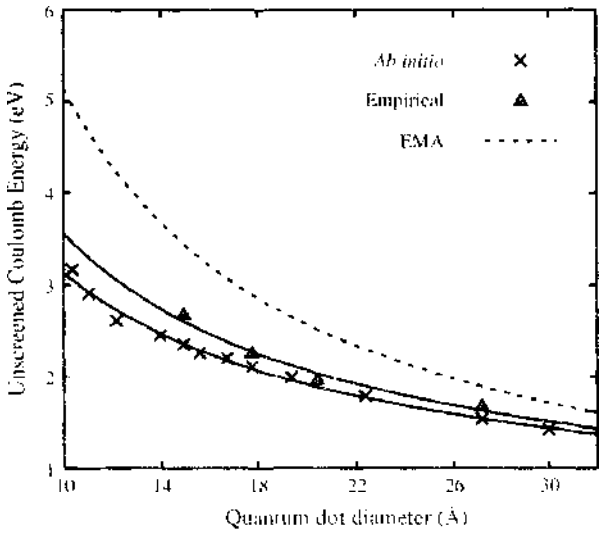


Figure 19. Unscreened exciton Coulomb energies as a function of the quantum dot diameter d (in Angstroms) calculated by effective mass approximation (dashed line), direct empirical pseudopotential calculations (Δ from Ref. [143]), and direct *ab initio* pseudopotential calculations (×) as explained in the text. The solid lines are power-law fits to the calculated data

distances. These are rather simple approximations, as the screening is different at different-length scales because of the wave vector dependence of $\tilde{\epsilon}$. For example, when $\mathbf{r}_1$ and $\mathbf{r}_2$ in Eq. (101) are very close to each other, there will be practically no screening, and $\tilde{\epsilon} \approx 1$. Typically, both the hole and electron wave functions are well localized toward the center of the quantum dot, and the screening will be reduced significantly, resulting in larger Coulomb energies compared to the case of using a single dielectric constant for all distances. Also, extracting the appropriate dielectric constant for these small systems is problematic.

One can improve on these approximations by explicitly using the wave functions as calculated by our pseudopotential approach and a realistic dielectric function that takes spatial variations of $\tilde{\epsilon}$ into account. To calculate the screening dielectric functions $\tilde{\epsilon}(\mathbf{r}_1, \mathbf{r}_2)$ of a particular quantum dot, one can proceed as follows: first, one applies spatially modulated electric fields at several wave vectors to calculate the q-dependent polarizability $\alpha(\mathbf{q})$ using a finite-field method. The q-dependent dielectric function $\tilde{\epsilon}(\mathbf{q})$ can be obtained using a dielectric sphere model [145]. The results for $\tilde{\epsilon}(\mathbf{q})$ of the $Si_{87}H_{76}$ quantum dot are shown in Fig. 20. After fitting the calculated $\tilde{\epsilon}(\mathbf{q})$ to a rational polynomial function of q and Fourier-transforming to real space [146], we obtained the dielectric function $\tilde{\epsilon}(r = |\mathbf{r}_1 - \mathbf{r}_2|)$. Implicitly, we are assuming spatial isotropy in writing $\tilde{\epsilon}(\mathbf{r}_1, \mathbf{r}_2) \approx \tilde{\epsilon}(r = |\mathbf{r}_1 - \mathbf{r}_2|)$. As shown in Fig. 20, the calculated $\tilde{\epsilon}(\mathbf{q})$ has a very sharp drop to ≈ 1 beyond $q = 0.2$ a.u., which corresponds roughly to the wave vector set by the linear dimension (or diameter) of this quantum dot. This sharp drop is typical for all quantum dots studied. In real space, this implies that the $e - h$ interaction is very inefficiently screened inside the dot, resulting in substantial excitonic Coulomb energies.

The resulting optical gaps $\varepsilon_g^{\text{opt}} = \varepsilon_g^{\text{qp}} - E_{\text{Coul}}$ along with the quasiparticle gaps and experimental absorption data [117] from Si:H nanocrystals are shown in Fig. 21. Although the calculated quasiparticle gaps are ~0.6–1.0 eV larger than the experimental absorption data, the calculated optical gaps are in very good agreement with the experimental data.

At this point, an interesting observation can be made about the good agreement of previous semiempirical calculations with experimental data [134–136]. In the above semiempirical approaches, it is the underestimate of both the excitation gap and the exciton Coulomb energies (through the use of a static dielectric constant of either the bulk or the quantum dot) that results in calculated values in agreement with experiment. As a matter of fact, the bare gaps of Refs. [135] and [136] without the exciton Coulomb energies are in better agreement with the experiment. The *ab initio* pseudopotential results demonstrate that the quasiparticle gaps in Si quantum dots are actually higher than previously thought, and the exciton Coulomb energies, because of the wave vector dependence of the dielectric response function $\epsilon(\mathbf{r}_1, \mathbf{r}_2)$, are higher than previously calculated, resulting in optical gap values that are in good agreement with the experimental absorption data.

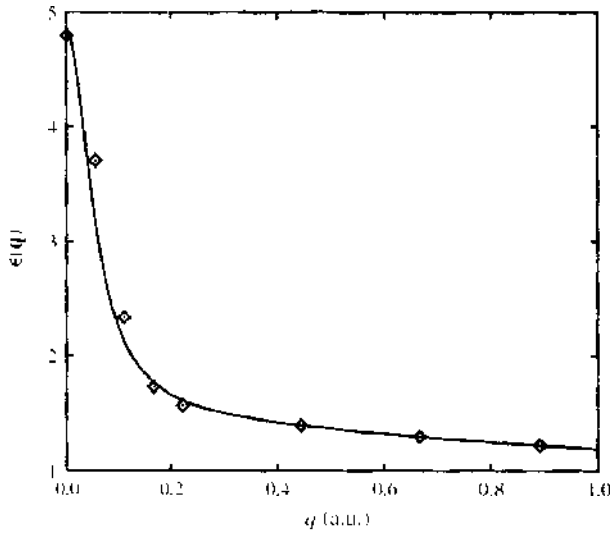


Figure 20. Wave vector dependence of the dielectric function for the Si_x-H_{76} quantum dot.

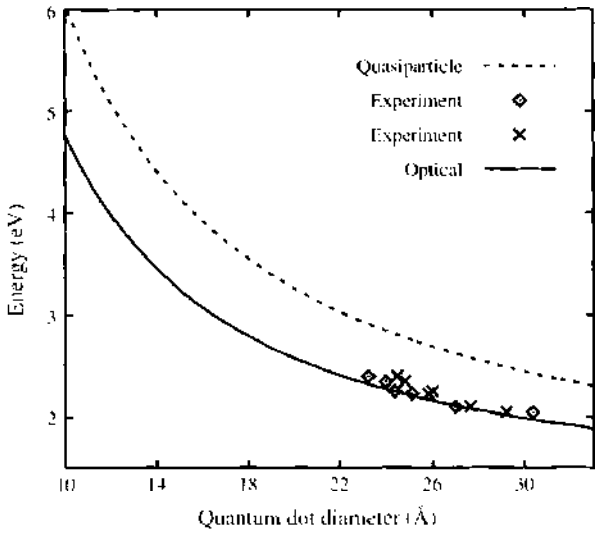


Figure 21. Calculated quasiparticle gaps (dotted line), optical gaps, and experimental absorption data from Si:H nanocrystals (× and ◇ from Ref. [117]) as a function of the quantum dot diameter d (in Angstroms). The two sets of experimental data (× and ◇) differ by the method to estimate the nanocrystal size.

TDLDA methods can be used to examine hydrogenated silicon quantum dots. For large clusters or quantum dots, the absorption spectra become essentially quasi-continuous, and it is incumbent on us to use care in defining the optical gap. In particular, a large number of low-intensity transitions exist near the absorption edge. Taken individually, the oscillator strengths of these transitions would be located far below the experimentally detectable limit. As a result, identifying the first allowed optical transition in the case of large clusters is not a trivial task.

As the size of clusters increases, the absorption gaps gradually decrease, and the discrete spectra for small clusters evolve into quasi-continuous spectra for silicon nanocrystals. Figure 22 demonstrates that oscillator strength of dipole-allowed transitions near the absorption edge decreases with increasing cluster size. This fact is consistent with the formation of an indirect band gap in the limit of bulk silicon [126].

Rather than associating the optical gaps with the individual transitions, one can define a procedure for fixing the optical gap $E_{\text{gap}}^{\text{opt}}$ via an integral of the oscillator strength.

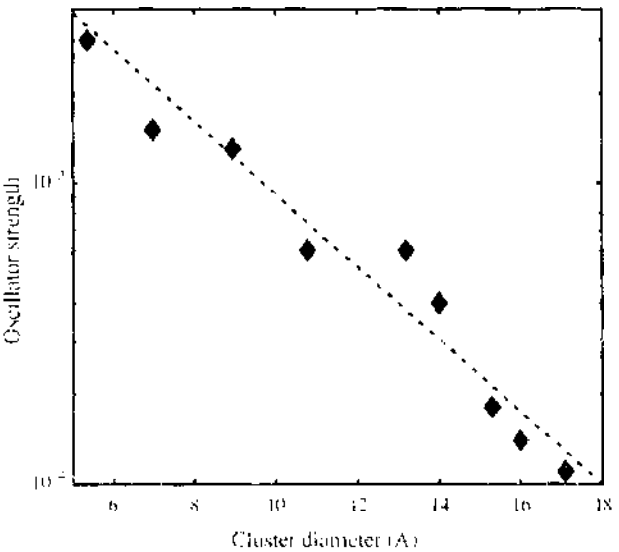


Figure 22. Oscillator strength for optical transitions in hydrogenated silicon clusters as a function of cluster size. The strength is determined by considering transitions near the gap. The dashed line is a linear fit.

In particular, the following prescription has been suggested [147] to define the gap

$$pF = \int_{0}^{E_{\text{opt gap}}^{(p)}} \sigma(\omega)d\omega \tag{102}$$

where F is the total optical cross section, $\sigma(\omega)$ is the optical cross section for a given frequency, and ω and p are some prescribed fractions of the total cross-section for the fixing the gap. For the photoabsorption gaps, a typical value of p might be 10^{-4}. This definition for the absorption gap does not affect the values of the optical gaps for small Si_nH_m clusters, as the intensity of their first allowed transitions is much higher than the selected threshold. An order of magnitude change in p does not typically change the gap size by more than ± 10 meV. At the same time, Eq. (102) offers a convenient way to evaluate optical gaps in large clusters.

The variation of the optical absorption gaps as a function of cluster size is shown in Fig. 23. Along with the TDLDA values, we include optical gaps calculated by the Bethe–Salpeter technique [121]. For very small clusters, SiH_4, Si_2H_6, and Si_3H_{12}, the gaps computed by the TDLDA method are close to the Bethe–Salpeter values, although for $Si_{10}H_{16}$ and $Si_{14}H_{30}$ our gaps are considerably smaller than the gaps calculated using the Bethe–Salpeter equation. At the same time, the TDLDA gaps for clusters in the size range from 5 to 71 silicon atoms are larger by ~1 eV than the gaps calculated by the Hartree–Fock technique, with the correlation correction included through the CI approximation (HF-CI) [127].

These differences are consistent with the fact that the Bethe-Salpeter calculations systematically overestimate and the HF-CI calculations of Ref. [127] underestimate the experimental absorption gaps. For example, for the optical absorption gap of Si_5H_{12} the Bethe-Salpeter, TDLDA, and HF-CI methods predict the values of 7.2, 6.6, and 5.3 eV, respectively, compared to the experimental value of 6.5 eV. However, it is not clear whether the gaps of Ref. [127] refer to the optically allowed or optically forbidden transitions, which may offer a possible explanation for the observed discrepancy. For large clusters, we find the TDLDA optical gaps to be in generally good agreement with the photoabsorption gaps evaluated by the majority of self-energy-corrected LDA [103, 126] and empirical techniques [124, 125, 136]. At present, the full TDLDA calculations for clusters larger than a few

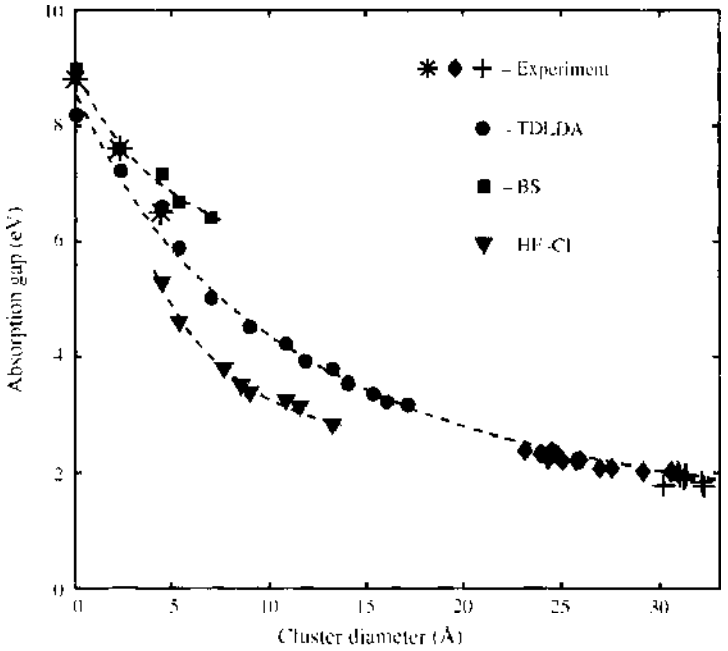


Figure 23. Variation of optical absorption gaps as a function of cluster diameter. Theoretical values shown in the plot include the gaps calculated by the time-dependent local density approximations (TDLDA) method (this work), by the Bethe–Salpeter technique (BS) [121] and by the Hartree–Fock method, with the correlation included through the configuration interaction approximation (HF-CI) [127]. Experimental values are taken from Refs. [116–118, 126]. The dashed lines are a guide to the eye.

nanometers can exceed the capabilities of most computational platforms. Nevertheless, the extrapolation of the TDLDA curve in the limit of large clusters comes very close to the experimental values for the photoabsorption gaps. Software and hardware advances should make a direct verification of this possible in the near future.

In determining the optical gaps within a linear response approach, only excitations with an induced dipole are incorporated. In real time methods, the induced dipole term is calculated directly [36–38]. Within our frequency domain description, two factors enter in the ascertaining the existence of an induced dipole: the existence of a transition energy and the corresponding oscillator strength. Within TDLDA, these terms can be obtained from Eq. (58) as Ω_I and $\mathbf{F}_I$. These terms must always be considered together when predicting optical properties, although sometimes this is not done [26].

In Fig. 24, we illustrate the lowest transitions without regard to oscillator strength for both LDA and TDLDA calculations. Transitions as defined by Eq. (102) are shown. For these transitions, the gap is defined when the oscillator strength assumes a value of at least 10^{-1} of the total optical cross-section. The main difference between LDA and TDLDA for these system is a strong blue-shift of the oscillator strength.

It should be noted that real time methods for TDLDA do not involve unoccupied eigenvalues [36–38]. In this formalism, the absorption spectrum evolves from taking the power spectrum of the instantaneous induced dipole. The resolution of an optical transition is determined by the length of the time integration [36–38]. Because the frequency domain method and the real time method should yield the same spectrum, "virtual transitions," that is, transitions that do not couple to the dipole, are not physically meaningful within frequency domain implementation of TDLDA.

7.1. Role of Oxygen in Silicon Quantum Dots

Porous and nanocrystalline silicon studied in experiments are prepared under a variety of surface conditions determined by the etching technique and external chemical environments employed. Only a fraction of published experimental data refers to "pure" hydrogenated silicon dots [117]. Other measurements are performed on partially oxidized nanocrystals [148, 149]. For many cases, a precise chemical composition of nanocrystalline surfaces is not known [118, 150, 151]

However, most calculations for optical absorption and emission in silicon dots do not take into account differences in structure and chemical composition of the dot surface. This creates an ambiguity in the interpretation of experimental data. Almost all *ab initio* and empirical simulations available in literature use silicon dots passivated with hydrogen [103, 121, 122, 125–127, 136, 152], although some notable exceptions exist [153, 154]. This limitation also is true for structural issues, where only a few systems have been examined for reconstructed surfaces [155, 156].

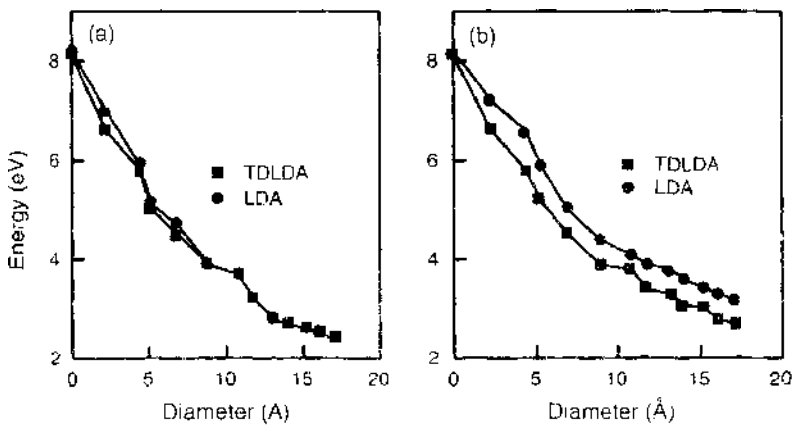


Figure 24. Gaps determined from local density approximation (LDA) and time-dependent LDA (TDLDA) for hydrogenated silicon clusters. (a) The gaps plotted without regard to the oscillator strength. (b) The gaps determined using the criteria from Eq. (102).

Theoretical calculations [103, 121, 122, 125-127, 136, 152] based on a quantum confinement model show general agreement with experimental measurements [117] for optical absorption in hydrogen-passivated silicon clusters. In contrast, experiments performed on oxidized samples often display photoluminescence with energies significantly below the values of optical gaps predicted by the confinement model for clusters in the same size range [148, 149]. This disagreement could be greater than 1 eV. It has been suggested that the onset of photoluminescence in silicon nanocrystals may be associated with the optical Stokes shift [157] and excitonic exchange splitting [158]. Although these effects could be significant in small silicon dots, it appears that neither the Stokes shift nor the excitonic exchange splitting alone could explain such a large disagreement between experiment and theory.

Recent experimental data present strong evidence that surface effects have a very substantial effect on the electronic and optical properties of nanocrystalline silicon. Specifically, Wolkin et al. observed a large redshift of photoluminescence in porous silicon after exposure to open air [119]. The study reported a shift of photoluminescence of the order of 1 eV for samples composed of crystallites smaller than 2 nm in size. The observed redshift has been attributed to surface oxidation of silicon nanocrystals. According to the interpretation proposed in Ref. [119], oxygen creates trapped electron and hole states on nanocrystalline surfaces. The trapped surface states reduce the effective size of the optical gap. This mechanism can explain the difference between the energy of the measured photoluminescence and theoretical predictions based on the quantum confinement model.

As a result of the very large number of possible configurations for oxidized silicon clusters, current studies are often limited to the case of a single oxygen atom attached to the cluster surface. Oxidized clusters were prepared from regular hydrogen-terminated spherical dots by replacing two hydrogen atoms on the surface with a single atom of oxygen, followed by relaxation of all interatomic forces. The model geometries for oxidized clusters are illustrated in Fig. 25.

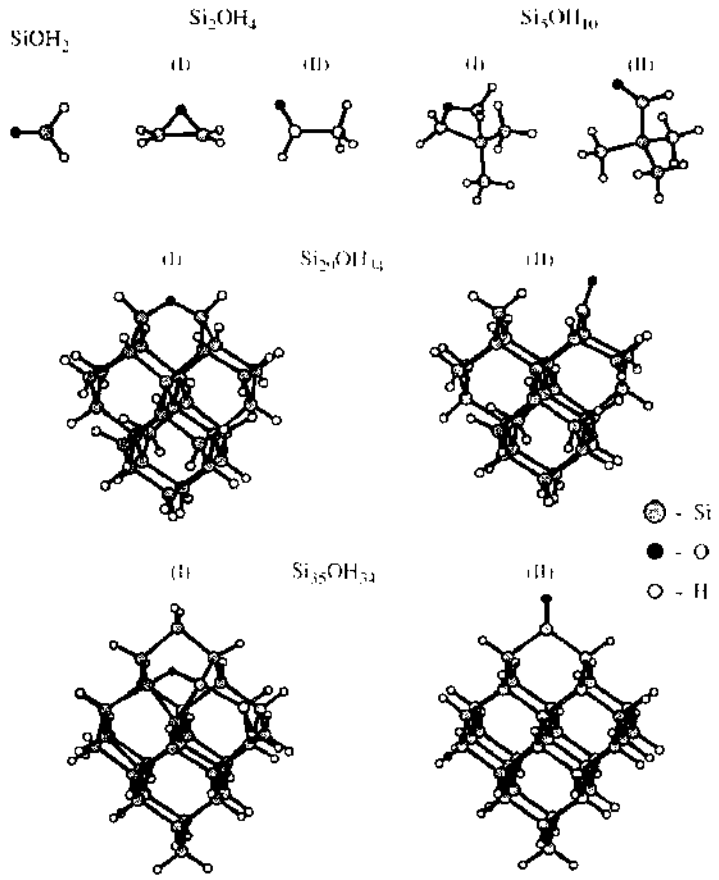


Figure 25. Model geometries for hydrogenated silicon with oxygen.

The calculated absorption spectra of oxidized silicon dots are shown in Figs. 26 and 27. In Fig. 26 the spectra of small oxidized and nonoxidized clusters are illustrated. The addition of oxygen creates new absorption bands in the region of lower transition energies. Optical excitations with higher energies are also affected by oxidation, although some intense absorption peaks observed in nonoxidized clusters (such as the peaks at 6.6 and 7.8 eV for Si_5H_{12}) appear to be only slightly shifted. Figure 27 shows the calculated spectra of the $Si_{29}OH_{34}$ and $Si_{35}OH_{34}$ clusters.

The overall change in the optical absorption spectra caused by the addition of a single oxygen atom is small larger clusters. To make the effect of oxidation more evident, we plot in Fig. 27 the differential spectra calculated as the difference in optical absorption of the same cluster before and after oxidation. Positive values of differential photoabsorption correspond to the new absorption peaks that appear only after oxidation. The differential absorption spectra for $Si_{29}OH_{34}$ and $Si_{35}OH_{34}$ clearly show the presence of low-energy optical transitions associated with surface oxygen. The calculated optical absorption gaps in oxidized and nonoxidized silicon dots are compared in Fig. 28. The TDLDA gaps for nonoxidized Si_nH_m clusters are adapted from our previous work [152]. The spectra are essentially quasicontinuous and exhibit a large number of low-intensity transitions near the absorption edge. As such, the effective optical gaps were evaluated at a very small but nonzero fraction of the complete electronic oscillator strength as in Eq. (102). The same criterion in defining the gap for silicon quantum dots was used for the oxidized silicon clusters. Fig. 28 demonstrates that surface oxidation reduces optical gaps in hydrogenated silicon clusters by as much as 1–2 eV. The change in the size of optical gaps is consistent with the redshift of photoluminescence observed in Ref. [119] and is likely responsible for the disagreement between experimental photoluminescence from oxidized silicon nanocrystals and theoretical estimates based on the quantum confinement model.

A surprising result of oxygen absorption is the small difference observed in the optical gaps between cluster isomers with Si=O and Si—O—Si bonds on the surface. At the same time, Figs. 26 and 27 reveal substantial differences in the shape of optical spectra for these clusters. One can understand this difference by examining the mechanism of the gap formation in two selected clusters: $Si_{35}OH_{34}$ (I) and (II). The order of electronic levels near the gap is illustrated for both isomers in Figs. 29 and 30, respectively. These diagrams represent simplified schemes that show only the dominant single-electron Kohn–Sham transitions

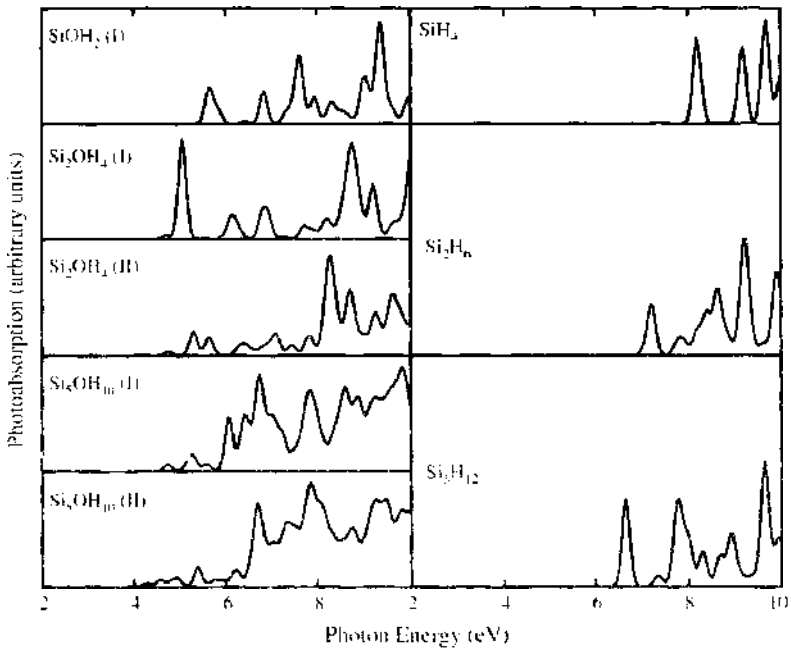


Figure 26. Left: calculated TDLDA absorption spectra of oxidized hydrogen-terminated silicon clusters. Right: time-dependent local density approximation spectra of nonoxidized clusters. All spectra were broadened by 0.1 eV using a Gaussian convolution.

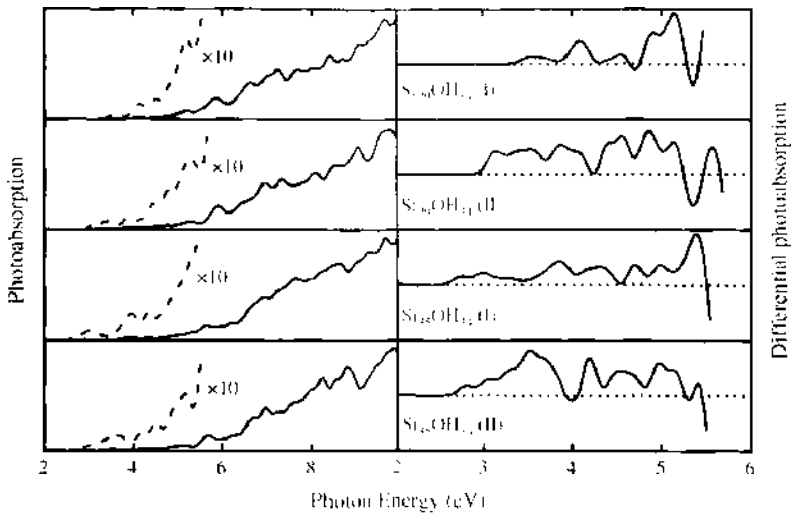


Figure 27. Left: calculated time-dependent local density approximation absorption spectra of $Si_{35}OH_{34}$ and $Si_{35}OH_{34}$ clusters. Right: difference in optical absorption between clusters with and without oxygen on the surface. All spectra are broadened by 0.1 eV.

within the TDLDA description and do not account for correlations among individual excitations. The energies of optical transitions shown in these figures correspond to one-electron singlet TDLDA excitations [32]. They differ from transition energies of the TDLDA optical spectra shown in Figs. 26 and 27 which correspond to collective electronic excitations. Nevertheless, the single-electron diagrams are useful for the qualitative analysis of optical transitions in oxidized silicon dots.

The authors of Ref. [119] proposed that photoluminescence in small oxidized silicon clusters occurs between the trapped electron and hole states, both of which are associated with the double Si=O bond on the cluster surface. Specifically, the trapped electron state is a p-state localized on silicon and the trapped hole state is a p-state localized on oxygen. Spatial distributions of electron densities for the lowest unoccupied molecular orbital and the highest occupied molecular orbital of the $Si_{35}OH_{34}$ (II) cluster plotted in Fig. 26 confirm that these states are indeed represented by p-states mainly localized on the silicon and oxygen

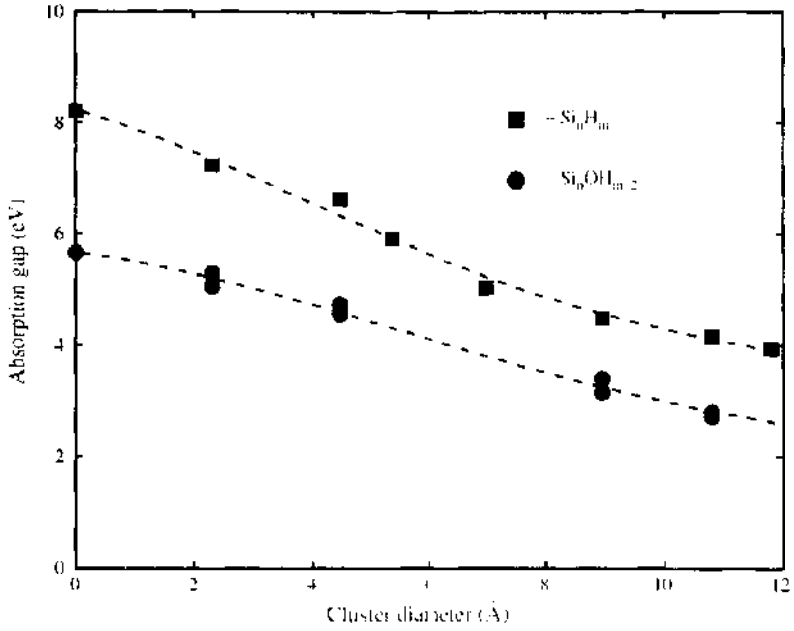


Figure 28. Comparison between the optical absorption gaps of regular and oxidized hydrogen-terminated silicon clusters. The gaps for SiH clusters are adapted from Ref. [152]. The dashed lines are a guide to the eye.

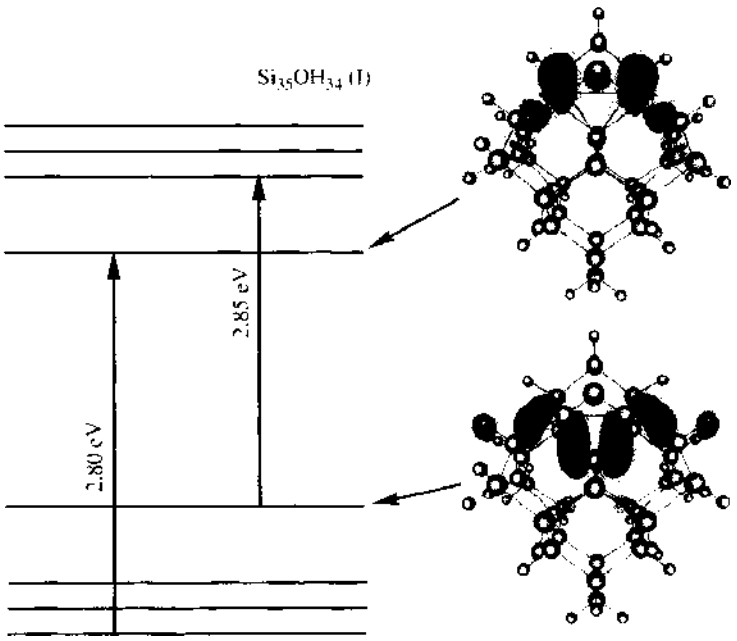


Figure 29. Schematic representation of electronic levels in the vicinity of the gap for $Si_{35}OH_{34}$ (I) clusters. Spatial distributions of electron densities are shown for the highest occupied molecular orbit and lowest unoccupied molecular orbit states.

atoms. However, the distributions of highest occupied molecular orbit and lowest unoccupied molecular orbit electron densities for the $Si_{35}OH_{34}$ (I) cluster shown in Fig. 27 reveal a different picture. The lowest unoccupied molecular orbit state is, for the most part, localized on two silicon atoms that form the Si—O—Si bonds. At the same time, the highest occupied molecular orbit state is not localized on the oxygen atom. Instead, this electronic state is spread among the layers of silicon atoms surrounding the Si—O—Si fragment. In both cases, the direct dipole transitions between the highest occupied molecular orbit and lowest unoccupied molecular orbit states are forbidden. The absorption edge for $Si_{35}OH_{34}$ (II) is formed

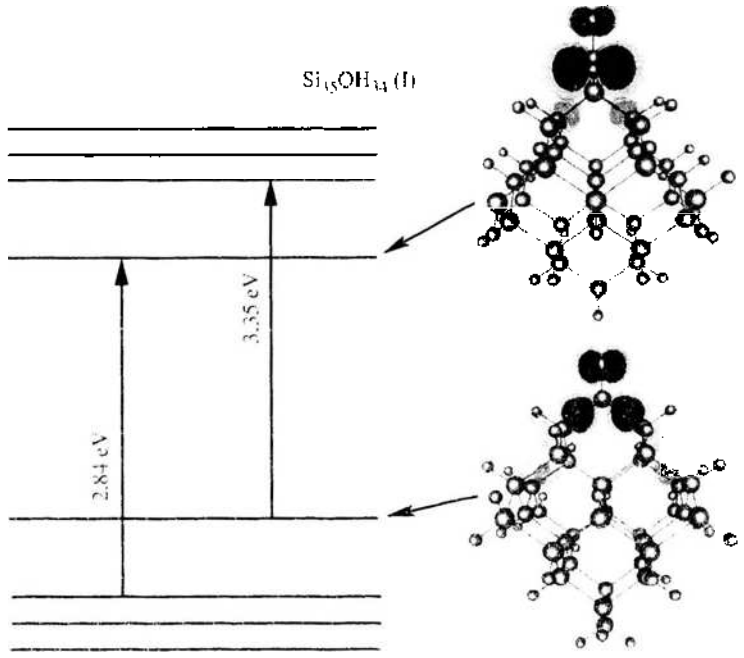


Figure 30. Schematic representation of electronic levels in the vicinity of the gap for $Si_{35}OH_{34}$ (II) clusters. Spatial distributions of electron densities are shown for the highest occupied molecular orbit and lowest unoccupied molecular orbit states.

mainly by transitions from lower occupied orbitals to the lowest unoccupied molecular orbit state. For this cluster, transitions from the highest occupied molecular orbit state to higher unoccupied orbitals do not contribute to optical absorption near the gap. For $Si_{35}OH_{34}$ (I), however, both of these types of electronic transitions are involved in the formation of the absorption edge.

Such calculations show that even a low concentration of oxygen on the surface can substantially alter the optical properties of silicon nanoclusters. However, experimental studies are not always limited to clusters with low oxygen content. Some limited studies have been performed on dots with a higher concentration of oxygen (e.g., $Si_{35}O_6H_{24}$). This cluster was prepared from the hydrogen-terminated dot $Si_{35}H_{36}$ by replacing 12 outer-shell hydrogen atoms with oxygen to form six double Si=O bonds at the positions symmetrically equivalent to that shown in Fig. 25 for $Si_{35}OH_{34}$ (II). The increase in oxygen coverage caused a further reduction of the absorption gap to 2.4 eV. This value was approximately 0.4 eV lower than the absorption gap for $Si_{35}OH_{34}$ (II), and almost 1.6 eV lower than the gap for the nonoxidized cluster $Si_{35}H_{36}$. The principal mechanism of gap formation for $Si_{35}O_6H_{24}$ appears to be similar to that for $Si_{35}OH_{34}$. The additional reduction of the absorption gap in the case of $Si_{35}O_6H_{24}$ could be explained by interactions among oxygen-induced electronic states. The absorption gap for $Si_{35}OH_{34}$ is reduced by the presence of localized oxygen-induced levels. In the limit of large clusters, the positions of these levels should be essentially independent of the cluster size. Because the gaps in silicon dots decrease with increasing cluster size as a result of diminishing quantum confinement, at some point the oxygen-induced states are expected to cross over the electronic levels from the body of the cluster. After this point, the oxygen-induced states would no longer be located inside the gap. Calculations indicate that depending on the fraction of oxygen coverage, the oxygen-induced states should not cross over the levels from the body of the cluster for silicon dots up to approximately 20–25 Å in diameter [153]. For larger dots, the overall effect of surface oxidation on the optical properties is likely to be less important.

7.2. Doping Quantum Dots

Electronic and optical properties of semiconductor nanostructures are strongly affected by quantum confinement because of the reduced dimensionality of these systems [159]. In nanocrystals or quantum dots, where motion of electrons (or holes) is limited in all three dimensions, quantum confinement results in a strong increase of the optical excitation energies when compared to the bulk. One expects that other electronic and optical properties such as the dielectric properties will be affected as well.

In bulk semiconductors, shallow donors (or acceptors) are crucial in determining the transport properties required to construct electronic devices. However, these properties are expected to be significantly altered in highly confined systems such as quantum dots. As a consequence, important questions exist as to whether dopants will continue to play a role similar to that in bulk semiconductors and on whether new applications such as quantum computation [160, 161] will become possible in the nano regime.

Experimental studies of shallow impurities in quantum dots, such as P in Si nanocrystals, have been slow to address such issues. In part, this is because of difficulties in the preparation of samples in a controllable manner (e.g., it is hard to ensure that a quantum dot contains only one impurity). For such reasons, only a few experimental studies have focused on the doping of quantum dots. These studies have used photoluminescence and electron spin resonance measurements, most of which have been performed on silicon quantum dots. Increasing the dopant concentration results in distinct changes in its photoluminescence properties such as suppression of the signal [162] and a blue-shift of photoluminescence maxima with decreasing particle size in heavily p-doped porous silicon [163]. It is also not clear whether or not the doping of Si nanocrystals provides a generation of free charge carriers in these systems [163, 164].

Electron spin resonance measurements are a popular tool for examining impurities in semiconductors and have recently been applied to these systems. Spin resonance experiments determine the hyperfine splitting (HFS) of the defective electron levels, which are

directly related to localization of the dopant electron density on the impurity site [165]. In Si nanocrystals with radii of 10 nm doped with P, a hyperfine splitting of 110 G has been observed [166]. This splitting is in sharp contrast to the bulk value of 42G. A size dependence of the HFS also exits in Si dots with radii around 50 nm [167], although in this case it is likely influenced by an asymmetrical shape of Si crystallites. Recently, a strong size dependence of the HFS of P atoms was observed in much smaller nanocrystallites with radii of 2–3 nm [168].

Theoretical studies of shallow impurities in quantum dots have also lagged relative to calculations for macroscopic systems. The large number of atoms and low symmetry hinder such studies. Some empirical studies have been performed for impurities in quantum dots [138, 169, 170]. These calculations involve various parameters, which are commonly assumed to have bulklike values. A common drawback in these studies is the use of a generic hydrogen-like potential to model the impurity atom.

The real space, *ab initio* pseudopotential density functional method [171] has been applied to the electronic properties of a single phosphorus impurity in a hydrogenated Si quantum dot containing hundreds of atoms [172]. The nanocrystals were modeled as spherical, bulk-terminated Si clusters whose surface is passivated by hydrogen atoms. One silicon atom is substituted by a phosphorus atom.

P, Si, and H atoms were modeled using *ab initio* Troullier–Martins pseudopotentials [63, 65]. Parameters for the Si and H pseudopotentials and other technical details are given elsewhere [171, 172].

Several substitutional geometries for the P atom have been explored (e.g., a P placed at the center of the dot, off center, and on the surface). No significant relaxations of the Si atoms were found in the vicinity of the P atom. The largest change in position occurred when the P was positioned off center. In this case, the P atom was shifted by about 0.1 a.u. in the outward direction.

In contrast to supercell approach, the real space method allows one to examine charged clusters in a straightforward manner [171]. Ionization energies I_d for P-doped nanocrystals and affinity energies A_p for pure Si nanocrystals have been determined using charged systems

$$I_d = E(n-1) - E(n) \tag{103}$$

$$A_p = E(n) - E(n+1) \tag{104}$$

where E is the ground state total energies of the n-, $(n+1)$-, and $(n-1)$-electron systems. The binding energy E_B for the donor atom can be calculated as a difference between these two quantities

$$E_B = I_d - A_p \tag{105}$$

This definition of the binding energy E_B corresponds to two separate processes. The doped dot is ionized (i.e., the electron is physically removed from the nanocrystal), and the affinity energy may be calculated by considering an isolated neutral dot of equal size and adding an electron. A similar approach has been used in the tight-binding calculations [138] in which the binding energy was calculated as a difference between the lowest conduction levels of the same crystallite with one excess electron both with and without impurity.

This definition of the binding energy for the donor atom can be contrasted with that for a bulk system, where this quantity is defined as the difference between the dopant electron level and conduction band continuum (i.e., the binding energy is equal to the ionization energy of the defect atom). In nanocrystals or quantum dots, such a definition is problematic because an electron being excited into an unoccupied state (below the vacuum level) will be confined by the physical size of the dot and will continue to interact strongly with the impurity atom.

The calculated ionization, affinity, and binding energy as functions of quantum dot radius R are shown in Fig. 31. The ionization energies for pure, hydrogenated Si nanocrystals are also given for comparison. The ionization and affinity energies for pure Si quantum dots have values close to those calculated recently for hydrogenated Ge nanocrystals. A surprising

feature in Fig. 31a is that the ionization energy I_d shows virtually no dependence on the size of the dot.

The dependence of $I_d(R)$ is different from the behavior of the ionization energy in Si quantum dots where this quantity is very large at small radii and gradually decreases, scaling as $R^{-1.1}$ to its bulk value. Although this dependence of the ionization energy on radius is weaker than R^{-2} law predicted by effective mass theory [173, 174], it is, nevertheless, a consequence of spatial confinement of electrons (holes) in quantum dots. It is surprising that this behavior is absent in the functional dependence of $I_d(R)$. The binding energy E_B, which scales as $R^{-0.8}$, is shown in Fig. 31b. These values are close to results of the tight-binding method [138], even though the ionization energy has a constant value in this range of sizes. Also plotted in Fig. 31b is the dependence of the "band gap" (i.e., the difference between the lowest level with single occupancy and the highest doubly occupied level) in P-doped systems. This quantity is blue-shifted with respect to the bulk values, where it should be approximately equal to Si band gap. Comparison with results for pure hydrogenated Si dots [140] of the same radius shows that it is smaller by about 10% than values of the highest occupied molecular orbit–lowest unoccupied molecular orbit gaps. The large values of the binding energy indicate that for dots in this size regime, the donors cannot be considered as shallow. This is largely a result of the weak screening present in quantum dots and the physical confinement of the donor electron within the dot.

The nature of the Si–P bond can be clarified by examining the charge density of the dopant electron $|\Psi(r)|^2$ for several dot sizes. In Fig. 32, we illustrate the charge profile for the case when the impurity is at the dot center. The density is plotted along [100] direction; results in other directions are similar. At all dots radii, the dopant wave function is strongly localized around the impurity site (i.e., the majority of the charge is within the P–Si bond length). From effective mass calculations [173, 174], it follows that the envelope wave function of the dopant electron is given by $j_0(\pi r/R) \propto \sin(\pi r/R)/r$. The calculated charge profile in Fig. 32 is at variance with this description. This difference in the spatial distributions can be attributed to the weaker screening in quantum dots. At these sizes, the dielectric constant is several times smaller than the bulk value [140, 144], giving rise to the increase of the effective electron-impurity potential and stronger localization of the electron around the defect atom.

Given the charge distribution of the dopant electron, one can evaluate the isotropic hyperfine parameter, which determines the contact interaction between the electron and defect nuclei. The method of Van de Walle and Blöchl [175] is used to extract the hyperfine parameters from pseudo-charge densities. The hyperfine parameters for a P atom positioned in the dot center are given in Fig. 33. At small sizes, the hyperfine parameter is very large because of the strong localization of the electron around impurity. As the radius increases, the value of A decreases. Our calculated results scale with radius R of the dot as $R^{-1.5}$

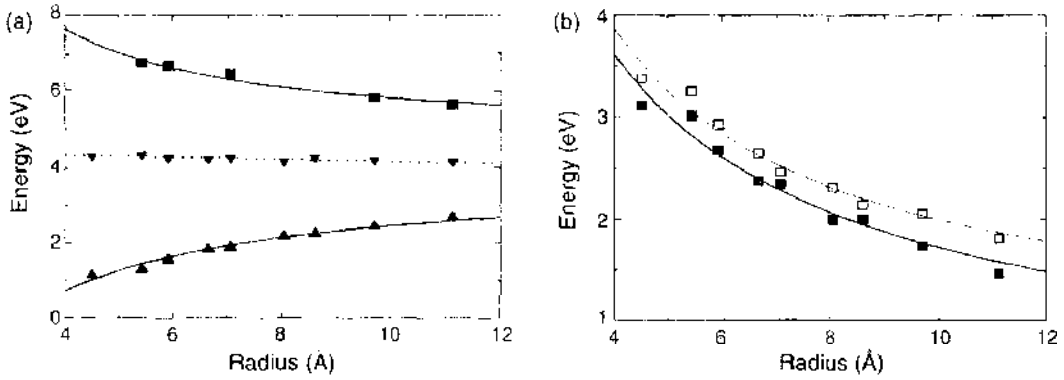


Figure 31. (a) Ionization energy I_d for phosphorus-doped nanocrystal (▼) and electron affinity A_p (▲) as a function of nanocrystal's radius R. Ionization energy for pure hydrogenated Si nanocrystals (■) is also shown. Solid lines are the best fits to calculated points, dotted line is a guide to an eye. (b) Binding energy E_B (■) and energy difference between defect level with single occupancy and highest occupied state with double occupancy (□) as a function of the dot's radius

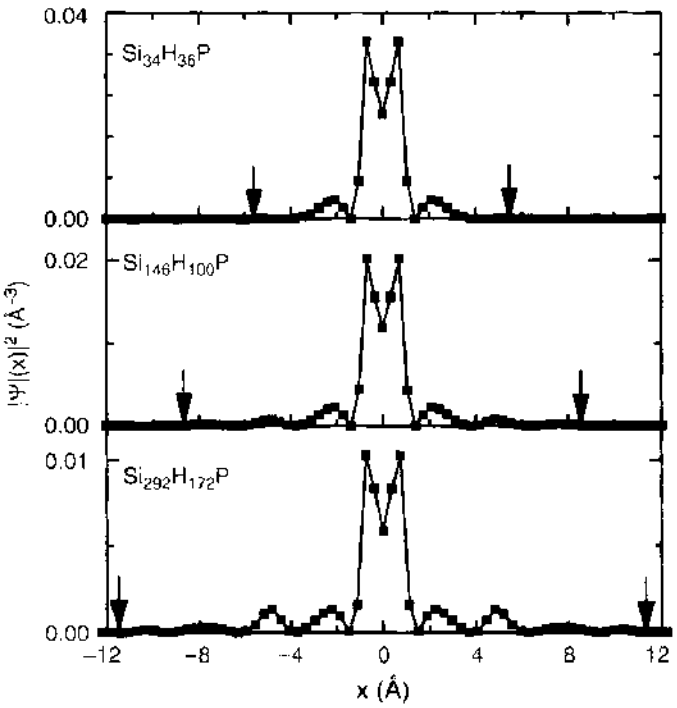


Figure 32. Charge density for the dopant electron along [100] direction. x/R is the x-coordinate normalized on the dot's radius R.

(effective mass theory gives R^{-3}). In Fig. 33, we also present the experimental data of Ref. [168]. The measured values of the hyperfine parameter falls on the best fit to calculated results; computational limitations prevent us from comparing directly to measured values.

The hyperfine values are not strongly dependent on the choice of the P site. Other sites have been tested by replacing one of the Si atoms in each shell with a P atom while retaining

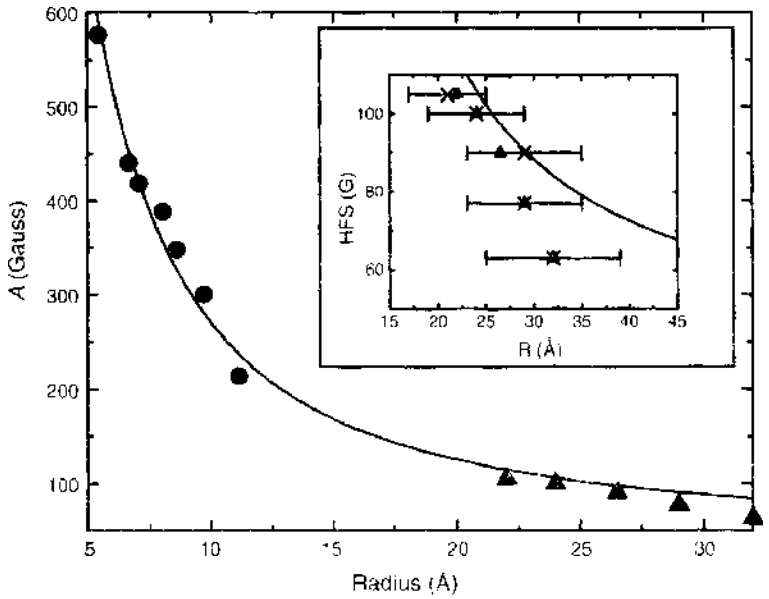


Figure 33. Calculated (•) and experimental (▲) isotropic hyperfine parameter A versus dot's radius R. The solid line is the best fit to calculations (bulk value of hyperfine parameter 42 G was used to obtain this fit). The inset shows experimental data of Ref. [162] together with the fit to results of calculations. Two sets of experimental points correspond to the average size of nanocrystals (×) and the size of nanocrystals (▲) estimated from comparison of photoluminescence energies for doped and undoped samples (for more details, see Ref. [162]).

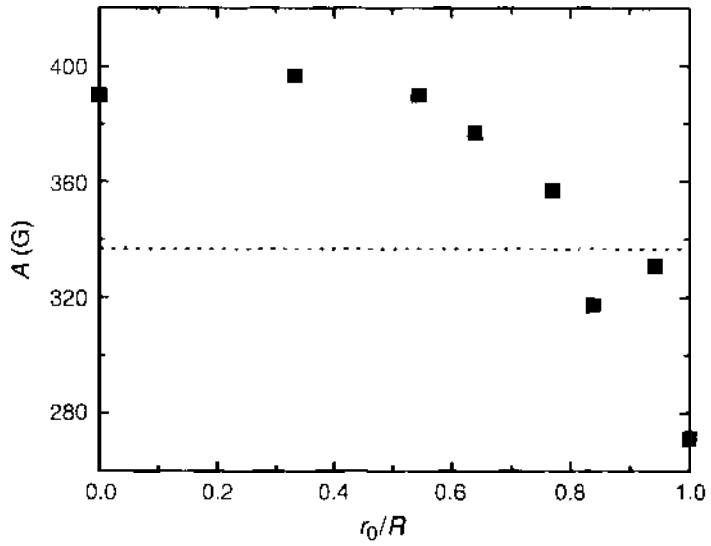


Figure 34. Isotropic hyperfine parameter A in $Si_{86}H_{76}P$ quantum dot as a function of the normalized position r_0/R of the P atom in the quantum dot. The dotted line is the average value of the hyperfine parameter.

the passivating hydrogen atoms. The ionization and binding energies were unchanged to within ~5%, independent of the impurity atom position [172].

The value of the isotropic hyperfine parameter also remains largely unchanged, save for the outermost layers of the dot. This behavior is demonstrated in Fig. 34, where the hyperfine parameter is plotted as a function of defect position in a representative dot $Si_{86}H_{76}P$. Near the surface, the P atom density becomes more delocalized and the hyperfine parameter shows a notable decrease in value from the value when P resides in the dot center. However, the average value of the hyperfine parameter over all sites is only about 15% lower than the value obtained when the P atom is at the center of the nanocrystal, which further increases agreement with experimental data.

ACKNOWLEDGMENTS

We would like to acknowledge support from the National Science Foundation, the United States Department of Energy, and the Minnesota Supercomputing Institute.

REFERENCES

1. L. T. Canham, *Appl. Phys. Lett.* 57, 1046 (1990).
2. J. M. Ziman, "Electrons and Phonons." Clarendon Press, Oxford, UK, 1960.
3. A. Haug, "Theoretical Solid State Physics." Pergamon Press, New York, 1972.
4. O. Madelung, "Introduction to Solid State Theory." Springer, Berlin, 1996.
5. M. Head-Gordon, *J. Phys. Chem.* 100, 13213 (1996).
6. L. H. Thomas, *Proc. Camb. Phil. Soc.* 23, 542 (1926).
7. E. Fermi, *Z. Phys.* 48, 73 (1928).
8. S. Lundqvist and N. H. March, "Theory of the Inhomogeneous Electron Gas." Plenum, New York, 1983.
9. C. Kittel, "Quantum Theory of Solids." Wiley, New York, 1987.
10. N. W. Ashcroft and N. D. Mermin, "Solid State Physics." Holt Rinehart, New York, 1976.
11. J. C. Slater, *Phys. Rev.* 81, 385 (1951).
12. J. C. Slater, "Quantum Theory of Molecules and Solids," Vol. 1–4. McGraw Hill, New York, 1974.
13. J. C. Slater, "Quantum Theory of Matter." McGraw Hill, New York, 1968.
14. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
15. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
16. D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* 45, 566 (1980).
17. J. M. Ziman, "Principles of the Theory of Solids." 2nd ed. Cambridge University Press, Cambridge, UK, 1972.
18. J. P. Perdew, K. Burke and Y. Wang, *Phys. Rev. B* 54, 16533 (1996).
19. J. R. Chelikowsky and S. G. Louie, Eds. "Quantum Theory of Materials." Kluwer, Boston, 1996.
20. E. K. U. Gross, J. F. Dobson, and M. Petersilka, "Density Functional Theory" (R. F. Nalewajski, Ed.). p. 81. Springer, Berlin, 1996

21. M. Petersilka, U. J. Gossmann, and E. K. U. Gross, *Phys. Rev. Lett.* 76, 1212 (1996).
22. V. R. Saunders and J. H. van Lenthe, *Mol. Phys.* 48, 923 (1983).
23. R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, *Mol. Phys.* 35, 771 (1978).
24. B. Bernu, D. M. Ceperley, and W. A. Lester Jr., *J. Chem. Phys.* 93, 552 (1990).
25. B. Bernu, D. M. Ceperley, and W. A. Lester Jr., *J. Chem. Phys.* 95, 7782 (1991).
26. A. J. Williamson, J. C. Grossman, R. Q. Hood, A. Puzder, and G. Galli, *Phys. Rev. Lett.* 89, 196803 (2002).
27. L. J. Sham and T. M. Rice, *Phys. Rev.* 144, 708 (1966).
28. L. Hedin, *Phys. Rev.* 139, A796 (1965).
29. M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* 34, 5390 (1986).
30. M. Casida, in "Recent Advances in Density-Functional Methods, Part I" (D. P. Chong, Ed.), p. 155. World Scientific, Singapore, 1995.
31. M. Casida, in "Recent Developments and Applications of Modern Density Functional Theory" (J. Seminario Ed.), p. 391, Elsevier, Amsterdam, 1996.
32. I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, *Phys. Rev. Lett.* 82, 1919 (1999).
33. K. Raghavachari, D. Ricci, and G. Pacchioni, *J. Chem. Phys.* 116, 825 (2002).
34. J. Jaramillo and G. E. Scuseria, *Theor. Chem. Acc.* 105, 62 (2000).
35. S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.* 314, 291 (1999).
36. K. Yabana and G. F. Bertsch, *Phys. Rev. B* 54, 4484 (1996).
37. K. Yabana and G. F. Bertsch, *Z. Phys. D* 42, 219 (1997).
38. K. Yabana and G. F. Bertsch, *Int. J. Quant. Chem.* 75, 55 (1999).
39. E. K. U. Gross and W. Kohn, *Phys. Rev. Lett.* 55, 2850 (1985).
40. E. K. U. Gross and W. Kohn, *Adv. Quant. Chem.* 21, 255 (1990).
41. G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* 74, 601 (2002).
42. L. Reining, V. Olevano, A. Rubio, and G. Onida, *Phys. Rev. Lett.* 88, 066404 (2002).
43. Y. H. Kim and A. Görling, *Phys. Rev. Lett.* 89, 096402 (2002).
44. J. P. Perdew, K. Burke, and Y. Wang, *Phys. Rev. B* 54, 16533 (1996).
45. A. D. Becke, *J. Chem. Phys.* 98, 5648 (1993).
46. M. E. Casida and D. R. Salahub, *J. Chem. Phys.* 113, 8918 (2000).
47. R. van Leeuwen and E. J. Baerends, *Phys. Rev. A* 49, 2421 (1994).
48. I. Vasiliev, *Phys. Stat. Sol. B* 233, 5 (2002).
49. G. D. Mahan, "Many-Particle Physics," Plenum, New York, 1981.
50. D. M. Ceperley, *Phys. Rev. B* 18, 3126 (1978).
51. J. P. Perdew and A. Zunger, *Phys. Rev. B* 23, 5048 (1981).
52. D. E. Beck, *Phys. Rev. B* 43, 7301 (1991).
53. I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, *Phys. Rev. B* 65, 115416 (2002).
54. W. R. Burdick, Y. Saad, L. Kronik, I. V. M. Jain, and J. Chelikowsky, *Comp. Phys. Comm.* 156, 22 (2003).
55. J. R. Chelikowsky and M. L. Cohen, "Handbook of Semiconductors" (T. S. Moss and P. T. Landsberg, Eds., 2nd ed. Elsevier, Amsterdam, 1992.
56. E. Fermi, *Nuovo Cimento* 11, 157 (1934).
57. D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* 43, 1494 (1979).
58. A. Zunger and M. L. Cohen, *Phys. Rev. B* 18, 5449 (1978).
59. G. P. Kerker, *J. Phys. C* 13, L189 (1980).
60. G. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* 26, 4199 (1982).
61. H. S. Greenside and M. Schlüter, *Phys. Rev. B* 26, 4199 (1983).
62. D. Vanderbilt, *Phys. Rev. B* 41, 7892 (1990).
63. N. Troullier and J. L. Martins, *Phys. Rev. B* 43, 1993 (1991).
64. S. G. Louie, S. Froyen, and M. L. Cohen, *Phys. Rev. B* 26, 1738 (1982).
65. L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* 48, 1425 (1982).
66. C. Y. Fong, "Topics in Computational Materials Science," World Scientific, Singapore, 1998.
67. A. Briley, M. R. Pederson, K. A. Jackson, D. C. Patton, and D. V. Porezag, *Phys. Rev. B* 58, 1786 (1997).
68. K. A. Jackson, M. R. Pederson, D. V. Porezag, Z. Hajnal, and Th. Frauenheim, *Phys. Rev. B* 55, 2549 (1997)
69. J. R. Chelikowsky and S. G. Louie, *Phys. Rev. B* 29, 3470 (1984).
70. R. W. Jansen and O. F. Sankey, *Phys. Rev. B* 36, 6520 (1987).
71. S. R. White, J. W. Wilkins, and M. P. Teter, *Phys. Rev. B* 39, 5819 (1989).
72. J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* 72, 1240 (1994).
73. J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, *Phys. Rev. B* 50, 11135 (1994).
74. J. R. Chelikowsky, N. Troullier, X. Jing, D. Dean, N. Biggeli, K. Wu, and Y. Saad, *Comp. Phys. Comm.* 85, 325 (1995).
75. E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Phys. Rev. B* 52, R5471 (1995).
76. F. Gygi and G. Galli, *Phys. Rev. B* 52, R2229 (1995).
77. G. Zumbach, N. A. Modine, and E. Kaxiras, *Solid State Commun.* 99, 57 (1996).
78. T. Ono and K. Hirose, *Phys. Rev. Lett.* 82, 5016 (1999).
79. Y. H. Kim, I. H. Lee, and R. M. Martin, *Comp. Phys. Comm.* 131, 10 (2000).
80. I. H. Lee, Y. H. Kim, and R. M. Martin, *Phys. Rev. B* 61, 4397 (2000).
81. T. L. Beck, *Rev. Mod. Phys.* 74, 1041 (2000).
82. J.-L. Fattebert and J. Bernholc, *Phys. Rev. B* 62, 1713 (2000).
83. B. Fornberg and D. M. Sloan, *Acta Numerica* 94, 203 (1994).

84. R. B. Morgan and D. S. Scott, *SIAM J. Sci. Comput.* 7, 817 (1986).
85. Y. Saad, A. Stathopoulos, J. R. Chelikowsky, K. Wu, and S. Öğüt, *BIT* 36, 563 (1996).
86. C. H. Tong, T. F. Chan, and C. C. J. Kuo, *SIAM J. Sci. Stat. Comput.* 13, 227 (1992).
87. B. N. Parlett, "The Symmetric Eigenvalue Problem," Prentice Hall, Englewood Cliffs, 1980.
88. Y. Saad, "Iterative Methods for Sparse Linear Systems," PWS Publishing, Boston, 1996.
89. R. Biswas and D. R. Hamann, *Phys. Rev. B* 34, 895 (1986).
90. R. Kubo, *Rep. Prog. Theor. Phys.* 29, 255 (1966).
91. H. Risken, "The Fokker-Planck Equation," Springer, Berlin, 1984.
92. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
93. R. Car and M. Parrinello, *Phys. Rev. Lett.* 60, 204 (1988).
94. N. Binggeli, J. L. Martins, and J. R. Chelikowsky, *Phys. Rev. Lett.* 68, 2956 (1992).
95. F. W. Kutzler and G. S. Painter, *Phys. Rev. B* 45, 288 (1992).
96. C. Rohlfing and K. Raghavachari, *J. Chem. Phys.* 96, 2114 (1992).
97. K. Raghavachari and V. Logovinsky, *Phys. Rev. Lett.* 55, 2853 (1985).
98. K. Raghavachari and C. Rohlfing, *J. Chem. Phys.* 94, 3670 (1990).
99. D. Deaven and K. M. Ho, *Phys. Rev. Lett.* 75, 288 (1995).
100. O. Cheshnovsky, S. H. Yang, C. L. Pettiet, M. J. Craycraft, Y. Liu, and R. E. Smalley, *Chem. Phys. Lett.* 138, 119 (1987).
101. D. Tomanek and M. Schluter, *Phys. Rev. Lett.* 56, 1055 (1986).
102. N. Binggeli and J. R. Chelikowsky, *Phys. Rev. B* 50, 11764 (1994).
103. S. Öğüt and J. R. Chelikowsky, *Phys. Rev. B* 55, R4914 (1997).
104. E. C. Honea, A. Ogura, C. A. Murray, K. Raghavachari, O. Sprenger, M. F. Jarrold, and W. L. Brown, *Nature*, 366, 42 (1993).
105. X. Jing, N. Troullier, J. R. Chelikowsky, K. Wu, and Y. Saad, *Solid State Comm.* 96, 231 (1995).
106. R. Fournier, S. B. Sinnott, and A. E. DePristo, *J. Chem. Phys.* 97, 4149 (1992).
107. R. Schäfer, S. Schlect, J. Woenckhaus, and J. A. Becker, *Phys. Rev. Lett.* 76, 471 (1996).
108. S. Baroni, P. Giannozzi, and A. Testa, *Phys. Rev. Lett.* 58, 1861 (1987).
109. X. Gonze, D. C. Allan, and M. P. Teter, *Phys. Rev. Lett.* 68, 3603 (1992).
110. J. L. Martins, J. Buttet, and R. Car, *Phys. Rev. B* 31, 1804 (1985).
111. I. Moullet, J. L. Martins, F. Reuse, and J. B. J. Buttet, *Phys. Rev. B* 42, 11598 (1990).
112. V. Bonacic-Koutecky, P. Fantucci, and J. K. Koutecky, *J. Chem. Phys.* 93, 3802 (1990).
113. V. Bonacic-Koutecky, P. Fantucci, and J. Koutecky, *Chem. Phys. Lett.* 166, 32 (1990).
114. G. Onida, L. Reining, R. W. Godby, R. del Sole, and W. Andreoni, *Phys. Rev. Lett.* 75, 181 (1995).
115. C. R. C. Wang, S. Pollack, D. Cameron, and M. M. Kappes, *Chem. Phys. Lett.* 93, 3787 (1990).
116. U. Itoh, Y. Toyoshima, H. Onuki, N. Washida, and T. Ibuki, *J. Chem. Phys.* 85, 4867 (1986).
117. S. Furukawa and T. Miyasato, *Phys. Rev. B* 38, 5726 (1988).
118. D. J. Lockwood, A. Wang, and B. Bryskiewicz, *Solid State Commun.* 89, 587 (1994).
119. M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allan, and C. Delerue, *Phys. Rev. Lett.* 82, 197 (1999).
120. T. Takagahara and K. Takeda, *Phys. Rev. B* 46, 15578 (1992).
121. M. Rohlfing and S. G. Louie, *Phys. Rev. Lett.* 80, 3320 (1998).
122. N. A. Hill and K. B. Whaley, *Phys. Rev. Lett.* 75, 1130 (1995).
123. C. Delerue, M. Lannoo, and G. Allan, *Phys. Rev. Lett.* 76, 3038 (1996).
124. C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 48, 11024 (1993).
125. L. W. Wang and A. Zunger, *J. Phys. Chem.* 100, 2394 (1994).
126. B. Delley and E. F. Steigmeier, *Phys. Rev. B* 47, 1397 (1993).
127. R. J. Baierle, M. J. Caldas, E. Molinari, and S. Ossicini, *Solid State Commun.* 102, 545 (1997).
128. A. Franceschetti, L. W. Wang, and A. Zunger, *Phys. Rev. Lett.* 83, 1269 (1999).
129. S. Öğüt, J. R. Chelikowsky, and S. G. Louie, *Phys. Rev. Lett.* 1999, 1270 (1999).
130. R. W. Godby and I. D. White, *Phys. Rev. Lett.* 80, 3161 (1998).
131. S. Öğüt, J. R. Chelikowsky, and S. G. Louie, *Phys. Rev. Lett.* 80, 3162 (1998).
132. I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, *Phys. Rev. B* 60, 8477 (1999).
133. C. Troparevsky, L. Kronik, and J. R. Chelikowsky, *Phys. Rev. B* 65, 33311 (2002).
134. D. J. Lockwood, *Solid State Comm.* 92, 101 (1994).
135. J. P. Proot, C. Delerue, and G. Allan, *Appl. Phys. Lett.* 61, 1948 (1992).
136. L. W. Wang and A. Zunger, *J. Phys. Chem.* 98, 2158 (1994).
137. N. A. Hill and K. B. Whaley, *Phys. Rev. Lett.* 76, 3039 (1996).
138. M. Lannoo, C. Delerue, and G. Allan, *Phys. Rev. Lett.* 74, 3415 (1995).
139. G. Allan, C. Delerue, M. Lannoo, and E. Martin, *Phys. Rev. B* 52, 11982 (1995).
140. S. Öğüt, J. R. Chelikowsky, and S. G. Louie, *Phys. Rev. Lett.* 79, 1770 (1997).
141. L. Brus, *J. Phys. Chem.* 90, 2555 (1986).
142. Y. Kayanuma, *Phys. Rev. B* 38, 9797 (1988).
143. A. Franceschetti and A. Zunger, *Phys. Rev. Lett.* 78, 915 (1997).
144. S. Öğüt, R. Burdick, Y. Saad, and J. R. Chelikowsky, *Phys. Rev. Lett.* 90, 127401 (2003).
145. I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, *Phys. Rev. Lett.* 78, 4805 (1997).
146. G. Srinivasan, *Phys. Rev. B* 178, 1244 (1969).
147. R. Schäfer and J. A. Becker, *Phys. Rev. B* 54, 10296 (1996).

148. S. Schuppler, S. L. Friedman, M. A. Marcus, D. L. Adler, Y. H. Xie, F. M. Ross, Y. J. Chabal, T. D. Harris, L. E. Brus, W. L. Brown, E. E. Chaban, P. F. Szajowski, S. B. Christman, and P. H. Citrin, *Phys. Rev. B* 52, 4910 (1995).
149. J. von Behren, T. van Buuren, M. Zacharias, E. H. Chimowitz, and P. M. Fauchet, *Solid State Comm.* 105, 317 (1998).
150. Y. Kanemitsu, H. Uto, Y. Masumoto, T. Matsumoto, T. Futagi, and T. Mimura, *Phys. Rev. B* 48, 2827 (1993).
151. V. Lehmann and U. Gösele, *Appl. Phys. Lett.* 58, 856 (1991).
152. I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, *Phys. Rev. Lett.* 86, 1813 (2001).
153. I. Vasiliev, J. R. Chelikowsky, and R. M. Martin, *Phys. Rev. B* 65, 121302 (2002).
154. A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, *Phys. Rev. Lett.* 88, 097401 (2002).
155. L. Mitas, J. Therrien, R. Twesten, G. Belomoin, and M. H. Nayfeh, *Appl. Phys. Lett.* 78, 1918 (2001).
156. I. Vasiliev and R. M. Martin, *Phys. Stat. Sol. (b)* 233, 5 (2002).
157. E. Martin, C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* 50, 18258 (1994).
158. T. Takagahara and K. Takeda, *Phys. Rev. B* 53, R4205 (1996).
159. A. D. Yoffe, *Adv. Phys.* 50, 1 (2001).
160. B. E. Kane, *Nature* 393, 133 (1998).
161. D. P. DiVencenzo, *Nature* 393, 113 (1998).
162. A. Mimura, M. Fujii, S. Hayashi, D. Kovalev, and F. Koch, *Phys. Rev. B* 62, 12625 (2000).
163. G. Mauckner, W. Rebitzer, K. Thonke, and R. Sauer, *Solid State Comm.* 91, 717 (1994).
164. D. Kovalev, H. Heckler, G. Polisski, and F. Koch, *Phys. Status Solidi B* 215, 871 (1999).
165. G. Feher, *Phys. Rev.* 114, 1219 (1959).
166. J. Müller, F. Finger, R. Carius, and H. Wagner, *Phys. Rev. B* 60, 11666 (1999).
167. B. J. Pawlak, T. Gregorkiewicz, C. A. J. Ammerlaan, and P. F. A. Alkemade, *Phys. Rev. B* 64, 115308 (2001).
168. M. Fujii, A. Mimura, S. Hayashhi, Y. Yamamoto, and K. Murakami, *Phys. Rev. Lett.* 89, 206805 (2002).
169. C. Y. Fong, H. Zhong, B. M. Klein, J. S. Nelson, *Phys. Rev. B* 49, 7466 (1994).
170. I. H. Lee, K. H. Ahn, Y. H. Kim, R. M. Martin, and J. P. Leburton, *Phys. Rev. B* 60, 13720 (1999).
171. J. R. Chelikowsky, *J. Phys. D: Appl. Phys.* 33, R33 (2000).
172. D. Melnikov and J. R. Chelikowsky, *Phys. Rev. Lett.*, 92, 046802 (2004).
173. L. E. Brus, *J. Chem. Phys.* 79, 4403 (1983).
174. L. E. Brus, *J. Chem. Phys.* 79, 5566 (1983).
175. C. G. Van de Walle and P. E. Blöchl, *Phys. Rev. B* 47, 4244 (1993).

Index

A

AB diblock copolymers, 108, 116, 145
Ab initio calculations, 26, 79
Ab initio methods, 2, 46, 63, 616, 701
Ab initio simulations, 60
Ab initio total energy calculations, 56
Absorbing boundary conditions (ABCs), 578, 587, 592, 594
Absorption frequency, 75
Absorption spectroscopy, 210
  hexatriene, 210
  jet-cooled butadiene, 210
  octatetraene, 210
Ad atoms, 451
ADIF code, 522
Adsorption
  on laterally structured substrates, 183
AFM cross-section image, 705
Aharanov-Bohm effect, 529
Al–Al intracluster bond length, 96
Al-based clusters, 95
  assembling of, 95
$Al_{12}C$ cluster dimer, 96
Aliasing, 778
Alkali clusters, 66, 74, 77, 83
  cationic, 75
  high valence impurities in, 83
Alkali–lead clusters
  simulation of assembling of, 94
Alkali metal clusters, 77, 87
  monovalent and divalent impurities in, 77
All-angle negative refraction (AANR), 548, 549, 579, 580
  metamaterial slab with, 580
All-electron
  calculation, 61
  computer codes, 61
  methods, 618
  potential 814
κ-Allumina, 627
  electrostatic repulsion, 627
  structure determination, 627
Alumina, 258
Alumina (γ and η), 624
  Cr atom adsorbed, 632
  two forms, 624
θ-Alumina transformation, 634
Aluminium clusters, 85
  impurities in, 85
AM1 method, 227, 494, 495
Amorphous phase, 412, 413, 420
Amorphous silicon, 408, 409
Amorphous state, 419
Amorphous zone, 412, 419
  bottom view of development, 413
AM1 quantum-chemical semiempirical method, 482
Analysis techniques, 620
Andreev reflection amplitude, 657
Andreev reflection and proximity effect, 648
Andreev reflection coefficient, 659
Andreev reflection process, 649
Andreev scattering probability, 661
Andreev tunneling, 672
Angular momentum, 58
  components, 60
  quantum number, 68
Anionic clusters, 67, 68
Anisotropic effects, 711
  anisotropic Green's functions, 712
Anisotropicity and shape, 727
Anisotropic media
  artificial, 544
  natural, 544
  polarization effects, 544
Anisotropic metamaterial
  cell containing nonspherical object, 545
Anisotropic metamaterial slab
  permittivity dependence, 546
  wavelength dependence of the transmission coefficients, 545
Anisotropic modulus, 314
Anisotropy, 412
Annealing process, 63
Anodic aluminum oxide (AAO), 263
  SEM micrographs, 264
Antiphase bonds, 524
AO functions, 618
Apex angles, 512–519, 524, 528, 533
  frequency histogram of the natural occurence, 517
  in Kværner's carbon-black process, 518

ISBN: 1-5888-050-0

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 8: Pages (853–880)

Approximate energy functional, 49
Arbitrary material, 560
Arbitrary vector, 721
Arc discharge method, 261
  SEM image of SWNT bundles, 261
Armchair tubes, 441, 443
  atomic chain of, 444
  changes in bond angles and bond lengths, 443
  structural changes in different stages, 442
AR-MWNT fracture surface, 295
Artificial electron configuration, 59
Artificial intelligence (AI), 538
Artificial media, 538
Artificial nacre, 289
Asperity
  cross-sectional view, 439
  diamond, 430–432
  penetration depth, 434
  radius, 431
  sizes, 430
Atomic and electronic coordinates, 799
Atomic arrangement, 63
Atomic density distribution, 90
Atomic distance
  between atoms, 406
Atomic force microscopy (AFM), 237
Atomic friction, 420
Atomic orientations, 435
Atomic-scale friction, 420
Atomic scale mechanism, 35
Atomistic approaches, 742
Atomistic model
  for calculating stresses, 400
Atomistic simulation, 450
Attenuation constants, 588
Au_{20}
  calculated ground state structure of, 68
Austin model-1 (AM1), 226
Austin model package (AMPAC), 461
Au substrate, 35
Au/titania system, 632
Autocatalysis, 471
Avogadro's number, 364
Axisymmetric cylinder model, 373
Azimuthal angle, 523

B

Back bond oxidation, 413
Ball-and-stick models, 831
  for hydrogenated silicon clusters, 831
Band diagram computation, 582, 585
Band diagrams, 559
Band model, 206
Band structure
  calculation, 583
  for E-polarization and H-polarization, 559
  of photonic crystal waveguide, 562
Band structure term, 522
Band-to-band dielectric function, 755
Band transition, 206
Bardeen-Cooper-Schrieffer (BCS) theory, 649
  derivation of, 650
Barker–Henderson recipe, 166
Bathochromic shifts, 205
Bead-spring model, 154, 163, 178
Bead-spring off-lattice model, 153
Beam theory, 309
Bending energy, 513
Bending/unbending transition, 183
p-Benzene dithiolate molecule, 238
Berendsen scheme, 446
Berendsen thermostat, 442–446
  scheme, 440
Bessel functions, 524
Bethe–Salpeter equation (BSE), 226
Bethe–Salpeter technique, 831
Betti's reciprocal theorem, 713
Bicapped antiprism structure, 825
Biharmonic equation, 717
Binary particle systems, 121
Binary polymer blends,
  behavior of nanorods in, 126
Binding energy (BE), 82, 84–85, 826
Biomass-derived alcohols, 630
Birch reduction, 494
Bisdithienothiophene (BDT), 227
Bit string, 600
Bloch theorem, 774
Bloch waves, 62
Bloch wave-vector, 662
Blonder, Tinkham, and Klapwijk (BTK), 657
B3LYP/3-21G
  of C_{72} isomers, 475, 476
  of C_{74} isomers, 477
  optimized structures, 476
B3LYP/3-21G~dz
  entropy, 490
  level, 488, 492
  optimized structures of Ca@ C_{72} isomers, 488
  optimized structures of Ca@ C_{74} isomers, 490, 491
  optimized structures of Ca@ C_{82} isomers, 491
B3LYP/6-31G*, 461, 475, 495
  computations, 474
  energetics, 477, 478, 489, 490, 492
  level, 494, 495
Body-centered cubic structure, 409
Boehmite, 618
Bogoliubov–de Gennes equations, 648, 649
  solutions of, 654
Bogoliubov–de Gennes Hamiltonian HBG, 653
Bogoliubov's self-consistent field method, 650
Bogoliubov–Valatin transformation, 654
Bohr magneton, 678
Bohr radius, 689
Boltzmann average, 161
Boltzmann constant, 10, 65, 88, 152, 756
Boltzmann distribution function, 625

Boltzmann factors, 462, 463
  simple, 463
Boltzmann weight, 156, 159, 162
Bond-fluctuation model, 153, 155
Bonding potentials, 165
Bond length, 407, 823
Bond-length alternation (BLA), 213
Bond-order potential, 6, 36
Born–Oppenheimer approximation, 2, 45, 519
Born–Oppenheimer products, 759
Born–Oppenheimer surface, 823
Boron nitride (BN), 259
Bose–Hubbard model, 680
Bottom-up approaches, 299
Boundary displacement effect, 398
Boundary methods, 580–583, 587
Boundary temperature effect, 398
Bowden–Tabor formula, 421
Bragg scattering, 608
Bravais lattice, 744
Brenner potential, 520
Brillouin zone, 62–63, 523, 524, 608, 662
Broken symmetry, 586
Brownian dynamics model, 105, 109
Brownian motion, 823
Brute-force approach, 590
BTK formula, 660
Buckling
  background, 449
  of diamond thin films, 449
  theory and simulation method, 450
Buckling pattern
  after releasing the biaxial stresses, 453
  after releasing the uniaxial stress, 451
  anisotropic effect, 453
  approaches, 453
  telephone-cord-like, 452
Buckyministerfullerene C60, 362
Buckypaper, 286
Bucky-shuttle system, 33
Bulk composite behavior, 387
Bulk energy, 66
Bulk Hamiltonian, 743
Bulk polymers, 208
  conductivity, 208
Burgers vector, 423, 433
Buried strained layer, 711
  illustration of, 711

C

Ca@C_{72} isomers, 488, 489
Ca@C_{74} isomers, 489
  structure, 489
Ca@C_{82} cages, 492
  symmetries, 492
Ca@C_{82} isomers, 491, 492
  structure–energetics relationships, 492
Cage separation energies, 482
Cahn-Hilliard (CH)
  approach, 108
  model, 109
  theory, 105
Cahn-Hilliard/Brownian dynamics (CH/BD)
  approach, 105
  calculation, 119, 129
  hybrid method, 117
  model, 105, 108, 116–117, 126
  simulations, 105, 119, 126
  technique, 116
Canonical ensemble, 67
Canonical free energy, 157
Canonical partition function, 160
Capillary condensation, 173
Capped double icosahedron, 89
Capsid protein (CA) helical polymorphism, 510
Capture energy, 84
Capture viscoelastic behavior, 111
Carbon black, 290
Carbon black-filled polymers, 255
Carbon chemistry
  in two and three dimensions, 510
Carbon clusters, 459
  basic topology, 463
  cyclic rings, 468
  energetics and thermodynamics, 463
  ion chromatography, 469
  monoatomic, 466
  small, 468
  stabilities of different dimensions, 465–468
Carbon fiber-epoxy composites, 256
Carbon nanocoils, 259, 264
  tensile loading of, 266
Carbon nanocones
  atomic and electronic structure, 519
  effect of substitutional atoms, 531
  formation, 509
  introduction, 507
  meta-stable structures, 526
  molecular mechanics, 519
  theoretical/computational methods and predictions, 519
Carbon nanotube/polymer composites, 363
  conformation of polymer matrix, 364
  experimental investigation of, 388
  interfacial bonding of, 363, 388
  mechanical behavior of, 377
  mechanical properties of, 390
Carbon nanotubes (CNTs), 10–13, 17, 19–20, 29, 34, 259
  annealed, 28
  armchair fluorinated, 26
  bone-shaped, 263
  characterization, 440
  chemically modified single-walled, 26
  chemical modification of, 22, 24
  comparison with other materials, 262
  computationally constructed, 17
  conductivity of, 27
  critical issues, 362

deformation, 35
density of states of, 26
direct covalent functionalization of, 22
discovery of, 362
disperse, 22
double-walled, 34
electron charge transfer, 28
electronic properties of, 26
equivalent continuum representation of, 306
fullerene structure, 362
gas interactions with, 28
gas sensors, 26
herring-bone, 519
hydrogen-terminated, 29
junctions, 17–19
local DOS of, 17
mechanical properties of, 19, 21
modeling, 441
multiwalled, 21, 24, 34
multiwalled nanotubes (MWNT), 259, 440
opened single-walled, 29
photomicrographs of, 279
potentials, 35
purified, 22
semiconducting single-walled, 26–27
single-walled, 12–14, 18–24, 26–32
single-walled nanotubes (SWNT), 259, 440, 532
structure of, 22
surface, 32
T-junction, 17–18
unit cell and chiral vector for, 259
walls, 31
Y-junctions, 17–18
Young's moduli of, 21
zigzag single-walled, 13
zipper-like mechanism, 23
Carbon-vapor pressure, 466, 467
saturated, 467
Carnahan-Starling equation, 107, 167
Car–Parinello density-functional calculations, 152
Car–Parrinello method, 823
Car–Parrinello molecular dynamics (CPMD), 520, 521
Car–Parrinello simulations, 521
Carr–Parinello technique, 9
Cartesian coordinates, 461, 541
vibrational analysis, 461
Cartesian coordinate system, 821
CASTEP (Cambridge Sequential Total Energy Package), 617
Catalytic chemical vapor deposition (CCVD), 389
Catalytic systems, 630
based on supported metal particles, 630
stability of, 634
Cathodic-arc carbon coatings, 450
Cationic alkali cluster, 75
C_{82} cage topologies, 482
C–C distance, 444
ccp oxygen lattice, 625
C_3:C_1 ratio, 494
CDM devices, 595
Center of mass distribution, 107
Central finite difference (CDS), 782
Ceperley–Alder functional, 811
Cerium oxide (ceria), 628
ceria for hydrogen, 628
properties, 628
$C_{60}F_{36}$ isomers, 494
C_1 isomer, 494
C_3 isomer, 494
$C_{60}H_{36}$, 494
isomers, 495
Chacon, Alvarellos, and Tarazona (CAT) functional, 56
Chain-stretching effects, 141
Charge distribution, 48
Charge-neutral systems, 5
Chemical passivity, 92
Chemical potential, 47, 161, 173
Chemical vapor deposition (CVD), 262
grown NTs in an epoxy, 287
Chemisorption, 152
Chiral elements, 547
Chiral media
artificial, 546
Chiral molecule
cell with a right-handed, 547
Chiral slab, 547
Chiral vector, 371
Chromia/alumina catalysts, 631
Chromium clusters, 622
Chromosomes, 600
Circuit theory, 669
Circumferential breathing frequency, 381
C_{2v} isomers
NMR patterns, 487
C_4 isomers
cyclic, 468, 469
energy differences, 470
linear, 468, 469
relative concentrations, 469
C_{78} isomers
^{13}C NMR spectra, 479
mole fractions, 480
symmetries, 479
C_{84} isomers, 482, 483
Classical continuum mechanics, 369
analytical modeling with, 369
Classical trajectory methods, 380
Clausius–Clapeyron equation, 467
Clay-filled polymers, 268
microstructures of, 268
Clay-nanocomposites, 288
partially intercalated and partially exfoliated, 314
Closed-shell configuration
40-electron, 95

Cluster-assembled solid, 44
  applications, 63–97
  electric polarizability of, 70
Cluster assembly, 92, 96
Cluster-cluster distance, 96
Cluster derived structures (CDS), 790
Clustering degree, 468
Cluster melting, 87
Cluster models, 616
Cluster radius, 70
Clusters, 822
  dynamical matrix, 826
  of aluminum group, 69
  structure, 822
  vibrational spectra, 826
Cluster shape, 89
  subshells and distortion of, 73
Cluster structure
  description of, 79
Cluster temperature, 88
Coarse-grained model, 152–155
Coefficient of thermal expansion (CTE), 300, 367
Coexistence chemical potential, 157
Cofacial dimmers, 227
Cohesive energy, 65, 95
Collective dipole resonance, 75
Collective electronic excitations, 74
Collective oscillation, 75
Complete active-space self-consistent field (CASSCF), 210
  wave-function, 210
Complex eigenvalue
  comparison of the frequency dependence, 591
  problems, 556
Complex periodic constant analysis
  with respect to number of PhC layers, 590
Composite applied strain, 282
Composite doped metamaterials (CDMs), 539, 577, 590, 595
  based sensors, 605
  optimization, 595, 596, 599
  selection of numerical methods, 578
  simulation of, 577
  structures, 539, 540, 577, 579, 590, 595, 596, 599, 601, 603, 605
Composite model, 334
  two models, 334
Composite modulus, 311
Composites, 254
  advantages of, 254
  isotropic properties of, 317
Compressed row storage (CRS)-format, 783
Computational chemistry, 459, 616
  periodic boundary conditions, 617
  three categories, 616
Computational methods, 2
Computational modeling, 616
Computation–observation agreement, 486
Conductance, 665
  two-probe differential, 665
Conductivity, 237
Cone-covered graphite aggregate
  FESEM images, 516
Cone formation, 510
Cone-helix growth model, 516
Cone shapes
  HRTEM images, 508, 509
Configurational bias algorithm, 156, 168
Configuration interaction (CI), 4, 211, 772
  calculations, 81
Configuration interaction including single excitations (CIS), 212
Confined systems, 822
  properties of, 822
Confining potential, 66–67
Conformational strain, 24
Conic shapes
  deviations from the ice-cream shape, 513, 514
  geometrical stability considerations, 513
  in graphite, 510
  topological stability considerations, 514
Conic tip, 518, 519
Conjugated organic polymers (COPs), 204
  conducting state of, 229
  optical properties of, 209
  properties of, 204
Conjugate gradient, 8
  method, 810
Constant-current surface, 530
Contact angle, 170
  macroscopic, 171
  of microscopic droplets, 170
Contact lengths
  the JKR and molecular dynamics analyses, 433
Contact size, 430, 433, 434
Contact width, 431
Continuum mechanics, 263, 400
Continuum modeling, 380
Continuum models, 523, 528, 529, 735
Contour map, 96
Control volume, 397, 421
  moving, 421–423
  size of, 398
Convenient method, 822
Conventional density-functional formalism, 807
Convex/concave angle, 527
Coordination numbers, 406
Copper
  temperature distribution in the atomic lattice, 427
Copper monocrystal, 422
Copper nanowhiskers
  atomic model, 447
  deformation, 446
  effects of atomic orientation and specimen shape, 447

effects of size, shape, and surface atomic structure, 448
modeling, 447
Cooperon propagator, 673
Cooper-pair box, 672
Cooper pairs, 649
Cooper pair tunneling, 675
Coordinate relaxation, 777
Copper specimen
distribution of dislocations in the subsurface, 425
CO_2 reduction, 623
schematic of, 623
Correlation function
direct, 167
Corundum, 619
Coulomb blockade, 671, 672
Coulomb energy, 76, 822
Coulombic empirical potentials, 31
Coulombic potentials, 29, 45
Coulomb interactions, 521
Coulomb repulsions, 239, 758
Coupler
optimized design, 572
CPU time, 522
Cr/alumina system, 631
Crank–Nicholson scheme, 162
Critical endpoint (CEP), 181
Critical slowing-down (CSD) phenomenon, 776
Crossover algorithm, 600
Crystalline material cools, 8
Crystallographers' cubic unit cell, 625
C_n structures
cyclic and linear, 470
relative concentrations, 470
C_{60} structures, 487
C_{2v} structure, 487
NMR lines, 487
Cubic close packing, 625
Cuboidal system geometry, 157
Cumelenic form, 444
Cumulative distribution function (CDF), 121, 123–124, 138, 140
Current density contours, 130
Current spin-density functional theory (CSDFT), 785
Cut-and-paste procedure, 525
for making a nanocone, 512
Cutoff energy, 618
Cutoff frequencies, 585, 586, 592
Cutting deformation, 421
Cutting regime, 423–427
CVD diamond films, 450
CVD-grown MWCNTs, 261
bimodal histogram of, 261
$C_{60}X_{36}$
C_3 isomer, 493, 494
schlegel diagrams of 15 isomers, 493
T isomer, 493, 494
Cyclic indentations, 417

D

DACAPO, 617
Dangling bonds, 745
Dark atoms, 423
Davidson procedure, 820
Davydov components, 224
Davydov splitting, 224
Debye frequency, 678
Debye's model, 399
Defect-free cell, 634
Deformation
characteristics, 409, 441
cross-sectional view, 435
molecular dynamics prediction, 435
of copper nanowhiskers, 446
plastic, 438
regimes, 436, 437
Deformation regimes, 423, 424, 429
cutting regimes, 424
no-wear regime, 424
transition diagram of, 424
Deformation zone, 424, 426
Delocalized orbitals, 239
Density correction, 781
linear integral equation, 781
Density distribution, 48
Density fluctuations, 156, 177
Density-functional approaches, 803
Density-functional calculations, 152
of clusters and cluster assembly, 43–97
quantum mechanical, 152
Density functional component, 142
Density functional methods, 55
Density functional theory (DFT), 2, 47, 49, 54, 60, 63, 79, 104, 108, 209, 460, 470, 520, 521, 525, 618, 774, 805
ab initio, 77
approach, 106, 461
calculational schemes, 92
calculation of excitation energies employing, 213
calculations, 4, 23, 29, 31–32, 36, 66, 68, 72, 77, 82, 85, 89
Car–Parinello, 152
codes, 4
exchange-correlation functionals in, 28, 79
formalism, 160
Kohn-Sham form of, 46, 52, 91
LDA calculations, 19
linear response, 70
MD simulations, 31
methods, 3
simulation, 54, 85, 91
simulations of STM images, 529, 530
snapshots of, 723
Tarazona, 107
Density gradient, 56
Density of states (DOS), 689

calculated local, 17
Density profiles, 176, 182
one-dimensional, 134–135
radially averaged, 172
two-dimensional, 133
Deposition technique, 264
Destabilizing effect, 790
Deterministic algorithms, 599
Deterministic optimizers, 596, 598, 599, 603, 604
Dewetting experiments, 178
Dewetting morphology, 175
Dewetting process
spinodal, 175
3D-FDTD
analysis, 608
of H.S experiment, 607
of W1-waveguide, 606
optimized power splitter, 611
simulation, 606, 607, 609
DFT algorithm, 522
DFT-LDA calculations, 19
DFT levels, 475, 492
DFT-MD simulations, 31
DFT orbitals, 215
band gap problem, 216
HOMO-LUMO gaps, 215
DFT simulations
of melting, 91
Diamond carbon coating
silicon (100) and (110) surfaces, 451
Diamond asperity, 430, 431, 432
Diamond–copper interactions, 433
Diamond–copper interface, 422
Diamond–copper sliding systems
nanofriction, 429
nanowear, 429
Diamond films, 449, 450
Diamond indenter, 410
Diamondoid covalent solids, 380
Diamond–silicon attraction, 436
Diamond–silicon sliding systems, 433
modeling, 433
Diamond slides, 424, 425, 427
Diblock composition, 133, 136–137
Diblock-confined particle systems, 121
Diblock copolymeric system, 122
Diblock copolymers, 106, 117, 120–122, 124, 293
architecture, 121
asymmetric filled, 145–146
domains, 118, 121
filled, 118
morphological studies, 132
nanoparticle-filled, 142–147
particles confined in, 118
periodic structure of, 145
properties of mixtures of, 116, 132
root mean-squared end-to-end distance of, 143
symmetric, 132, 143
symmetric filled, 144
symmetric unfilled, 143
systems, 118
Diblock entropic free energy, 107
Diblock matrix, 133
Dielectric constant, 763
Dielectric functions, 837
Dielectric lens
with relative permittivity, 553
Diffuse jellium model (DJM), 71
Diffusion coefficient, 90
Diffusion equation
modified, 162–163
DIIS-method, 783
Dimethylformamide (DMF), 279
Dipole frequency, 75
Dipole polarizability, 70
Dirac equation, 524
Direct approach
for photonic crystal waveguide analysis, 587
Direct covalent functionalization, 22
Discontinuous molecular dynamics (DMD), 302
Discrete modeling, 377
Displacement step, 441, 445, 446
Dissociation
barrier, 76
channel, 6
Dissociative adsorption, 640
of NO and CO, 640
Distribution function
end-segment, 162
Divergence theorem, 698
2D MMP
model, 610
optimized power splitter, 611
Domain methods, 578, 579, 585
Donohue's description, 406
Dopants, 554
P-Doped systems, 847
Doping, 539, 595
Doping quantum dots, 845
Doping shifts, 206
Double grid technique, 778
Double peak phenomenon, 584
Double-walled carbon nanotube, 385
truss model of, 385
Doubly percolating composites, 131
Downhill movements, 8
Droplet-like model, 97
Drying transition, 173
Dulong–Petit's model, 399
Dynamical electric dipole polarizability, 74
Dynamic correspondence principle, 312
Dynamic density functional theory (DDFT), 132
Dynamic mechanical analysis (DMA), 273
Dynamic polarizability dating back, 75
Dynamic polarizability tensor, 809

E

Edge dislocations
  determination, 423
EERM model, 325
  nanotube waviness ratio, 325
Effective core potential (ECP), 488
Effective electron–ion interaction, 46
Effective external potential, 50
Effective mass approximation (EMA), 735, 736, 836
Effective mass theory (EMT), 523, 529
  Hamiltonian, 523, 524
Effective reinforcing modulus (ERM) model, 324
Efficient stress transfer, 129
Efficient time stepping methodology, 114
Egg-box effect, 778
Eigenfrequencies, 556
Eigenproblem solvers, 783
  solving, 818
Eigenvalue, 556, 585
  equations, 653
  problems, 556, 557, 562, 582
  search function, 588, 589
  solver, 584, 585, 588, 592
Eigenvalue estimation technique (EET), 585
Eilenberger equation, 668
Einstein's model, 399
Elastic energy, 111
  ...ity theory, 269
Elastic ...odulus, 313, 448
Elastic–plastic deformation process, 448
Elastic shell model, 370
  schematic of, 370
Elastic shell theory, 330
Electric dipole moments, 81
Electric dipole polarizabilities, 70–71
Electric-field operator, 236
Electroabsorption spectroscopy, 225
Electrochromism, 206
Electroluminescence, 206
  interchain and intrachain, 227
Electromagnetic field, 551
Electromechanical coupling factor, 717
Electron affinities (EAs), 77, 205
Electron density, 4, 47–48, 88
  average bulk, 78
  ground state, 48
Electron density distribution, 47, 49, 76
Electron distribution, 47–48
Electronegativity, 72
Electron–electron interactions, 46, 49, 60, 66
Electron–electron repulsion, 4
  classical, 4
Electron energy loss spectroscopy (EELS), 636
Electron-hole Coulomb, 762
Electronic band gaps, 539
Electronic charge distributions, 81
Electronic conductance, 237
Electronic energy, 4
Electronic shells, 77
  closing, 70
  effects, 66
  in clusters, 63–64
  in large clusters, 66
  structure, 67
Electron–ion interaction, 58–59, 618
  two choices of pseudopotentials, 618
Electron–ion interaction energy, 60
Electron kinetic energy, 48–49, 52–53
Electron–nuclear attraction, 4, 6
Electron-positron system, 791
Electron pseudowavefunctions, 46
Electron spin resonance (ESR)
  pulse, 458
Electron wavefunction, 4, 47, 52
Electrospinning, 264
Electrospun polymer nanofibers, 264
Electrostatic potential, 47, 64
Electrostatic self-energy, 48
Elliptic operator, 821
Embedded atom method (EAM)
  approach, 6
  potentials, 34
Embedding energy, 6
Empirical bond-order potential, 19
Empirical methods, 5
Empirical potential models, 617
Empirical pseudopotential method (EPM), 745
Empirical tight binding (ETB), 736
  method, 743
Empty core potential, 58
EMT equation, 523
  for one-electron wave function, 523
EM theory, 759
Endohedral metallofullerenes, 488
Energy-dependent Andreev, 660
Enthalpic interactions, 107
Epoxy-nanoclay composite, 289
Epoxy-silica nanocomposites, 276
Equiaxed nanoparticulate reinforcement, 274, 258
Equilibrate dense polymer liquids, 156
Equilibrium constants, 466
Equilibrium structure, 95
Equivalent-continuum modeling, 386
Equivalent continuum models, 305
  recent application, 306
Equivalent-truss model, 387
Eshelby's general formalism, 695
  for shape effects, 695
Eshelby tensors, 310, 696
Ethene, 211
  through decapentaene, 212
Euler buckling formula, 369
Euler-Lagrange equations, 49, 56, 724
Euler's polyhedral theorem, 463
Euler's theorem, 463
Evanescent modes, 565, 590–593
  model truncation, 591

Evans–Hoover thermostat, 444, 445
Evans–Hoover thermostatting scheme, 445, 446
Evaporation energy, 69, 84
Evolutionary optimizations, 601
Evolutionary strategies (ESs), 599, 601, 602
Exact diagonalization (ED), 786
Exchange-correlation, 52
  effects, 71
  energy, 51–53, 521, 618
  functional, 52, 55
  hole, 52
  potential, 61, 64, 77
Excitation energies of oligomers, 221
  in liquid and solid solutions, 221
Excite collective oscillations, 74
Exciton binding energy, 206, 765
Excitons, 761
Exciton transition energy, 755
Expanded graphite (EG), 291
Extended X-ray absorption fine structure
  spectra (EXAFS), 636
  equations of, 774
External body forces, 692
External potential, 829

F

Fabrication process, 539
Face-centred cubic (FCC) lattice, 397
Far-field radial density, 308
FD and FI codes, 578, 579
FENE
  interactions, 157
  potential, 163
Fermi commutation relations, 650
Fermi–Dirac energy distribution, 522
Fermi–Dirac function, 659
Fermi distribution, 661
  of incoming quasi-particles, 661
Fermi energy, 12, 659
Fermi golden rule, 672
Fermi level, 12, 95, 523, 527, 528, 804
Fermi velocity, 523, 525, 668
Fermi wavelength scale, 668
Ferromagnet, 654
  dispersion curves, 655, 656
Ferromagnet/superconductor interface, 657
Fiber-reinforced polymer composites, 363
Fictitious boundaries, 582, 583
Fictitious eigenstrain, 349
Field emission (FEM), 509
  properties, 526
Filled polymeric systems, 126
Filling carbon nanotubes, 29
Filter
  band-pass and band-stop, 573
  photonic crystal, 575
  RLC, 575
  simple bandpass photonic crystal, 574, 575
Fine-grained lattice models, 155
Finite computational domain, 114
Finite difference approximation, 113
Finite difference model (FDM), 105, 113, 772
  advantages, 775
  electrical properties, 113
Finite difference time domain (FDTD), 105
  calculation, 143
  method, 105, 114
  optical properties, 114
  simulation, 114–115, 143
  technique, 105
Finite element approaches, 319
Finite element method (FEM), 128, 319
  simulations, 128
Finite-element (FE) techniques, 772
  advantages, 775
Finite-field method, 837
Finitely extensible nonlinear elastic (FENE)
  potential, 154
First-order optical transitions, 759
First-order properties, 208
First-principles methods, 617
First-principles quantum mechanical methods,
  640
First-principles thermodynamics, 629
Fissioning cluster, 76
  electron density of, 76
Fitness definitions, 601
Fitness function, 599, 603, 604
Flat graphene sheet, 523, 532
Flattening of the tetrahedron
  in the diamond cubic structure, 407
Floquet modes, 541–545, 547, 548, 588
  negative-order, 548
  zero-order, 542
Floquet theory, 544
Flory–Huggins model, 166
Flory-Huggins parameter, 107, 152
Fluctuation–dissipation relations, 110
Fluctuation–dissipation theorem, 822
Fluid–fluid correlations, 168
Fluorescence excitation, 225
Flux transformer, 682
Fock-Darwin energies, 788
Focused ion beam (FIB), 332
  technology, 605
Folded spectrum method, 746
Force-constant matrix, 460
Force-displacement relationship, 399
Force field model, 759
Force minimization, 7
Fourier basis, 617
Fourier coefficients, 55
Fourier transform, 56
Fourier transform algorithm, 62
Fourier transformations, 705
Frank–Condon principle, 210
Frank–Condon shift, 763
Frank index, 524

Free-electron gas, 803
  Hartree–Fock exchange in, 804
Free-electron model, 803
Free-energy barrier, 170
Free-energy density, 165
Free energy functional, 108–109, 160, 163–164
Frequency selective surface (FSS), 538, 540–542, 545
  gratings, 540, 542
Friction, 426
  transition, 430
Frictional behavior, 426, 428
Frictional force, 420, 429
  relationship with contact length, 429
Frictional model, 303
Frictional stress, 430, 431
  versus contact width for indentation depths, 431
Friction coefficient, 420, 421, 426
  variation, 428
Frictionless motion, 429
Friedel oscillations, 55
Fries Kekulé structures, 514, 515, 525
Frozen orbital approximation, 214
Full-approximation storage method (FAS), 776
Fullerene cages, 471
  enumerations, 471
Fullerene cones, 508–510, 513
  schematic figures, 513
  STM image, 508, 509
Fullerenes
  cages, 473, 474
  C_{72} fullerenes, 475
  C_{74} fullerenes, 475, 476
  C_{82} fullerenes, 482
  C_{72} isomers, 475, 476
  C_n isomers, 472, 473
  C_{32} structures, 473
  derivatives, 493
  endohedral, 460
  Fermi level, 511
  halogenated, 493
  heptagon-containing structures, 473
  higher, 475
  hydrogenated, 493
  isomerizations, 471
  isomers, 461, 462, 472
  Kekulé structure, 514
  missing, 475
  multishell, 514
  quantum-chemical calculations, 460
  quasi-, 473, 474
  shapes, 513, 514
  single-walled, 513
  smaller, 472
  structures, 461
  synthesis, 471, 481, 486
Fullerite crystal, 95
Fully coupled model, 717
  comparison, 718
Functional integral, 161
Fundamental-measure theory (FMT), 108

G

GaAs/AlGaAs system, 707
  measured transport spectrum, 788
  strain contours, 708
Gaussian
  behavior, 162
  chain model, 162–163, 168, 178
  distributions, 157
  orientation, 321
  orthogonal ensemble, 678
  potential barrier, 68
  type functions, 3
  white noise, 109
Gauss-Seidel method, 776
Gauss theorem, 696
Generalized gradient approximation (GGA), 4, 53, 56, 71, 521, 532, 618, 784
  parameterization, 28
Generalized minimum residual (GMRES)-method, 782
Generalized multipole techniques (GMT), 581, 584
Generalized Penn's model (GPM), 765
General public license (GPL), 773, 793
Genetic algorithms (GAs), 599–602, 824
  advantages, 824
  microGAs, 601, 602
Genetic operations, 599
Genetic programming
  basic principle, 600
Geometrical surface structures, 640
Geometry optimizations, 211, 461
Gerade and ungerade states, 224
Gibbs energies, 462, 492
Gibbs free energy, 465, 533, 630
Ginzburg–Landau
  effective free energy, 680
  type equation, 680
Glass transition temperature, 175, 270
  molecular dynamics calculation, 301
  physical aging, 271
Global electronic structure, 621
Global minimum energy crystal structure, 8
Global round-off error, 398
Global truncation error, 398
Golden rule, 74
Good fit arrangements, 516
Gorkov equation, 668
Gradient approximation, 57
Gradient energy contributions, 109
Gradient expansion, 57
Gradient vector, 603
Grafting density, 181
  high, 182
  intermediate, 181
  low, 182

Grain boundaries, 424
Gram-Schmidt orthogonalization, 777
Grand canonical
ensemble, 156–157, 161, 170, 172–173
free energy, 174
reservoir, 170
simulations, 157–158
Graphene cell, 378
bond structures and corresponding energy terms, 378
Graphene hexagonal lattice, 259
Graphene sheet, 305, 511, 512
unit cell of, 305
Graphite, 269
Graphite cones, 507, 513, 516, 519
FESEM images, 516
SEM and FESEM images, 517
tubular, 518
Graphite nanoplatelet (GNP), 291
Grating, 541–544, 553, 555
consisting of single layer of dielectric rods, 543
consisting of two layers of dielectric rods, 544
perfect, 555
relative permittivity, 542
sinusoidal, 541, 542
with single layer of dielectric rods, 541
with two layers of dielectric rods, 541
Green's function, 527
Green's tensors, 713
Grid-based real-space methods, 773
Ground state density, 51
Ground state energy, 51–52
calculation, 63
functional, 51
Ground-state geometries, 824
Ground-state structures, 822
Guaranteed Reduction Pulay (GRPulay)
method, 781
Guided waves, 585
G-vectors, 62
GW absorption spectrum, 830

H

Hadamard transformation, 681
Half-space solution, 709
Halpin-Tsai equations, 309
Halpin-Tsai (HT) methods, 309, 317
Halpin-Tsai relations, 317
parameters, 317
Hamaker constant, 152, 154, 176–177, 179, 181
Hamann–Schluter–Chiang [57] potentials, 816
Hamiltonian, 3, 49–50, 62, 159
Kohn-Sham, 63
Hamiltonian matrix, 820
Hard-sphere pressure, 167
Harmonic vibrational analysis, 488
Hartree approximation, 800
Hartree equation, 801
Hartree-exchange correlation kernel, 781
Hartree-Fock
approximation, 50, 52, 801
eigenvalues, 804
electronic energy, 802
kinetic energy, 50
level, 52
method, 476
orbital energies, 802
orbitals, 214
Schrödinger equation, 802
theory, 772
Hartree–Fock Moller–Plesset (HF-MP)
approach, 4
energy eigenvalues, 54
theory, 4
Hartree–Fock self-consistent-field (HF SCF), 460
Hartree potentials, 71, 654, 745, 774
Hartree term, 521
Hartree wave function, 800
H chemisorption, 756
Heavyside step function, 168
Heine-Abarenkov potential, 58
Heisenberg transistor, 676
schematic drawing, 676
Helical cone growth, 515
cone-helix growth model, 516
Hellmann-Feynman
force, 522
theorem, 63, 822
Helmholz equation, 781
Heptagon, 473, 475, 488
Heptagon–pentagon pairs, 17
Herringbone structure, 226
Hertzian contact, 432
Hertzian prediction
of contact area, 410
Hertz theory, 433
Hessian matrix, 460, 603
Hexagonal nanowhisker specimen
atomic model, 447
HF approach, 464
HF/4-31G
approach, 484
computations, 474, 481, 484
energetics, 481, 485
Higher fullerenes, 475
Higher-order compact discretizations, 782
Highest occupied molecular orbital (HOMO), 73, 208, 623
High molecular weight (HMW), 289
High-pressure CO conversion process (HiPCO), 262
High-resolution electron microscopy (HREM), 620
Hilbert space, 682
Histogram extrapolation techniques, 158
Histogram-reweighting techniques, 159
HIV-1 virus
core shell, 510

Hohenberg-Kohn theorem, 51, 525
Hole-phonon matrix elements, 758
Holes, 740
  heavy-hole (hh) and light-hole (lh), 740
Holey photonic crystal structure
  SEM image, 560
Holy grail, 624
HOMO energies, 515
HOMO–LUMO gap, 65, 68, 73, 91–94, 96–97, 622
  substantial, 86
Homooligomers, 227
Hooke's law, 269, 693
Hückel shells, 515, 525
Hückel theory, 216, 511
  Hamiltonian, 511
  modified, 511
Hückel tight-binding scheme, 238
Hückel-type calculations, 459
Hund's rule, 786
Hybrid FE-micromechanics model, 328
Hybridization, 510, 511
  sp^2, 510
  sp^3, 510
Hybrid modeling methods, 322
Hybrid systems, 648
Hydrogen-aluminum-spinel formula, 626
Hydrogenated silicon clusters, 838
  from local density approximation (LDA), 840
  model geometries, 841
  oscillator strength, 838
  time-dependent LDA (TDLDA), 840
Hydrogen spinel, 618
Hydrogen-terminated
  diamond tips, 36
  nanocrystals, 735
Hydrostatic pressure, 409
Hyperfine splitting (HFS), 845
Hz-field, 608

I

Icosahedral
  cluster, 69, 87
  structure, 69, 85, 89
Icosahedron, 90
Ideal gas law, 630
Identity tensor, 310, 693
Implant loosening, 391
InAs pyramidal-shaped quantum dots, 698
  strain components, 703
Inclusion-inclusion interactions, 727
Inclusion theory, 726
  bonding conditions of inclusion, 726
  coated inclusions, 726
Indentation, 402, 409, 414–420
  cycles, 415
  cyclic, 417
  effect of water, 415
  elastic, 411
  inelastic, 411
  in water, 416
  length of bond, 407
  load-displacement curve, 411
  maximum indentation, 409
  phase transformations, 406
  repeated, 418
  residual, 416
  second and third, 417, 419, 420
  variation of key quantities, 412
Indentation force
  versus contact area, 418
Indenter
  diamond, 419
Indenter–silicon interface, 416
Index materials
  negative, 548–550
INDO/SCI method, 226
Inelastic deformation, 434
Infinitesimal strain tensor, 693, 722
InGaAs/GaAs quantum dots, 701
  Shin et al.'s finite element model, 701
Inorganic semiconductors, 229
InP triangular wire, 707
  strain contours, 707
Integral time step, 445
Interaction
  energies, 2
  forces, 2
Interatomic force, 400
Interatomic potentials
  n-body, 401
Interchain interactions, 224
  theory, 225
Interface free energy, 159
Interface potential, 170–171, 175–179
  effective, 173, 179
Interface tension, 169
Interfacial
  friction model, 369
  shear stress, 367
  tension, 157, 174
Intermediate neglect of differential overlap (INDO), 226
Internal light source (ILS) technique, 606
Interphase behavior, 314
Intertube corrugation, 330
Intrinsic stress, 450
  on diamond-like carbon film, 450
Ion core, 814
Ion deposition, 23
Ionic charge distributions, 81
Ionization effects, 80
Ionization potentials (IPs), 66, 68, 72, 83, 85, 205
IPR cages, 464, 471, 475, 484
  isomerism of C_{78}, 479
IPR fullerenes, 464, 472
IPR isomers
  of C_{78}, 480

of C_{90}, 487
of C_{92}, 487
of C_{94}, 487
of C_{96}, 487
of C_{98}, 488
IPR rule, 525, 528
IPR structures, 471–473, 475, 476, 478
cage isomerism, 478
of C_{80}, 480
of C_{82}, 492
of C_{84}, 483
of C_{86}, 483
of C_{90}, 485
of $Ca@C_{74}$, 489, 491
topological symmetries D_2 and T_d, 478, 479
Irreducible Brillouin zone (IBZ), 557, 558, 559, 582
one-dimensional, 561, 562
Ising universality class, 174
two-dimensional, 174–175
Isokinetic thermostatting, 440
Isolated atomic clusters, 43
Isolated pentagon rule (IPR), 460, 514
cage isomerism, 460
energetics, 464
Isolated pentagon rule set
C_{76}, 478
C_{78}, 479
C_{80}, 480
C_{82}, 482
C_{84}, 482
C_{86}, 483
C_{88}, 483
C_{90}, 485
C_{92}–C_{98}, 487
Isomeric cone tips
relative stability of, 525
Isomeric fullerene systems, 483
Isomeric stability problem, 467
Isomerization transitions, 88–89
Isotropic model, 718
Isotropic-nematic ordering, 110

J

Jahn–Teller deformation, 480
Jahn–Teller distortion, 73, 475, 480, 484, 485, 515
Jahn-Teller effect, 69, 460, 484
pseudo, 484
Jahn–Teller theorem, 478
Jellium background densities, 77
Jellium calculations, 79
Jellium model, 65–66, 73
Jellium-on-jellium model, 77–78
Jellium sphere, 77
JKR theory, 410, 411, 421, 431, 432
Josephson arrays, 678
Josephson junctions, 648
Josephson-junctions arrays (JJAs), 678
Jump-to-contact (JC), 35

K

Kekulé structures, 511, 514
Fries, 514, 515
Kelly-Tyson model, 370
Kerker pseudopotential, 816
Kinetic energy (KE), 826
functionals, 55, 88, 91
Kirchoff's law, 113
Kleinman-Bylander form, 618
Knudson diffusion, 32
Kochetov and Osipov model, 528
Kohn-Sham (KS)
approach, 51
band structure calculation, 54
calculation, 59, 89
density functional calculation, 46
density matrix, 808
effective potential, 57, 59
eigenvalues, 52, 54
energy eigenvalues, 54, 62
equation, 4–5, 52, 60, 70–71, 74, 79, 88, 521, 773
formulation, 64, 91
Hamiltonian, 63, 521
molecular dynamics simulations, 88
one-electron energies, 63
one-electron orbitals, 70
one-electron wavefunctions, 61
orbitals, 54, 62, 73, 521, 522, 789
potential, 51, 53, 56, 70, 74
procedure, 55, 91
scheme, 50, 52, 54
wavefunctions, 63
Koopman's theorem, 54, 214
Krenchel expression, 319
Krylov techniques, 776
Kvaerner's carbon-black process, 509, 512
measured apex angles, 518

L

La/Alumina system, 635
Lagrangen multipliers, 9
Lagrange undetermined multiplier, 48, 50
Lamé constants, 721
Lamellar mesophase, 134
Lanczos method, 777
Landauer–Büttiker formula, 659
Landauer formula, 237
Langevin dynamics, 825
Langevin equation, 109–110, 823
Langevin molecular dynamics, 822
Lanthanum, 635
undoped and La-doped, 636, 637
Laplace operator, 162

Laplace's equation, 105, 514
Laplace transform, 711
Laplacian term, 819
Large ring defects, 530
Laser vaporization processes, 261
Lattice Boltzmann model, 112
Lattice space
  reciprocal, 557, 558
Lattice spring model (LSM), 105, 128, 138
  calculations, 141
  dynamic, 106
  mechanical properties, 111
  network, 110
  purely elastic, 112
  simulation, 119, 128, 136–137
  technique, 128
  viscoelastic, 112
Lattice vectors, 557
  reciprocal, 557
Layer-by-layer (LBL) assembly, 389
LCAO expression, 523, 529
Leapfrog algorithm, 9
Leapfrog principle, 514
Lee-Yang-Parr's (LYP) correlation functionals, 214
Left-handed
  materials, 549
  media, 538, 548
  medium, 549
Lennard–Jones (LJ)
  chains, 167, 178
  interaction, 109, 154, 160, 165, 177
  model, 155
  monomeric fluids, 164, 166, 174
  monomeric reference fluid, 166
  monomers, 166
  particles, 165
  potential, 6, 25, 29, 32, 163–164, 166, 302, 416
  potential function, 378
  potential parameters, 416, 417
  type, 154
  units, 178
Leveraging experiments, 620
Leveraging infrared spectroscopy, 641
Li_nAl clusters
  electronic structure of, 83
Lindhard screening function, 55
Linear combination of atomic orbitals (LCAO), 3, 61, 511, 620
Linear elasticity theory, 111
Linear low density polyethylene (LLDPE), 275
  dynamic mechanical properties of, 275
Linear response theory, 55
Li_nPb clusters
  calculated ground state geometries of, 83
Lippman-Schwinger equation, 238
Liquid drop model, 66
Liquid–vapor coexistence, 156, 159, 174
  chemical potential, 173
  value, 170
Liquid–vapor critical point, 159
Liquid–vapor interface, 152, 158, 170–171, 177, 178
  free energy, 158
  tension, 173
Lithium clusters, 67
Load-displacement curves, 411, 412, 417, 419
  of repeated indentations, 418
  of second indentation cycle, 415
Loading–unloading process, 417
Local charge neutrality, 77
Local density approximation (LDA), 4, 52–53, 64, 92, 521, 618
  diffuse jellium model (DJM), 71
  exchange-correlation potential, 71
  spin polarization (LSDA), 52–53
Local density functional, 178
Local density of states (LDOS), 526–530
  apical magnetic flux, 529
  to n-membered rings, 528
Local pseudopotentials
  first-principles, 60
Local spin density approximation (LSDA), 214
Loss modulus, 272
Loss parameter, 608, 609
Löwdin orthogonalization procedure, 742
Lower-density droplet, 787
Lowest-unoccupied molecular orbital (LUMO), 73, 208, 622
Low molecular weight (LMW), 289
Low pressure chemical vapor deposition (LPCVD), 735
Luminescence, 227
LUMO energies, 515
Luo and Daniel approach, 314

M

Macroscopic drops dewets, 175
Macroscopic material properties, 546
Madelung constant, 6
Madelung summation techniques, 806
Magic
  clusters, 65, 622
  numbers, 64, 67, 75, 77, 87
Many-body interactions, 6
Many-body theory, 208
Marconi's time, 538
Marcus theory, 238
Masterbatch nylon 6, 289
Materials modeling, 377
  two kinds of techniques, 377
Matrix equation, 810
Matrix–fiber stress transfer mechanism, 370
Maximum-density droplet (MDD), 786
  MDD-window limits, 787
Maxwell's curl equations
  discretization of, 114
Maxwell's equations, 105, 114

I Maxwell solvers, 598
I Maxwell theory, 549
I Mean-field approximation, 108
I Mean field theory of superconductivity, 652
key feature, 652
I Mean spherical approximation (MSA), 166
I Meissner effect, 677
I Melting temperature, 87, 90
I Melting transition, 87–88
I Mesoscopic Josephson junctions, 675
I Mesoscopic physics, 647
I Metallocarbohedranes, 622
I Metallofullerenes, 459
endohedral, 488
isomerism, 460
stability computations, 488
I Metal oxide materials, 625
I Metamaterials, 538–540, 548, 550, 555, 580, 611
anisotropic effect, 546
coating, 580
I Method of auxiliary sources (MAS), 581, 584
computation, 584
I Method of fictitious sources (MFS), 584
I Method of moments (MoM), 579, 580, 581
I Microcracking, 403
I Microelectromechanical systems (MEMS), 396
I MicroGA, 601, 602
table-based, 602, 603
I Micromechanical dislocation model, 429
I Micromechanical simulation, 142
I Micromechanics modeling, 310
I Microphase separation, 124
I Microscopic monomer density, 160
I MIIKA package, 773
schematic illustration, 773
I Mills-Nixon effect, 514
I MINDO, 459
I Minimum energy configuration, 112
I Mini-stopband, 608
I MMA monomer, 282
I MMP code, 583, 584
I MM simulations, 520, 525
I MNDO heats
formation per atom, 465
I MNDO (modified neglect of differential overlap) method, 217, 464
I Mode gaps or mini-stopbands (MSBs), 608, 6609
I Model pseudopotentials, 57
I Model truncation lines, 593
I Mode matching technique, 566, 567, 571, 592, 5593
waveguide discontinuities, 593
I Modified embedded atom method (MEAM) approach, 6
I Modified neglect of differential overlap (MNDO), 460, 466, 479, 482
I Molecular binding energies, 53
I Molecular dynamic modeling, 299
I Molecular dynamics analysis, 396, 397, 399, 4123, 431, 433, 453
Molecular dynamics model
for nanoindentation in water, 416
of silicon, 397
of sliding processes, 421
Molecular dynamics modeling
of materials, 396
Molecular dynamics (MD) simulations, 8–9, 18, 20, 24–25, 30, 32, 34–36, 72, 87–88, 96, 368, 420, 428–434, 440, 444, 448
constant energy, 88
electron-beam irradiation in, 19
equilibrium, 30
hydrogen-terminated diamond surface in, 34
nonequilibrium, 32
Molecular mechanics force, 385
Molecular mechanics (MM) methods, 616
Molecular mechanics simulation, 364
Molecular memory
based on an endohedral fullerene, 458
fullerene-based, 458
Molecular orbital package (MOPAC), 461
Molecular structural mechanics, 384
major steps, 384
Molecular transistors, 458
Molecular wire bridge, 237
Molecules
chiral, 546
curves, 546
axis of, 545, 547
relative permittivity, 545–547
Mole fractions, 462
Moller–Plesset (MP)
calculations, 82
many-body perturbation theory, 79
perturbation theory MP2, 210, 213
theory, 4
Molybdenum nanoclusters, 640
interaction of N2, 640
Monkhorst-Pack scheme, 618
Monodisperse oligomers, 208
polyenes, 208
thiophene, 208
Monomer density, 162
Monomeric molar fraction, 467
Monomeric reference system, 165
Monomer–monomer interactions, 155
Monomer number density profile, 168
Monomer–substrate interactions, 179
Monomer–wall
attraction, 172
potential, 179
Monte Carlo approach, 520
Monte Carlo integration, 599
Monte Carlo lottery, 156
Monte Carlo (MC) methods, 304
dynamic problems, 304
Monte Carlo sampling, 159
Monte Carlo (MC) simulations, 10, 32, 89–90, 154, 156, 158, 166, 168–169, 175, 177, 183
for thin polymer films, 151–199

grandcanonical, 169
tight-binding, 89
wetting transition in, 170
Monte Carlo statistical method, 331
Montmorillonite, 255
structure of, 267
Mori-Tanaka method, 310
derivation of, 334
schematic of, 334
Mori-Tanaka micromechanical model, 283
Mori-Tanaka modulus, 311
illustration of, 311
Morphological transition, 183
Morse pair potential, 401
Morse potential, 405, 422, 447
cutoff radius, 447
parameters, 405, 422
Mott insulator, 679
MPB code, 583
MPG code, 583, 584
Mulliken population analysis (MPA), 232
Multiasperity contact sliding
mechanics model, 438
Multiasperity sliding, 437
configuration I, 438
configurations II and III, 437
modeling, 437
parallel asperities, 437
spherical diamond asperities, 437
Multichain partition function, 160
Multi-complex-origin-source method, 593
Multigrid (MG) algorithms, 775
Multiple multipole program (MMP), 581, 584, 585
computation, 584
eigenvalue solver, 588
Multipole expansions, 583
Multireference configuration interaction (MRCI), 210
Multiscale methods, 329
Multiscale modeling, 299, 383
Multisource methods, 592
Multiwalled coiled carbon nanotubes (MCCNTs), 389
Multiwalled nanotubes (MWNT), 440
Mutation
rates, 600
single-bit, 603
MWNT-PMMA nanocomposites, 285
MWNT-polycarbonate system, 281
frequency domain data, 313
MWNT-reinforced polymers, 283
for glassy modulus of, 283
M_nX clusters, 85

N

Na clusters, 74, 79, 88
chemical hardness of, 73
polarizability of, 71
Na_N^+ clusters
melting temperature of, 87
Nambu space, 668
Nanocatalysts, 620
Nanoclay-polymer composites, 288
Nanoclay-polymer nanocomposites, 257
Nanocomposites, 106, 254
applications, 255
cylindrical phase of, 136
lamellar phase of, 133
macroscale behavior of, 254
macroscopic properties, 111–116
micromechanical studies, 137
model, 316
multifunctional properties, 331
structure of, 106
thermo-mechanical properties, 331
Nanocones, 507, 511
comparison of simulated STM images, 530
formation energies, 532
four configurations, 526
hydrogen desorption curve, 509
nucleation theory, 532
simulated STM image, 530
two-dimensional free-standing structures, 527
with one (a–c) and two (d–f) pentagons, 531
with substitutional atoms (b, c, d, f), 531
Nanocrystals, 620
quantization effects, 622
quantum confinement in, 833
Nanodeformation, 405
Nanodevices, 648
Nanodewetting, 176
Nanodroplets
on layered substrates, 175
Nanoelectromechanical systems (NEMS), 396
Nanofibers, 239
Nanofibers electrospun, 259
Nanofiller elements, 256
Nanohorns, 518
Nanoindentation, 34, 36, 399, 401–403, 410, 415, 417, 438, 453
displacement field, 404
modeling, 403
on monocrystalline silicon, 402
simulation, 431
Nanoindentation with water
cross-sectional view of simulation, 417
Nanomaterials, 1
Nanomechanical modeling, 298
Nanomechanics, 254, 299
Nanometer-scale systems
application, 10
Nanoparticle-filled diblock copolymers
optical properties of, 142–147
relation of morphology, 142
structure of, 142–147
Nanoparticles, 620
influences their chemistry, 620

Nanophysics, 788
in axial symmetry, 788
Nanopipettes, 518, 519
Nanoplatelet reinforcement, 267, 288
partially exfoliated nanoplatelets, 314
Nanoprocessing operation, 397
Nanoreinforced materials, 254
critical issues, 256
Nanoreinforced polymers, 274
experimental results, 274
interphase region, 292
viscoelastic behavior, 292
Nanorod/polymer blend system, 127
Nanorods, 126, 258
electrical properties, 130
in binary polymer blends, 126–132
mechanical properties, 128
properties of mixtures of, 132
volume fraction of, 127–128, 130–131
Nanoscale, 798
Nanoscale defects, 791
positron states, 791
Nanoscale friction model, 368
Nanoscale reinforcement, 258
Nanoscale reinforcement fillers, 258
types of, 258
Nanoscale rods, 126
Nanosilica-reinforced polymers, 293
Nanospheres, 258
Nanotechnology, 538, 624
Nanotribology
introduction, 420–421
Nanotube, 254
controlled orientation of, 280
different chiral vectors, 259
energy plots during the pullout of, 367
potential applications, 390
properties, 260
single-walled, multi-walled or nanotube bundle, 256
snapshots, 366
Nanotube/epoxy resin composite system, 364
Nanotube–matrix interfacial shear mechanism, 370
Nanotube-polymer composites, 297
equivalent-continuum modeling, 307
relaxation spectra of, 297
Nanotube/polymer composite system, 388
equivalent-continuum modeling of, 388
Nanotube-polymer interface, 281
load transfer, 281
Nanotube-reinforced polymers (NRPs), 278
current issues in nanomechanics, 329
electrical and thermal properties, 286
mechanical properties of, 282
nanotube dispersion, 279
processing of, 278
Young's modulus predictions, 237
Nanotube reinforcement, 258, 277
Nanotubes' walls, 23
chemically functionalize, 23
Nanowhisker, 446–449
Na_mPb clusters, 84
closed-shell structure of, 84
Natural optimization algorithms, 599
Navier's equations, 693
Nb/α-Alumina system, 632
Nearest-neighbor atoms, 6
Negative illumination, 551
Negative index lens, 553
Negative index materials, 548–550, 553
disk, 552
Negative index slab, 550
Negative refraction, 548
Neighbors against time
graph, 406
Neutral cluster, 72
Newtonian
atoms, 403, 423, 441, 444, 447
copper atoms, 430
dynamics, 396, 826
equations, 63
fluid, 270
Newton-like optimizer, 569
Newton's constitutive law, 368
Newton's equation, 9
Ni tip, 35
indention of, 35
NN system, 665
schematic picture, 665
Noble metal clusters, 75
Nonbulk interphase polymer, 292
experimental evidence of, 292
Non-dilute composite, 349
Nondimensional indentation depth
versus lubrication/contamination, 426
versus sliding speed, 426
versus tip radius, 426
Nondimensional parameters, 422
Nonlinear effects, 714
Nonlinear optical properties, 208
Nonlinear truss rod, 385
load-displacement curve, 385
Nonlocal effects, 723
Nonlocal kinetic energy functional, 56
Non-monotonic variation, 63
Nonoverlapping/low-energy chain conformations, 156
Nonselective interactions, 134
Nonuniform mismatch strains, 726
Normal-metal/superconductor interface, 659
Normal-mode diffusion, 29
Normal-superconductor-normal (NSN) transistor, 671
No-wear regime, 423, 424, 426
Nozzle tube temperature, 65
NRP systems, 323
NSN transistor, 674
NS system, 666
non-monotonic behaviors, 670
panorama on results, 670

schematic picture, 666
NT phase, 325
NT-polymer composites, 330
load transfer, 330
NT structure, 303
Nuclear magnetic resonance (NMR)
pulse, 458
Nuclear–nuclear repulsion, 2
Nucleation puzzle, 532
Nucleation theory, 798
for nanocones, 532
Numerical methods, 381
Numerical optimization, 596, 598
deterministic optimizers, 596
Numerical strain concentration tensor (NSCT)
method, 325
Nylon-clay systems, 322

O

Oblate distortion, 86
Occupied valence orbitals, 83
Octahedral sites, 626
Octatetraene, 211
Octithiophene (8-T), 224
Octupole deformation, 91
Off-axis orientations, 321
Off-lattice model, 155
Ohmic current, 679
Ohm's law, 113, 669
Oligothiophene, 225
One-electron excitation energies, 54
Ono-Hirose scheme, 778
Open port approach, 594
Open-shell electronic configuration, 69
Optical coupler
improved, 570
Optical gaps, 760
Optical spectra, 830
Optimal polymer matrix, 106
Optimal polymer–particle interactions, 106
Orbital angular momentum, 64
π-Orbital axis vector (POAV) analysis, 511
Orbital-dependent potential, 53
Orbital-free (OF)
approach, 54
calculations, 91
energy functional, 88
method, 60, 87
scheme, 60
simulations, 87, 89–91
Orbital-free total energy simulations, 60
Orbitals, 511
Organic light emitting diodes (OLED)s, 206
Ornstein–Zernike equations, 166
Orthogonality wiggles, 46
Oscillator strengths, 755, 810
Out-of-plane scattering, 605–608
loss, 605–607
Overbinding effect, 443
Oxidized clusters, 841
Oxidized nanocrystal, 735
Oxygen penetration
effect of, 413

P

Packing fraction, 167
Padé approximant, 56
Pair correlation function, 118, 167
Pair potentials, 87, 401, 404
Pairwise functional, 6
Parallelepiped dot, 699
Parallel hard cubes (PHC), 108
Parallel hard rectangles (PHR), 108
Parameterization, 743
Pariser-Parr-Pople approximation (PPP), 217
Parity effects, 671
Parrinello-Rahman approach, 380
Particle-based model, 163
Particle center distribution, 135, 138
Particle centers, 133
Particle correlation function, 118
Particle density, 133
profiles, 137
Particle distribution function, 138
Particle entropic contributions, 107
Particle-hole
excitations, 75
space, 668
symmetry, 661
transitions, 78
Particle-in-a-box (PIB), 617
Particle–matrix interface, 119
Particle–particle interactions, 110
Particle–polymer interactions, 135
Particle swarm optimization (PSO), 602
Particle-to-diblock volume ratio, 107
Particle volume fraction, 107, 125
Partition function, 107, 163
single-chain, 162
Pauli matrices, 523
Pauli principle, 801
Pb clusters, 84
PBG lithographic tuning, 606
Pb_4Li_4 clusters, 95
PbNa alloy, 98
Pb_2Na_2 clusters, 94, 98
Peapod-in-nanotube structures, 33
Peapods, 458
Peapod structure, 33
Peierls distortions, 12, 233
Penning-trap, 68
Penn's model, 763
Pentagon/hexagon rings, 472
Pentagon–pentagon junctions, 472, 473, 475
Pentagon road model, 532
Pentagon rule energetics, 464

Pentagons, 509, 511–514, 518, 519, 523, 525–533
Percolating cluster, 119
Percolating nanorods, 131
Percolation threshold, 119
Percolation type process, 286
Percus–Yevick closure, 167
Perdew's 86 (P86) correlation functionals, 214
Perdew-Zunger parametrization, 811
Perfect lens, 549, 550, 553, 580
  Ray model of, 550
Perfectly matched layer (PML), 114–115
  boundaries, 115
  media, 115
  structure, 115
Perfect photonic crystals, 555, 583, 584
  idealization, 555
  original lattice space of, 558
  symmetry considerations, 557
Periodic boundary conditions force, 803
Periodic symmetry
  structures with, 541
Periodic waveguide
  problems, 586
  structures, 585, 586
Permutational isomers, 89
Perturbation theory, 764
  first-order, 70
  second-order, 46, 55
  thermodynamic, 166–167
Perturbed clusters, 70
Perturbing periodic potential, 57
Petal superstructures, 529, 530
  origin, 529
  STM images, 529
Phase coexistence, 158
  equal-weight-rule of, 158
Phase-coherent conductor, 659
Phase-separating
  blend, 126, 128
  polymer blend, 131
  system, 129–130
  system corral, 131
Phase separation dynamics, 105
Phase transformation, 405, 406, 438
  in silicon, 408
Phenomenological 2D model, 606
Phenomenological two-dimensional modeling
  of planar photonic crystal devices, 605
Phonon-assisted transitions, 757
Photoabsorption cross section, 74, 89
Photocatalysis, 616, 638
Photodetachment spectroscopy, 68
Photoelectron spectrum, 622
Photoemission spectra, 824
Photoexcitation, 623
Photoluminescence (PL), 606, 734
  decay, 756
Photonic atom, 554
Photonic band gaps (PBGs), 543, 544, 553, 554, 560, 586, 605, 608
  materials, 106
Photonic crystals (PhCs), 539, 542, 543, 544, 548, 553–555, 559–577, 586
  analysis of structures, 577
  antenna, 568
  band diagram computation, 582
  cells, 539
  circular rods, 583
  classes of problems, 555
  coupling energy into, 568
  coupling energy out of, 568
  devices, 605, 606
  dispersion relation, 562
  filter consisting of four coupled cavity resonators, 575
  filtering T-junction, 577
  filters, 573–575
  frequency characteristics and geometry, 597
  hole-type, 563, 588
  improving in-coupling of light, 608
  modification of waveguide termination, 569
  perfect crystals, 555
  planar, 605
  positions and radii of two rods, 570
  power divider, 575–577, 595
  power divider search area, 597
  power divider sensitivity analysis, 604
  power splitter, 611
  radiating waveguide, 589
  reciprocal lattice space, 558
  resonators, 574
  rods, 568
  rod-type, 563, 587
  simple bandpass filter, 574
  simple open waveguide, 570
  structures, 554, 556, 577, 579, 605, 606
  three-dimensional, 557, 558
  two-dimensional, 558, 559
  walls, 563, 588
  waveguide bending, 570
  waveguides, 555, 559–564, 567–572, 578, 586, 594
  waveguide termination, 568, 569
  with different dopants, 554
Photons, 554
Physioabsorbing surfaces, 316
Physisorption, 152
Piezoelectric constant, 715
Piezoelectricity, 715
  effect around any dot, 716
  effect around spherical dot, 715
Piezoelectric material, 717
Piezoelectric pseudopotential, 716
Piezoelectric tensor, 715
Planar photonic crystal
  bends and power splitters, 609
  devices, 605
Planar photonic crystal tapers
  modeling, 607
Planck's constant, 799
Plane wave approximation (PWA), 583, 584

Plane-wave coefficients, 780
Plane wave method, 819
Plane wave refraction
  at negative index material slab, 548
Plane waves, 581
  expansions, 581, 584
  time-pulsed, 582
Plasmon effects, 546
Plasmon peak, 78
Plasmon resonances, 546, 556
Platonic solid, 624
Ploughing, 424, 425, 436
  regimes, 424, 425, 426, 428, 429
PM3-computed energy, 495
PMMA, 275
  function of nanosilica loading in, 275
PM3 methods, 494
Point-force formalism, 712
Poisson's equation, 47–48, 701
Poisson's ratios, 111, 441, 446, 693
Polarization
  effects, 544
  incident wave, 546
  of incident wave, 547
  orthogonal, 545
  p, 545
  s, 543–545
Polaron model, 232
Polarons, 226
  bipolaron, 230
Polar surface terminations, 629
Polyacetylene (PA), 204
  photocurrent, 206
Polyamide-nanoclay composite, 288
Polyaniline (PAn), 204, 286
Pólya's enumeration theorem, 472
Polycarbonate (PC), 293
Polydiacetylene (PDA), 226
Polyenes, 209, 227
  ab initio calculations, 209
  HOMO-LUMO gaps of, 216
  TDDFT excitation energies for, 214
  theoretical excitation energies, 209
  two low-lying singlet excited states, 209
Polyene spectra, 217
Polyethylene (PE) composite systems, 368
Poly(ethylenedioxythiophene) (PEDOT), 204
Polyfuran (PFu), 236
Polyhedral cages, 459
Polyhydroxyaminoether (PHAE), 281
Polyimide, 288
Polyimide-clay hybrid, 290
  permeability coefficient of, 290
Polyisothianaphthene (PITN), 204
Polymer brush
  wetting on, 180
Polymerization index, 132
Polymer lattice, 229
Polymer liquids,
  one-component, 156
  wetting of, 170
Polymer matrix composites, 254
Polymer nanocomposites, 329
  multiscale modeling strategy for, 331
  strength and toughness modeling, 329
Polymer/nanoparticle composites, 103
  heterogeneous, 105
Polymer-NT unit cell, 306
Polymer-particle coupling, 108
Polymer–particle interactions, 105, 117, 122, 134
Polymer-processing simulation, 111
Polymer reference interaction site model (P-RISM) theory, 167
Polymers, 227
  articles available in the literature, 335–348
  comparison between theory and experiment, 227
  evolution of excitation energies, 228
  polymer mobility, 269
  thermosetting and thermoplastic, 362
Polymer–substrate interactions, 107
Polymethylthiophene (PMeT), 228
Poly(m-phenylenevinylene-co-2,5-dioctoxy-p-phenylenevinylene) (PmPV), 286
Polyparaphenylene (PPP), 204
Poly(paraphenylenevinylene) (PPV), 232
Poly(phenyleneethynylene) (PPE), 279
Polypyrrole (PPy), 205
Polystyrene (PS), 175, 277
  comparison of, 278
Polythiophene (PT), 204
Pop-out phenomenon, 403
Popping atoms, 35
Porous silicon (PS), 734
Position vector, 8
Positive charge density, 64
Posteri theory, 807
Potential energy contributions, 50
Potential energy surface (PES), 44, 642, 822
Potential function, 401
Potential well, 64
  effective spherical, 64
Power dividers, 575–577
  intuitive design, 576
  three most crucial rods, 598
  T-shaped structure, 575, 576
Power reflection, 610
Power spectrum, 827
Power transmission, 610
Poynting vector, 543, 549
  intensity of the time average, 552
  time-average, 561
Poynting vector field, 550, 552, 567, 569–571, 580
  for single layer photonic crystal waveguide, 573, 589
Preconditioning "interior" eigenvalues, 821
Preconditioning techniques, 821
Predictor-corrector algorithm, 9

Pre-preg lamina, 281
Pressure volume (PV), 629
Prewetting coexistence, 171, 174
Prewetting critical point, 174
Prewetting line, 174
  toward coexistence, 174
Printed circuit boards (PCBs), 540
Priori, 807
Probabilistic optimizers, 599, 602
  deterministic search, 599
  random search, 599
Probability distribution, 157–160, 162, 183
Projector-augmented wave (PAW), 636, 778
Prolongation operation, 776
Prony coefficients, 297
Prony series representation, 271
Propagation constant, 563, 585–590
Proximal probe microscopes, 34
Proximal probe tip indentation, 36
Proximity effect, 650
Pseudoatoms
  valence electron density of, 56
  wavefunctions, 60
Pseudo-direct, 748
Pseudo-elastic state, 271
Pseudoion potential, 58
Pseudomolecular mechanics approach, 305
Pseudo-orbitals, 60
Pseudopotentials (PP), 46, 57, 61–62, 79, 760, 812
  first-principles nonlocal, 60
  local, 88
  method, 745
  model, 57, 812
  nonlocal, 66
  nonlocal Pb, 83
  normconserving, 58, 60
  scheme, 60
  solid, 813
Pseudo-vectors, 549
Pseudo–wave function, 61, 814
Pt/alumina system, 632
Pulay corrections, 772
Pullout energy, 366
  fiber-pullout model, 372
  pullout of the nanotube, 367
  three terms, 366
Pullout force, 376
Pyracylene transformation, 471
Pyramidal dot, 699
  geometry of, 699
  radial and axial strains, 702
Pyrrole oligomers, 223
  theoretical excitation energies, 223

Q

Quantized colloids, 620
Quantum chemical calculations, 152
  outline, 460
Quantum computation (QC), 680
Quantum computer
  concept, 459
Quantum computing, 458
Quantum confinement (QC), 735, 746
  model, 841
Quantum descriptions of matter, 798
Quantum dot–quantum dot interaction, 728
Quantum dots (QDs), 689, 698, 738
  configurations of, 695
Quantum dots and wires, 691
  effect of shape, 694
  impurities in, 788
  of arbitrary shape, 695
  review of strain field calculations, 691
  simple illustrative example, 692
  size effects, 727
  two physical mechanisms, 720
Quantum mechanical (QM)
  approaches, 2–3
  equation, 799
  information, 2
  laws, 798
  methods, 3, 5
  modeling, 616
  principles, 2
  scattering matrix, 660
  wave function, 2
Quantum Monte Carlo calculations, 760
Quantum number, 64
  principal, 66
Quantum phase model (QPM), 679
Quantum statistical methods, 208
Quantum wires (QWRs), 689, 703, 736
  configurations of, 695
Quasi-classical Green's function theory, 650, 668
Quasi-classical theory of superconductivity, 668
Quasicontinuous energy bands, 66
Quasioctahedral, 626
Quasi-particle excitations, 648
Qubit–qubit coupling, 682
Qubits, 681
Quick and dirty modelers, 590, 591
Quiniod structures, 222
Quinoid form, 229

R

Radar technology, 538
Radial atomic density distribution, 90
Radiating modes, 566
Radio frequency (RF) regime, 538
Raman experiment, 825
Raman spectroscopy, 260, 281
  technique, 390
Random-matrix theory, 667
Random number, 603
Random phase approximation (RPA), 212

Random vector, 602
Rayleigh models, 541, 588
Rayleigh quotient, 777
Rayleigh-quotient minimization multigrid (RQMG) method, 773, 776
  with Galerkin conditions, 783
Rayleigh-Ritz variational principle, 50
Real space algorithms, 820
Real-space approaches, 775
Real-space multigrid methods, 775
Real-space superlattice, 62
Real-space tool, 784
Reciprocal lattice vector, 819
Reciprocal-space situations, 62
Recrystallization, 412
Reentrance effect, 670
Reference electron density, 56
Reference hypernetted chain (RHNC), 166
Reflected waves
  compensation scheme for, 594
Reflection coefficients, 594–597, 605
Reflectionless tunneling, 670
Refraction index, 548, 550, 608
Reinforcement efficiency, 121
Relative concentrations
  fullerene isomers, 462
  of $Ca@C_{72}$ isomers, 489
  of $Ca@C_{74}$ isomers, 490, 491
  of $Ca@C_{82}$ isomers, 492
  of $C_{60}F_{36}$ isomers, 494
  of $C_{60}H_{36}$ isomers, 495
  of C_4 isomers, 469
  of C_6 isomers, 470
  of C_{72} isomers, 477
  of C_{74} isomers, 478
  of C_{80} isomers, 481
  of C_{86} isomers, 484
  of C_{88} isomers, 485
  of C_{90} isomers, 486
Relative density, 164
  of transformed silicon versus indentation depth, 419
Relative elastic fields, 139
Relative local strain field, 121, 124, 139
Relative normal stress field, 140
  contours, 129
Relative potential energies, 463
Relative root mean square (RMS)
  bond length fluctuation, 88–89
Relative stress field, 140
Relaxation spectrum, 272
Relaxation steps, 441, 445, 446
Representative volume element (RVE), 125, 307
  effective isotropic properties, 308
  schematic of three types of, 382
  two-dimensional, 125
Residual deformation
  in silicon, 416
Residual stress ratio, 453
Resonators, 556
Response-iteration methods, 781
Restricted active-space SCF (RASSCF) method, 210
Restriction operation, 776
Reverse-mapping technique, 306
Rigid indenter simulations, 35
Rigid-rotor and harmonicoscillator (RRHO) approximation, 462
Ring-stacking model, 532, 533
RLC filters, 601
Rod distribution, 130
Rod–rod interaction, 110
Rod–rod repulsion, 110, 127
Rolling graphene model, 371
Rotational–vibrational partition function, 476
Rudimentary methods, 310
Rutile band structures, 634
Rydberg character, 210

S

Saddle points, 108
  approximation, 161, 163
  values, 165
SAM1
  computations, 484, 486
  energetics, 485
  levels, 473–475
  lowest-energy structures, 486
  methods, 461, 483
  optimized structures, 474
  quantumchemical method, 487
  semiempirical method, 483
Scaled velocity, 398
Scale effect, 429, 430
  of contact size on friction transition, 429
Scanning near-field optical microscopes (SNOMs), 549
Scanning probe microscopy (SPMs), 332
Scanning tunneling microscopy (STM), 237, 624
Scattering formalism, 650
Scattering matrix, 566
Scattering problem, 555, 556, 593, 658
SCF/DFT
  approach, 104–105
  calculation, 114–115, 138–139
  grid of points, 114
  LSM approach, 132
  methodology, 132, 137, 142
  model, 137–138, 147
  morphological studies, 105
  simulation, 115, 138
  structure, 106
Schrödinger equation, 4, 49, 654, 799
  Mehrstellen-discretized, 782
  N-electron, 49
  radial, 59
  single particle, 50–52
  time-dependent, 520

time-independent, 3
Screw dislocations, 530
model in graphite, 531
Search direction vector, 7–8
Search spaces, 596, 598, 599, 604
Sector angles, 512
Seitz radius, 811
Selective interactions, 132
Self-assembled monolayers (SAMs), 222
Self-consistency loop, 821
Self-consistent field (SCF)
calculations, 160, 166, 168–169, 175, 178–180
theory, 174
Self-consistent field/density functional theory
approach, 106
Self-consistent field technique, 156
Self-consistent field theory (SCFT), 104, 106–107, 169
for thin polymer films, 151–197
Self-interaction corrections (SIC), 71
Semiconductors, 554
Semicore electrons, 61
Semi-coupled model, 717
comparison, 718
Semiempirical methods, 5, 216, 461
Semiempirical quantum-chemical methods, 487
Semi-infinite domain, 709
illustration of, 709
Sensitivity analysis, 604, 605
of 90-degree photonic crystal bend, 605
Sexithiophene (6-T), 224
differences, 225
Shape deformations, 76
Sharp-kink approximation, 177, 179
Sharp melting transition, 89
Shear banding and necking, 448
evolution process of, 449
Shear bands, 448, 449
Shear modulus, 265
Shell-closing numbers, 74, 77
Shell clusters
closed, 65–66, 69, 72–73
Shell effects, 66, 72
in cluster fragmentation, 75
Shell index, 66
Shell oscillations, 67
Shift-and-invert techniques, 821
Si bulk unit cell, 755
Si_{11} cluster, 827
absorption spectrum, 830
dynamic matrix calculation, 827
excitation energies of, 832
Power spectrum of, 828
simulation for, 828
time-dependent density-functional formalism, 830, 833
$Si_{860}H_{76}P$ quantum dot, 849
isotropic hyperfine parameter, 849
$Si_{877}H_{76}$ quantum dot, 837
absorption spectra of, 843
wave vector dependence, 837
Silica, 258
Silica-polyimide, 306
interfacial conditions, 306
Silica-polyimide nanocomposites, 308
radial density profiles, 308
Silicate clays, 267
Silicon
effect of water, 415
mechanical properties, 415
monocrystalline, 415, 417, 453
phase transformation, 438
volume of transformed, 418
Silicon and germanium clusters, 829
static dipole moments, 829
Silicon atoms
comparison of the total number, 415
transformed from diamond cubic to other structures, 419
Silicon monocrystals, 403, 415
subsurface microstructure, 434
Silicon nanoelectronic devices, 735
Silicon nanoindentation
in loading–unloading cycle, 410
Silicon on insulator (SOI) layer systems, 605
Silicon–oxygen coordination, 413
Silicon quantum wires and dots, 734
different forms, 734
electronic structure calculations, 735
optical and charging properties, 735, 749
SiO_2 layers, 735
Silicon specimen, 403, 405
subsurface microstructure, 403
variation of average temperature, 420
Silicon workpiece
through center of asperities, 438
Silver di-anionic clusters, 68
Simple weighted density functional (MGM), 174
Simplified interaction-free functional (MGM), 180
Simulated annealing (SA), 602
methods, 822
Simulation temperature, 397
Si nanodots, 761
oxygen role, 761
Single-chain partition function, 162
Single-electron methods, 212
Single-electron transistor (SET), 671, 672
Singles configuration interaction (SCI), 226
Single-walled carbon nanotubes (SWCNT), 12–13, 303, 440
calculated band structure of, 12
dissolution of, 23
electronic properties of, 10, 14
from a polystyrene matrix, 303
functionalization of, 23
semiconducting singlewalled, 28
Single-walled nanotube/epoxy resin composite system, 365
interaction energies of, 366

molecular model of, 365
Singly charged magic fragment, 75
Sinusoidal buckling pattern, 451
Si–O bonds, 414
$Si_{35}OH_{34}$ (1) clusters, 844
principal mechanism, 845
schematic representation, 844
Si quantum dot, 834
atomic structure of, 834
role of oxygen in, 840
Slab
relative permittivity and permeability, 546, 550, 551
Slater determinants, 50, 762, 801
Slater orbital, 61
Slater-type functions, 3
Sliding
copper atomic lattice, 423
friction, 368
interactions, 437
interface, 424
interfacial, 421
multiasperity, 437
noncontact, 426
regime transition, 426
speed, 425
three-body, 421
three-body contact, 433–437
two-body, 421
two-body contact, 421, 433–436
variation, 435
zones, 428
Sliding forces, 426
variation, 428
Sliding simulation
with diamond asperities of different tip radius, 430
with indentation depth, 432
Sliding systems
diamond–copper, 422, 435
diamond–silicon, 433, 435
methods of modeling and analysis, 422
no-wear, 424
Slithering snakelike movements, 156
Small carbon clusters, 468
Smaller fullerenes, 472
Smart animals, 602
S-matrix elements, 661
Sodium clusters, 74, 79
chemical hardness of, 73
neutral, 82
optimized geometries of, 80–82
polarizability of, 71
singly charged, 91
Sodium dodecyl sulfate (SDS), 279
Sodium quantum dots, 790
on Cu(111), 790
pseudopotential compared, 818
Solid cohesive energies, 53
Solid state physics community, 617
three categories, 617
Solid-state polarization, 221
Soliton, 232
Sommerfield model, 57
Spatial inhomogeneity, 162
$sp^3d^2s^*$ model, 744
Specific energy terms, 386
Specific polymer configurations, 160
Specimen types, 447, 448
hexagonal, 447, 448
Spectroscopic gap, 677
Spectroscopic techniques, 225
Spherical jellium model (SJM), 64, 66, 67–68
Spherical nanoparticles
properties of mixtures of, 116
Spherical quantum dot, 693
radial strain tensor component, 713
schematic of, 693
Spherical sodium clusters, 67
Spheroidal inclusion model, 711
Spin-down electron densities, 52
Spin–orbit interaction, 744
Spin-polarization, 670
function, 811
Spiral algorithm, 471
Square well
potentials, 58
rounded, 64
Stability function, 74
Standard rule of mixtures, 318
Standing wave ratio (SWR) analysis, 594
Static dipole polarizabilities, 70
Statistical-mechanical calculations
outline, 460
Statistical segment length, 162
Steady-state behaviour, 421
Steepest descent method, 776
Steering effect, 610
Stillinger-Weber potentials, 701
Stochastic forces, 825
Stochastic time evolution, 672
Stoner model, 650
Stone–Wales rearrangements, 473
Stone–Wales transformation, 471–473
Storage modulus, 272
Strain-displacement law, 696
Strain distributions in QD/QWRs, 701
atomistic versus continuum calculations of, 701
Strain energy, 20
Strain gradient theory, 727
Strain–stress relation, 582
Strain tensor, 692
Stress analysis, 399, 401
Stress concentrations, 140
Stresses
biaxial, 452
Cauchy, 452, 453
compressive, 45
frictional, 430, 431
intrinsic, 450

residual, 451
shear, 433, 435
shrinkage, 447
thermal, 450
tensile, 451
uniaxial, 452
Stress perturbations, 129
Stress ratio, 451
Stress relief patterns, 450
Stress–strain behavior, 446
Stress–strain curves, 441, 445
with Tersoff and Tersoff–Brenner potentials, 442
Stress–strain relation, 441
Stress–strain relationship
of Si(100) and Si(110) under a biaxial tension, 452
of silicon under a uniaxial tension, 451
true, 452
Stress tensor, 401
Stress vector, 400, 401
on atomic scale, 400
Stretching frequency, 381
Stroh formalism, 713
Structural polymers, 263
mechanical properties, 263
Subband's energy, 751
Substrate
structural changes, 414
Sum over states (SOS) scheme, 236
Supercell approach, 561, 564, 586–588
for photonic crystal waveguides, 561
Supercell method, 775, 819
Supercell model, 562
Supercells, 631
Supercell structure, 67
Superconducting
flux quantum, 675
nanocircuits, 681
nanoparticles, 677
Superconducting nanostructures, 647
characteristic, 647
properties of, 647
Superconducting quantum interference device (SQUID), 675
Superconductivity, 668, 671
granular films, 679
mechanisms, 671
Superconductor, 656
Superconductor-normal-superconductor (SNS) transistors, 677
Superconductor-superconductor-superconductor (SSS) transistors, 677
I–V characteristic of, 677
Superlattice geometry, 61
Superposition law, 238
Super-resolution, 549, 551
lens, 551
microscopes, 549
Supershell
effect, 67
node, 67
oscillations, 67
Supramolecular networks, 126
formation of, 126
Surface adhesion, 411
Surface elastic effects, 721
Surface elastic modulus, 722
Surface free-energy difference, 159
Surface gradient operator, 721
Surface/interface energy effects, 720
Surface melting temperature, 90
Surface passivation, 756
Surface projection tensor, 721
Su-Shrieffer-Heeger (SSH) Hamiltonian, 230
SWNT-epoxy nanocomposites, 286
SWNT-reinforced polymers, 285
for glassy (elastic) modulus of, 285
Sword-insheath mechanism, 263
Symmetric power splitter, 611
Symmetry operations, 461
System configuration, 10
π-Systems, 204
theoretical approach, 206

T

Tamm-Dancoff approximation, 212
Tangent hard chains, 167
Taper model
2D, 609
Taper structures
top view on staggered PhC, 609
TB Hamiltonian, 523, 527, 528
TB interactions, 523
TBMD simulation, 522
Temperature conversion, 399
Temperature conversion models
comparison, 400
Tensile modulus, 318
wall layers, 372
Tersoff–Brenner (T-B) potential, 300, 440–444, 446
Tersoff energy, 404
Tersoff potential, 399, 402, 404, 409, 416, 441–443, 451, 520, 701
curve, 443
parameters, 404, 405
Tert-butyl-capped polyenes, 222
Tetracapped trigonal prism structure, 825
Tetragonal body-centered crystal structure, 409
Tetrahedral sites, 626
Theoretical gas phase calculations, 221
Thermodynamic equilibrium
between carbon gas phase and graphite, 466
treatment, 496
Thermodynamic gas–liquid equilibrium, 467

Thermodynamic gas–solid equilibrium, 467
Thermodynamic perturbation theory (TPT1), 165–167
  scheme, 166
Thermoset system, 271
Thermostat atoms, 443
Thermostat schemes
  influence of, 444
Thermostatting methods, 440
Thieno[3,4-b]-thiophene (TT), 220
Thin film buckling analysis
  simulation model, 450
Thiophene oligomers, 218, 222
Third-law entropies, 628
Thomas-Fermi
  approach, 54
  description, 803
  equation, 47, 49
  kinetic energy functional, 50
  method, 49
  model, 47, 53
  result, 55
Thouless energy, 670
Three-body contact sliding, 433–437
  material removal via adhering, 437
Three-way catalysis (TWC), 628
Tight-binding (TB)
  calculations, 18
  formalism, 667
  method, 5, 617, 742
  model, 522
Time-dependent density functional theory (TDDFT), 4, 74, 772, 783, 807
  calculations, 75
Time-dependent Hartree-Fock (TDHF) theory, 212
Time-dependent modulus, 272, 315
Time-domain methods, 590
Time-independent Schrödinger equation (TISE), 617
Time marching algorithm, 105
Time-reversal symmetry, 678
Time steps
  different types of errors, 398
  integration, 398
Time-temperature superposition, 273
  application, 273
Timoshensko beam, 380
β-Tin, 402, 406–409
  phase, 414
β-Tin silicon, 411, 439, 440
  atoms in structures, 408
  crystal structure, 408
  in configurations II and III, 439
  phase, 439
Tip–surface interactions, 34
Tip–surface separations, 35
Titania, 258
Titanium oxide, 638
  optical properties, 638
  qualitative differences, 638
Tool–workpiece interactions, 417
Top-down approaches, 309
Topological defects, 526
Topological generations
  enumerations and generations, 470
  of cages, 470
Torsional twisting frequency, 381
Total dielectric function, 755
Traditional mixing schemes, 780
Transcendental equation, 381
Transferability, 618
  limits, 618
Transitional aluminas, 618
  computed material densities, 619
Transition-mode diffusion, 29
Translational entropy, 180
Transmission electron microscope (TEM), 260, 402
Transport mechanisms, 649
Transverse kinetic energy, 659
Transverse loss modulus, 313
Transverse modulus, 313
  two-phase frequency domain, 313
Tricritical point (TCP), 181
Troullier–Martins formalism, 818
Troullier–Martins pseudopotentials, 62, 846
True ground-state energy, 3
Truncation boundary, 587
Truss mechanics, 305
Truss rod model, 385
Tuning parameter, 580
Tunneling rates, 671
Two-body contact sliding, 433–436
Two-dimensional electron gas (2DEG), 785
Two-dimensional quantum dots, 785
Two-level model, 763

U

Ultimate Jellium model, 76–77, 789
  for a breaking nanowire, 789
  snapshots from a simulation of nanowire breaking, 789
  types of nanocontact stabilization mechanisms, 790
Ultimate tensile strength (UTS), 448
Ultrahigh molecular weight polyethylene (UHMWPE), 390
Ultrasoft potentials, 60
Ultrasoft pseudopotentials of Vanderbilt USP, 618
Ultraviolet spectra, 205
Undulating and interacting wires, 756
Unit-cell techniques, 620
Unloading
  bond lengths and bond angles, 446
Unsaturated metal atoms, 629
Uphill movements, 8
Urethane acrylate polymer, 280

Ujsaki equation, 668
UV spectroscopy, 210

V

Valence electron, 67
densities, 52, 88
wavefunctions, 46
Valence energy eigenvalue, 58–59
Valence force field (VFF), 701
Valence Hamiltonian (VEH) technique, 217
Valence pseudo-orbital, 57, 59
Valence radial wavefunction, 59
Valence state properties, 819
Valence wavefunctions, 46
Vanadium, 640
Vanadium pentoxide, 641
types of surface exposed oxygen, 641
Vanderbilt's scheme, 60
van der Waals
attraction, 152, 514
bonding, 12
interactions, 175–176, 178, 227, 260
theory, 167
Van Hove singularities, 16
vant Hoff equation, 467
Vapor-grown carbon fibers, 287
Vapor–liquid–solid (VLS), 734
Variational density functional, 55
Velocity scaling method, 440
Verlet algorithm, 9
Verlet's method, 399
Vibrational contribution, 630
Vibrational frequencies, 642
two approximations, 642
Vibrational modes, 825
of a cylinder, 380
Vibration theory, 369
Vienna *ab initio* simulation package (VASP), 617
Virtual transitions, 840
Viscoelasticity, 254
introduction to, 269
molecular theory of, 270
of the interphase region, 257
viscoelastic phases, 257
volume fraction, 257
Viscoelastic materials
mechanical response of, 271
Viscoelastic time-dependent modulus, 271
Volume fraction, 107, 123, 133
combined, 125
profiles, 142–146
Von Weizsäcker functional, 55
Vortex clusters, 786
Vortex solutions, 788
types of vortices, 788
Vosko-Wilk-Nusair (VWN) [389] correlation, 214

W

Wave equation, 380
Wavefunction, 3, 49–50, 63
electronic, 79
Kohn-Sham, 63
one-electron, 61
Waveguide bend, 610, 611
frequency response, 610
phenomenological 2D model, 610
Waveguide discontinuities, 555, 564–568, 577, 578, 590–593
S matrix computation, 593
Waveguide modes
excitation, 592
Waveguides
abrupt transitions, 566
abromatic 90-degree bend in a PhC defect, 574
analysis of discontinuities, 565
bend, 610
bending, 570
coupling of dielectric slab and defect, 569
cylindrical, 562–564, 585
defect, 563
design of different types, 564
discontinuities, 564–568, 577, 590
in photonic crystals, 560
light coupling, 571
main purpose, 564
modification of termination, 569
numerical treatment of discontinuity, 567
optimized design of open photonic crystal, 571
other, 563
output ports, 594, 596
periodic, 585, 586
PhC, 563, 564, 567–572, 578, 586
radiating PhC, 589
simple open photonic crystal, 570
single-layer PhC, 589
single-mode, 578, 592, 593
slab, 568–572
slab bend improved by small PhC patch, 573
structures, 560, 585, 595
TE-polarized, 607
termination, 568, 569
transition obtained by tapering, 566
transitions, 566
W1, 607
W3, 607
Wave propagation, 585
Wavevector, 62
Wear
deformation regimes, 423
diagram, 436
mechanisms, 423
play, 420
process, 436
regimes, 435, 436

Weighted density approximation (WDA), 71, 165
Weight function, 159
Weighting function, 56, 165, 167–169
Welded CNT system, 19
Wertheim's perturbation theory, 165
Wetting phase diagram, 175, 181
Wetting transition, 172–174, 177
  first-order, 170–171, 173
  in Monte Carlo simulations, 170
  temperature, 170–171
White noise spectrum, 822
Wick's theorem, 651
Williams-Landis-Ferry (WLF) representation, 274
W metal, 640
Work–tool interactive potential, 401

X

X_α method, 806
X-ray photoelectron spectroscopy (XPS), 254
XY-model, 678

Y

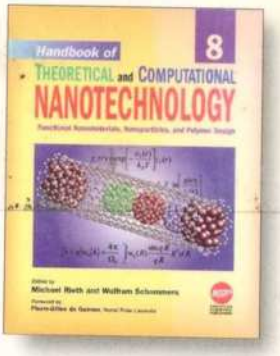
Yee cell, 114–115
Young's equation, 171, 173, 175, 180
Young's modulus, 20–22, 33–34, 111–112, 120–121, 124–126, 130, 140, 142, 262, 441, 446, 698
  global, 124, 128, 141

Z

Z-contrast scanning transmission electron microscopy (Z-STEM), 636
Zeeman energy, 650
Zeolites, 29, 263, 629
Zero bias anomaly (ZBA), 670
Zero-field results, 786
Zero-frequency Lindhard function, 781
Zero-order efficiencies, 545
Zero phonon transitions, 757
Zero-point energy
  vibrational, 463
Zero-point motion, 786
Zigzag tubes, 441, 443, 445, 446
  atomic chain, 444
  deformation, 445
  structural changes in different stages, 443
Zinc-blende semiconductors, 744
ZINDO, 475
  entropy, 478
  method, 491
Zipped structures, 518, 519
Zirconia, 628

Handbook of THEORETICAL and COMPUTATIONAL NANOTECHNOLOGY

Edited by **Michael Rieth** and **Wolfram Schommers**



The future applications of nanotechnology in high-tech industries require deep understanding of the theoretical and computational aspects of all kinds of materials and devices on a nanometer scale. ***Handbook of Theoretical and Computational Nanotechnology*** is the first single reference source ever published in the field that offers such an unified approach, covering all of the major topics dealing with theory, modeling, design, and simulations of nanostructured materials and nanodevices, quantum computing, computational chemistry, physics, and biology, nanomechanics, nanomachines, nanoelectronics, nanoprocesses, nanomagnetism, nanooptics, nanomedicines, nanobiotechnology, etc. This 10-volume handbook provides the first ideal introduction and an up-to-date survey of the fascinating new developments and interdisciplinary activities in the whole field presented by scientists working in different subject areas of science, engineering, and medicine. This handbook is the most profound publication on this topic-the first treatment of computational nanotechnology. This outstanding handbook, presented by the world's leading scientists, is the most significant academic title ever published in this research field. This handbook has been divided into 10 thematic volumes by documenting computational treatment of nanomaterials and nanodevices.

Volume 1: Basic Concepts, Nanomachines, and Medical Nanodevices
Volume 2: Atomistic Simulations—Algorithms and Methods
Volume 3: Quantum and Molecular Computing, Quantum Simulations
Volume 4: Nanomechanics and Multiscale Modeling
Volume 5: Transport Phenomena and Nanoscale Processes
Volume 6: Bioinformatics, Nanomedicine, and Drug Design
Volume 7: Magnetic Nanostructures and Nanooptics
Volume 8: Functional Nanomaterials, Nanoparticles, and Polymer Design
Volume 9: Nanocomposites, Nano-Assemblies, and Nanosurfaces
Volume 10: Nanodevice Modeling and Nanoelectronics

KEY FEATURES

- The World's first handbook ever published in the field of theoretical and computational nanotechnology.
- The first comprehensive reference dedicated to all disciplines of science, engineering, and medicine.
- Most up-to-date reference source drawing on the past two decades of pioneering research.
- About 140 Review chapters written by world leading scientists familiar with the current trends of nanotechnology.
- Over 8,000 pages written by 265 authors from 30 countries, truly international.
- 26,000 references, 4124 figures, 374 tables, and thousands of mathematical equations and formula.
- Clearly written, self-contained, timely, authoritative, and most comprehensive contributions.
- Extensive cross-refereeing in each chapter provides reader with a broader range of knowledge.
- Multidisciplinary reference source for scientists, engineers, biologists, medical experts and related professionals.

READERSHIP

This handbook is an invaluable reference source for scientists, engineers, and biologists working in the field of theoretical and computational nanotechnology. The handbook is intended for a broad audience working in the fields of quantum chemistry, physics, biology, materials science, electrical and electronics engineering, mechanical engineering, optical science, ceramic and chemical engineering, device engineering, aerospace engineering, computer science and technology, information technology, bioinformatics, biotechnology, medical sciences, medicine, surface science, and polymer science and technology.

ISBN 1-58883-050-0
9 781588 830500 90000

ASP **AMERICAN SCIENTIFIC PUBLISHERS**
Los Angeles, California, USA

Printed in the United States of America